

Atypical DNA replication
in *Haloferax volcanii*
in the absence of replication origins

Victoria Emily Smith, MSci (Hons)

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

December 2021



University of
Nottingham

UK | CHINA | MALAYSIA

Abstract

Replication origins are the sites of initiation of DNA replication. Origins are universal and have been assumed to be essential. However, the halophilic archaeon *Haloferax volcanii* is able to survive in the absence of origins, in fact growing faster than its wild-type counterpart. Replication in the absence of origins has been proposed to depend on replication-dependent replication (RDR), due to the requirement in the origin-deleted *H. volcanii* mutant for the recombinase protein, RadA.

In the work presented here, treatment of origin-deleted *H. volcanii* mutants revealed a tolerance of the PolB-specific inhibitor, aphidicolin, which increased with each origin deletion. While this suggests that originless strains may have a reduced requirement for the Family B DNA polymerase PolB, the replicative polymerases, PolB and PolD were found to be essential in both origin-containing and origin-deleted strains.

During eukaryotic DNA replication, the Cdc45 protein forms the CMG replicative helicase complex, in conjunction with MCM helicase and GINS. Archaeal species, including *H. volcanii*, encode counterparts of the MCM and GINS components of the CMG complex. Studies in *Thermococcus kodakarensis* have revealed that the GINS-associated nuclease (GAN), a RecJ-family protein, adopts the role of Cdc45. *H. volcanii* encodes four RecJ proteins (RecJ1-4). Phylogenetic, genetic and biochemical analysis carried out here suggests that RecJ1 acts as GAN but the role of RecJ2 remains unknown, while RecJ3 and RecJ4 have roles alongside the DNA repair protein Hef. All *recJ* genes were found to be dispensable in *H. volcanii*, with the exception of *recJ2*; overexpression of RecJ2 from an ectopic site was not sufficient to compensate for *recJ2* deletion from the wild type locus. Phenotypic analysis of *recJ* mutants in *H. volcanii* has shed light on the possible functions of these proteins, however questions remain around their specific roles.

Previous data has shown an increased requirement for essential replicative helicase MCM in the absence of DNA replication origins. GINS is essential in eukaryotes and the same is assumed for archaea. It was determined here that deletion of GINS is not possible in *H. volcanii*, but strains with inducible *ginS* alleles suggested that the requirement for GINS does not match that of its fellow CMG complex member MCM.

The interplay between DNA polymerases, replication origins and the CMG complex warrants further work. Differential usage of polymerases and CMG proteins in the presence or absence of origins could provide critical information on the mechanisms of both canonical and recombination-dependent DNA replication in archaea.

Acknowledgements

Well, that was a dramatic four years, wasn't it...?

Firstly, I would like to thank my supervisor Thorsten Allers for talking me into this project and for his ongoing advice, but also for tolerating me when I had a yearly breakdown about how everything was going to fail.

Thank you to the Nottingham BBSRC DTP for hiring both myself and Carl Aston onto the 2017 PhD cohort; of all the things I didn't expect to gain during my PhD, a husband is top of the list.

Thank you to my husband, Carl, for making the PhD experience enjoyable and constantly being my hype man (and also for being the provider of wine and hugs). A paragraph of acknowledgements will never be enough to explain how much you helped me through the last four years, I love you.

A big shout out to Stephen Gray (and his screaming goat) for keeping me sane/matching my insanity during the last year of the PhD. I am so glad I met you and look forward to seeing you and your lab thriving!

Thank you to my Mum, Dad and brothers for supporting and encouraging me, and for valiantly pretending to understand what archaea are ('they're like our ancestors but are also like bacteria'). Thanks to my gene bud and Marrow mate for listening to my complaints and telling me how capable I am, I appreciate you both hugely.

Wahooo! Who would have guessed reading and writing would pay off?!

- Homer Simpson

COVID-19 Impact Statement

Prior to providing a statement on the impact COVID-19 had on my own PhD, I would like to take this opportunity to think of those who have lost their lives, or the lives of loved ones to this disease. I have found the entire pandemic harrowing, have struggled to comprehend the huge numbers of people this has and will affect forever, and have nothing but respect for the time and effort the NHS and the plethora of other individuals that have dedicated to overcoming this. The pandemic impacts me daily and I am a different person because of it.

The COVID-19 pandemic caused the closure of the Allers' lab on 10th March 2020; this caused a sudden halt on all ongoing laboratory work, only overcome when the lab reopened on 29th July 2020. During this time, I was able to dedicate my time to writing my thesis and had the opportunity to perform bioinformatic analyses that would likely not have been completed had we not been based at home. Upon return to labs, social distancing meant a shift system was implemented; from 29th July 2020 until my exit from labs in July 2021, I only had access to labs from 1.30pm onwards (for 9 of 10 working days; every other Friday was a full day for one shift), limiting the time I had available for lab projects, with a major impact on certain day-long protocols which could only be performed every other week. Additionally, resources were scarce, and orders had severe time delays. However, with time, I was able to adapt and work around this and appreciate the positive impact it had on my multitasking and time-keeping abilities.

One of the largest hurdles to overcome following the COVID shutdown was the lack of radioactivity; from its cessation in March 2020, radioactive work did not resume until late January 2021. *Haloferax* genetics projects require radiation for confirmation of strain generation, due to their polyploid nature. The lack of this resource impacted the majority of my ongoing projects and instead, I had to focus my time away from these and generate ideas for new projects that would not rely on strain generation. This was somewhat tricky but forced me to ask the questions I wanted to answer in a different way and taught me some important lessons about research.

While I could easily frame this list negatively, I also want to highlight that the COVID pandemic also taught me a number of lessons that I would not have learnt without it. I am proud of finishing a PhD during a pandemic, and believe I adapted well to overcome this.

Abbreviations

5-FOA	5-fluoroorotic acid
AAA+	ATPase associated with diverse cellular activities
ACL	Allosteric control loop
ADP	Adenosine 5'-diphosphate
Amp	Ampicillin
arCOG	Archaeal cluster of orthologous genes
ATP	Adenosine 5'-triphosphate
BER	Base excision repair
BIR	Break-induced repair
BLAST	Basic Local Alignment Search Tool
bp	Base pair(s)
COG	Cluster of orthologous genes
crRNA	CRISPR-derived RNA
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRi	CRISPR interference
DMSO	Dimethyl sulfoxide
D-loop	Displacement loop (DNA-based)
DNA	Deoxyribonucleic acid
DNAP	DNA polymerase
dNTP	Deoxynucleotide
DPBB	Two-double-psi- β -barrel
DS	Downstream
DSB	Double strand break
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
DUE	DNA unwinding element
E	Elution
EDTA	Ethylenediaminetetraacetic acid
GFP	Green fluorescent protein
GGR	Global genome repair
HE	Holoenzyme
HJ	Holliday junction
His	Histidine
HR	Homologous recombination
HTH	Helix-turn-helix
IDR	Intrinsically disordered region
kb	Kilobase(s)
LB	Lysogeny broth
Leu	Leucine
Mb	Megabase(s)
MMC	Mitomycin C
MMEJ	Microhomology-mediated end joining
MMR	Mismatch repair
MW	Molecular weight
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
OB	Oligonucleotide binding
OD	Optical density

ORB	Origin-recognition box
ORC	Origin recognition complex
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
PDB	Protein database
PEG	Polyethylene glycol
PIP	PCNA-interaction protein motif
PMSF	Phenylmethanesulfonylfluoride
R-loop	RNA-based displacement loop
RDR	Recombination-dependent replication
rf	Relative fluorescence
RF	Recombination frequency
RNA	Ribonucleic acid
RNAP	RNA polymerase
rRNA	Ribosomal RNA
SAMP	Small archaeal modifier protein
SDS	Sodium dodecyl sulfate
ssDNA	Single-stranded DNA
SSPE	Saline sodium phosphate EDTA
Strep	Streptavidin
SW	Salt water
TAE	Tris acetic acid EDTA
TBE	Tris borate EDTA
TCA	Trichloroacetic acid
TCE	2,2,2-Trichloroethanol
TCR	Transcription-coupled repair
TE	Tris EDTA/transformation efficiency
TEMED	Tetramethylethylenediamine
TEX	Terminator endonuclease
Thy	Thymidine
TLS	Translesion synthesis
Trp	Tryptophan
TTS	Transcription termination site
US	Upstream
Ura	Uracil
UV	Ultraviolet light
v/v	Volume per volume
w/v	Weight per volume
WH	Winged helix
WT	Wild type
X-gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside

Table of Contents

Abstract.....	i
Acknowledgements	iii
COVID-19 Impact Statement	iv
Abbreviations.....	v
Chapter 1: Introduction.....	1
1.1 Archaea.....	1
1.1.1 <i>Haloferax volcanii</i> as a model organism	7
1.2 DNA replication	19
1.2.1 Bacterial replication	27
1.2.2 Eukaryotic replication.....	34
1.2.3 Archaeal replication.....	43
1.3 Alternative methods of DNA replication	48
1.4 CRISPR interference (CRISPRi)	53
1.5 Aims and Objectives	57
Chapter 2: Materials and Methods.....	59
2.1 Materials.....	59
2.1.1 Strains	59
2.1.2 Plasmids	65
2.1.3 Oligonucleotides	73
2.1.4 Chemicals and Enzymes	79
2.1.5 Media and Solutions	79
2.2 Methods.....	83
2.2.1 General <i>Escherichia coli</i> Microbiology.....	83
2.2.2 General <i>Haloferax volcanii</i> Microbiology.....	84
2.2.3 DNA Extraction	85
2.2.4 Nucleic Acid Manipulation.....	87
2.2.5 Genetic Manipulation of <i>Haloferax volcanii</i>	91
2.2.6 Genotype Screening.....	94
2.2.7 Phenotype Screening.....	96
2.2.8 Protein Expression and Purification.....	104
Chapter 3: Cellular requirement for replicative DNA polymerases PolB and PolD in <i>Haloferax volcanii</i>	109
3.1 Background	109
3.2 Aims and Objectives	119
3.3 Results	120
3.3.1 Altered requirement for PolB in the absence of origins.....	120
3.3.2 Deletion of <i>polB1</i> (HVO_0858).....	124
3.3.3 Attempted deletion of <i>polD2</i> (HVO_0065).....	133
3.3.4 Assessing the altered requirement for PolB and PolD expression using inducible promoters.....	137

3.3.5 CRISPR interference (CRISPRi) against replicative polymerases	166
3.4 Discussion.....	175
3.5 Conclusion.....	180
Chapter 4: Genetic analysis of <i>recJ</i> genes in <i>Haloferax volcanii</i>	181
4.1 Background.....	181
4.2 Aims and Objectives.....	190
4.3 Results	191
4.3.1 In silico analysis of RecJ proteins in Archaea.....	191
4.3.2 Characterisation of strains deleted for <i>recJ1</i> , <i>recJ3</i> and/or <i>recJ4</i> ..	205
4.3.2.1 Dependence of $\Delta recJ1$ mutant on <i>radA</i> expression.....	220
4.3.2.2 Replication profiling of <i>recJ</i> mutants	226
4.3.3 Genetic interactions with <i>recJ</i> genes in <i>H. volcanii</i>	229
4.3.6.1 Deletion in combination with Hef and Hjc	231
4.3.6.2 Deletion in combination with origins of replication	240
4.3.6.3 Deletion in combination with RNase HII (<i>rnhB</i>) and Flap endonuclease 1 (<i>fenI</i>)	246
4.3.6.4 Deletion in combination with Hel308.....	265
4.3.4 Confirming the essentiality of <i>recJ2</i> in <i>H. volcanii</i>	271
4.3.4.1 Purification of RecJ2.....	289
4.3.5 Identification of RecJ protein-protein interactions using affinity purification	296
4.3.5.1 Purification of RecJ1.....	309
4.3.5.2 Purification of RecJ3.....	311
4.3.5.3 Purification of RecJ4.....	312
4.3.6 Confirmation of RecJ protein-protein interactions using Split-GFP	316
4.4 Discussion.....	334
4.5 Conclusion.....	345
Chapter 5: Deciphering the role of the CMG replicative helicase complex in <i>Haloferax volcanii</i>	346
5.1 Background.....	346
5.2 Aims and Objectives.....	360
5.3 Results	361
5.3.1 Deletion of <i>ginS</i>	361
5.3.2 Assessing the requirement for <i>ginS</i> expression using inducible promoters.....	363
5.3.3 Deletion of <i>ginS</i> incorporating <i>priS</i> transcription termination site	378
5.3.4 Identification of CMG complex members through co-purification	384
5.3.5 Investigating the increased requirement for MCM helicase in absence of origins	398
5.4 Discussion.....	406
5.5 Conclusion.....	412
Chapter 6: Discussion	413

Chapter 7: References.....	421
Chapter 8: Appendix.....	450

Chapter 1: Introduction

The research subject of this thesis is the archaeon *Haloferax volcanii*. During the explanation of the phylogeny of archaea, please note that *H. volcanii* is a member of the Euryarchaeota. The biology of this model organism will be discussed in more detail following an overview of the Archaea.

1.1 Archaea

Prior to the advent of nucleotide sequencing, organisms were often classified based on visible characteristics, including cell morphology and growth conditions. Pioneering work by Carl Woese and George Fox used RNA analysis techniques to redefine the classification of prokaryotes into two distinct domains: Bacteria and Archaea (then named Eubacteria and Archaeobacteria) (Woese and Fox, 1977).

Small subunit ribosomal RNA (16S rRNA) is a slow-evolving and ubiquitously expressed component in all free-living organisms. Its strong sequence conservation allows single base pair changes to dictate phylogenetic mapping, whereby 16S rRNA acts as a molecular chronometer. These results revealed a new structure for the tree of life containing three domains of life: Eukarya, Bacteria and Archaea (**Figure 1.1**). Further study of this newly defined domain showed whilst previously misclassified as Bacteria, Archaea were more similar to Eukarya at the genetic level (Woese *et al.*, 1990). This was subsequently supported by the discovery of several molecular features shared only between eukaryotes and archaea, including histone proteins and multiple origins of replication. The hypothesis that archaea and eukaryotes shared a common ancestor, subsequent to the split from the bacteria, led to the definition of archaea and eukaryotes as ‘sister’ groups (Pace, 1997).

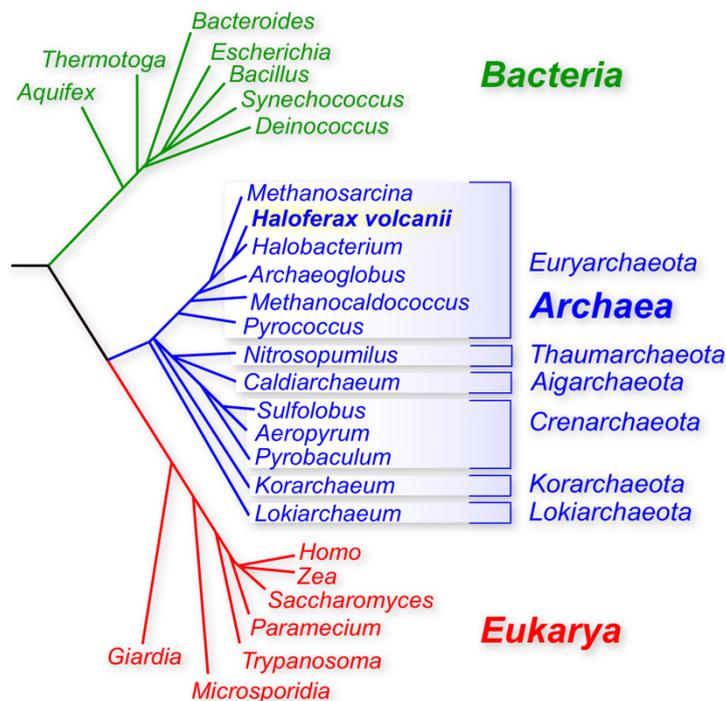


Figure 1.1: Three domain representation of the tree of life

Phylogenetic tree based on 16S rRNA sequencing, defining three domains of life; Bacteria, Archaea and Eukarya. Adapted from (Allers and Mevarech, 2005).

Whilst this work was met with resistance in the scientific community, subsequent 16S rRNA sequencing was able to further define the archaeal domain into two distinct phyla: Euryarchaeota and Crenarchaeota (Winker and Woese, 1991). Advances in sampling and sequencing protocols over the past decades have provided the means to collect genomic data from a variety of archaeal lineages, increasing the number of defined archaeal phyla and superphyla. Together Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota were shown to be part of a monophyletic group, the TACK superphylum (Guy and Ettema, 2011). It is believed the split within the TACK superphylum occurred before the split of Euryarchaea and Crenarchaea (Auchtung *et al.*, 2006, Elkins *et al.*, 2008). The DPANN superphylum is a more recent definition, first proposed in 2013, defined according to the first phyla identified: Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaea (Dombrowski *et al.*, 2019, Rinke *et al.*, 2013). Since then additional phyla have been grouped within the DPANN superphylum, including Woesearchaeota, Pacearchaeota and Altiarchaeota (Spang *et al.*, 2017).

Recent metagenomics studies have brought Carl Woese's three domain tree (as portrayed in **Figure 1.1**) into question; new data argues for a tree of life that in fact clusters into two primary domains, Bacteria and Archaea, where Eukaryotes are a direct descendent of Archaea, from within the TACK superphylum (**Figure 1.2**) (Embley and Williams, 2016, Guy and Ettema, 2011, Koonin and Yutin, 2014). This proposal is based on comparison of the presence of core components of eukaryotes (eukaryotic signature proteins), including:

- ubiquitin signalling
- cytoskeletal structures
- trafficking machinery
- methods of RNA interference

Archaeal species have now been identified carrying the above components, however a single archaeal species carrying all eukaryote-like components is yet to be identified (Eme *et al.*, 2017) (**Figure 1.3**). For example, archaea have been shown to utilise a system of small archaeal modifier proteins (SAMPs) in a process named SAMPylation, which shares similarities with the eukaryotic ubiquitin-proteasome system (Maupin-Furlow, 2013).

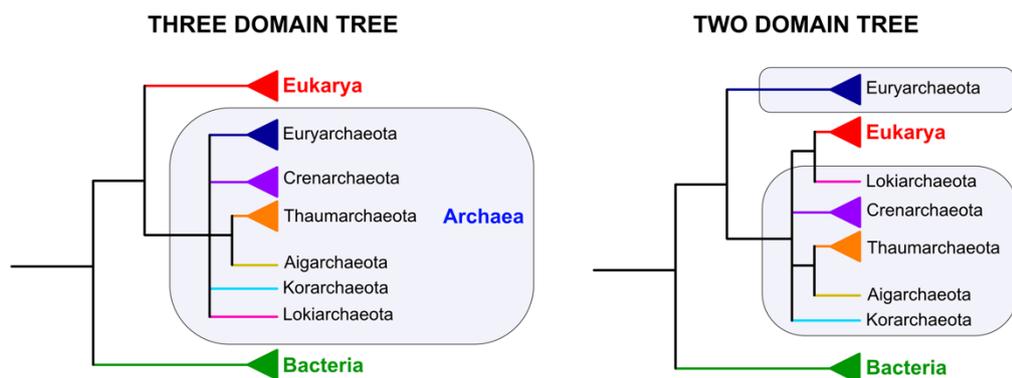


Figure 1.2: Three domain and two domain representation of the phylogenetic trees of life. The three-domain tree divides cellular life into three separate major groups: Eukarya, Archaea (within shaded box) and Bacteria. In the three-domain model, Eukarya are shown to have a common prokaryotic ancestor with Archaea. The two-domain tree shows Eukarya nested within the Archaea, with Lokiarchaeota mapping phylogenetically as its closest relative. Therefore, in the two-domain model, the eukaryal ancestor was already an archaeal species. Adapted from (Embley and Williams, 2016).

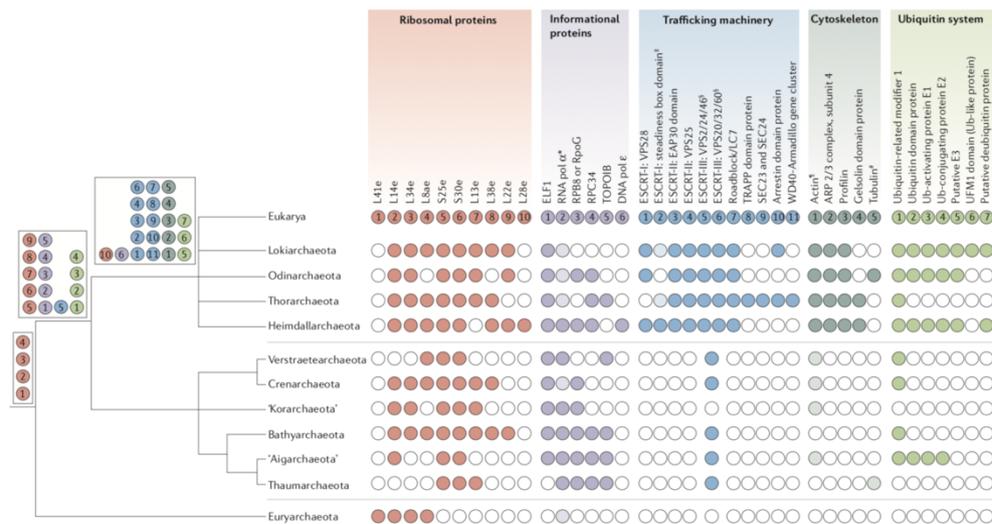


Figure 1.3: Analysis of eukaryotic signature proteins (ESPs) present in the first eukaryotic common ancestor. Homologues of ESPs in various archaeal lineages are represented by filled circles. The origin of each ESP is indicated on a schematic tree of life on the left, with Eukarya represented as the top branch. Time of emergence of each ESP is predicted from left to right across each group of proteins. The lowest grouping (Euryarchaeota) represents the model organism of this thesis, *Haloferox volcanii*. This group also contains model organisms of the genus *Pyrococcus* and *Thermococcus*. Model genus *Sulfolobus* falls within the grouping of Crenarchaeota. Figure taken from (Eme *et al.*, 2017).

Additionally, some archaeal species have been found to carry core cell division machinery homologous to the ESCRT (Endosomal Sorting Complex Required for Transport) machinery used for membrane abscission in vertebrate cells (Samson *et al.*, 2017, Samson *et al.*, 2008). The presence of these key eukaryotic components within Archaea provides strong support for the hypothesis that Eukaryotes evolved from Archaea. In addition, TACK species generally have defined cell cycle phases, as in eukaryotes, while studies on euryarchaeal species suggest less regulation of the cell cycle occurs (Samson *et al.*, 2017). Numerous euryarchaeal species have been shown to carry multiple genome copies (polyploidy), which contrasts with members of the TACK superphylum that have been characterised; all show a chromosome copy number oscillation of one to two during their cell cycles (Lundgren *et al.*, 2008).

Sequencing of previously uncultivated archaea has revealed additional lineages that fall within the archaeal clade Lokiarchaeota (Zaremba-Niedzwiedzka *et al.*, 2017, Spang *et al.*, 2015). Metagenomic profiling since

defined further archaeal phyla, including Thorarchaeota, Odinararchaeota and Heimdallarchaeota, which together form the Asgard superphylum (Zaremba-Niedzwiedzka *et al.*, 2017, Seitz *et al.*, 2016). Of the Asgard archaea, Lokiarchaeota represent the closest ancestor to Eukaryotes when mapped phylogenetically, providing support for the two-domain model of the tree of life. Members of this superphylum were shown to encode proteins previously thought to be eukaryotic-specific, including a homologue to replicative DNA polymerase Pol- ϵ (Zaremba-Niedzwiedzka *et al.*, 2017). Asgard archaea live in a wide range of environments, and as such have been shown undergo a versatile range of metabolic activities, including carbon fixation, nucleotide salvaging pathways, phototrophy, nitrogen cycling and sulphur cycling (MacLeod *et al.*, 2019). They have also been shown to be able to utilise atmospheric nitrogen or nitrate as nitrogen sources. This wide range of metabolic processes may be important evidence for the two-domain model; at the origin of eukaryotes it would be expected there would be a metabolic symbiosis event (Spang *et al.*, 2019).

This two-domain hypothesis is not without criticism: it has been shown that removal of a single protein sequence from the metagenomic data breaks the strong relationship seen between Lokiarchaeota and Eukaryotes (Da Cunha *et al.*, 2017). Metagenomics is known to be problematic and contamination can be an issue and thus this data alone could not be argued to be proof of a close relationship between Archaea and Eukaryotes.

Recent *in vivo* cultivation of Asgard archaeon *Prometheoarchaeum syntrophicum* has tempered the power these criticisms held (Imachi *et al.*, 2020). Sequencing of the genome of cultured cells confirmed the previously observed close relationship between Lokiarchaeota and Eukaryotes, with the species carrying a high number of eukaryote-like genes. The cells are morphologically complex and have unique protrusions that are long and often branching: these structures are proposed to be key for eukaryogenesis. More specifically, they predict the E³ model of entangle-engulf-endogenise (Imachi *et al.*, 2020), whereby the host archaeon uses its protrusions to catch and engulf surrounding bacterial species leading to endosymbiosis and formation of the first primitive eukaryotic cell.

Archaea are often thought of extremophiles due to the harsh environments in which they live, including extremes of temperature, acidity, salinity and alkalinity. While this generalisation can be applied to numerous archaeal species, some archaea are found in non-extreme environments. Archaeal species have been identified in relatively 'normal' environments, including soils, fresh-water sediments and the human gut, where they co-habit with bacterial species (Khelaifia and Raoult, 2016, Chaban *et al.*, 2006).

Genome organisation

In spite of their shared ancestry with eukaryotes, archaea exhibit many bacterial characteristics. Physically, archaea are similar to bacteria in that they are single-celled organisms lacking cellular organelles and a nuclear envelope. At the genetic level archaea, like bacteria, possess circular double stranded DNA (dsDNA) chromosomes. Depending on species, they may have single or multiple origins of replication (Wu *et al.*, 2014b), which give rise to bidirectional replication forks. They have relatively small genomes (~2-4 Mb, compared to ~3,000 Mb in humans) and their genes are often organised into an operonic layout, as seen in bacteria (Olsen and Woese, 1997).

In eukaryotes, large genome sizes require extensive packaging for compaction, which is accomplished by histones and various chromatin-binding proteins. This compaction affects nuclear processes including replication and transcription by altering the availability of DNA to proteins, along with providing the genome with protection from damage. There is no common chromatin protein shared by all archaeal species; some have been shown to have histones as in eukaryotes, while other species may contain various archaea-specific chromatin-binding proteins (Reeve, 2003). Archaeal histones are a simplified version of the eukaryotic counterparts; archaea have homotetrameric histones, in comparison with hetero-octameric eukaryotic histones (Reeve *et al.*, 1997). An archaeal-specific chromatin-binding protein, Alba, is found in various phyla including crenarchaea. Its commonality in crenarchaeal species, which lack histones, implicate Alba in DNA packaging. There is also evidence for Alba having a role in protection from nuclease damage (Peeters *et al.*, 2015).

In general, archaea appear to have a higher gene density than bacterial or eukaryotic genomes (Koonin and Wolf, 2008), suggesting archaeal genomes are more compact. This is due to the cumulative effect of minor differences in protein lengths and intergenic region length when compared to bacterial counterparts (Koonin and Wolf, 2008). However, it is worth noting that when comparing differences between bacterial and archaeal genomes, this is minor compared to the vast difference in genome layouts when comparing both archaea and bacteria to eukaryotic species.

Information processing

Similarities between archaea and bacteria, including phenotypic appearance and genome organisation, suggested a close relationship between archaea and bacteria. Early work by Wolfram Zillig on archaeal RNA polymerase (RNAP) showed an evolutionary link not between archaea and bacteria, but between archaea and eukaryotes (Huet *et al.*, 1983) (**Figure 1.4**).

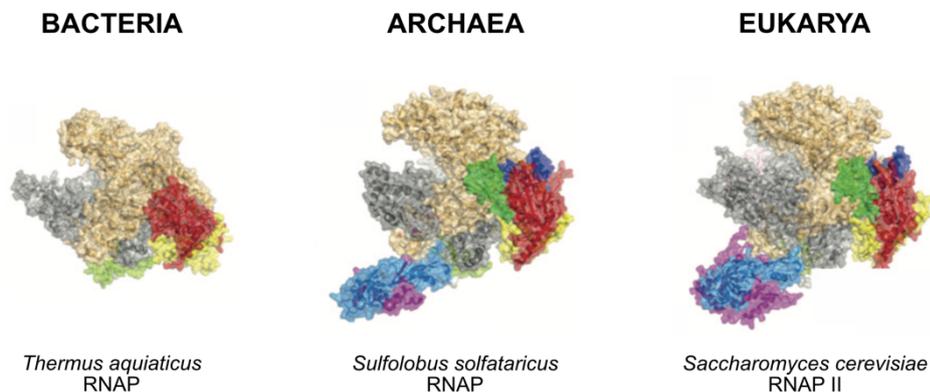


Figure 1.4: Comparison of RNA polymerase structures from the three domains of life. Homologous subunits are colour-coded, highlighting the similarities between all three, with an emphasis on similarity between archaeal and eukaryotic RNA polymerase structures. Adapted from Albers *et al.*, 2013.

This discovery undermined the previously held view regarding complex transcription machinery only being found in relatively ‘complex’ eukaryotes compared to the more basic machinery found in ‘simple’ prokaryotes. This result prompted further investigations into other central enzymes in archaea, including DNA polymerases and topoisomerases (Albers *et al.*, 2013), to assess for similarities between archaeal and eukaryotic information processing systems. It is now widely accepted that the information processing genes of archaea, involved in major processes such as DNA replication, transcription and repair, are generally more similar to the eukaryotic counterparts than the generally simplified system in bacteria (Barry and Bell, 2006). However, it is worth noting some mechanisms in archaea are much simpler than the eukaryotic counterpart, and some carry only the bacterial homologues (Ishino *et al.*, 2013).

1.1.1 *Haloferax volcanii* as a model organism

Haloferax volcanii is a halophilic euryarchaeon that was first isolated from the hypersaline Dead Sea (Mullakhanbhai and Larsen, 1975). It is a well-established model organism for the study of DNA replication, homologous recombination and repair, with numerous genetic tools available (Allers and

Ngo, 2003, Allers *et al.*, 2004, Perez-Arnaiz *et al.*, 2020). Growth of *H. volcanii* is relatively simple to maintain; it grows aerobically at 45°C and at a salt concentration of 1.7 - 2.5 M NaCl and can grow in both rich and minimal media. It has a generation time of ~ 2 hours and can form colonies on agar plates within 4-5 days at 45°C.

H. volcanii is yet to have shown a defined cell cycle, however this may be due to its high ploidy making it hard to define (Lindås and Bernander, 2013). Instead, the organism carries out concurrent rounds of replication. This means at any time, a locus found at a replication origin may have fired, extending replication forks in both directions, and be priming to fire again while the terminus region remains unreplicated. With this example, the origin would have a copy number at the origin of 3, while the terminus region would have a copy number of 1 (ratio 3:1). If concurrent replication was occurring and all origins were to fire simultaneously, the copy number would only ever double from origin to terminus (ratio 2:1). In fact, deep sequencing replication profiles have provided evidence for non-synchronous origin firing, showing a maximum:minimum copy number ratio of >2:1 (Hawkins *et al.*, 2013b).

Genome structure

The wild type *Haloferax volcanii* genome (strain DS2) consists of five separate circular components (Hartman *et al.*, 2010): the main chromosome (2.85 Mb), mega-plasmids pHV1 (85 kb), pHV3 (438 kb), pHV4 (636 kb) and small plasmid pHV2 (6.4 kb). Laboratory strain H26 has been experimentally cured of plasmid pHV2 (Wendoloski *et al.*, 2001), which inadvertently resulted in the integration of pHV4 onto the main chromosome via recombination between two identical insertion sequence (IS) elements (Hawkins *et al.*, 2013a, Wendoloski *et al.*, 2001), giving an increased chromosome size of 3.5 Mb (**Figure 1.5**). The complete genome sequence for *H. volcanii* is readily available (Hartman *et al.*, 2010). Like almost all haloarchaea, the organism is highly polyploid, with up to 20 genome copies per cell, and has a GC content of 65% (Breuert *et al.*, 2006).

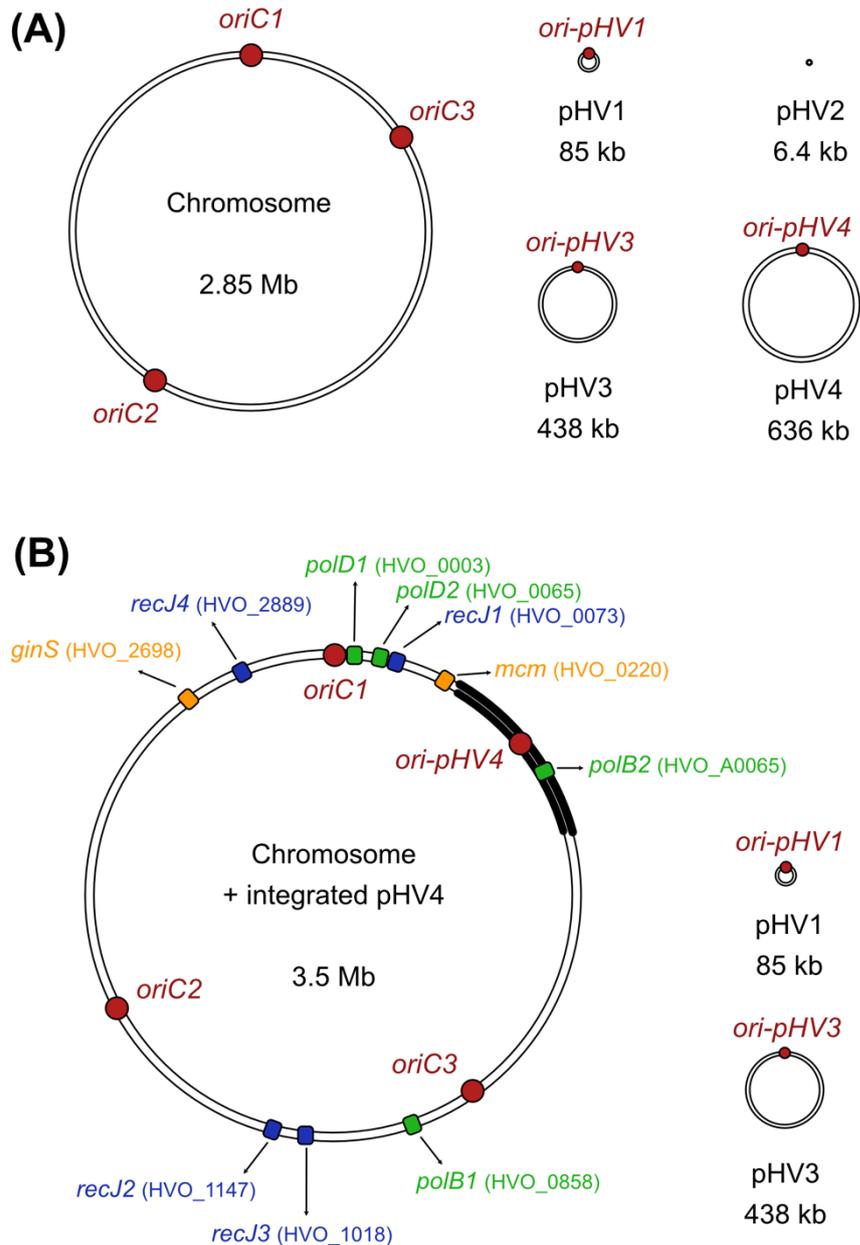


Figure 1.5: Genome architecture of *Haloferax volcanii* and loci of key genes used in this study. (A) Genome structure of wild type *Haloferax volcanii* strain DS2, containing the main chromosome, mega-plasmids pHV1, pHV3, pHV4 and small plasmid pHV2. (B) Genome structure of laboratory strain H26, where mega-plasmid pHV4 has been integrated onto the chromosome (thick black line) and pHV2 has been deleted. Genes of interest in this study have been annotated and coloured according to the chapter in which they are studied in this thesis (Chapter 3 = green, Chapter 4 = blue, Chapter 5 = orange). Locations are approximate and not to scale.

Deep sequencing-based marker frequency analysis has defined six replication origins in wild type *H. volcanii*, with each being adjacent to a gene for an Orc1/Cdc6 initiator protein (Hawkins *et al.*, 2013a). In the laboratory strain the main chromosome has four origins due to the integration of megaplasmid pHV4, named *oriC1*, *oriC2*, *oriC3* and *ori-pHV4*, along with plasmids pHV3 and pHV1 having their own discrete replication origins, *ori-pHV3* and *ori-pHV1* (Figure 1.5B).

Cell structure

Haloferax volcanii cells are disc-shaped and are surrounded by a glycoprotein surface (S-) layer, the subunits of which are held together by divalent cations such as Mg^{2+} (Cohen *et al.*, 1991). The cells are pigmented red by the production of carotenoids (Rosenshine *et al.*, 1989, Fineran, 2019) (Figure 1.6). They range in size from 1-3 x 2-3 μm and are between 0.4 – 0.5 μm in thickness (Mullakhanbhai and Larsen, 1975). *H. volcanii* are rod-shaped cells (Pohlschroder and Schulze, 2019). Growth occurs by elongation of the cells, with reproduction via binary fission (Mullakhanbhai and Larsen, 1975). Interestingly they are also capable of gene transfer using a natural mating system (Mevarech and Werczberger, 1985, Shalev *et al.*, 2017, Makkay *et al.*, 2020).

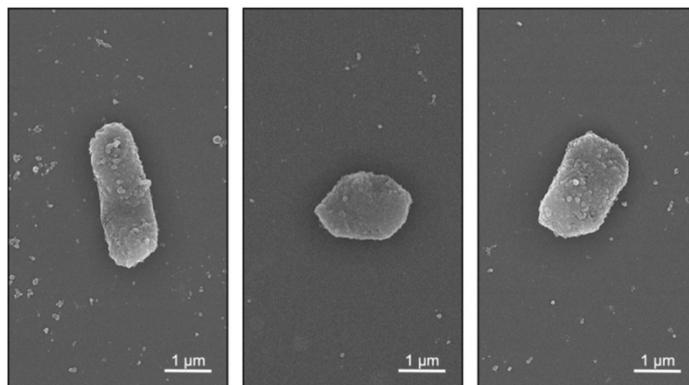


Figure 1.6: Scanning electron microscopy showing morphology of *Haloferax volcanii* cells. Image taken from (Wörtz *et al.*, 2022).

H. volcanii does not have a cell wall, as seen in bacteria with peptidoglycan walls, however its single-layer glycoprotein surface layer (S-layer) can be glycosylated. This acts as a structure to maintain rigidity. The S-layer coincidentally also provides the cells with preventative measure against DNA uptake but can be intentionally removed by chemicals or detergents (e.g., EDTA used for removal of S-layer in *H. volcanii* transformation) (Cline *et al.*, 1989b).

The membrane of *H. volcanii* is typical of those found in archaea. While bacteria and eukaryotes typically use fatty acid side chains in their membranes, archaeal membranes consist of ether-linked isoprenoid lipids (**Figure 1.7**) (Albers and Meyer, 2011). Generally, bacteria and eukaryotes will use ester linkages and fatty acid side chains, however, ether linkages have been seen in some bacterial membranes (Lombard *et al.*, 2012), and isoprenoid side chains have been observed in all three domains of life (Sojo *et al.*, 2014). Archaea and bacteria also differ in the composition of the phospholipid headgroup. Bacteria (and eukaryotes) utilise a glycerol-3-phosphate (G3P) headgroup, while archaea will use glycerol-1-phosphate (G1P) in their membranes.

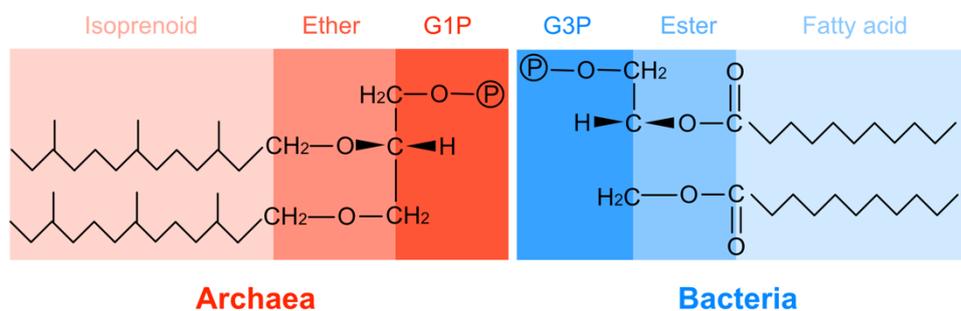


Figure 1.7: Membrane lipid composition in Archaea and Bacteria. Archaeal membrane lipids (red) are composed of isoprenoid side chains, ether-linked to a glycerol-1-phosphate (G1P) backbone. Bacterial membrane lipids (blue) are typically composed of fatty acid chains, ester-linked to a G3P backbone. Figure adapted from (Sojo *et al.*, 2014).

H. volcanii have adapted to live in high salt by utilising a ‘salt-in’ mechanism (Siglioccolo *et al.*, 2011). Most halophilic bacterial species use a ‘salt-out’ strategy, whereby they exclude salt from the cytoplasm and accumulate high levels of organic solutes, such as glycerol, to create an osmotic balance with the environment (Oren, 2008, Christian and Waltho, 1962). By contrast, the salt-in method allows the internal salt concentration to be maintained at the same salinity as the environment (Oren *et al.*, 2002); in the case of *H. volcanii* this is $\sim 1.7 - 2.5$ M NaCl. While the salt-out mechanism does not require protein modification to deal with the high concentration of solutes, the high internal salt concentration associated with the salt-in mechanism means cytoplasmic proteins must adapt to fold in the presence of high ionic concentrations (Siglioccolo *et al.*, 2011). They typically have a large number of acidic residues and a negatively charged external surface, while hydrophobic residues are confined to the core of the protein (for an example, see **Figure 1.8** showing the structure of proliferating cell nuclear antigen,

PCNA. Please note, the role of PCNA will be discussed further in section 1.2) (Winter *et al.*, 2009). This outer negative charge is thought to help with protein solubility; however, it is also likely to cause protein denaturation at low salt conditions. This must be accounted for when attempting protein purification from halophilic organisms (see *Genetic tools for manipulation*). The salt-in strategy also means *Haloferax* species will struggle to adapt to low salt conditions and therefore cannot grow in low salt media (Oren, 2008).

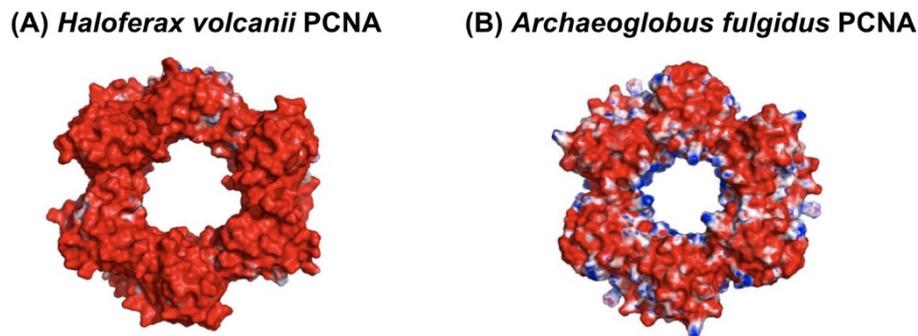


Figure 1.8: Comparison of proliferating cell nuclear antigen (PCNA) sliding clamp proteins in halophilic (A) and non-halophilic (B) archaea. Red denotes acidic/negative residues, blue denotes positive/basic residues. (A) *Haloferax volcanii* PCNA contains highly negatively charged, acidic surface residues. (B) *Archaeoglobus fulgidus* PCNA contains positively charged, basic residues. Adapted from (Winter *et al.*, 2009).

Genetic tools for manipulation

Haloferax volcanii is a well-established model organism within the field of archaeal genetics. There are multiple selectable markers targeting amino acid or nucleotide biosynthesis that are utilised for genetic manipulation of the organism (**Table 1.1**). The promoter of ferredoxin gene *p.fdx*, isolated from close relative *Halobacterium salinarium*, will ensure strong expression of the selectable markers *in vivo* to allow for direct selection of transformants (Gregor and Pfeifer, 2005, Pfeifer *et al.*, 1993).

Table 1.1: Common selection markers used in the genetic manipulation of *Haloferax volcanii*. * indicates the most commonly used selection markers.

Gene name	Selection	Reference
<i>pyrE2</i> *	Uracil biosynthesis	(Bitan-Banin <i>et al.</i> , 2003)
<i>trpA</i> *	Tryptophan biosynthesis	(Allers <i>et al.</i> , 2004)
<i>leuB</i>	Leucine biosynthesis	(Allers <i>et al.</i> , 2004)

<i>hdrB*</i>	Thymidine biosynthesis	(Ortenberg <i>et al.</i> , 2000)
<i>metX</i>	Methionine biosynthesis	(Leigh <i>et al.</i> , 2011)
<i>argH</i>	Arginine biosynthesis	(McMillan <i>et al.</i> , 2018)
<i>lysA</i>	Lysine biosynthesis	(McMillan <i>et al.</i> , 2018)

Transformation of *H. volcanii* is relatively easy; EDTA treatment removes the surface layer (through chelation of divalent magnesium ions), after which polyethylene glycol (PEG 600) aids uptake of DNA by cells (Cline *et al.*, 1989a). *H. volcanii* encodes a restriction endonuclease, *mrr*, which targets *dam*-methylated DNA for degradation. Numerous bacterial species use *dam* methylation, whereby the adenine of 5'-GATC-3' gains a methyl group in newly synthesised DNA. Degradation of *dam* methylated DNA thus acts as a defence mechanism for *H. volcanii* against incoming foreign DNA. Therefore, transformations into *H. volcanii* must be performed using plasmid DNA lacking *dam* methylation, or *mrr* must be deleted to allow transformation with *dam*⁺ DNA (Allers *et al.*, 2010). While transformation with linear DNA is possible, this reduces the efficiency approximately 100-fold (Delmas *et al.*, 2009). The ease of transformation can be utilised, along with selectable markers (including *pyrE2* and *trpA*; **Table 1.1**), to carry out gene deletions using the pop-in/pop-out methodology (**Figure 1.9**) (Bitan-Banin *et al.*, 2003).

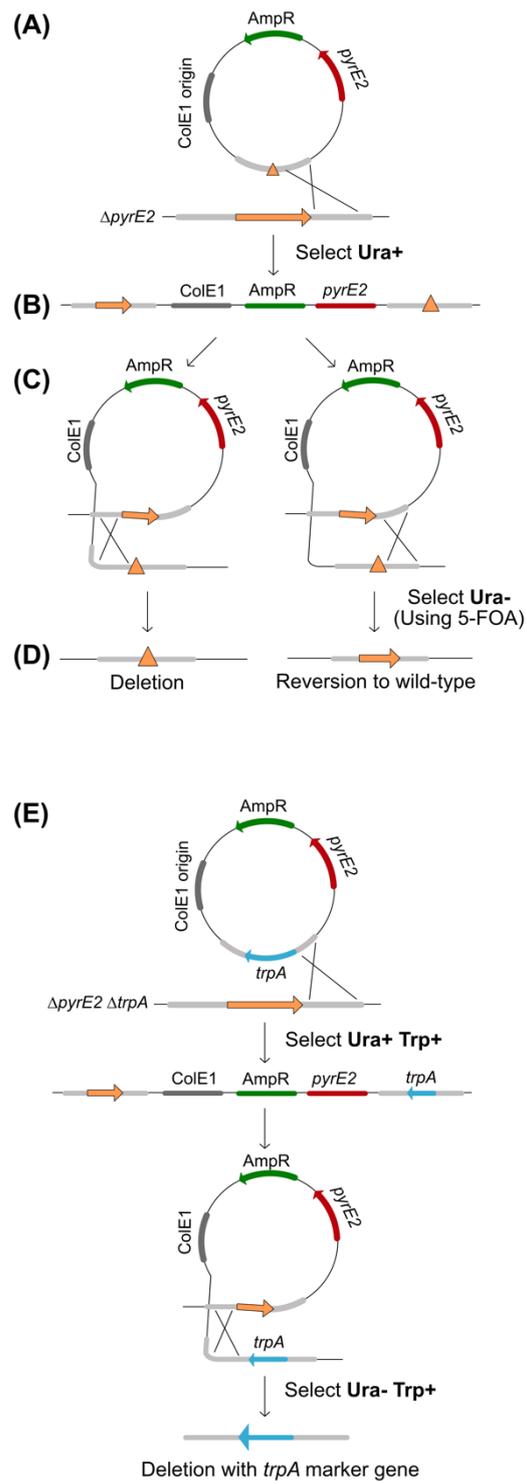


Figure 1.9: Gene deletion utilising the pop-in/pop-out method. (A) A $\Delta pyrE2$ strain is transformed with a *pyrE2*-marked deletion construct. (B) Pop-ins are selected by their ability to grow on media lacking uracil (*ura+* phenotype). (C) Pop-out can be in the upstream (left) or downstream (right) orientation, resulting in the loss of the plasmid backbone (including *pyrE2*). The loss of *pyrE2* in pop-outs is selected for by plating on 5-FOA. (D) The gene is deleted (left) or reverts to wild type (right). (E) A *trpA* marker can be used in deletion constructs to allow for direct selection of deletion pop-out candidates.

The pop-in/pop-out method of gene deletion relies on the *pyrE2* marker for selection of uracil auxotrophic/prototrophic cells. Strains deleted for *pyrE2* ($\Delta pyrE2$) are transformed with plasmid DNA containing the *pyrE2* marker and a deletion construct for the targeted gene (with an internal *trpA* marker if the gene is considered difficult to delete). These ‘pop-in’ candidates are able to grow on media lacking uracil. Relieving the uracil selection will allow the pop-out recombination event to occur, whereby the strain will either lose the deletion construct and revert to wild-type or gain the deletion construct in place of the wild-type gene. This pop-out event leads to loss of the *pyrE2* marker and this can be selected for by plating cells on media containing 5-fluoroorotic acid (5-FOA); cells that are able to synthesise uracil (and therefore have not successfully undergone the pop-out event) will convert 5-FOA into toxic compound 5-fluorouracil. Use of the *trpA* marker within the gene deletion construct allows for a further level of selection, where plating on media lacking tryptophan should select for the *trp*⁺ candidates carrying the deletion construct over the unmarked *trp*⁻ wild type background.

H. volcanii lacks an active copy of β -galactosidase and therefore lacks *lacZ* reporter activity. However, the introduction of active β -galactosidase from close relative species *Haloferax alicantei* (gene *bgaH*) allows for blue-white screening using X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside) (Holmes and Dyall-Smith, 2000). GFP is a common reporter in many species, but due to the high internal salt concentration of *H. volcanii* the protein is not well tolerated. However, a salt-tolerant GFP variant has been produced through targeted mutations allowing solubilisation (namely Phe99Ser, M153Thr and Val163Ala) (Leigh *et al.*, 2011, Reuter and Maupin-Furlow, 2004). Salt-stable GFP allows the fluorescent tagging of components within *H. volcanii* for utilisation in visualisation and flow cytometry studies.

Protein purification from *Haloferax volcanii* must account for the high salinity present within cells, whereby proteins are adapted to fold correctly at high ionic concentrations. Therefore, expression of *H. volcanii* proteins in common overexpression systems, such as *Escherichia coli* or insect cells, is not a viable option. Salt stable poly-histidine (His), streptavidin (StrepII) and tandem histidine-streptavidin (7xHis 2xStrepII) tags have been adapted for use in *H. volcanii*, allowing co-immunoprecipitation and protein pulldown experiments (Allers, 2010). Constructs are readily available encoding either 6x and 7xHis tags; increased numbers of histidine moieties are favoured to enhance affinity of the pulldown. An octa (8x)His tag has not yet been developed for use in *H. volcanii*.

H. volcanii is known to contain a well-documented Clustered Regularly Interspaced Short Palindromic Repeats with CRISPR-associated proteins (CRISPR-Cas) system (Maier *et al.*, 2015b), utilised as defence mechanism

against foreign invaders. An understanding of the type I-B CRISPR system of *Haloferax volcanii* has allowed the development of a system utilising the cell's own CRISPR system as a means to interfere with gene expression and translation (CRISPR interference; CRISPRi) (Stachler and Marchfelder, 2016). This system is discussed in further detail in *Section 1.4*.

Gene expression in *H. volcanii* can be modulated by the use of a tryptophan-inducible promoter, *p.tnaA*. The *p.tnaA* promoter is native to *H. volcanii* and normally regulates the tryptophanase gene, *tnaA*. As tryptophan is an energetically costly amino acid for cells to make, the process of degrading it using tryptophanase is tightly controlled. The specificity of tryptophan as an inducer for this promoter is high, making this promoter ideal for control of other genes (Large *et al.*, 2007, Allers, 2010). This system can be utilised for both inducible gene expression from episomes or can be integrated onto the main chromosome in place of a gene's native promoter. A mutated version of the promoter, *p.tnaM3*, contains a mutation in the promoter's TATA box (T>G at position -26; **Figure 1.10**) that reduces the expression level by ~50% (Braun *et al.*, 2019). This low activity promoter ensures there is reduced leaky expression when the gene is not being actively induced by the addition of tryptophan.

```
tnaA      1  CCTGCCGATTACTTCACATTCGCGGACCTATTGCGCA  37
tnaAM3   1  CCTGCCGATGACTTCACATTCGCGGACCTATTGCGCA  37
```

Figure 1.10: TATA box of *tnaA* and *tnaM3* mutant promoters. Point mutation from T in wild type *tnaA* to G in *tnaM3* promoter is highlighted in white.

Where there are practical difficulties to placing genes under an inducible promoter (e.g., where gene expression is essential), additional selection using the *hdrB* marker will allow for an extra level of selection (**Figure 1.11**). For example, essential gene *radA* was able to be placed under the tryptophan-inducible promoter in originless strains (Hawkins *et al.*, 2013a), where its expression is essential, using additional *hdrB* selection and mapping of pop-in orientations. Through identification of the orientation of integration, it can be predicted which pop-out will occur with *hdrB* selection. By screening for upstream (US) pop-in events, it is known only a downstream (DS) pop-out event would give rise to colonies both resistant to 5-FOA (*ura*⁻) as well as able to synthesise thymidine (*hdrB*⁺) (**Table 1.2**).

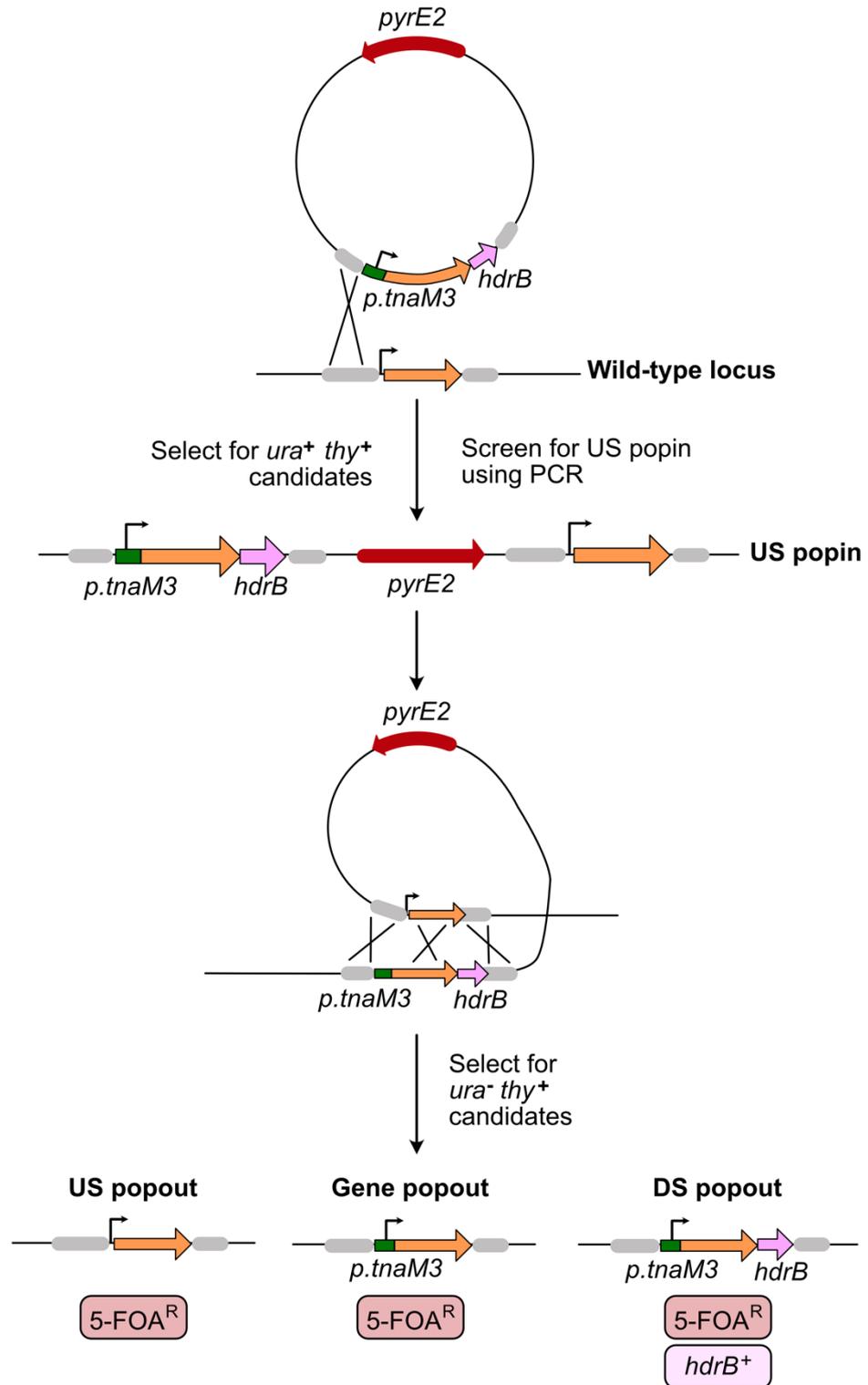


Figure 1.11: Method for integration of inducible promoter using additional *hdrB* selection. Screening of pop-in candidates for orientation by PCR will allow prediction of pop-out events. Upstream (US) pop-in candidates should be selected for pop-out. Following the second recombination event, only a downstream (DS) pop-out should retain both the promoter and *hdrB* marker. Therefore, selective media containing 5-FOA and lacking thymidine should allow for direct selection of the correctly integrated promoter.

Table 1.2: Possible outcomes of pop-in and pop-out events of different orientations for an inducible gene with added *hdrB* selection. Only an upstream (US) pop-in followed by a downstream (DS) pop-out or a DS pop-in followed by US pop-out give the inducible gene and *hdrB* cassette. However, the DS pop-in gives rise to two products with *hdrB* selection and thus only selection of *thy*-5-FOA^R candidates of an US pop-in would give the correct product.

	US popin	Gene popin	DS popin
US popout			
Gene popout			
DS popout			

1.2 DNA replication

The accurate and timely replication of deoxyribonucleic acid (DNA) is a ubiquitous requirement of all cells, whereby the entire genome needs to be copied once per cell division (DePamphilis and Bell, 2011). Regardless of the specific components used, DNA replication can be broken down into three stages: initiation, elongation and termination (DePamphilis and Bell, 2011, Dewar and Walter, 2017).

To ensure correct inheritance of a complete genetic complement when a cell divides, diverse mechanisms have developed which ensure both temporal and spatial co-ordination of replication. Errors are not well tolerated during DNA replication and if fixed as mutations can lead to genetic diseases, including cancer (Hanahan and Weinberg, 2000, Haber, 2013, DePamphilis and Bell, 2011).

The replicon model, first postulated by Jacob, Brenner and Cuzin, proposed that a positive *trans*-acting ‘initiator’ protein would activate initiation of replication by binding a nearby *cis*-acting ‘replicator’ sequence within the genome of model bacterium *Escherichia coli* (**Figure 1.12**) (Jacob *et al.*, 1963). In this model, genetic sequences located within the genome are required for successful initiation of DNA replication, where the localised unwinding of double stranded DNA (dsDNA) allows loading of replication machinery, including the replicative helicase. The helicase actively unwinds DNA, allowing access of the replication machinery to single stranded DNA (ssDNA) to use as a template.

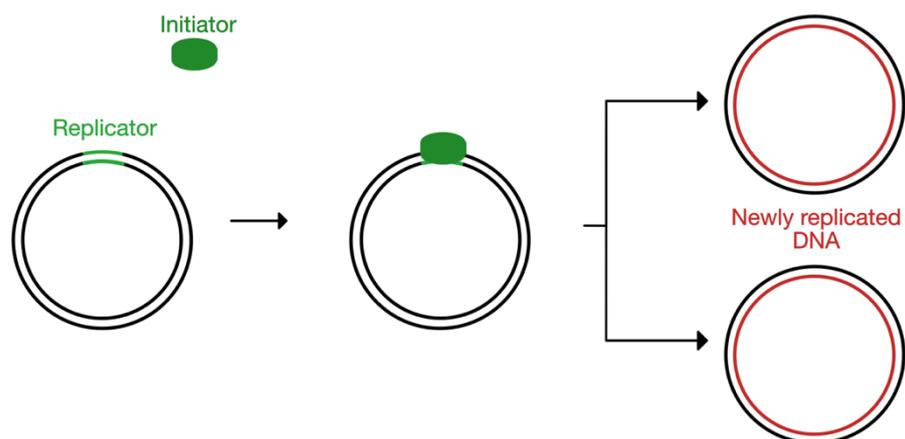


Figure 1.12: Jacob, Brenner and Cruzin Replicon model of DNA replication. A *trans*-acting initiator protein binds the *cis*-acting replicator sequence to initiate loading of replisome components and replication of the chromosome. Adapted from (Jacob *et al.*, 1963).

The replicon model, whilst simple, proved helpful in gaining understanding of how DNA replication initiates in bacteria, but was found to be less relevant

to the more complicated replication mechanism found in eukaryotes. Replication is now widely accepted to initiate at ‘origins of replication’ (Leonard and Mechali, 2013).

Bacteria, Eukaryotes and Archaea carry various replication fork proteins, with some broad conservation between domains. All require recruitment of a replicative helicase to allow loading of further replisome proteins, including sliding clamps, primases and ssDNA-binding proteins. A comparison of these key proteins across the three domains of life can be found in **Table 1.3**. As an example, **Figure 1.13** shows an active archaeal replication fork.

Table 1.3: Comparison of chromosomal DNA replication in the three domains of life. ssDNA – single-stranded DNA; dsDNA – double-stranded DNA; MCM – mini-chromosome maintenance; ORC – origin recognition complex; PCNA – proliferating cell nuclear antigen; RFC – replication factor C; RPA – replication protein A; SSB – single-stranded DNA-binding protein. Figure adapted from (Kelman and Kelman, 2004), (Kelman and White, 2005) and (DePamphilis and Bell, 2011).

	Role	Bacteria	Eukaryotes	Archaea
Chromosome	Encodes genetic information	Linear or circular	Linear	Circular
Replication origin(s)	Site of replication initiation	Single	Multiple	Single or multiple
Origin recognition	Binds to origin to begin replication	DnaA	ORC	Cdc6/Orc
Helicase loader	Loads the replicative helicase	DnaA, DnaC	ORC, Cdc6, Cdt1	Cdc6/Orc
Replicative helicase	Unwinds dsDNA to allow replisome access	DnaB	MCM(2-7)	MCM
ssDNA-binding protein	Binds and protects ssDNA from degradation and prevents secondary structures	SSB	RPA	SSB or RPA
Primase	Synthesises short RNA primers	DnaG	Pol α /Primase	Primase
Clamp loader	Loads and unloads the sliding clamp	γ -complex	RFC	RFC
Sliding clamp	Keeps DNA polymerase associated with the replisome	β -clamp	PCNA	PCNA
Replicative polymerase	Copies DNA 5'-3' at high processivity and high fidelity	Pol-III	Pol ϵ , Pol δ	PolB, PolD
Removal of primers	Removal of RNA primers at start of leading strand and throughout lagging strand	Pol-I, RNase H	Fen1, RNase H	Fen1, RNase H
Lagging strand maturation	Ligation of Okazaki fragments	DNA ligase	DNA ligase I	DNA ligase
Topoisomerase	Changes DNA supercoiling near the replication fork	Type I & II, reverse gyrase	Type I & II	Type I & II, Topo VI

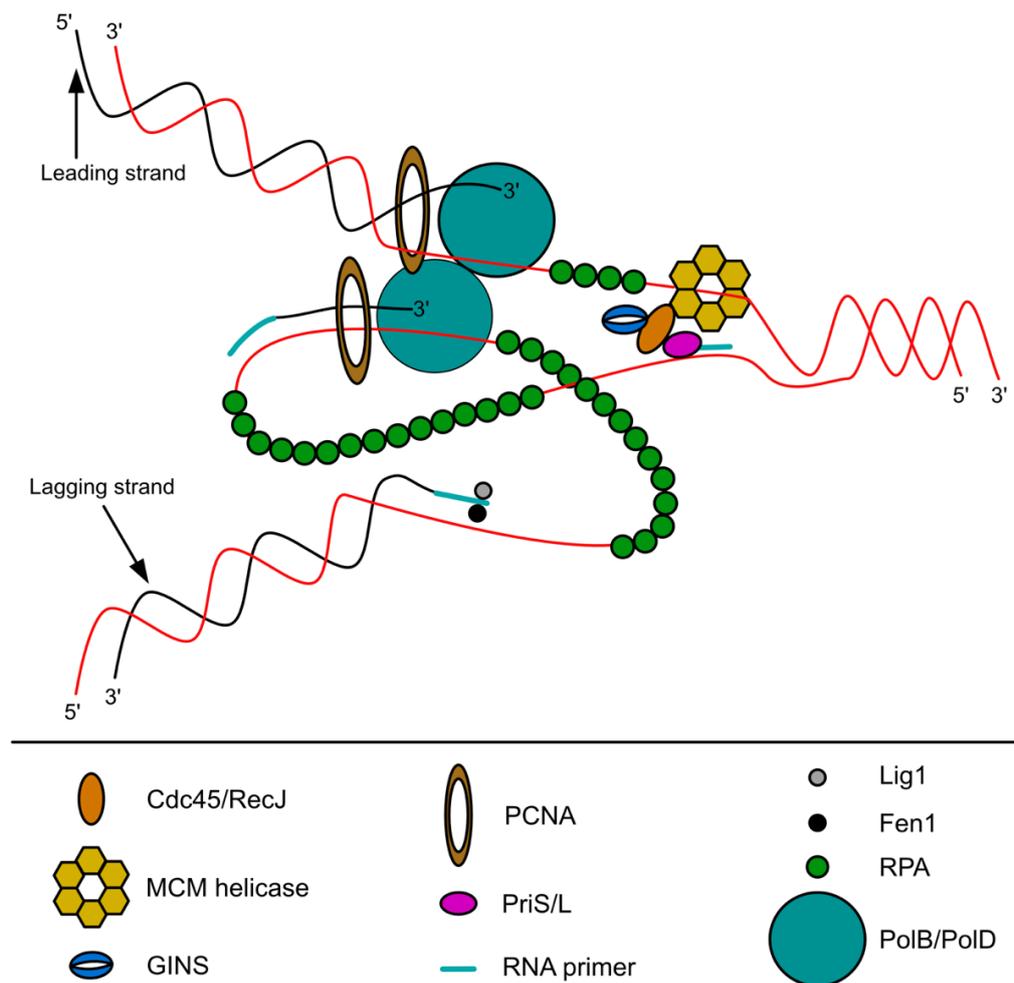


Figure 1.13: The active replication fork in archaea. The archaeal replisome moves along the chromosome as replication progresses. Proliferating cell nuclear antigen (PCNA) is loaded onto DNA by replication factor C (RFC; not shown). PCNA encircles DNA and firmly attaches DNA polymerases to the leading and lagging strand templates. The heterodimeric primase PriSL synthesises short RNA primers for lagging strand extension by replicative polymerases PolB and PolD. RNA primer removal and Okazaki fragment maturation are carried out by Fen1 and Lig1 respectively. GINS interacts with MCM and Cdc45 to stimulate the helicase activity of MCM. Single stranded DNA (ssDNA; unwound by MCM) is protected by ssDNA-binding protein RPA to prevent degradation. Figure adapted from Perez-Arnaiz *et al.*, 2020.

Origins of replication

An origin of replication is the defined zone where initiation of DNA replication will begin the process of doubling the entire chromosome; origins are *cis*-acting sequences that control the replication of DNA in all organisms. Effectively, origins are a redefinition of the previously mentioned ‘replicator’ element as defined in Jacob & Brenner’s replicon model (Jacob *et al.*, 1963). While origins in higher eukaryotes are poorly defined, the core components of an origin of replication have been defined in bacteria, yeast and archaea. These include the origin recognition box (ORB), where the initiator will bind, and one or multiple DNA unwinding elements (DUEs), where DNA unwinding initially occurs (Leonard and Mechali, 2013, Kowalski and Eddy, 1989). DUEs have been shown to be AT-rich sites, facilitating the easy unwinding of DNA. A-T base pairs are inherently less stable as they rely on two hydrogen bonds, while G-C base pairs have three.

Replication in bacteria is initiated at a single origin of replication, while eukaryotes will have multiple origins per chromosome (Leonard and Mechali, 2013). Depending on species, archaea can have a single or multiple origins of replication per chromosome (e.g. *Pyrococcus abyssi* vs *Haloferax volcanii* respectively) (Norais *et al.*, 2007, Myllykallio and Forterre, 2000). Origins of replication were previously believed to be essential for life, however deletion of all replication origins on the main chromosome of *Haloferax volcanii* is possible, with originless cells growing 7.5% faster than wild type (Hawkins *et al.*, 2013a). Origin deletion has also been observed in *Escherichia coli* following mutation of an RNase HI gene, however originless replication is not well tolerated: there was a major effect on cell growth and viability (Kogoma, 1997). Mechanisms for replication in the absence of origins will be discussed further in **Section 1.3**.

The temporal control of DNA replication by origins ensures the maintenance of the correct number of chromosome copies (ploidy) of an organism. Incorrect maintenance of ploidy can lead to aneuploidy (incorrect chromosome copy number) and genome instability, including chromosomal rearrangements, all of which can be detrimental to cells (DePamphilis and Bell, 2011, Kogoma, 1997).

Semi-conservative DNA synthesis

DNA replication was primarily studied in *E. coli* prior to the discovery and in-depth study of archaea. These early experiments were able to define the basic mechanisms of DNA replication; principally how the structure allows for the duplication of the genome and how the two strands are inherited.

Chapter 1: Introduction

Replication results in two identical daughter duplex copies of the parental DNA. Depending on the method of inheritance used the daughters could receive:

- The original duplex along with a new duplex (conservative replication)
- The parental DNA could be randomly distributed between daughter duplexes (dispersive replication)
- The original strand could act as a template for synthesis, pairing with a newly synthesised strand in the product duplex (semi-conservative replication)

Following the publishing of the double helix structure of DNA (Watson, 1953), Watson & Crick, among others, hypothesised each single strand would act as a template for synthesis during the next round of replication. In 1958, Meselson & Stahl utilised 'heavy' isotopic nitrogen to label DNA. This allowed the definition of DNA replication as semi-conservative through the use of density gradient centrifugation, whereby the heavy DNA was followed through divisions when provided with normal 'light' nitrogen as a source (Meselson and Stahl, 1958). This showed each heavy duplex was split between the two 'light' daughter products, forming a band of an intermediate weight (**Figure 1.14**).

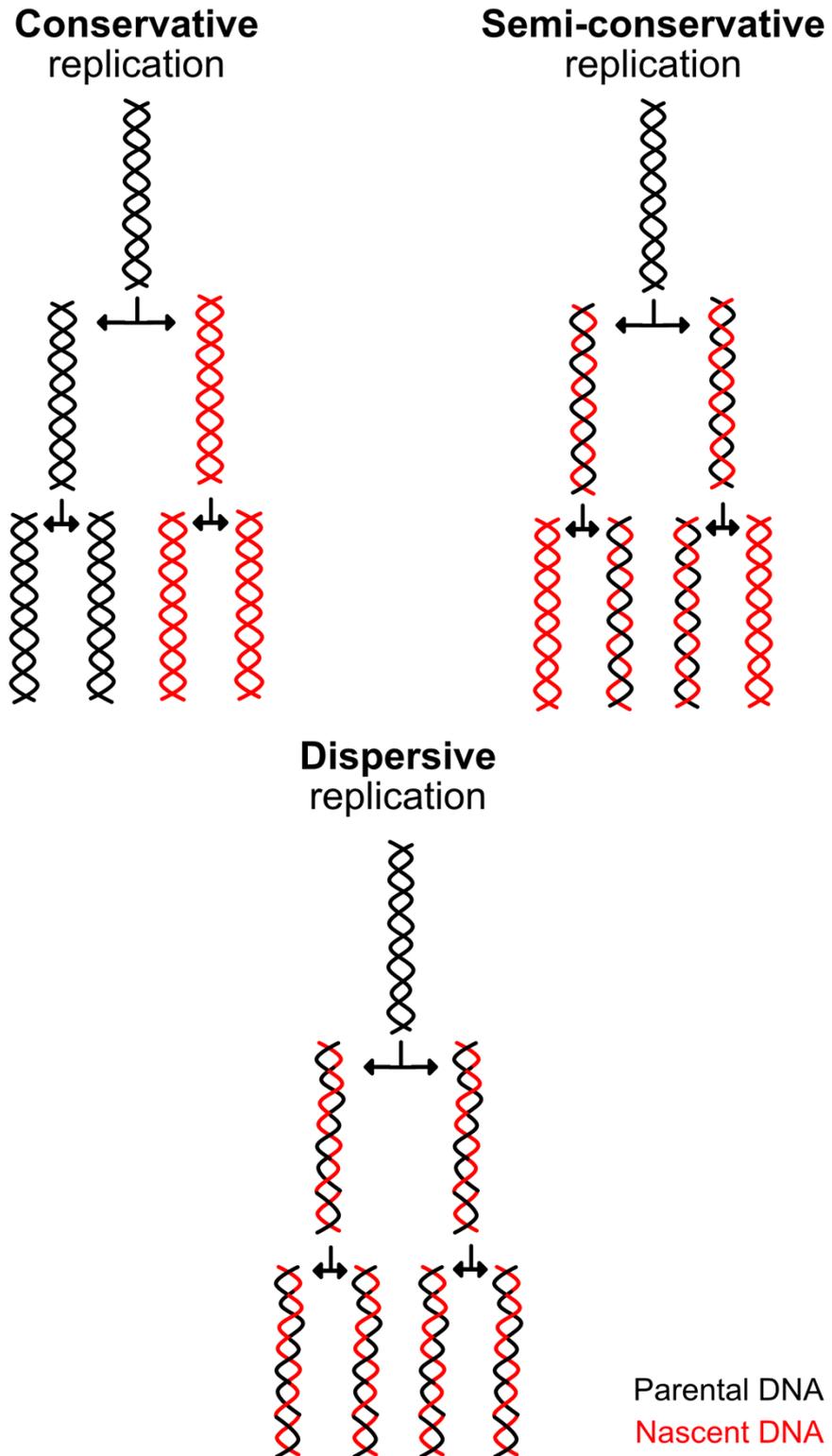


Figure 1.14: A summary of the three postulated models of DNA replication. During replication, the double stranded helix (black) is unwound and acts as a template for synthesis of the new daughter strand (red). Following work by Meselson & Stahl, replication is now widely accepted to be as a result of semi-conservative replication. Figure based on information from (DePamphilis and Bell, 2011).

Initiation

A specific sequence of events must occur to initiate DNA synthesis, whereby two replication forks (and their associated replisome) will move away from the origin of replication bidirectionally, each synthesising both leading (continuous) and lagging (non-continuous) strands to fully replicate the DNA. A summary of steps required for replication initiation is as follows:

- The initiator protein binds to the origin of replication
- The initiator will remodel DNA at the origin to an open structure (the 'replication bubble'), allowing recruitment of the replicative helicase
- The replicative helicase will continue to unwind DNA to allow access of further replication components, including primases and polymerases
- Primase will then prime replication by producing short RNA primers for extension by DNA polymerases
- DNA synthesis will begin bidirectionally away from the origin

Tight regulation of the initiation of DNA replication ensures all daughter cells receive a complete and correct complement of the genome. If initiation occurs too frequently, or not enough, this can lead to major problems, including genome instability and aneuploidy (Hanahan and Weinberg, 2000, O'Donnell *et al.*, 2013, Michel and Bernander, 2014).

Loading of the replicative helicase is a rate limiting step in replication (DePamphilis and Bell, 2011); even once a replisome forms at an activated origin of replication, the replication fork cannot progress along the DNA without the active unwinding activity of the helicase (DePamphilis and Bell, 2011).

Elongation

Following the activation of the replicative helicase, a bidirectional replication bubble is formed, and the remaining members of the replisome are recruited to establish *bona fide* replication forks. Elongation is where established replication forks replicate the DNA bidirectionally using semi-conservative DNA synthesis. All domains of life share key principles and functional divisions of DNA synthesis with the proteins being used by organisms differing (Burgers and Kunkel, 2017, Kelman and Kelman, 2014). While bacterial DNA replication machinery has been defined extensively, some questions still remain regarding the more complex machinery found in Eukaryotes and Archaea.

Termination

Termination of DNA replication is where converging replication forks meet and resolve, allowing for correct chromosome segregation at the completion of DNA synthesis. Any catenated (linked DNA) structures linking sister chromatids must be resolved to ensure segregation and cell division can occur. In comparison to initiation and elongation, termination is relatively poorly studied, despite its frequency matching that of initiation events (Berezney *et al.*, 2000). It is essential that converging DNA forks do not pass one another and continue replication, as this leads to over-replication and major genomic instabilities.

In order to terminate replication, five events specific to termination must occur:

- Resolution of topological stress caused by overwinding at the terminus site
- Convergence of replication forks (sometimes referred to as encounter)
- Disassembly and recycling/degradation of replication components to prevent re-replication
- Completion of DNA synthesis through gap filling
- Solving of any catenated structures remaining in DNA, ensuring chromosome segregation can occur

It remains unknown whether these events must occur in a specific order, are sequence-specific or are stochastic events that occur when two replication forks meet (Dewar and Walter, 2017).

1.2.1 Bacterial replication

Initiation

Bacteria have a single origin of replication (named *oriC* in model bacterium *Escherichia coli*) with the vast majority of bacterial species having a single circular dsDNA chromosome. There are two groups standing as exceptions to this rule: some bacteria, for example *Vibrio cholerae* (Trucksis, 1998), carry two chromosomes, and few species (including *Streptomyces* and *Borrelia*) carry linear chromosomes (Volf, 2000). Initiation at *oriC* establishes bidirectional replication forks that move away from the origin until meeting at defined termination zones.

Origin recognition

Replication origin *oriC* is adjacent to bacterial initiator gene *dnaA* (in *E. coli*) (Hwang and Kornberg, 1992). The proximity of these two genetic loci fits well with the pre-established replicon model hypothesis. DnaA is a member of the AAA+ (ATPases Associated with diverse cellular Activities) protein family (Messer, 2002). DnaA binds as a monomer to *oriC* at sequence-specific DnaA boxes (also called origin recognition elements or OREs) (DePamphilis and Bell, 2011, Leonard and Mechali, 2013). Its C-terminal DNA binding domain, the helix-turn-helix (HTH) motif, can fit within the major groove of DNA at the DnaA box (Erzberger *et al.*, 2002, Duderstadt and Berger, 2013). DnaA boxes are found at multiple locations on the bacterial chromosome, with many clustering within close proximity of *oriC*. DnaA binding is co-operative, with each monomer binding a single DnaA box, forcing the duplex to bend, melting the two strands to give a ssDNA ‘bubble’ (**Figure 1.15**) (Schaper and Messer, 1995). DnaA-ADP and DnaA-ATP are both capable of binding DnaA boxes (with equal affinity), however only the ATP-bound form is capable of remodelling the DNA unwinding element (DUE) at *oriC*, meaning DnaA-ATP is required to provide access for helicase loading by helicase loader DnaC (DePamphilis and Bell, 2011).

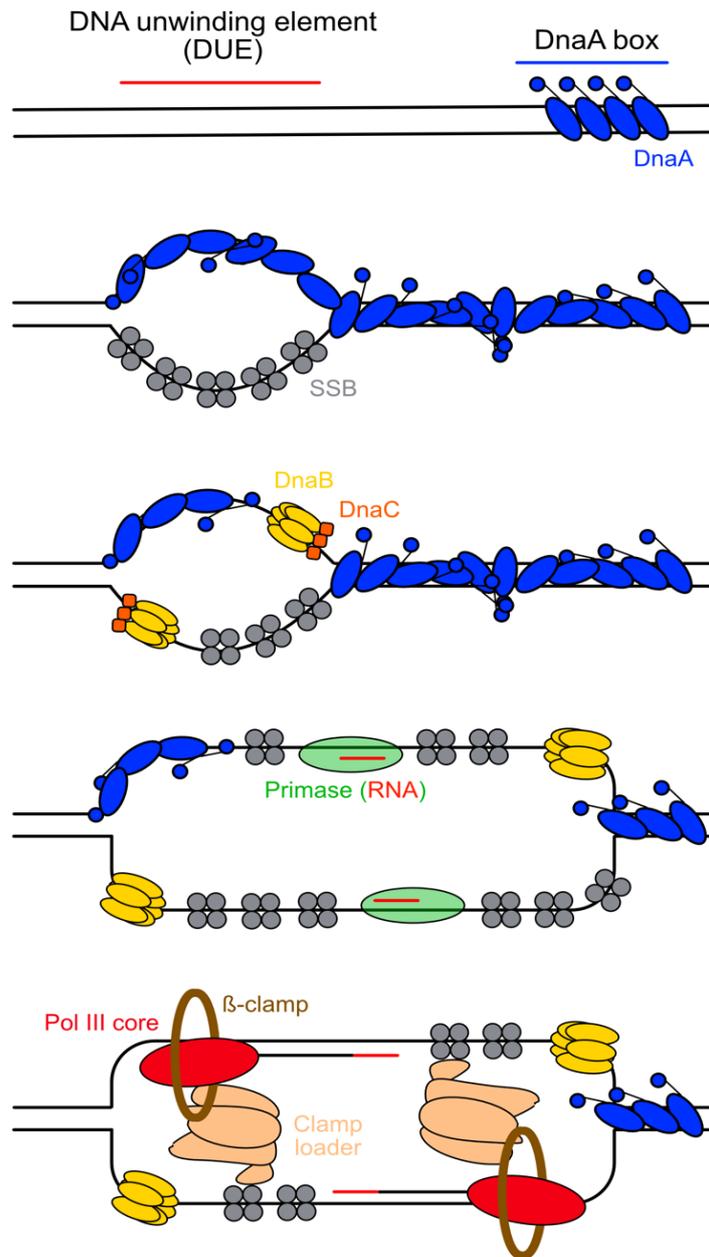


Figure 1.15: Bacterial origin recognition and replisome loading in *Escherichia coli*.

DnaA binds sequence-specific DnaA boxes at *oriC* with high affinity. The co-operative nature of DnaA binding leads to the unwinding of DNA at the AT-rich DNA unwinding element (DUE). DnaC helicase loader, assisted by initiator DnaA, will then load replicative helicase DnaB onto ssDNA at the DUE. DnaG primase is recruited, which allows the release of DnaC, activation of DnaB and synthesis of RNA primers. The remaining replication components are recruited, and two replisomes move away from the origin bi-directionally. Figure omits lagging strand fragments for simplicity. Figure adapted from (Wegrzyn *et al.*, 2016).

The ability to replicate in bacteria is dictated at the level of the DnaA:origin interaction. The concentration of DnaA and accessibility of origins and DnaA boxes will determine whether the origin can be remodelled and therefore whether a replisome can be established (Leonard and Grimwade, 2009). Commitment to replicate is therefore growth-dependent, as opposed to cell cycle-dependent as seen in higher eukaryotes (among others) (Ausiannikava and Allers, 2017). In rapidly growing bacteria, it is possible to re-fire origins before the genome is fully replicated. This has been seen in strains of *E. coli* where the division time is 20 minutes, while it takes a full 40 minutes to fully replicate the genome (Leonard and Grimwade, 2009).

Replisome loading

Due to the unstable nature of ssDNA, single-stranded DNA binding proteins act quickly to cover any ssDNA to protect it from degradation or chemical modification. In bacteria, this job is carried out by single-stranded DNA binding protein (SSB), which binds ssDNA and forms a nucleoprotein filament to prevent degradation and prevent secondary structure formation. In *E. coli*, SSB binds DNA as a homotetramer via its oligosaccharide / oligonucleotide-binding fold (OB-fold), interacting with other SSB molecules through its acidic C-terminal domain (Raghunathan *et al.*, 1997).

Once DnaA has successfully catalysed strand melting, helicase loader DnaC (also an AAA+ ATPase) (Makarova and Koonin, 2013) acts as a chaperone to load homohexameric ring helicase DnaB onto the lagging strand (Fang *et al.*, 1999, Wegrzyn *et al.*, 2016). Stimulation from ATP allows the DnaC:DnaB complex to interact with ssDNA, which will in turn trigger ATP hydrolysis, giving DnaC-ADP (Davey *et al.*, 2002, Mott *et al.*, 2008). This conversion allows activation of DnaB as a 5'–3' helicase, as DnaC-ATP inhibits DnaB helicase activity (Davey *et al.*, 2002, Mott *et al.*, 2008). Only a single molecule of DnaB is loaded per replication fork (Barry and Bell, 2006). Following helicase loading and dsDNA unwinding, further replication components are recruited (including primase, DNA polymerases and sliding clamp β -clamp), and a replisome is formed (**Figure 1.14**). Two replication forks will be formed per origin, one moving in either direction, thus at the origin two helicases are loaded and two replisomes are assembled adjacent to one another.

Elongation

DNA polymerases are incapable of initiating DNA synthesis *de novo*; they require a 3' hydroxyl-primed template to commence elongation (Kelman and Kelman, 2014). Primase is a DNA-dependent RNA polymerase that synthesises short (~8-12 nucleotide) primers which are then used by DNA

Chapter 1: Introduction

polymerase to prime DNA synthesis. On the leading strand, continuous synthesis means the requirement for primase is low (single priming event), however the lagging strand needs repeated priming to form the Okazaki fragments. In bacteria, primase is a single subunit protein, named DnaG (Rowen and Kornberg, 1978). DnaG works with replicative helicase DnaB and Pol-III to unwind the DNA template and simultaneously synthesise regularly spaced primers for Okazaki fragments to ensure correct replication of the lagging strand (Corn and Berger, 2006, Kelman and Kelman, 2014).

The main replicative DNA polymerase (DNAP) in *E. coli* is the Pol-III holoenzyme; two copies of this enzyme are responsible for the synthesis of both the leading and the lagging strands simultaneously. The Pol-III core, consisting of α , ϵ and θ subunits, is tightly associated with the replisome, held in place by the homodimeric β -clamp. The β -clamp is assembled by the clamp loader, named γ -complex, which binds both Pol-III and DnaB helicase (Indiani and O'Donnell, 2006, O'Donnell *et al.*, 2013). The β -clamp acts to increase the processivity of the polymerase and ensure its continued engagement with the template. As the lagging strand is repeatedly primed by DnaG primase, the clamp loader continually loads new circular β -clamps onto the template for use by the lagging strand polymerase. The α subunit of Pol-III is a family C polymerase (Ito and Braithwaite, 1991) responsible for nascent DNA synthesis, while subunits ϵ and θ have 3'-5' exonuclease activity and stabilise the Pol-III core respectively. At least two $\alpha\epsilon\theta$ cores are maintained per bacterial replisome; one for the leading strand and one for the lagging strand (Johansson and Dixon, 2013).

Pol-I, a family A polymerase, is also an important replicative polymerase in bacteria. Its primary role is in processing Okazaki fragments following lagging strand synthesis. When Pol-III meets a new RNA primer it will dissociate and find a new primer end, leaving a gap between each pair of fragments. Pol-I acts to excise the RNA primers at the 5' end of each Okazaki fragment, but will also prime DNA synthesis from the 3' end found behind it (Allen *et al.*, 2011). Thus, Pol-I will leave a 3'-OH and 5'-phosphate which can be acted upon by DNA ligase to give a continuous DNA strand product.

In low-GC Gram negative bacteria, the α subunit of Pol-III has evolved to gain different functions, meaning this subset of organisms instead use Pol III-like polymerases Pol C and DnaE (discussed in more detail in *Chapter 3*).

Termination

Bacterial chromosomes carry defined termination zones flanked by 'pause' sites. These act to ensure DNA replication only occurs once and in a single direction. In *E. coli*, these termination zones are defined by *Ter* DNA

sequences (*TerA-J*) which are bound by terminus site-binding protein Tus in a specific orientation (Berghuis *et al.*, 2015). The replication fork will proceed unhindered past 5 *Ter* sites in one orientation but cannot proceed past *Ter* sequences from the opposite direction. *Ter* sequences act as potent and polar replication fork barriers, causing the fork to stall as a locked complex within the termination zone (**Figure 1.16**) (Duggin *et al.*, 2008). This halt has been linked to inactivation of DnaB and is dependent on Tus (Bastia *et al.*, 2008). However, this is not fully understood, as Tus is known to be non-essential and deletion has no impact on growth phenotype (Roeklein *et al.*, 1991). This suggests a key role for orientation-specific *Ter* sequences in preventing re-replication.

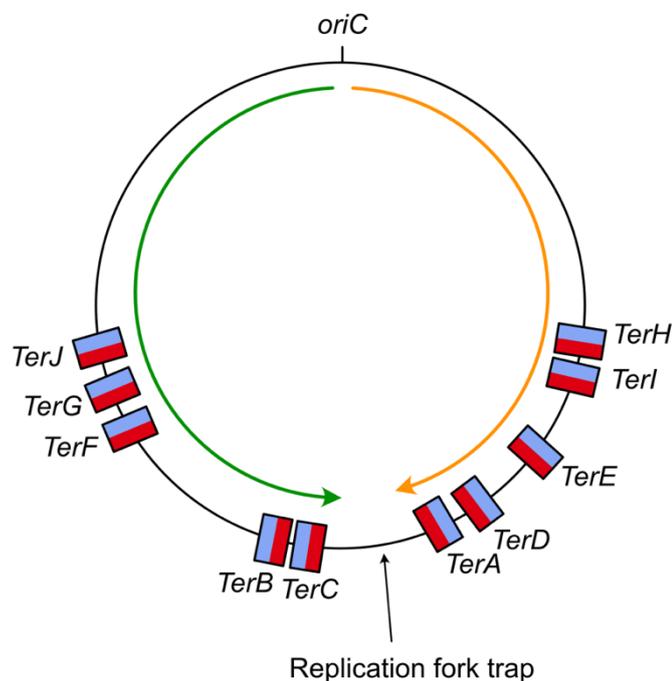


Figure 1.16: Location and orientation of *Ter* sites on the *E. coli* chromosome. Turquoise represents the permissive face of the *Ter* sequence, whereby an incoming replication fork can pass through. The red face represents the nonpermissive face of the *Ter* sequence, which blocks an oncoming replication fork. The green and orange arrows represent the two replisomes firing from *oriC* and ending in the replication fork trap zone. Adapted from (Berghuis *et al.*, 2015).

Unwinding of the DNA duplex will lead to overwinding of the helix ahead of the replication fork, leading to formation of supercoils. If not resolved, this increased torsional stress would prevent the DNAP from proceeding along the DNA. Topoisomerase enzymes act to correct the topology of DNA so replication can proceed unimpeded (Wang, 1996). Type I

topoisomerases will cut the DNA on one strand, relax DNA by passing the other strand through the nick, and religate the duplex. Type I topoisomerases do not require ATP, due to the removal of DNA supercoils being energetically favourable (Lodish *et al.*, 2000). Type II topoisomerases cut both strands of the duplex, allow an unbroken duplex to pass through, before religating. Unlike, type I, type II topoisomerases are dependent on ATP for their activity (Wang, 1996, Champoux, 2001).

The removal of supercoils in bacteria is carried out by type II topoisomerases, in *E. coli* named Topoisomerase IV (TopoIV) and DNA gyrase. DNA gyrase acts to relieve positive supercoils formed as a product of DNA unwinding while TopoIV resolves pre-catenanes allowing fork convergence to occur and be resolved successfully (Espeli *et al.*, 2003, Hiasa and Marians, 1994). Following topoisomerase activity, RecG translocase and nuclease-helicase RecBCD (or in the absence of RecB, nuclease SbcCD) are free to resolve the overlapping DNA sequences at the terminus to give a resolved product for dissolution and chromosome segregation (Beattie and Reyes-Lamothe, 2015, Rudolph *et al.*, 2013, Dimude *et al.*, 2018).

A known issue with chromosome circularity is that should an odd number of crossover events occur between sister chromosomes, a chromosome dimer will be generated, where sister chromosomes are fused to one another (Lesterlin *et al.*, 2004). These dimers must be separated to ensure each cell receives one complete genetic complement upon division. In bacteria, Xer site-specific recombinases act at specific loci in the terminus region, named *dif*, to resolve any chromosome dimers into monomers (Sherratt, 2003). In *E. coli*, FtsK translocase is required for XerCD-mediated recombination at *dif* sites. FtsK anchors itself at the mid-cell division site and acts to bring the two *dif* sites of *E. coli* to the mid-cell for synapsis. FtsK then stimulates catalysis by XerD, leaving sister chromatids as monomers for segregation (Aussel *et al.*, 2002, Lesterlin *et al.*, 2004). This link between terminal *dif* loci, FtsK and Xer suggests a strong coupling for replication, segregation and division in bacteria.

Some bacterial species, for example *Borrelia*, have linear chromosomes. Therefore, their method of termination differs from that of circular chromosomes explained above. Instead, they utilise a telomere resolvase, ResT, which binds replicated telomere (*rTel*) sites and resolves the dimer junction into two covalently closed hairpin ends (Chaconas and Kobryn, 2010).

1.2.2 Eukaryotic replication

Due to the additional complexity of eukaryotic genomes, initiation of DNA replication in eukaryotes is more complicated when compared to bacteria. Archaeal genomes are similar to those of bacteria, however at the level of the replication machinery the enzymes utilised by archaea are more similar to those of eukaryotes.

Eukaryotic genomes are made up of multiple linear chromosomes, carrying numerous origins of replication per chromosome (O'Donnell *et al.*, 2013, Burgers and Kunkel, 2017). Eukaryotes require larger numbers of replication origins due to increased genome size, but also for time efficiency; the speed at which eukaryotic replication forks progress is 20x slower than those of bacteria (O'Donnell *et al.*, 2013). A recent study has comprehensively mapped DNA replication initiation events across the human genome using single-molecule techniques and shown that initiation events are randomly distributed across broad initiation zones that are utilised stochastically across a population of cells. While they possess origins, it was confirmed that there were no well-defined initiation sites or co-ordination of replication initiation between cells (Wang *et al.*, 2021).

The large size of eukaryotic genomes requires considerable compaction to fit into cells; this is achieved by wrapping DNA around histone proteins to form filaments, which will then further loop to condense the genome. This genome compaction has implications for the replication of DNA, whereby coiled DNA must be unwound for replication to occur and be rewound following completion. Histone chaperone and remodeller proteins act to aid this process, removing histones ahead of the replication fork and reloading them to recently replicated strands respectively (Tyler, 2002). Compaction of the genome also has implications for transcription, where the extent of DNA coiling will determine which genes are accessible to the transcription machinery.

The fastest mechanism for replicating the multiple chromosomes in eukaryotes would be to simultaneously fire all origins. While this does occur in some situations, namely during the rapid division phase in the early *Xenopus* embryo and in *Plasmodium* schizogony (Stanojcic *et al.*, 2017), most somatic cells do not fire all origins simultaneously. Instead, cells will fire designated origins during S-phase, with some firing early while others delay firing until late in S-phase.

Chapter 1: Introduction

The choice of which origins fire at what time is not fully understood, but it is known to be affected by numerous factors (Barry and Bell, 2006), including:

- transcriptional status
- chromatin structure
- cell type
- developmental stage

Initiation

Origin recognition

The additional levels of complexity for regulating initiation of replication in eukaryotes is apparent when comparing initiator proteins. While bacteria only rely on a single initiator protein, DnaA, eukaryotes utilise the origin recognition complex (ORC), a six-protein complex that acts specifically at origins (**Figure 1.17**) (Li and Stillman, 2012). Orc1-5 proteins contain a C-terminal winged helix (WH) domain that acts to contact DNA at the origin of replication (Bleichert *et al.*, 2015). ORC will bind to origins in G1-phase, with binding being dependent on ATP (Bell and Stillman, 1992), however binding of ORC alone will not cause initiation. At S-phase ORC, together with regulator Cdc6 and licensing factor Cdt1, acts to load the replicative helicase minichromosome maintenance (MCM2-7) to form the pre-replicative complex (pre-RC) (Cocker *et al.*, 1996, Bell and Dutta, 2002). MCM helicase is a heterohexameric helicase, encoded by six paralogous genes (MCM2-7), all part of the AAA+ superfamily of ATPases (Makarova and Koonin, 2013, Makarova *et al.*, 2012). These proteins assemble onto the leading strand, forming heterohexameric rings with N-termini forming the N-tier and C-termini forming the C-tier (Li and O'Donnell, 2018). Unlike bacteria loading a single helicase per replication fork, many MCM molecules will associate with the replisome in eukaryotes (Kelman *et al.*, 2020). MCM has even been shown to associate with non-replicating DNA by immunofluorescence, suggesting MCM may also play a function from a distance (Madine, 1995).

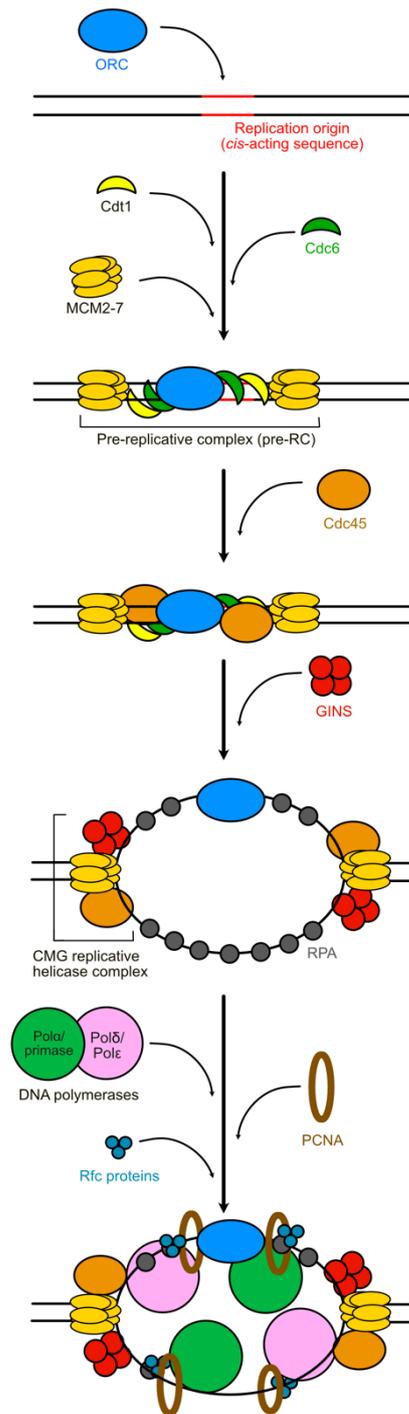


Figure 1.17: Eukaryotic origin recognition and replisome loading in eukaryotes. ORC (Origin Recognition Complex) binds the origin of replication during G1 phase and recruits Cdc6 and Cdt1 at S phase. ORC:Cdc6:Cdt1 are then able to recruit the replicative helicase, MCM (minichromosome maintenance), forming the pre-RC (pre-replicative complex). Cdc45 associates with the pre-RC, followed by GINS, forming the CMG complex (Cdc45:MCM:GINS). Final components required for replication are then recruited to the origin and replication will be primed and initiate away from the origin bidirectionally. Figure adapted from (Sawa, 2008).

Replisome loading

To proceed with replication, the pre-RC must be activated; this allows for a further level of regulation in eukaryotes. Following successful recruitment of MCM to the origin, ORC, Cdc6 and Cdt1 are no longer required and will dissociate. The activation of the helicase requires recruitment of two further proteins, Cdc45 and GINS, which together with MCM form the CMG complex (Cdc45-MCM-GINS; **Figure 1.17**). Protein kinases cyclin-dependent kinase (CDK) and Dbf4-dependent kinase (DDK) control the activation of the pre-RC into a full replisome, namely by activating the replicative helicase (Tanaka and Araki, 2013). In yeast, Cdc45 is regulated by DDK while GINS is regulated by CDK (Takeda and Dutta, 2005).

Cdc45 (Cell cycle division protein 45) is a well-established member of the CMG complex, however its function has been difficult to pin down. The N-terminus of Cdc45 carries a DHH phosphoesterase superfamily domain, providing a structural similarity to bacterial RecJ, a 5'–3' ssDNA exonuclease that functions in recombination and repair (Pellegrini, 2017, Simon *et al.*, 2016). However, Cdc45 has since lost its catalytic activity and thus is predicted to play only a structural role at the replisome. GINS (named after Japanese numbers 5, 1, 2 and 3: *go ichi ni san*) is a four-subunit protein in eukaryotes, consisting of subunits Sld5, Psf1, Psf2 and Psf3 (Choi *et al.*, 2007). Although the exact molecular function of GINS remains unknown, evidence suggests it acts as a molecular linker mediating assembly of the replisome around MCM helicase (Pacek *et al.*, 2006, Calzada *et al.*, 2005). GINS is also known to associate with further replisome components, including Pol- α (primase), Pol- δ (lagging strand polymerase) and Pol- ϵ (leading strand polymerase), placing it at the heart of the active replisome (Pacek *et al.*, 2006, Makarova *et al.*, 2012).

Following the formation of the CMG complex, binding of Mcm10 stabilises the CMG complex and activates MCM by directly stimulating its helicase activity (Lõoke *et al.*, 2017). The binding of MCM switches from dsDNA to ssDNA, allowing unwinding to occur (Evrin *et al.*, 2009). At this point further replication components can be loaded, such as DNA polymerases, primases and clamp proteins. Akin to bacteria, the ssDNA exposed upon activation of MCM helicase is bound by replication protein A (RPA), a heterotrimeric ssDNA-binding protein carrying structural and functional analogy to bacterial SSB (O'Donnell *et al.*, 2013). This fully established replisome will then commence replication moving away from the origin bidirectionally in a 5'–3' direction.

Higher eukaryotes appear to not rely on sequence specificity for delineation of the origin, with the structure of the chromatin instead defining where initiators are capable of binding. They also differ from bacterial systems

in that the initiator protein complex ORC does not appear to actively melt the dsDNA at origins; instead, it is thought ORC acts to recruit MCM2-7 to dsDNA to begin unwinding prior to replication initiation (Evrin *et al.*, 2009, Heller *et al.*, 2011).

Elongation

The replicative polymerases of eukaryotes are more complex than their bacterial counterpart, relying on three separate family B polymerases; Pol- α , Pol- δ and Pol- ϵ (Burgers, 2009). These polymerases are all multi-subunit enzymes, containing a catalytic core subunit alongside various accessory domains. In eukaryotes, DNA primase is a heterodimer containing a regulatory PriL/p58 subunit and a catalytic PriS/p48 subunit (Arezi and Kuchta, 2000). This heterodimer interacts *in vivo* with both the Pol- α catalytic subunit and the B subunit to form the active Pol α -primase complex (Frick and Richardson, 2001). Pol α -primase complex is vital during initiation of replication at origins, along with priming Okazaki fragment synthesis on the lagging strand. The primase component of Pol α -primase will synthesise ~10-15 nucleotides of RNA that are then further elongated by Pol- α to produce a RNA:DNA hybrid of ~40 nucleotides (Frick and Richardson, 2001, Kelman and Kelman, 2014). Pol- α has low processivity and lacks proofreading activity, meaning this short priming synthesis event is error-prone; Pol- δ is known to act to correct errors made by Pol- α (Pavlov *et al.*, 2006). Pol- δ and Pol- ϵ carry out the bulk synthesis of DNA, extending primers and short DNA tracts synthesised by Pol α -primase complex. Both Pol- δ and Pol- ϵ possess exonuclease activities (Burgers, 2009).

In eukaryotes proliferating cell nuclear antigen (PCNA) acts as the clamp and shares structural similarity to bacterial β -clamp protein (O'Donnell *et al.*, 2013). PCNA is a trimeric ring that encircles dsDNA and can travel bidirectionally along it (Kelman and Kelman, 2014). PCNA must be assembled on the DNA by pentameric replication factor C (RFC). RFC subunits also share homology to the bacterial clamp loader counterpart, at both structural and sequence level (Garg and Burgers, 2005). However, the details of interactions between clamp loader and replication components are less clear in eukaryotes than the well-established relationships in bacteria between γ -complex and polymerases/helicases. Association with clamp protein PCNA usually occurs through a sequence-specific PCNA-interaction protein (PIP) motif (Warbrick *et al.*, 1998).

Pol- ϵ is a highly processive polymerase, likely due to a small domain in its catalytic subunit allowing it to encircle DNA (Hogg *et al.*, 2014). In contrast, the intrinsic processivity of Pol- δ is much lower. When both Pol- ϵ

and Pol- δ associate with PCNA, they increase their processivity; Pol- ϵ is already highly processive, so this increase is minimal, while Pol- δ activity is vastly enhanced by PCNA binding to a level where the processivity of Pol- ϵ and Pol- δ is comparable (Chilkova *et al.*, 2007, Burgers and Kunkel, 2017).

Current understanding places Pol- ϵ as the leading strand polymerase and Pol- δ as the lagging strand polymerase (Miyabe *et al.*, 2011, Pursell *et al.*, 2007). Pol- ϵ is a strong candidate as leading strand polymerase for two main reasons:

- (i) Pol- ϵ interacts with several components of the CMG complex, which is found at the leading strand, ensuring its position at the heart of the active replisome (Sengupta *et al.*, 2013, Sun *et al.*, 2015)
- (ii) Pol- ϵ is intrinsically highly processive and its 3' exonuclease activity is capable of repairing its own replication errors increasing its fidelity (Langston *et al.*, 2014, Burgers and Kunkel, 2017)

Lagging strand polymerases must be capable of strand displacement for efficient Okazaki fragment maturation; Pol- ϵ lacks efficient strand displacement activity and thus is a poor candidate for the lagging strand polymerase, while Pol- δ has efficient strand displacement activity (Ganai *et al.*, 2016, Maga *et al.*, 2001, Garg *et al.*, 2004). When Pol- δ reaches the 5' end of the subsequent Okazaki fragment, its strand displacement activity generates a small 5' flap, predominantly only one nucleotide in length with the flap acting as a molecular brake on the polymerase (Stodola and Burgers, 2016). Flap endonuclease FEN1 remodels and cuts the flaps and the remaining nick is ligated by DNA ligase (Liu *et al.*, 2004). It is of note that each of the Okazaki fragment maturation proteins, Pol- δ , FEN1 and DNA ligase, contain one or more PIP motifs, suggesting interaction with PCNA. Here, the 'molecular toolbelt' of PCNA:Pol- δ :FEN1:DNA ligase acts to ensure Okazaki fragment maturation occurs processively and without the need for enzyme dissociation (Indiani *et al.*, 2005).

Termination

Unlike bacteria, eukaryotes do not have defined termination sites. Replication of their multi-origin linear chromosomes terminates when replisomes travelling in different directions converge. Having complex multi-origin chromosomes, it is important that termination occurs accurately to avoid genomic rearrangements or introduction of mutations. Generally, eukaryotic termination occurs midway between origins, with more active origins generally allowing for better mapping of termination sites (McGuffee *et al.*, 2013).

Chapter 1: Introduction

While *E. coli* termination is sequence-dependent on *ter* sites, the same does not seem to apply to eukaryotes; changing the timing of origin firing alters the termination site in a predictable manner, suggesting the timing of initiation events defines the termination zone, not sequence (McGuffee *et al.*, 2013, Greenfeder and Newlon, 1992). An overview of eukaryotic termination is represented in **Figure 1.18**.

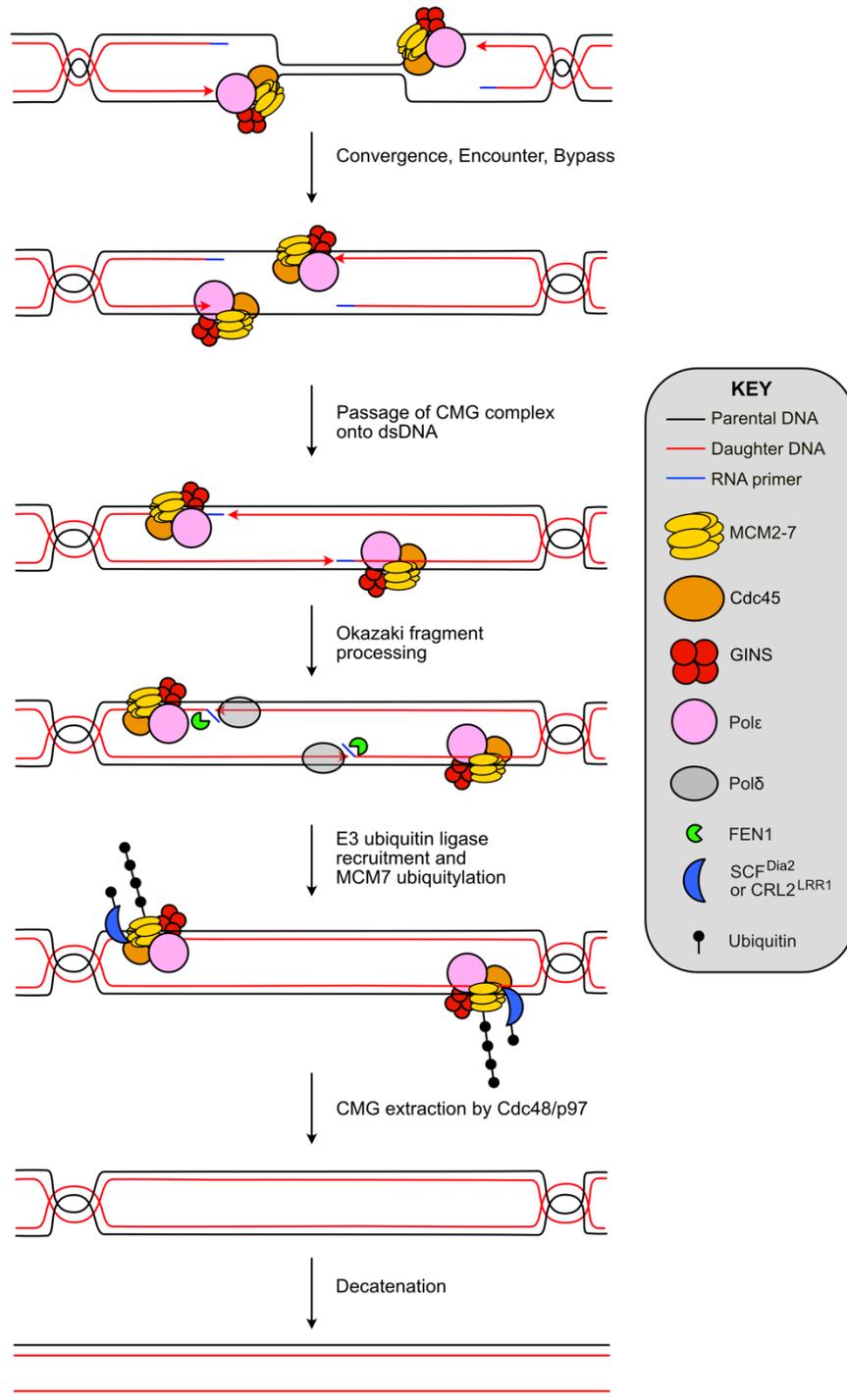


Figure 1.18: Overview of replication termination in eukaryotes. Once most of the chromosome has been replicated, forks come close to one another leading to supercoil formation in unreplicated DNA, causing convergence. There is no fork stalling upon encounter of forks, suggesting CMG complexes can bypass one another. They then pass over the ssDNA-dsDNA junction at the Okazaki fragment and translocate on dsDNA. The last Okazaki fragment is processed by Polδ and FEN1. CMG encircling dsDNA is targeted for polyubiquitylation of its MCM7 subunit by SCF^{Dia2} or CRL2^{Lrr1}. Ubiquitylated MCM7 is extracted from DNA by ATPase Cdc48/p97. Finally, catenanes remaining in the DNA are removed. Adapted from (Dewar and Walter, 2017).

Reconstitution of budding yeast replication *in vitro* suggests initiation and elongation are rapid processes, while termination is inefficient by comparison (Devbhandari *et al.*, 2017). However, termination in *Xenopus* egg extracts suggests DNA synthesis does not slow at termination (Dewar *et al.*, 2015). The lack of pausing suggests no steric clash between overlapping replisomes. This may be due to the positioning of the CMG complex only on the leading strand, meaning both replisomes can pass one another without major collision. Pausing would decrease replisome stability and therefore be dangerous to the cell, so arrangement of the CMG complexes to pass each other is beneficial.

The polymerase responsible for the final synthesis (gap filling) still remains a controversial topic. However, in both *Caenorhabditis elegans* and *Xenopus laevis*, Pol- ϵ has been shown to interact with the post-replication CMG complex, but not Pol- δ (Sonneville *et al.*, 2017).

Removal of the CMG replicative helicase complex from DNA is a key step in eukaryotic replication termination, as the CMG complex is central to the active replisome and thus needs to be removed efficiently following the completion of replication. However, this step occurs late, after gap filling has been completed; this is probably to prevent premature replisome disassembly (Dewar *et al.*, 2015). Therefore, the CMG complex will be bound to dsDNA prior to its removal. In budding yeast, it has been shown polyubiquitylation of MCM helicase subunit MCM7, and CMG unloading, are dependent on E3 ubiquitin ligase SCF^{Dia2} (Skp, Cullin, E-box containing complex associated with Digs into Agar 2) (Maric *et al.*, 2014). In vertebrates, E3 ubiquitin ligase CRL2^{Lrr1} (Cullin RING Ligase 2 associated with Leucine Rich Repeats 1) was identified as the E3 ligase responsible for MCM7 ubiquitylation and CMG unloading (Dewar *et al.*, 2017, Sonneville *et al.*, 2017). Ubiquitylated MCM7 has been shown to be deubiquitylated and recycled by Cdc48/p97 (Fullbright *et al.*, 2016), and thus the same would be predicted for the remainder of the polyubiquitylated CMG complex. If this is true, the binding of the CMG complex to dsDNA may trigger the E3 ligase to polyubiquitylate MCM7 to ensure dissociation where they are no longer required (Dewar and Walter, 2017). This ubiquitin-mediated pathway suggests termination of replication may be as highly regulated as initiation (where phosphorylation dictates each step) and thus further work is needed to dissect the specifics of eukaryotic termination.

On completion of unwinding and replicating the parental duplex, daughter molecules will be catenated to one another. Any pre-catenanes present would also be converted into catenanes (Ullsperger, 1995). These would be present behind the fork where lack of access of topoisomerases ahead of the fork during replication means they were converted into pre-catenanes (Gambus, 2017). The details of chromosome resolution remain under

consideration for eukaryotes, however Topoisomerase II (TopoII; Type II topoisomerase) is known to be essential for the termination of replication forks (Bailey *et al.*, 2015, Goto and Wang, 1984). Inactivation of TopoII leads to stalling in G2, with a complete failure to terminate replication and high numbers of hypercatenated molecules (Baxter and Diffley, 2008). This provides strong evidence TopoII is required for decatenation of daughter plasmids, however it has been shown TopoII is not required for initial fork convergence or DNA ligation (Dewar *et al.*, 2015). Further work is required to pinpoint the chain of events leading to successful decatenation and segregation of chromosomes in eukaryotes.

1.2.3 Archaeal replication

Initiation

Archaea can use single or multiple origins to initiate DNA replication, and thus combine features of both bacteria and eukaryotes. The speed of replication also reflects this, whereby on average replication forks progress 10x faster than those in eukaryotes (O'Donnell *et al.*, 2013). Similar to bacterial and eukaryotic origins, archaeal replication origins contain a region of AT-rich sequence (DNA unwinding element; DUE), surrounded by several origin recognition boxes (ORBs) that are recognised and bound by the archaeal initiator protein Orc1/Cdc6. Archaeal origins are sequence-specific, as in bacteria (and yeast), however initiation proteins share homology with those in eukaryotes (Ausiannikava and Allers, 2017).

Origin recognition

The Orc1/Cdc6 initiator protein found in archaea (also denoted as Orc or Cdc6 in literature) is likely an ancestral form of both Orc1 and Cdc6 proteins found in higher eukaryotes. Orc1/Cdc6, like eukaryotic ORC, has a C-terminal winged helix (WH) domain to contact the DNA at the origin (Barry and Bell, 2006). In eukaryotes, Orc1 is a subunit of the known ORC initiator complex, while Cdc6 is thought to act as a helicase loader. Orc1/Cdc6 has previously been implicated in initiation of replication at origins (Ausiannikava and Allers, 2017), however the known role of Cdc6 in eukaryotes provides potential for an additional role of archaeal Orc1/Cdc6 in helicase loading.

There is no current evidence for Orc1/Cdc6 causing melting of dsDNA at the origin, instead it is believed to recruit MCM to the origin to mark it for replisome loading (Samson and Bell, 2016). In this regard Orc1/Cdc6 shares an evolutionarily conserved role with the eukaryotic ORC complex (Samson *et*

al., 2013). The gene encoding Orc1/Cdc6 is nearly always found adjacent to the origin at which it binds. Such pairings allow independent control of each origin along with reducing competition between multiple initiators acting at multiple origins (Samson *et al.*, 2013).

It is important to note that not all Orc1/Cdc6 genes play a role in replication initiation: for example, *H. volcanii* has nine Orc1/Cdc6 genes on the main chromosome, while there are only three origins present (Hartman *et al.*, 2010). The majority of the additional Orc1 genes have been acquired for LGT and no longer function in DNA replication. Orc1/Cdc6 is not the exclusive archaeal initiator protein: *Sulfolobus* species use protein WhiP (Winged-helix initiator Protein) for initiation of origin *oriC3* (Robinson and Bell, 2007).

Replisome loading

Orc1/Cdc6 binding is hypothesised to allow the loading of minichromosome maintenance replicative helicase MCM (homologous to eukaryotic MCM2-7) (Wu *et al.*, 2014b). The archaeal *mcm* gene encodes a single subunit which will form a homo-hexameric ring that is able to surround the leading DNA strand. These MCM rings can form doublets, with the double hexamer known to be more active as a helicase compared to the single ring form (Fletcher *et al.*, 2005).

Akin to eukaryotes, the loading of MCM is believed to trigger the formation of a CMG complex (Cdc45-MCM-GINS), whereby proteins Cdc45 and GINS are recruited to the origin (Makarova *et al.*, 2012). However, unlike eukaryotes which only possess helicase activity when MCM is associated with members of the CMG complex, homo-hexamers of MCM in archaea show helicase activity *in vitro* without association with Cdc45 and GINS (Sakakibara *et al.*, 2009a). GINS in archaea is a dimer of dimers, either a heterodimer of GINS51 and GINS23, or a homodimer of GINS51 (Makarova *et al.*, 2012). While archaea do not encode a *bona fide* Cdc45, the N-terminus of eukaryotic Cdc45 contains a DHH phosphoesterase domain and thus archaeal RecJ proteins (part of the DHH phosphoesterase superfamily) are thought to be the archaeal orthologues of eukaryotic Cdc45 (Makarova *et al.*, 2012, Pellegrini, 2017). It is predicted that, as in eukaryotes, formation of the CMG complex allows local unwinding of DNA and recruitment of replication components to form a full replisome. Archaeal replisomes will then translocate 3'-5' along DNA as replication progresses (Barry and Bell, 2006).

Elongation

Similar to eukaryotes, activation of the archaeal CMG complex and unwinding of DNA triggers the loading of ssDNA-binding protein to protect ssDNA from degradation or modification. Depending on species, the SSB protein in archaea may be bacterial (SSB-like) or eukaryotic (RPA-like). The SSB of crenarchaeal species *Sulfolobus solfataricus* is bacterial-like in domain structure (Wadsworth and White, 2001), but at the structural level carries OB folds comparable to that of human RPA (Kerr *et al.*, 2003), while the euryarchaeon *Pyrococcus abyssi* encodes a heterotrimer showing direct homology to eukaryotic RPA (Komori and Ishino, 2001).

The archaeal replicative polymerases, and the replisome as a whole, are in most aspects a simplified version of the eukaryotic system. All studied species of archaea contain at least one copy of PolB, a family B polymerase with catalytic and proofreading activities contained within a single polypeptide (Johansson and Dixon, 2013). All phyla apart from Crenarchaeota also contain the uniquely archaeal family-D polymerase, PolD.

Archaea can carry both bacterial-like and eukaryotic-like primases. Bacterial DnaG-like primases in archaea, however, have gained roles in RNA degradation, with eukaryotic-like primases in archaea acting in replication (Li *et al.*, 2010). However, no Pol- α or B subunit homologues have been identified in archaeal genomes, suggesting a Pol α -primase-like system is not used by archaea (Lao-Sirieix *et al.*, 2005). Archaeal replicative primase is a two-subunit complex consisting of a small catalytic subunit (PriS/p41) and a large regulatory subunit (PriL/p46) (Kelman and Kelman, 2014). Fusion events of PriS and PriL have been seen within nanoarchaeal genomes (Raymann *et al.*, 2014). Unlike bacterial and eukaryotic primases, which can only utilise ribonucleotides, archaeal primases have been shown to be capable of both RNA and DNA synthesis, with DNA synthesis reaching lengths of several kilobases, allowing for the definition of some archaeal primases as non-canonical DNA polymerases (Lao-Sirieix and Bell, 2004, Galal *et al.*, 2012). It is therefore predicted there will be a handoff event from the primase to the replicative polymerase, similar to that seen in eukaryotic Pol α -primase complex following short error-prone synthesis by primase and Pol- α . It is also worth noting the similarity between archaeal PriS and family X polymerases: family X polymerases are involved in replication, repair and recombination but are largely absent from archaea (Barry and Bell, 2006, Rodriguez *et al.*, 2019). It could be possible that the DNA synthesis ability of PriSL is due to PriS having gained an additional role as a family X-like polymerase, acting in DNA repair and recombination in archaea.

As in bacteria and eukaryotes, archaeal replisomes also recruit a clamp protein to increase efficiency and processivity of the replication process. Archaea, like eukaryotes, use a homologue of PCNA clamp protein that is loaded by homologues of eukaryotic RFC clamp loader (Kelman and Kelman, 2014). Crenarchaeal PCNA is heterotrimeric, as in eukaryotes, however PCNA from other archaeal species form homotrimers. The ring structures formed by both homo- and hetero-trimers in archaea are directly comparable and share similarity with both the bacterial β -clamp and eukaryotic PCNA structures (Pan *et al.*, 2011a, Kelman and Odonnell, 1995). Structurally, archaeal PCNA forms a trimeric ring through head-to-tail interaction of three monomers. PCNA is often thought of as a ‘molecular toolbelt’ whereby it can interact with numerous proteins that may be required at the replisome, including polymerases (both replicative and translesion), ligases and flap endonucleases (Barry and Bell, 2006).

Archaeal PolB is relatively well studied and thermophilic high-fidelity PolB species have been exploited for use in PCR applications (Ishino and Ishino, 2014). Archaeal PolB also possesses a unique damage sensing mechanism not observed in eukaryotic family B polymerases, whereby it will scan DNA ahead of the catalytic domain for incorrect incorporation of deaminated bases that have escaped repair by uracil-N-glycosylase (Greagg *et al.*, 1999, Connolly, 2009). This causes stalling of the polymerase four bases ahead of the incorrect base to allow repair, although the details remain unknown. Contrarily, a recent study characterising the PolB protein of hyperthermophilic and radioresistant species *Thermococcus gammatolerans* showed its polymerase was able to bypass uracil in DNA (Zhang *et al.*, 2020). Both this species and *Sulfolobus solfataricus* family B polymerases have also been shown to be stalled by the presence of an abasic site in DNA (where a base in DNA has lost its purine or pyrimidine base, either spontaneously or due to DNA damage) (Zhang *et al.*, 2020, Gruz *et al.*, 2003).

Archaea-specific DNA polymerase PolD is a heterodimer, consisting of DP1 and DP2 subunits. DP1 is a small subunit with 3'-5' exonuclease activity, while large subunit DP2 is the catalytic polymerase. PolD has been shown to interact with PCNA, as PolB does (Tori *et al.*, 2007). Interaction of PolD with PCNA boosts its processivity (Madru *et al.*, 2020). Recent work has revealed that PolD connects primase to the archaeal replisome before interacting with PCNA (Oki *et al.*, 2021). This suggests roles for PolD in both initiation and elongation processes of DNA replication. It is thought to be one of the main replicative polymerases, supported by the fact PolD, but not PolB, is essential in species including *Thermococcus kodakarensis* (Cubonova *et al.*, 2013).

Arguments have been made that PolB may act as the leading strand polymerase while PolD is the lagging strand polymerase, however the

evidence for this is not without disagreement and requires further work. This hypothesis is due to the requirement for the lagging strand polymerase to carry out repeated priming of Okazaki fragments. Both PolB and PolD have the ability to extend RNA primers, however PolD is more efficient at extension, and thus is more likely to fulfil the requirement for lagging strand synthesis (Greenough *et al.*, 2015). However, PolD lacks RNA displacement activity, whereby it halts 4 nucleotides upstream of the next Okazaki fragment, and thus is unlikely to be involved in Okazaki fragment processing (Greenough *et al.*, 2015). PolB is capable of efficient strand displacement of RNA, which would be required to remove the primers associated with Okazaki fragments on the lagging strand. Further arguments for the roles of archaeal polymerases during genome replication will be discussed in further detail in *Chapter 3*.

As in eukaryotes, FEN1 is known to be involved in Okazaki fragment processing in archaea, but RNaseHIII and RecJ protein GAN have also been implicated in flap removal (Burkhart, 2017, Henneke, 2012). In *T. kodakarensis* either GAN or FEN1 and RNaseHIII are necessary for viability, but the interactions at the replication fork in wild-type cells are yet to be defined (Burkhart, 2017).

Termination

As with eukaryotes, archaeal termination of replication is not well studied in comparison to both initiation and elongation. Archaea, akin to bacteria, have circular chromosomes. However, archaea do not carry the defined termination zones seen in bacterial chromosomes (Duggin *et al.*, 2011, Hawkins *et al.*, 2013a). Instead, termination of replication appears to occur in termination ‘zones’, where converging replication forks meet randomly as in eukaryotes (Duggin *et al.*, 2011). This can be seen where marker frequency analysis maps termination zones as sharp ‘canyons’ in bacteria, compared to broad valleys of termination in archaea. Termination of replication in *Sulfolobus* is seen to be asynchronous (Lundgren *et al.*, 2004) and thus the rate of initiation and number of origins may dictate where termination occurs.

Sulfolobus solfataricus encodes a single Xer homologue and has a corresponding *dif* site, akin to that utilised in bacteria for resolution of chromosome dimers prior to segregation. However, *dif* is located outside of the widely termed ‘termination zones’ and therefore it is likely that termination and segregation are less tightly linked processes at least in this archaeal species when compared to bacteria (Duggin *et al.*, 2011).

1.3 Alternative methods of DNA replication

Recombination-dependent replication

Replication origins were previously believed to be a requirement for cellular life, whereby deletion of replication origins leads to impaired growth or cell death (Ogawa *et al.*, 1984). Work in *E. coli* showed that replication could be primed through the formation of either a D-loop (displacement loop; where ssDNA invades the dsDNA duplex) or an R-loop (RNA displacement loop; where an RNA-DNA hybrid is formed, such as by aborted transcription). These structures can then be remodelled to give a replication fork from which canonical DNA synthesis can proceed (Kogoma, 1997). Where an R-loop is utilised, RNase HI must be mutated; if present in its wild-type form, RNase HI would act to degrade any RNA:DNA hybrids formed (Kogoma, 1997, Ogawa *et al.*, 1984). Both RNA- and DNA-mediated mechanisms were shown to require the strand exchange protein, RecA (Masai and Arai, 1996), suggesting the manner in which the genome is replicated in the absence of an origin is dependent on homologous replication (recombination-dependent replication; RDR).

The T4 bacteriophage was shown to utilise RDR prior to the studies in *E. coli*. It utilises both D- and R-loops depending on the phase of replication (Luder and Mosig, 1982). Unlike canonical replication using origin melting and replisome loading, T4 viruses are thought to initiate replication through the formation and extension of R-loops at the origin (Miller *et al.*, 2003). The origins of T4 bacteriophage consist of a middle-mode promoter (where middle refers to the stage of transcription: early, middle or late) and a downstream AT-rich DNA unwinding element (DUE) (Carles-Kinch and Kreuzer, 1997). Priming transcripts are initiated when transcription factor MotA binds middle promoters present at the origin. Transcripts from these origin-positioned promoters then form persistent R-loops within the DUE region. The 3' end of the RNA is then processed by RNase H to give a 3' hydroxyl group for extension by DNA polymerase (Carles-Kinch and Kreuzer, 1997, Belanger and Kreuzer, 1998). As extension of these 'priming' loops occurs, the process becomes dependent on recombination proteins and converts to RDR (Kreuzer and Brister, 2010). This forcing open of the DNA duplex by the R-loop allows for loading of the replication components (including primers for lagging strand priming) onto DNA and thus sets up a *bona fide* replication fork for genome duplication (Kreuzer and Brister, 2010). Until recently, only viruses were thought to routinely utilise a method of replication dependent on recombination.

Genetic manipulation of *H. volcanii* has revealed that the four origins on the main chromosome (of the lab strain H26; *ori-C1*, *-C2*, *-C3* and *-pHV4*)

are dispensable; cells lacking origins remain viable, in fact growing 7.5% faster than the wild type (**Figure 1.19**) (Hawkins *et al.*, 2013a).

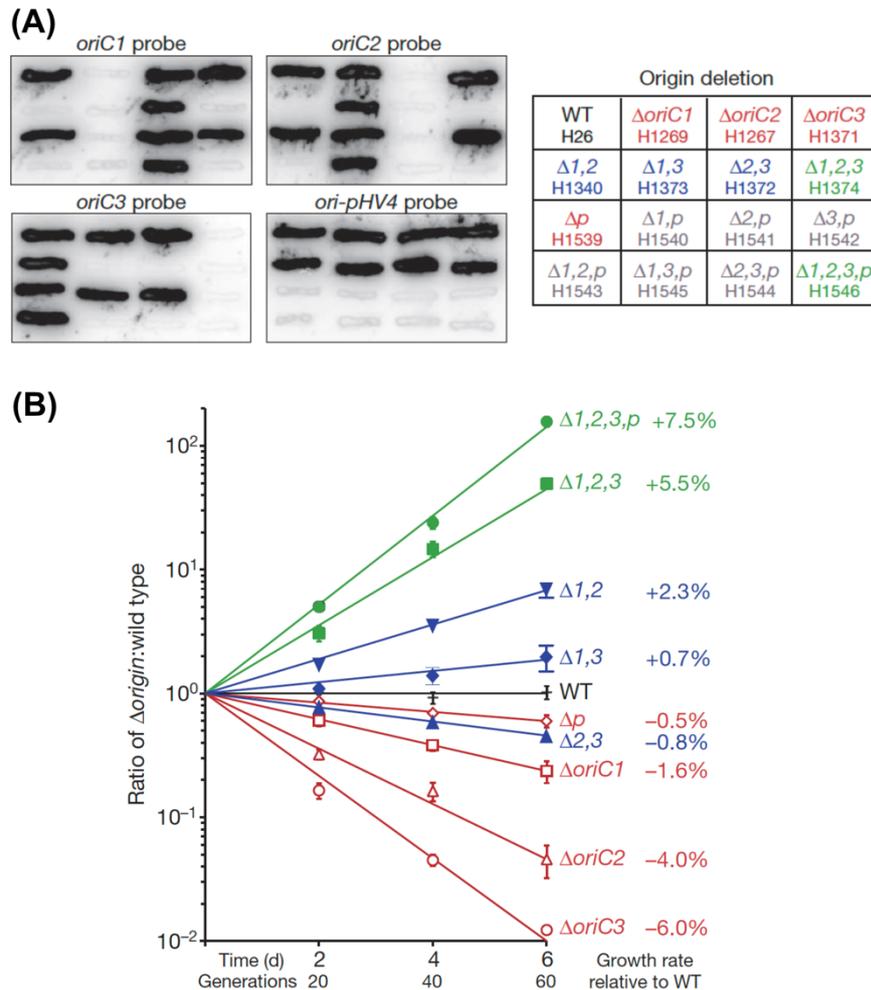


Figure 1.19: Deletion of replication origins is viable in *H. volcanii*. (A) Deletion strains were confirmed by hybridization with origin-specific probes (WT, wild type; ‘p’, ori-pHV4). (B) Growth competition assays show the viability of origin deletion strains; strains deleted for all main chromosomal origins grow 7.5% faster than WT. Adapted from Hawkins *et al.*, 2013.

Previous studies have shown that deletion of RadA (archaeal homologue of bacterial RecA and eukaryotic Rad51; recombinase) renders cells incapable of performing homologous recombination (Woods and Dyll-Smith, 1997). The inability to delete the gene encoding RadA in an originless *H. volcanii* background led to the conclusion that these cells replicate in a manner dependent on homologous recombination (i.e., RDR) (Hawkins *et al.*, 2013a, Michel and Bernander, 2014). Since the strain still encodes RNase HI, it is unlikely to replicate via persistent R-loops, however this has not been proven experimentally to date. It is also noteworthy that use of replication origins in the wild type strain has been shown to differ with growth rate, where

the wild type strain can show an MFA profile similar to a strain lacking origins (Thorsten Allers, personal communication).

Similar work has since shown that originless replication is possible without deleterious effects in both *Thermococcus kodakarensis* (Gehring *et al.*, 2017) and *T. barophilus* (Moalic, 2021). It could be imagined that RDR should occur with relative ease in polyploid organisms, such as *H. volcanii*, *T. kodakarensis* and *T. barophilus*, due to the high genome copy number. Availability of multiple genome copies would allow the strand invasion of homologous DNA, which could subsequently recruit replication machinery, establish a replication fork and migrate along the homologue to synthesise DNA.

Most cyanobacterial species possess more than one copy of their chromosomes (Griese *et al.*, 2011). Recent work has shown some cyanobacterial species with high ploidy are able to replicate asynchronously from multiple discrete sites along the chromosome in a DnaA-independent manner (Ohbayashi *et al.*, 2020). These species show a flat replication profile through marker frequency analysis, akin to that of *H. volcanii*, suggesting these cyanobacterial species are utilising recombination to replicate their genomes. It was shown that where *dnaA* was lost, there was an associated increase in chromosomal copy number per cell, directly suggesting ploidy is inherently linked to the ability to replicate away from origins (Ohbayashi *et al.*, 2020).

This link between high ploidy and successful originless replication using RDR may explain why organisms with low ploidy, including *E. coli* and eukaryotes, have an apparent essentiality for replicating using origins; without numerous homologues to invade, they would have severe problems carrying out RDR. The mechanisms for replication and the machinery utilised to carry out this mode of replication is yet to be determined.

Break-induced replication

Break-induced replication (BIR) is a form of RDR whereby DNA replication is primed by a single-ended DNA double strand break (DSB). As in homologous recombination (HR), the DSB will be resected to produce a 3' ssDNA overhang. This overhang will search for homology and invade a homologous chromosome. Usually during HR, both DNA ends will invade the homologue, but during BIR only a single DNA end will invade to form a D-loop. The invading 3' end will then be extended unidirectionally, establishing a one-way replication fork.

The unusual mechanism of BIR generally leads to a conservative method of inheritance (**Figure 1.20**), whereby the migrating D-loop acts like the leading strand, with the nascent leading strand DNA acting as the template for lagging strand synthesis (Donnianni, 2013, Saini *et al.*, 2013). BIR can also result in semi-conservative inheritance, whereby cleavage of the invading D-loop intermediate and ligation to the donor duplex gives rise to a stable replication fork (Llorente *et al.*, 2008). Interestingly, unlike RDR in bacteria and viruses, eukaryotic BIR is able to take place in a Rad51 recombinase-independent manner, where it instead utilises Rad52 (Malkova *et al.*, 1996, Ira and Haber, 2002).

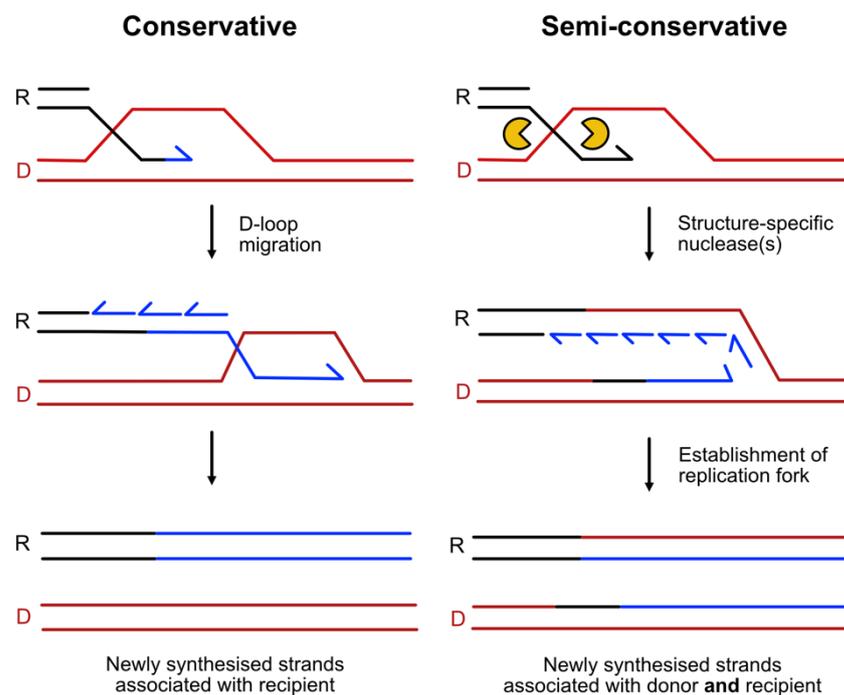


Figure 1.20: Semi-conservative and conservative inheritance as a product of break induced replication (BIR)

Conservative inheritance will occur if the displaced nascent DNA strand acts as the template for lagging strand synthesis; both nascent strands will segregate with the recipient. If the invading structure is cleaved (red arrows) by structure-specific nucleases, a typical replication fork will be established, and inheritance of nascent DNA will be semi-conservative. R = recipient; D = donor. Adapted from (Donnianni, 2013).

BIR is an atypical mode of DNA replication. Unlike replication initiating at origins, BIR does not require origin recognition complex proteins (ORC) (Lydeard, 2010). Instead, it primes invasion using only the resected DNA end. The replisome associated with BIR also differs; in eukaryotes

(namely budding yeast), Pif1 helicase is essential (Saini *et al.*, 2013). While previously, MCM was thought to not be required, recent evidence suggests MCM likely works in tandem with Pif1 to enhance processivity (Drissi *et al.*, 2018).

Work on HR-dependent fork restart in *Schizosaccharomyces pombe* has also revealed that Pol- δ acts as the main polymerase in BIR, in place of canonical leading strand polymerase Pol- ϵ (Miyabe *et al.*, 2015) and synthesis will continue for ~30 kb before maturing the canonical Pol- δ /Pol- ϵ configuration (Naiman *et al.*, 2021). Polymerase usage sequencing also revealed that Pol- α is not used significantly in HR-restarted replication forks, suggesting instead Pol- δ is subsequently filling gaps in the lagging strand (Naiman *et al.*, 2021).

Leading and lagging strand synthesis is asynchronous in BIR; a single stranded tail extends from the migrating bubble (D-loop), which is likely due to the late priming of lagging strand synthesis. The availability of this ssDNA is likely to accumulate lesions and thus increases chance of mutations when utilising BIR for replication (Saini *et al.*, 2013). Generally, BIR is associated with high levels of mutagenesis, including increased levels of genome rearrangements and copy number variations, many of which mimic genomic lesions annotated for cancerous cells (Kramara *et al.*, 2018). It has, however, been shown that eukaryotes have mechanisms to attenuate the use of BIR *in vivo*: Mus81-mediated mechanisms prevent long extensions of BIR, halting the extending D-loop at a single DNA nick, allowing merging with an approaching replication fork (Mayle, 2015). However, it is worth noting that HR-restarted replication forks are capable of progressing uninterrupted through a replication barrier that would block a canonical replication fork (Naiman *et al.*, 2021).

1.4 CRISPR interference (CRISPRi)

CRISPR-Cas (Clustered Regulatory Interspaced Short Palindromic Repeats – CRISPR associated) is a system found in bacteria and archaea, used as a defence mechanism against foreign genetic elements (for recent review, see (Barrangou and Marraffini, 2014)). Multiple CRISPR systems exist, ranging in genetic composition as well as sequence and locus architecture (Makarova *et al.*, 2018). However, while they vary greatly, all CRISPR-Cas systems have maintained 2 components: 1) the Cas proteins and 2) the CRISPR-derived RNA (crRNA). CRISPR loci give rise to crRNAs, the spacer sequence of which should match sequences of foreign invaders that have been previously acquired. The crRNA maturation and activity are critical, along with its interaction with several Cas proteins. Cas proteins act to ensure the production of crRNAs is possible: they acquire and edit incoming DNA into the CRISPR locus to increase the number of recognised spacer sequences for future crRNA production, thereby increasing the organism's immune memory. Cas proteins also make up the Cascade complex, involved in recognising crRNA:foreign-DNA molecules for degradation (Maier *et al.*, 2019).

CRISPR systems are grouped into two classes (Class I/Class II) and six major types (Type I-IV), which are then further subdivided into various subtypes (e.g. I-A-F and I-U) (Makarova *et al.*, 2018). Archaea predominantly use three subtypes of CRISPR-Cas Class I type: I-A, I-B and I-D (Volf, 2000). It is worth noting the CRISPR-Cas systems in archaea are very diverse, with a number of species encoding more than one CRISPR-Cas system (Vestergaard *et al.*, 2014).

Halobacteriales have found to be homogenous with regards to the CRISPR systems they encode: they only encode type I-B and type I-D systems (Maier *et al.*, 2017). The type I-B CRISPR-Cas system utilised by *Haloferax* species has previously been characterised in detail (Maier *et al.*, 2013, Stoll *et al.*, 2013, Maier *et al.*, 2019). This system is composed of eight Cas proteins (Cas1-5, Cas6b, Cas7 and Cas8b) and three constitutive CRISPR loci (P1, P2 and C). Each CRISPR locus has a leader sequence containing associated promoter sequences. This drives transcription of the CRISPR locus, giving rise to a crRNA precursor (pre-crRNA) comprising a number of spacers separated by identical repeat sequences (Maier *et al.*, 2019). This immature crRNA precursor is then processed by Cas6 endonuclease to give a pool of mature crRNAs, each carrying an individual spacer sequence (**Figure 1.21**).

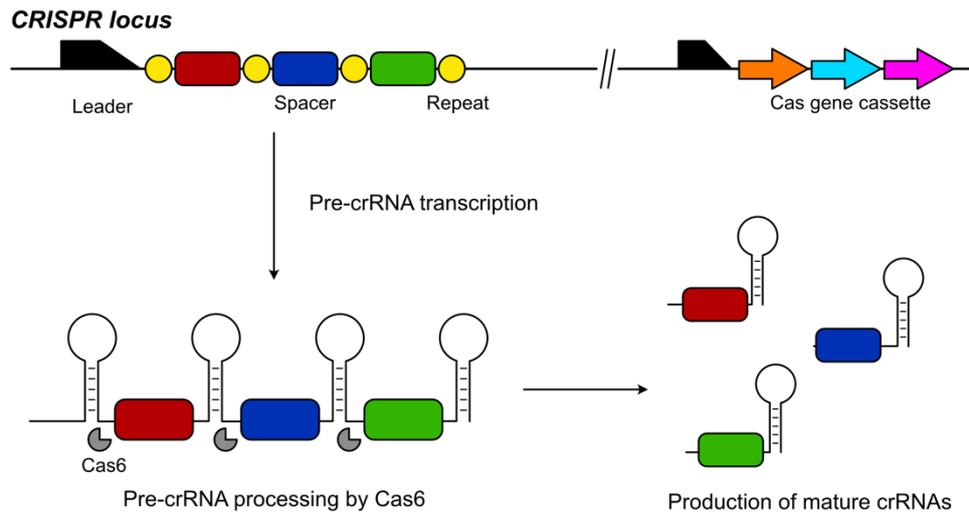


Figure 1.21: Production of mature crRNAs in type IB systems.

Spacer sequences are flanked by repeat sequences. The leader sequence is upstream of spacer and repeat sequences and contains promoter elements to lead transcription of the long precursor, the pre-crRNA. Endonuclease Cas6 cleaves the pre-crRNA within the repeat sequence, generating a pool of crRNAs each targeting an individual spacer sequence. Adapted from (Maier *et al.*, 2019).

During invasion, the invading genetic element will release a fragment of DNA into the cell, which will be recognised as foreign and degraded by the cell. These degraded DNA fragments are known as protospacers. Where protospacers are flanked by a protospacer adjacent motif (PAM), Cas proteins (namely Cas1, Cas2 and Cas4) will act to integrate this protospacer sequence into the organism's CRISPR locus in a process named adaptation.

Within the spacer sequence of a mature crRNA will be a 5' seed sequence (~10 bp in length), preceded by a 5' handle. The spacer portion of the crRNA will base pair to a complementary sequence in the selected host DNA, named the protospacer. The seed sequence is required to base pair perfectly for the successful interaction between crRNA and target DNA. The 5' handle is essential and cannot be deleted, while the 3' handle can be removed completely without affecting the efficiency of recognition (Maier *et al.*, 2015b). The crRNA:target DNA interaction will trigger recruitment of the Cascade complex, a multiprotein complex containing numerous Cas proteins (in *H. volcanii* Cas5, Cas6 and Cas7), some of which are capable of binding DNA. However, for recognition of the target site and for Cascade binding to occur a short sequence motif, named protospacer adjacent motif (PAM), must be present (Mojica *et al.*, 2009). The PAM is a short motif found in the DNA sequence of the organism being targeted. The loading of Cascade marks the

Chapter 1: Introduction

DNA for the recruitment of CRISPR-associated nuclease, Cas3, which will degrade the DNA highlighted by the binding of crRNA and therefore Cascade.

The understanding of CRISPR I-B systems has allowed for the development of a method for gene interference exploiting its own machinery against itself, namely CRISPR interference (CRISPRi) (Stachler and Marchfelder, 2016). *Haloflex* species are able to recognise six different types of PAM, providing an increased choice in target: an advantage for using these organisms (Stachler and Marchfelder, 2016).

During CRISPRi, crRNAs will be designed to specifically recruit the Cascade complex to target the promoters of selected gene sequences, with the aim of Cascade binding blocking access of RNA polymerase (RNAP) to these sites. Cascade binding would usually lead to the recruitment of Cas3 and subsequent degradation of the targeted sequence (**Figure 1.22**). To ensure the gene expression is affected without degradation of the promoter sequence, *cas3* must be deleted from strains used for CRISPRi. In order to increase chances of CRISPRi being efficient within *Haloflex* cells, it is critical that Cascade complexes are readily available to bind the synthetic crRNA once it has bound its target DNA. Wild-type *Haloflex volcanii* have 51 endogenous crRNAs encoded at three CRISPR loci (P1, P2 and C). Cas6 is the endonuclease involved in processing the pre-crRNA transcript into mature crRNAs. By deleting *cas6* in CRISPRi strains, these native transcripts will not be processed into mature crRNAs. Instead, by engineering an already mature crRNA with selected spacer sequence, it would mean the crRNA added would be the only viable crRNA in the strain.

1.5 Aims and Objectives

Many proposals have been put forward regarding the components of DNA replication in archaea, usually based on *in vitro* data, however the specifics of archaeal replication are yet to be defined fully. With respect to *Haloferax volcanii*, little is known about the specifics of its replication and whether the method of replication utilised differs in the absence of origins. Break-induced replication in yeast has been shown to require a different template as well as different replication machinery for replication (e.g., Pif1 helicase in place of canonical helicase MCM). It could therefore be hypothesised the same may be true of originless replication in *H. volcanii*. If there are no defined start sites (origins) and replication is primed throughout the genome, it is possible that the method of priming (and the associated replisome then formed) is wholly different to the origin-based system where ORC proteins control access of replication machinery to DNA.

H. volcanii, as a euryarchaeon, encodes both family B and family D polymerases, which are predicted to be the main replicative polymerases. However, three species have now been reported where deletion of PolB was possible: *Thermococcus kodakarensis* (Cubonova *et al.*, 2013), *Thermococcus barophilus* (Birien *et al.*, 2018) and *Methanococcus maripaludis* (Sarmiento *et al.*, 2013). It is worth testing the assumption that both PolB and PolD are essential in *H. volcanii*, or that they both play a major role in the replication of its genome. The presumed essentiality of both PolB and PolD in *H. volcanii* will be assessed using inhibitors and deletions will be attempted. Should both polymerases prove essential, their requirement in wild type and originless strains will be assessed by placing the genes under a tryptophan-inducible promoter. Attempts will also be made to create strains and plasmids to inhibit their expression using CRISPRi, targeting the CRISPR machinery of *H. volcanii* against the promoter of its own polymerases.

H. volcanii is known to encode all components of the replicative helicase CMG complex: Cdc45 (RecJ in archaea), MCM and GINS. While MCM and GINS are each encoded by a single gene in *H. volcanii*, the situation regarding Cdc45 is fundamentally more complicated. Archaea do not carry a *bona fide* Cdc45 protein; instead, the homology between DHH superfamily RecJ proteins and the ex-nuclease DHH-like Cdc45 places RecJ family proteins as the likely member of the CMG complex in archaea. However, in *H. volcanii*, there are four RecJ genes. Of these, three (RecJ1, RecJ3 and RecJ4) have previously been shown to be dispensable, with only RecJ2 being essential for viability. It has yet to be determined which, if any, of these RecJ proteins play a role in DNA replication. It would also be of interest to characterise the RecJ mutant strains and deduce the function of these proteins *in vivo*. This

Chapter 1: Introduction

study aims to characterise the RecJ proteins of *H. volcanii* and elucidate their roles. Work will also be undertaken to prove the supposed essentiality of RecJ2 through construction of an ectopic inducible-RecJ2 allele.

As mentioned previously, it is possible that originless replication proceeds in a different manner to canonical origin-based replication. For example, it has been reported that there is an increased requirement for MCM helicase in *H. volcanii* in the absence of origins (Marriott, 2017). It remains under question whether this is due to an increased requirement for the helicase alone or for the CMG complex as a whole. The use of tryptophan-inducible promoters will allow this question to be answered. This study will aim to define the components of canonical and non-canonical replication in *H. volcanii* and decipher whether there are any changes in the methods of replication utilised when strains are deleted of chromosomal origins.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Strains

Haloferax volcanii strains

Table 2.1: *Haloferax volcanii* strains used in this study

Strain	Grandparent	Genotype	Notes
H26	DS70	$\Delta pyrE2$	Constructed by TA (Allers <i>et al.</i> , 2004)
H37	DS70	$\Delta pyrE2 \Delta leuB$	Constructed by TA (Allers <i>et al.</i> , 2004)
H53	H26	$\Delta pyrE2 \Delta trpA$	Constructed by TA (Allers <i>et al.</i> , 2004)
H164	H142	$\Delta pyrE2 bgaHa-Bb leuB-Agl \Delta trpA$	Constructed by TA
H282	H195	$\Delta pyrE2 bgaHa-Bb leuB-Agl \Delta trpA \Delta hdrB \Delta hjc$	Constructed by SH
H364	H195	$\Delta pyrE2 bgaHa-Bb leuB-Agl \Delta trpA \Delta hdrB \Delta hef$	Constructed by ZD
H730	H92	$\Delta leuB \Delta hdrB$	Constructed by TA
H4743	H53	$\Delta pyrE2 \Delta trpA \Delta rnhB::trpA+$	
H5382	H4743	$\Delta pyrE2 \Delta trpA \Delta rnhB$	
H5381	H588	$\Delta pyrE2 bgaHa-Kp \Delta trpA \Delta fen1$	
H2117	H164	$\Delta pyrE2 bgaHa-Bb leuB-Agl \Delta trpA \Delta hel308::trpA+$	Constructed by RG-M (Gamble-Milner, 2016)
H4361	H2117	$\Delta pyrE2 bgaHa-Bb leuB-Agl \Delta trpA \Delta hel308$	
H1530	H91	$\Delta pyrE2 \Delta hdrB$	Constructed by TA
H1608	H1591	$\Delta pyrE2 \Delta hdrB \Delta oriC1 \Delta oriC2 \Delta oriC3 \Delta ori-pHV4$	Constructed by TA

Δmrr strains			
H4045	H53	$\Delta pyrE2 \Delta trpA \Delta mrr$	Constructed by LM
H4598	H1804	$\Delta pyrE2 \Delta trpA \Delta oriC1 \Delta oriC2 \Delta oriC3 \Delta ori-pHV4-2 \Delta mrr$	
H4691	H4045	$\Delta pyrE2 \Delta trpA \Delta mrr \Delta hdrB$	
H4695	H4598	$\Delta pyrE2 \Delta trpA \Delta oriC1 \Delta oriC2 \Delta oriC3 \Delta ori-pHV4-2 \Delta mrr \Delta hdrB$	
H4829	H4691	$\Delta pyrE2 \Delta mrr \Delta hdrB$	
H4832	H4695	$\Delta pyrE2 \Delta oriC1 \Delta oriC2 \Delta oriC3 \Delta ori-pHV4-2 \Delta mrr \Delta hdrB$	
Originless strains			
H1340		$\Delta pyrE2 \Delta trpA \Delta oriC1 \Delta oriC2$	Constructed by TA
H1460		$\Delta pyrE2 \Delta trpA \Delta oriC2 \Delta oriC3$	Constructed by TA
H1462		$\Delta pyrE2 \Delta trpA \Delta oriC1 \Delta oriC3$	Constructed by TA
H1464		$\Delta pyrE2 \Delta trpA \Delta oriC1 \Delta oriC2 \Delta oriC3$	Constructed by TA
H1804		$\Delta pyrE2 \Delta trpA \Delta oriC1 \Delta oriC2 \Delta oriC3 \Delta ori-pHV4-2$	Constructed by KP (Hawkins <i>et al.</i> , 2013a)
CRISPRi strains			
H4385	H53	$\Delta pyrE2 \Delta trpA \Delta cas3::trpA+$	Constructed by RL
H4387	H1804	$\Delta pyrE2 \Delta trpA \Delta oriC1 \Delta oriC2 \Delta oriC3 \Delta ori-pHV4-2 \Delta cas3::trpA+$	Constructed by RL
H4606	H4385	$\Delta pyrE2 \Delta trpA \Delta cas3::trpA+ \Delta cas6$	
H4608	H4387	$\Delta pyrE2 \Delta trpA \Delta oriC1 \Delta oriC2 \Delta oriC3 \Delta ori-pHV4-2 \Delta cas3::trpA+ \Delta cas6$	
H4739	H4606	$\Delta pyrE2 \Delta trpA \Delta cas3::trpA+ \Delta cas6 \Delta mrr$	
H4741	H4608	$\Delta pyrE2 \Delta trpA \Delta oriC1 \Delta oriC2 \Delta oriC3 \Delta ori-pHV4-2 \Delta cas3::trpA+ \Delta cas6 \Delta mrr$	

Chapter 2: Materials and Methods

H4886	H4739	$\Delta pyrE2 \Delta trpA \Delta cas3::trpA+$ $\Delta cas6 \Delta mrr \Delta leuB$	
H4888	H4741	$\Delta pyrE2 \Delta trpA \Delta oriC1 \Delta oriC2$ $\Delta oriC3 \Delta ori-pHV4-2$ $\Delta cas3::trpA+ \Delta cas6 \Delta mrr \Delta leuB$	
$\Delta polB1$ strains			
H4625	(H4598)	$\Delta pyrE2 \Delta trpA \Delta oriC1 \Delta oriC2$ $\Delta oriC3 \Delta ori-pHV4-2 \Delta mrr$ $polB+::[\Delta polB::trpA+ pyrE2+]$	Pop-in of pTA2300
H4627	(H4598)	$\Delta pyrE2 \Delta trpA \Delta oriC1 \Delta oriC2$ $\Delta oriC3 \Delta ori-pHV4-2 \Delta mrr$ $polB+::[\Delta polB::trpA+ pyrE2+]$	Pop-in of pTA2307
$\Delta polD2$ strains			
H4953	(H4045)	$\Delta pyrE2 \Delta trpA \Delta mrr$ $polD2+::[\Delta polD2::trpA+$ $pyrE2+]$	Pop-in of p2367
H4954	(H4045)	$\Delta pyrE2 \Delta trpA \Delta mrr$ $polD2+::[\Delta polD2::trpA+$ $pyrE2+]$	Pop-in of p2368
H4955	(H4598)	$\Delta pyrE2 \Delta trpA \Delta oriC1 \Delta oriC2$ $\Delta oriC3 \Delta ori-pHV4 \Delta mrr$ $polD2+::[\Delta polD2::trpA+$ $pyrE2+]$	Pop-in of p2367
H4956	(H4598)	$\Delta pyrE2 \Delta trpA \Delta oriC1 \Delta oriC2$ $\Delta oriC3 \Delta ori-pHV4 \Delta mrr$ $polD2+::[\Delta polD2::trpA+$ $pyrE2+]$	Pop-in of p2368
RecJ mutant strains			
H3929	H164	$\Delta pyrE2 bgaHa-Bb leuB-Agl$ $\Delta trpA \Delta recJ1::trpA+$	Constructed by RL (Lever, 2019)
H3931	H164	$\Delta pyrE2 bgaHa-Bb leuB-Agl$ $\Delta trpA \Delta recJ3::trpA+$	Constructed by RL (Lever, 2019)
H3932	H164	$\Delta pyrE2 bgaHa-Bb leuB-Agl$ $\Delta trpA \Delta recJ4::trpA+$	Constructed by RL (Lever, 2019)
H4273	H164	$\Delta pyrE2 bgaHa-Bb leuB-Agl$ $\Delta trpA \Delta recJ1 \Delta recJ3$ $\Delta recJ4::trpA+$	Constructed by RL (Lever, 2019)
H5286	H364	$\Delta pyrE2 bgaHa-Bb leuB-Agl$ $\Delta trpA \Delta hdrB \Delta hef$ $\Delta recJ1::trpA+$	

Chapter 2: Materials and Methods

H5297	H364	<i>ΔpyrE2 bgaHa-Bb leuB-Agl</i> <i>ΔtrpA ΔhdrB Δhef</i> <i>ΔrecJ3::trpA+</i>
H5306	H364	<i>ΔpyrE2 bgaHa-Bb leuB-Agl</i> <i>ΔtrpA ΔhdrB Δhef</i> <i>ΔrecJ4::trpA+</i>
H5269	(H282)	<i>ΔpyrE2 bgaHa-Bb leuB-Agl</i> <i>ΔtrpA ΔhdrB Δhjc recJ3+::</i> <i>[ΔrecJ3::trpA+]</i>
H5270	(H282)	<i>ΔpyrE2 bgaHa-Bb leuB-Agl</i> <i>ΔtrpA ΔhdrB Δhjc recJ4+::</i> <i>[ΔrecJ4::trpA+]</i>
H5291	H282	<i>ΔpyrE2 bgaHa-Bb leuB-Agl</i> <i>ΔtrpA ΔhdrB Δhjc</i> <i>ΔrecJ1::trpA+</i>
H5282	H1804	<i>ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC2</i> <i>ΔoriC3 Δori-pHV4-2</i> <i>ΔrecJ1::trpA+</i>
H5294	H1804	<i>ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC2</i> <i>ΔoriC3 Δori-pHV4-2</i> <i>ΔrecJ3::trpA+</i>
H5303	H1804	<i>ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC2</i> <i>ΔoriC3 Δori-pHV4-2</i> <i>ΔrecJ4::trpA+</i>
H5404	H5382	<i>ΔpyrE2 ΔtrpA ΔrnhB</i> <i>ΔrecJ1::trpA+</i>
H5406	H5382	<i>ΔpyrE2 ΔtrpA ΔrnhB</i> <i>ΔrecJ3::trpA+</i>
H5408	H5382	<i>ΔpyrE2 ΔtrpA ΔrnhB</i> <i>ΔrecJ4::trpA+</i>
H5387	(H5381)	<i>ΔpyrE2 bgaHa-Kp ΔtrpA Δfen1</i> <i>recJ1+::[ΔrecJ1::trpA+]</i>
H5400	H5381	<i>ΔpyrE2 bgaHa-Kp ΔtrpA Δfen1</i> <i>ΔrecJ3::trpA+</i>
H5402	H5381	<i>ΔpyrE2 bgaHa-Kp ΔtrpA Δfen1</i> <i>ΔrecJ4::trpA+</i>
H5288	H4361	<i>ΔpyrE2 bgaHa-Bb leuB-Agl</i> <i>ΔtrpA Δhel308 ΔrecJ1::trpA+</i>
H5301	H4361	<i>ΔpyrE2 bgaHa-Bb leuB-Agl</i> <i>ΔtrpA Δhel308 ΔrecJ3::trpA+</i>
H5309	H4361	<i>ΔpyrE2 bgaHa-Bb leuB-Agl</i> <i>ΔtrpA Δhel308 ΔrecJ4::trpA+</i>

<i>ΔginS</i> strains			
H4730	(H4045)	<i>ΔpyrE2 ΔtrpA Δmrr</i> <i>ginS+::[ΔginS::trpA+]</i>	Pop-in of pTA2335
H4732	(H4598)	<i>ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC2</i> <i>ΔoriC3 Δori-pHV4-2 Δmrr</i> <i>ginS+::[ΔginS::trpA+]</i>	Pop-in of pTA2335
H5038	(H4045)	<i>ΔpyrE2 ΔtrpA Δmrr</i> <i>ginS+::[ΔginS::trpA+]</i>	Pop-in of pTA2437
H5039	(H4598)	<i>ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC2</i> <i>ΔoriC3 Δori-pHV4-2 Δmrr</i> <i>ginS+::[ΔginS::trpA+]</i>	Pop-in of pTA2437
H5039	(H4045)	<i>ΔpyrE2 ΔtrpA Δmrr</i> <i>ginS+::[ΔginS::trpA+]</i>	Pop-in of pTA2439
H5040	(H4598)	<i>ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC2</i> <i>ΔoriC3 Δori-pHV4-2 Δmrr</i> <i>ginS+::[ΔginS::trpA+]</i>	Pop-in of pTA2439
Inducible promoter strains			
H5081	H730	<i>ΔleuB ΔhdrB ΔpyrE2::p.tnaM3- recJ2::hdrB+</i>	
H5082	H730	<i>ΔleuB ΔhdrB ΔpyrE2::7xHis</i> <i>2xStrepII p.tnaM3-recJ2::hdrB+</i>	
H5131	(H5081)	<i>ΔleuB ΔhdrB ΔpyrE2::p.tnaM3- recJ2::hdrB+</i> <i>recJ2+::[ΔrecJ2::leuB+]</i>	
H5132	(H5082)	<i>ΔleuB ΔhdrB ΔpyrE2::7xHis</i> <i>2xStrepII p.tnaM3-recJ2::hdrB+</i> <i>recJ2+::[ΔrecJ2::leuB+]</i>	
H5235	H37	<i>ΔleuB ΔpyrE2::p.tnaA-recJ2</i>	
H5237	H37	<i>ΔleuB ΔpyrE2::7xHis 2xStrepII</i> <i>p.tnaA-recJ2</i>	
H5250	(H5235)	<i>ΔleuB ΔpyrE2::p.tnaA- recJ2+::[ΔrecJ2::leuB+]</i>	
H5252	(H5237)	<i>ΔleuB ΔpyrE2::7xHis 2xStrepII</i> <i>p.tnaA-recJ2+::[ΔrecJ2::leuB+]</i>	
H5017	H4829	<i>ΔpyrE2 Δmrr ΔhdrB p.tnaM3- ginS</i>	
H5383	H4832	<i>ΔpyrE2 ΔoriC1 ΔoriC2 ΔoriC3</i> <i>Δori-pHV4-2 Δmrr ΔhdrB</i> <i>p.tnaM3-ginS</i>	
H5312	H4829	<i>ΔpyrE2 Δmrr ΔhdrB 7xHis</i> <i>2xStrepII-p.tnaM3-ginS</i>	

Chapter 2: Materials and Methods

H5385	H4832	$\Delta pyrE2 \Delta oriC1 \Delta oriC2 \Delta oriC3$ $\Delta ori-pHV4-2 \Delta mrr \Delta hdrB$ 7xHis 2xStrepII- <i>p.tnaM3-ginS</i>	
H1904	H1530	$\Delta pyrE2 \Delta hdrB$ <i>p.tnaM3-</i> <i>mcm::hdrB+</i>	Constructed by TA (Marriott, 2017)
H1911	H1608	$\Delta pyrE2 \Delta hdrB \Delta oriC1 \Delta oriC2$ $\Delta oriC3 \Delta ori-pHV4$ <i>p.tnaM3-</i> <i>mcm::hdrB+</i>	Constructed by TA (Marriott, 2017)
Protein expression strains			
H5199	H3929	$\Delta pyrE2$ <i>bgaHa-Bb leuB-Ag1</i> $\Delta trpA$ 7xHis 2xStrepII- <i>recJ1</i>	
H5200	H3931	$\Delta pyrE2$ <i>bgaHa-Bb leuB-Ag1</i> $\Delta trpA$ 7xHis 2xStrepII- <i>recJ3</i>	
H5313	H3932	$\Delta pyrE2$ <i>bgaHa-Bb leuB-Ag1</i> $\Delta trpA$ 7xHis 2xStrepII- <i>recJ4</i>	
H5109	(H26)	$\Delta pyrE2$ { <i>p.fdx::</i> ^N GFP-spacer Nov ^R } { <i>p.fdx::</i> ^C GFP-spacer Mev ^R }	Constructed by PP
H5110	(H26)	$\Delta pyrE2$ { <i>p.fdx::</i> ^N GFP-spacer Nov ^R } { <i>p.fdx::</i> spacer- ^C GFP Mev ^R }	Constructed by PP
H5111	(H26)	$\Delta pyrE2$ { <i>p.fdx::</i> spacer- ^N GFP Nov ^R } { <i>p.fdx::</i> ^C GFP-spacer Mev ^R }	Constructed by PP
H5112	(H26)	$\Delta pyrE2$ { <i>p.fdx::</i> spacer- ^N GFP Nov ^R } { <i>p.fdx::</i> spacer- ^C GFP Mev ^R }	Constructed by PP
H5334	(H26)	$\Delta pyrE2$ { <i>p.fdx::</i> ^N GFP- <i>hel308</i> Nov ^R } { <i>p.fdx::</i> <i>recJ3-</i> ^C GFP Mev ^R }	Constructed by AD
H5335	(H26)	$\Delta pyrE2$ { <i>p.fdx::</i> ^N GFP- <i>hel308</i> Nov ^R } { <i>p.fdx::</i> <i>recJ4-</i> ^C GFP Mev ^R }	Constructed by AD
H2047	H1606	$\Delta pyrE2 \Delta trpA \Delta mrr$ <i>Nph-pitA</i> <i>cdc48-ct</i>	Constructed by TA (Wardell <i>et</i> <i>al.</i> , 2017)
H2962	H2047	$\Delta pyrE2 \Delta trpA \Delta mrr$ <i>Nph-pitA</i> <i>cdc48d-Ct ginS+</i> -StrepII tag	Constructed by HM (Marriott, 2017)
H3628	H2962	$\Delta pyrE2 \Delta trpA \Delta mrr$ <i>Nph-pitA</i> <i>cdc48d-Ct ginS+</i> -StrepII tag 6xHis- <i>mcm+</i>	Constructed by HM (Marriott, 2017)

H4614 H2047 $\Delta pyrE2 \Delta trpA \Delta mrr Nph-pitA$
 $cdc48-ct$ 7xHis 2xStrepII-*ginS*⁺

() signifies parent strains were generated for construction of this strain, but daughter strains were not able to be generated. [] signifies presence of an integrated plasmid. TA = Thorsten Allers, MM = Moshe Mevarech, KP = Katarzyna Ptasinska, AM = Anita Marchfelder, RL = Rebecca Lever, DA = Darya Ausiannikava, ZD = Zhenhong Duan, RG-M = Rebecca Gamble-Milner, PP = Patricia Perez, AD = Ambika Dattani

Escherichia coli strains

Table 2.2: *Escherichia coli* strains used in this study

Strain	Genotype	Notes
XL1 Blue MRF'	<i>endA1, gyrA96</i> (NalR), <i>lac</i> [F' proAB <i>lacIqZ</i> Δ M15 <i>tn10</i> (TetR)], Δ (<i>mcrA</i>)183, Δ (<i>mcrCB-hsdSMR-mrr</i>)173, <i>recA1, relA1, supE44, thi-1</i>	Standard cloning strain for blue-white screening using pBluescript-based plasmids. Tetracycline resistant, restriction endonuclease and recombination deficient. <i>dam</i> methylase + positive. From Stratagene.
N2338 (GM121)	<i>F-</i> , <i>ara-14, dam-3, dcm-6,</i> <i>fhuA31, galK2, galT22, hsdR3,</i> <i>lacY1, leu-6, thi-1, thr-1, tsx-78</i>	<i>dam- dcm-</i> mutant for preparing unmethylated DNA for transformation of <i>mrr</i> ⁺ strains of <i>Haloferax volcanii</i> . From RG Lloyd.

2.1.2 Plasmids

Table 2.3: Plasmids used in this study

Name	Use	Notes
pTA131	Derivative of pBluescript, for making deletions in $\Delta pyrE2$ backgrounds	Constructed by TA (Allers <i>et al.</i> , 2004)
pTA298	Contains <i>trpA</i> under control of the ferredoxin promoter <i>p.fdx</i>	Constructed by TA (Lestini <i>et al.</i> , 2010)

Genomic clones

pTA44	Genomic clone of <i>leuB</i>	Constructed by TA (Allers <i>et al.</i> , 2004)
-------	------------------------------	--

pTA49	Genomic clone of <i>trpA</i>	Constructed by TA (Lam <i>et al.</i> , 1990, Allers <i>et al.</i> , 2004)
pTA193	Genomic clone of <i>polB1</i>	Constructed by TA
pTA327	Genomic clone of <i>polD2</i>	Constructed by SH
pTA1912	Genomic clone of <i>recJ1</i>	Constructed by RL (Lever, 2019)
pTA1905	Genomic clone of <i>recJ2</i>	Constructed by RL (Lever, 2019)
pTA1913	Genomic clone of <i>recJ3</i>	Constructed by RL (Lever, 2019)
pTA1882	Genomic clone of <i>recJ4</i>	Constructed by RL (Lever, 2019)
pTA1716	Genomic clone of <i>ginS</i> gene	Constructed by HM (Marriott, 2017)
pTA1246	Genomic clone of inteinless <i>polB1</i>	Constructed by TA
pTA2129	Genomic clone of <i>polD2</i> with <i>pyrE2</i> marker	pTA327 with added <i>pyrE2</i> marker
Inducible gene replacement constructs		
pTA1451	Cloning vector for placing genes under the tryptophan-inducible promoter <i>p.tnaM3</i> , with added <i>hdrB</i> selection	Constructed by TA (Braun <i>et al.</i> , 2019)
pTA2096	Cloning vector for placing genes under the tryptophan-inducible promoter <i>p.tnaM3</i> , with 5' 7xHis 2xStrepII tags and added <i>hdrB</i> selection	Constructed by RL (Lever, 2019)
pTA2181	Intermediate plasmid in generating <i>p.tnaM3-polB1-hdrB</i> construct	Contains full-length inteinless <i>polB1</i>

Chapter 2: Materials and Methods

pTA2225	<i>p.tnaM3-polB1::hdrB+</i> gene replacement construct	Contains full-length inteinless <i>polB1</i>
pTA2182	Intermediate plasmid in generating <i>p.tnaM3-polD2-hdrB</i> construct	Contains full-length <i>polD2</i>
pTA2290	pTA2182 lacking <i>BamHI</i> restriction site, removed by blunt-ended cloning	Contains full-length <i>polD2</i>
pTA2295	<i>p.tnaM3-polD2::hdrB+</i> gene replacement construct	Contains full-length <i>polD2</i>
pTA2382	<i>p.tnaM3-polB1::hdrB+</i> gene replacement construct with additional US sequence	Contains full-length <i>polB1</i>
pTA2394	<i>p.tnaM3-polD2::hdrB+</i> gene replacement construct with additional US sequence	Contains full-length <i>polD2</i>
pTA2422	Derivative of pTA131 with <i>pyrE2</i> under control of its native promoter	Constructed by TA
pTA2515	<i>p.tnaM3-polB1</i> gene replacement construct	Contains truncated <i>polB1</i> allele
pTA2518	<i>hdrB+::p.tnaM3-polB1</i> gene replacement construct	Contains truncated <i>polB1</i> allele
pTA2554	Empty vector for generation of truncated alleles for integration of <i>p.tnaM3</i> at large genes	
pTA2555	Intermediate plasmid in construction of <i>hdrB+::p.tnaM3-polD2</i> construct	Contains truncated <i>polD2</i> allele
pTA2560	<i>hdrB+::p.tnaM3-polD2</i> gene replacement construct	Contains truncated <i>polD2</i> allele
pTA2478	Intermediate plasmid in construction of <i>p.tnaM3-recJ2::hdrB+</i>	
pTA2481	Intermediate plasmid in construction of 7xHis 2xStrepII <i>p.tnaM3-recJ2::hdrB+</i>	
pTA2498	Construct for replacement of <i>pyrE2</i> with <i>p.tnaM3-recJ2::hdrB+</i>	
pTA2499	Construct for replacement of <i>pyrE2</i> with 7xHis 2xStrepII <i>p.tnaM3-recJ2::hdrB+</i>	

Chapter 2: Materials and Methods

pTA2546	Construct for replacement of <i>pyrE2</i> with <i>p.tnaM3</i> cassette	
pTA2553	Construct for replacement of <i>pyrE2</i> with <i>p.tnaA</i> cassette	
pTA2561	Construct for replacement of <i>pyrE2</i> with <i>p.tnaA-recJ2</i>	
pTA2563	Construct for replacement of <i>pyrE2</i> with 7xHis 2xStrepII <i>p.tnaA-recJ2</i>	
pTA2361	Intermediate plasmid in construction of <i>p.tnaM3-ginS::hdrB+</i>	
pTA2365	<i>p.tnaM3-ginS::hdrB+</i> gene replacement construct	
pTA2465	Intermediate plasmid in construction of 7xHis 2xStrepII <i>p.tnaM3-ginS::hdrB+</i>	
pTA2479	7xHis 2xStrepII <i>p.tnaM3-ginS</i> gene replacement construct	
pTA2524	7xHis 2xStrepII <i>p.tnaM3-ginS::hdrB+</i> gene replacement construct	
pTA2421	Intermediate plasmid in construction of 7xHis 2xStrepII <i>p.tnaM3-mcm::hdrB+</i>	
pTA2426	7xHis 2xStrepII <i>p.tnaM3-mcm</i> gene replacement construct	
CRISPR constructs		
pMA-telecRNA19	CRISPRi construct containing a strongly inducible crRNA cassette for spacer insertion	Courtesy of AM (Maier <i>et al.</i> , 2015b)
pTA232	Episomal vector carrying <i>leuB</i> selection and <i>pHV2</i> origin	Constructed by TA (Allers <i>et al.</i> , 2004)
pTA2227	pMA-telecRNA19 with replacement of crRNA spacer with anti-polB1 sequence #1	Anti-polB1 #1: cgggtgcggttcgcggaacgc cggggtttttagcc
pTA2228	pMA-telecRNA19 with replacement of crRNA spacer with anti-polB1 sequence #2	Anti-polB1 #2: gcggaacgccggggttttagc cgcgccccaag
pTA2293	pMA-telecRNA19 with replacement of crRNA spacer with anti-polB1 sequence #3	Anti-polB1 #3: ccgcgccccaagcgattgcc atgacgcagacggg

Chapter 2: Materials and Methods

pTA2334	pMA-telecRNA19 with replacement of crRNA spacer with anti-polD1 sequence #2	Anti-polD1 #3: ggcatccttttgccgcgtgtcgc gcactccgggtg
pTA2251	pTA232 with crRNA cassette with anti-polB1 sequence #1	
pTA2252	pTA232 with crRNA cassette with anti-polB1 sequence #2	
pTA2305	pTA232 with crRNA cassette with anti-polB1 sequence #3	
pTA2342	pTA232 with crRNA cassette with anti-polD1 sequence #2	
Deletion constructs		
pTA73	Deletion construct for <i>leuB</i>	Constructed by TA (Allers <i>et al.</i> , 2004)
pGB68	Deletion construct for <i>pyrE2</i>	Courtesy of MM (Bitan-Banin <i>et al.</i> , 2003)
pTA155	Deletion construct for <i>hdrB</i>	Constructed by GN (Allers <i>et al.</i> , 2004)
pTA1150	Deletion construct for <i>mrr</i> endonuclease	Constructed by TA (Wardell <i>et al.</i> , 2017)
pTA2291	First step in construction of deletion construct for <i>polB1</i>	
pTA2294	Deletion construct for <i>polB1</i>	
pTA2300	Deletion construct for <i>polB1</i> with <i>trpA</i> marker running in the opposite orientation to <i>polB1</i>	
pTA2307	Deletion construct for <i>polB1</i> with <i>trpA</i> marker running in the same orientation to <i>polB1</i>	
pTA436	Deletion construct for <i>polD2</i>	Constructed by SH
pTA2217	Deletion construct for <i>polD2</i> with insertion of linker sequence integrating <i>Bam</i> HI and <i>Nde</i> I sites	
pTA2367	Deletion construct for <i>polD2</i> with <i>trpA</i> marker running in the same orientation to <i>polD2</i>	

Chapter 2: Materials and Methods

pTA2368	Deletion construct for <i>polD2</i> with <i>trpA</i> marker running in the opposite orientation to <i>polD2</i>	
pTA131-updo(cas3)	Deletion construct for <i>cas3</i>	Courtesy of AM (Stachler and Marchfelder, 2016)
pTA131-updo(cas6)	Deletion construct for <i>cas6</i>	Courtesy of AM (Stachler and Marchfelder, 2016)
pTA1958	Deletion construct for <i>recJ1</i> with <i>trpA</i> selection	Constructed by RL (Lever, 2019)
pTA1951	Deletion construct for <i>recJ2</i>	Constructed by RL (Lever, 2019)
pTA1960	Deletion construct for <i>recJ3</i> with <i>trpA</i> selection	Constructed by RL (Lever, 2019)
pTA1997	Deletion construct for <i>recJ4</i> with <i>trpA</i> selection	Constructed by RL (Lever, 2019)
pTA1775	Deletion construct for <i>rnhB</i>	Constructed by DA
pTA1166	Deletion construct for <i>rpa1</i> containing <i>trpA</i> with no promoter	Constructed by AS (Stroud <i>et al.</i> , 2012)
pTA2329	Deletion construct for <i>rnhB</i> with promoterless <i>trpA</i> marker	
pTA2484	Deletion construct for <i>recJ2</i> with <i>leuB</i> selection	
pTA1254	Deletion construct for <i>hel308</i>	Constructed by TA
pTA2335	Deletion construct for <i>ginS</i> with C-terminal 100 bp	Constructed by RL (Lever, 2019)
pTA2184	Intermediate in construction of deletion construct for <i>ginS</i> (no <i>trpA</i> marker)	Constructed by RL (Lever, 2019)
pTA2315	Intermediate in construction of deletion construct for <i>ginS</i> (no <i>trpA</i> marker) with C-terminal 100 bp	Constructed by RL (Lever, 2019)
pTA2432	Intermediate construct in construction of pTA2439	
pTA2433	Intermediate construct in construction of pTA2437	
pTA2439	Deletion construct for <i>ginS</i> introducing premature stop codon	

pTA2437 Deletion construct for *ginS* encoding the first alpha helix of *ginS*

Protein expression and purification constructs

pTA2090	6xHis- <i>recJ1</i> -StrepII gene replacement construct	Constructed by RL (Lever, 2019)
pTA2091	6xHis- <i>recJ3</i> -StrepII gene replacement construct	Constructed by RL (Lever, 2019)
pTA2095	6xHis- <i>recJ4</i> -StrepII gene replacement construct	Constructed by RL (Lever, 2019)
pTA2564	6xHis- <i>recJ3</i> -StrepII gene replacement construct	Constructed by RL (Lever, 2019)
pTA1771	For the expression of N-terminally tandem 7xHis 2xStrepII tagged genes under the control of their native chromosomal promoter	Constructed by TA (Wardell <i>et al.</i> , 2017)
pTA2390	7xHis 2xStrepII- <i>recJ1</i> gene replacement construct	
pTA2392	7xHis 2xStrepII- <i>recJ2</i> gene replacement construct	
pTA2551	7xHis 2xStrepII- <i>recJ3</i> gene replacement construct	
pTA2648	7xHis 2xStrepII- <i>recJ4</i> gene replacement construct	
pJAS-NGFP-Nterm	For construction of N-terminal ^N GFP fusions, Nov ^R	Courtesy of FP (Winter <i>et al.</i> , 2018)
pJAS-NGFP-Cterm	For construction of C-terminal ^N GFP fusions, Nov ^R	Courtesy of FP (Winter <i>et al.</i> , 2018)
pWL-CGFP-Nterm	For construction of N-terminal ^C GFP fusions, Mev ^R	Courtesy of FP (Winter <i>et al.</i> , 2018)
pWL-CGFP-Cterm	For construction of C-terminal ^C GFP fusions, Mev ^R	Courtesy of FP (Winter <i>et al.</i> , 2018)
pTA2586	N-terminal ^N GFP- <i>recJ1</i> construct, Nov ^R	
pTA2587	C-terminal ^N GFP- <i>recJ1</i> construct, Nov ^R	
pTA2588	N-terminal ^C GFP- <i>recJ1</i> construct, Mev ^R	

Chapter 2: Materials and Methods

pTA2589	C-terminal ^C GFP- <i>recJ1</i> construct, Mev ^R	
pTA2592	N-terminal ^N GFP- <i>recJ2</i> construct, Nov ^R	
pTA2593	C-terminal ^N GFP- <i>recJ2</i> construct, Nov ^R	
pTA2617	N-terminal ^C GFP- <i>recJ2</i> construct, Mev ^R	
pTA2594	C-terminal ^C GFP- <i>recJ2</i> construct, Mev ^R	
pTA2595	N-terminal ^N GFP- <i>recJ3</i> construct, Nov ^R	
pTA2618	C-terminal ^N GFP- <i>recJ3</i> construct, Nov ^R	
pTA2619	N-terminal ^C GFP- <i>recJ3</i> construct, Mev ^R	
pTA2620	C-terminal ^C GFP- <i>recJ3</i> construct, Mev ^R	
pTA2601	N-terminal ^N GFP- <i>recJ4</i> construct, Nov ^R	
pTA2602	C-terminal ^N GFP- <i>recJ4</i> construct, Nov ^R	
pTA2603	N-terminal ^C GFP- <i>recJ4</i> construct, Mev ^R	
pTA2604	C-terminal ^C GFP- <i>recJ4</i> construct, Mev ^R	
pTA1791	N-terminal 6x His and C-terminal StrepII tagged <i>ginS</i> gene replacement construct	Constructed by LM
pTA1663	N-terminal 6x His and C-terminal StrepII tagged <i>mcm</i> gene replacement construct	Constructed by HM (Marriott, 2017)
pTA2260	<i>ginS</i> coding sequence placed under 7xHis 2xStrepII tag in cloning vector pTA1771	
pTA2289	N-terminally tandem 7xHis 2xStrepII tagged <i>ginS</i> gene replacement construct	

TA = Thorsten Allers, HM = Hannah Marriott, MM = Moshe Mevarech, AM = Anita Marchfelder, SH = Sam Haldenby, GN = Greg Ngo, DA = Darya Ausiannikava, AS = Amy Stroud, FP = Felicitas Pfeifer

Plasmid Construction

All plasmids constructed in this project were verified by restriction digest and sequencing. All plasmid numbers/names listed were grown in *E. coli* XL1-Blue cells. All figures containing plasmid diagrams will contain the following abbreviations, as listed in **Table 2.4**.

Table 2.4: Plasmid content abbreviations

Abbreviation	Definition
<i>AmpR</i>	Ampicillin resistance gene, <i>E. coli</i>
<i>ColE1</i>	Replication origin, <i>E. coli</i>
<i>f1 (+) ori</i>	Replication origin, <i>E. coli</i>
<i>lacZ</i>	β -galactosidase used for blue/white selection, <i>E. coli</i>
MCS	Multiple cloning site
<i>p.lac</i>	Promoter for <i>lacZ</i> , <i>E. coli</i>
<i>p.fdx</i>	Ferredoxin promoter, <i>H. volcanii</i>
<i>p.tnaA</i>	Tryptophan-inducible promoter, <i>H. volcanii</i>
<i>p.tnaAM3</i>	Tryptophan-inducible promoter with reduced activity, <i>H. volcanii</i>
<i>hdrB</i>	Thymidine biosynthesis, <i>H. volcanii</i>
<i>trpA</i>	Tryptophan biosynthesis, <i>H. volcanii</i>
<i>pyrE2</i>	Uracil biosynthesis, <i>H. volcanii</i>
His ₆	Hexahistidine tag
StrepII	Streptavidin-binding tag
His ₇ -2xStrepII	Tandem Heptahistidine-Streptavidin-binding tag
<i>t.L11e</i>	L11e rRNA terminator, <i>H. volcanii</i>
<i>t.syn</i>	Synthetic terminator, <i>H. volcanii</i>

2.1.3 Oligonucleotides

Table 2.4: Oligonucleotides used in this study

Name	Sequence (5'-3')	Use
CRISPR oligonucleotides		
<i>polB1#1antifwd</i>	ACGCCGGGGTTTTTAGCCACCGA TATTGGTATGGC	Used to construct pTA2227
<i>polB1#1antirev</i>	TTCCGCGAACCGCACCCGGCTTC AACTACCGATCA	Used to construct pTA2227
<i>polB1#2antifwd</i>	ATTAGCCGCGCCGCGAAGCCG ATATTGGTATGGC	Used to construct pTA2228
<i>polB1#2antirev</i>	AAACCCCGGCGTTTTCCGCGCTTC AACTACCGATCA	Used to construct pTA2228
<i>polB1#3antifwd</i>	TGCCATGACGCAGACGGGACCG ATATTGGTATGGC	Used to construct pTA2293

Chapter 2: Materials and Methods

<i>polB1#3antirev</i>	ATCGCTTCGGCGGCGCGGGCTTC AACTACCGATCA	Used to construct pTA2293
<i>polD1#1antifwd</i>	ATCCTTTTTGCCGCGTGTACCGA TATTGGTATGGC	
<i>polD1#1antirev</i>	GCCGAAACCCAGAACGTTGCTTC AACTACCGATCA	
<i>polD1#2antifwd</i>	TGTCGCGCACTCCGGGTGACCGA TATTGGTATGGC	Used to construct pTA2334
<i>polD1#2antirev</i>	CGCGGCAAAAAGGATGCCGCTT CAACTACCGATCA	Used to construct pTA2334
<i>polD1#3antifwd</i>	ACGCCGGCGCGCATCGTCACCGA TATTGGTATGGC	
<i>polD1#3antirev</i>	CTCCAGTGGCACACCCGGGCTTC AACTACCGATCA	
Oligonucleotides for generating and screening inducible gene constructs		
<i>polBNdeF</i>	GCGATTcatATGACGCAGACGGGT CTGACCG	Used to construct pTA2181
<i>polBBamR</i>	CTCGGgGATcCGGCCGGTCACAT G	Made by TA. Used to construct pTA2181
<i>dp2fwdNde</i>	GCTGTAcataTGC GCGAGGAGGAA ACCCGG	Used to construct pTA2182
<i>dp2revBam</i>	C GACTCGCggAtCCTGGAGGAAA AACCGACCGC	Used to construct pTA2182
<i>dp2USF</i>	TGGCAGCCCCACCCGTTGCCTTC AGG	Used to screen for presence of <i>p.tnaM3-polD2</i>
<i>dp2intR</i>	CGATGCGACGGAAGTACCGGGT TTCC	Used to screen for presence of <i>p.tnaM3-polD2</i>
<i>polB1USKpnF</i>	GTGGGTaCcCCGCGTCGTAGAAC ACC	Used to construct pTA2515
<i>polBintBamR</i>	CTCCgGaTcCCCCTTCGAGGGGAG CGCG	Used to construct pTA2515
<i>pfdxhdrBHindF</i>	TCGGCaaGcTtCCGTGGATAAAAC CCC	Used to construct pTA2518
<i>pfdxhdrBEcoR</i>	ATCAAGgaaTTCTAGAGTTACTCA TCGG	Used to construct pTA2518
<i>pBSF2</i>	TTAAGTTGGGTAACGCCAGGG	Made by TA. Used to screen for presence of <i>hdrB+::p.tnaM3- polB1</i> and to construct pTA2560

Chapter 2: Materials and Methods

<i>polR</i>	GAAGTCGTAGTTCGGCAACG	Made by TA. Used to screen for presence of <i>hdrB+::p.tnaM3-polB1</i>
<i>RBDX1</i>	CGTAATACGACTCACTATAGGGCG	Made by SH. Used to construct pTA2554, pTA2553, pTA2546, and to screen for presence of <i>hdrB+::p.tnaM3-polD2</i>
<i>ptnaM3rev_Not</i>	ACgCGgCcGCGTCATATGCGC	Used to construct pTA2554
<i>polD2USF</i>	CTGGTTTTACGACCAGACGGACCAGG	Used to construct pTA2555
<i>dp2intNotR</i>	GGCGgcCGcCGCCGTTAGCGAG	Used to construct pTA2555
<i>dp2USHindR</i>	ATAaGcttCCGTCACCGCGTCG	Used to construct pTA2560
<i>dp2intR</i>	CGATGCGACGGAAGTACCGGGTTCC	Used to screen for presence of <i>hdrB+::p.tnaM3-polD2</i>
<i>recJ2fwdNde</i>	ATAACcatATGTCCGTGAGCCCCG CCG	Used to construct pTA2478 and pTA2561
<i>recJ2revDSBam</i>	CAGGaTCcCACGCCGGCTCATCG GCGCACCTCCC	Used to construct pTA2478 and pTA2561
<i>recJ2PciI_F</i>	CCTCGAAGAATAACTacATGTCC GTGAGCCCCGC	Made by RL (Lever, 2019). Used to construct pTA2481
<i>recJ2Eco_R</i>	GGGgaaTtCGCAGGGTCGCACGCC GGC	Used to construct pTA2481
<i>tsynBglR</i>	CACGaGATctCGCCGAAAAATGC GATGGTCC	Used to construct pTA2553 and pTA2546
<i>ginSfwdNde</i>	TGACcatATGAACGTGGACGACCT CAGGAGCG	Used to construct pTA2361
<i>ginSrevDSBam</i>	AGAAGGaTCcTTCTCAGTCGAGTC GCTCGGC	Used to construct pTA2361

Chapter 2: Materials and Methods

<i>gintagGBspF</i>	TGACA ^{tc} ATGAACGTGGACGACC TCAGGAGCG	Made by HM. Used to construct pTA2465
<i>gintagGEcoR</i>	AGAAGGG ^{gaa} TTCTCAGTCGAGTC GCTCGG	Used to construct pTA2465
<i>priSintF</i>	GAGATGGAAGAAGACGCCGCGC TCTCG	Made by HM. Used to screen for presence of <i>p.tnaM3-ginS</i> and tagged <i>ginS</i>
<i>ginSintR</i>	TCCGCGGTCTCGATTTCTGTCGGT GAGCCG	Made by HM. Used to screen for presence of <i>p.tnaM3-ginS</i> and tagged <i>ginS</i>
<i>mcmtagGNcoF</i>	CTGTG ^{Cc} ATGGCGCAGGCCCCCC AGAACC	Made by HM. Used to construct pTA2421
<i>mcmtagGEcoR</i>	GCGCA ^{GaAtt} CGAGGACGGCTCAA GTCGCGCG	Used to construct pTA2421
Oligonucleotides for generating and screening tagged gene constructs		
<i>delrecJ3NsiR</i>	GCTTTACAA ^{At} GCATCTCGCGTGC GCGGCC	Made by RL (Lever, 2019). Used to construct pTA2564
<i>recJ1NcoI_F</i>	CGTGAGTACCACACC ^c ATGGACG GACCCGTCCCC	Made by RL (Lever, 2019). Used to construct pTA2387
<i>recJ1Eco_R</i>	GCG ^{gaa} TTCTCGGGGAGAACCCGGT CGACG	Used to construct pTA2385
<i>recJ2PciI_F</i>	CCTCGAAGAATAACT ^{ac} ATGTCC GTGAGCCCCGC	Made by RL (Lever, 2019). Used to construct pTA2385
<i>recJ2Eco_R</i>	GGG ^{gaa} TtCGCAGGGTCGCACGCC GGC	Used to construct pTA2385
<i>recJ3BspHI_F</i>	CCTAGCGGG ^{Atc} ATGAGCGACGA GCACGCCGGGG	Made by RL (Lever, 2019). Used to construct pTA2388
<i>recJ3Bam_R</i>	TCG ^{GAtc} CTGAGAGTCGAACGCCG GCTTACG	Used to construct pTA2388
<i>recJ4NcoI_F</i>	GCTCAACG ^{Cc} ATGGATTGGATTA CGCACGAGGAGG	Made by RL (Lever, 2019). Used to construct pTA2386

Chapter 2: Materials and Methods

<i>recJ4Bam_R</i>	AGGA tc CTGGATTCTGGCTTAGAAC TGCTCG	Used to construct pTA2386
<i>RecJ1_F_BspHI</i>	CACAC tc AT Ga ACGGACCCG	Used to construct pTA2586, pTA2587 and pTA2589
<i>RecJ1_R_Kpn</i>	CGGT acc CGATTAGTCCGCG	Used to construct pTA2586 and pTA2588
<i>RecJ1_R_Blp</i>	CGGT gctCagc GAGTCCGCGTTTTTC AGCC	Used to construct pTA2587
<i>RecJ1_F_Bam</i>	TACC ggAtCc ATGGACGGACCCGT CC	Used to construct pTA2588
<i>RecJ1_R_Bam</i>	GGTCG ggatcc AGTCCGCGTTTTCA GC	Used to construct pTA2589
<i>RecJ2_F_BspHI</i>	ATAACT tc AT Ga CCGTGAGCC	Used to construct pTA2592, pTA2593 and pTA2594
<i>RecJ2_R_Kpn</i>	CCGCAG GtaCc CACGCCGGCTCAT CG	Used to construct pTA2592 and pTA2617
<i>RecJ2_R_Blp</i>	ACG Ctcagc GATCGGCGCACCTCCC	Used to construct pTA2593
<i>RecJ2_F_Bam</i>	AGAAT ggaTcc ATGTCCGTGAGCC	Used to construct pTA2617
<i>RecJ2_R_Bam</i>	GCACGC gGatc CATCGGCGCACC	Used to construct pTA2594
<i>RecJ3_F_BspHI</i>	GGGAT tc AT Ga GCGACGAGCACGC C	Used to construct pTA2595, pTA2618 and pTA2620
<i>RecJ3_R_Kpn</i>	AGAGT gGtACc CCGGCTTACGCC	Used to construct pTA2595 and pTA2619
<i>RecJ3_R_Blp</i>	AACG Ctcagc CACGCCGTCTCGAC AGC	Used to construct pTA2618
<i>RecJ3_F_Bam</i>	AGCG GatCc ATGAGCGACGAGCA CG	Used to construct pTA2619
<i>RecJ3_R_Bam</i>	AACGC gGatcc ACGCCGTCTCGA CAGC	Used to construct pTA2620
<i>RecJ4_F_BspHI</i>	CAACG tc AT Ga ATTGGATTACGCA CG	Used to construct pTA2601, pTA2602 and pTA2604

Chapter 2: Materials and Methods

<i>RecJ4_R_Kpn</i>	TCTGGtacCGGCTTAGAACTGC	Used to construct pTA2601 and pTA2603
<i>RecJ4_R_Blp</i>	TGGAgctcagcGAGAACTGCTCGGC GG	Used to construct pTA2602
<i>RecJ4_F_Bam</i>	TGCTCggatCcATGGATTGGATTAC GC	Used to construct pTA2603
<i>RecJ4_R_Bam</i>	GATTgGatccAGAACTGCTCGGCG GC	Used to construct pTA2604
<i>fdx-prom-F</i>	GAAGCCGAACTCTGCAGTGATG	Made by PP. Used to screen split GFP candidates.
<i>recJ2intR</i>	CGGCGAAGCCGCGGACGGTTTCG AACGGCC	Used to screen RecJ2 split GFP candidates.
<i>recJ4intR</i>	GCTCACGCCGGCGGGCTTCGTCG TCTTGCC	Used to screen RecJ4 split GFP candidates.
<i>gintagGBspF</i>	TGACAtcATGAACGTGGACGACC TCAGGAGCG	Made by HM. Used to construct pTA2260
<i>gintagGEcoR</i>	AGAAGGGgaaTTCTCAGTCGAGTC GCTCGG	Used to construct pTA2260
Oligonucleotides for generating deletion constructs		
<i>polBUSKpnF</i>	GGGGTAcCCCCAGCGGGTTTCCG GGTCC	Used to construct pTA2291
<i>polBUSClaR</i>	GCAATCGaTTCGGCGGCGCGGCT AAAAACCC	Used to construct pTA2291
<i>polBDSBamF</i>	CCGGatccATCACCGAGTAATGAA ACTATATTCG	Used to construct pTA2294
<i>polBDSXbaR</i>	ATGtCtAGaGCGGCGCGCTCGTTC GGC	Used to construct pTA2294
<i>MluLinkF</i>	CGCGGCATATGGGATCCCC	Used to construct pTA2217
<i>MluLinkR</i>	CGCGGGGATCCCATATGC	Used to construct pTA2217
<i>priSKpn_F</i>	CACGGtACcACGACCCAGCGCGT CCTCCGAACCG	Used to construct pTA2432 and pTA2433
<i>ginSintNde_R</i>	CGCGcAtaTGCTGTAGGCTGTCTT TCTaTCGCTCC	Used to construct pTA2432
<i>ginSint2Nde_R</i>	TCGTcatAtGACTaGCGGAGGTGCT GTAGG	Used to construct pTA2433
Oligonucleotides for screening deletion candidates		

Chapter 2: Materials and Methods

<i>rps15intF</i>	AGAAGGTCACCACCATCCTCGAG GAGAACG	
<i>HvoLeuB_R</i>	GAAGGAGTCCACCGCGGTCCG	Made by AD
Oligonucleotides for probes		
<i>mrrF</i>	TGGGCGTTCAGGCGAAGC	Made by TA
<i>mrrR</i>	CGGGTGAGCGACCAGCGG	Made by TA
<i>recJ1F</i>	TTCTTTCACCCACTGGAGGC	Made by RT
<i>recJ1R</i>	CTCCGGTTCGGTCTCAACG	Made by RT
<i>recJ2probeF</i>	GCACGACACCGAAGGAAGACAC CGACC	Made by RL (Lever, 2019)
<i>recJ2probeR</i>	GGCGAGGACAGGTCACAGGAGA GCG	Made by RL (Lever, 2019)
<i>recJ3probeF</i>	CGCCTATCTCCAGTCTCTCGGTC GGC	Made by RL (Lever, 2019)
<i>recJ3probeR</i>	GCTCGGCAGGCGCTTGAACAGGT GTCG	Made by RL (Lever, 2019)
<i>recJ4probeF</i>	CGAGTACCACTACTTCACCCGTC GCCCCG	Made by RL (Lever, 2019)
<i>recJ4probeR</i>	CCTGATACGACTGGTAGTAGGCT TCC	Made by RL (Lever, 2019)
<i>cas6F</i>	TCAGTCACTCGCCCGTGGGAAGCG TTTTGTCCG	
<i>cas6R</i>	ATAGAATTAGCGCTCGATGCCGT TGCTGATGC	

Restriction sites engineered into the oligonucleotide are marked in red. Mismatches against the template are marked as lower-case letters. TA = Thorsten Allers, HM = Hannah Marriott, SH = Sam Haldenby, PP = Patricia Perez, AD = Ambika Dattani, RT = Richard Thompson

2.1.4 Chemicals and Enzymes

All enzymes were purchased from New England Biolabs (NEB) and all chemicals from Sigma, unless otherwise stated. Enzymes were used following the manufacturer's instructions.

2.1.5 Media and Solutions

Haloferax Media

Media are sterilised using an autoclave for 1 minute at 121°C. Liquid media are stored at room temperature in the dark, while solid media plates are stored at 4°C in sealed bags to prevent desiccation. Plates are dried for at least 30 minutes before use.

30% salt water (SW):

4 M NaCl, 148 mM MgCl₂·6H₂O, 122 mM MgSO₄·7H₂O, 94 mM KCl, 20 mM Tris.HCl pH7.5.

18% salt water (SW):

Made with 30% SW, 3 mM CaCl₂.

CaCl₂ added after autoclaving.

Trace elements:

1.82 mM MnCl₂·4H₂O, 1.53 mM ZnSO₄·7H₂O, 8.3 mM FeSO₄·7H₂O, 200 µM CuSO₄·5H₂O.

Filter sterilised and stored at 4°C.

Hv-Min Salts:

0.4 M NH₄Cl, 0.25 M CaCl₂, 8% v/v of trace element solution.

Stored at 4°C.

Hv-Min carbon source:

10% DL-lactic acid Na₂ salt, 8% succinic acid Na₂ salt·6H₂O, 2% glycerol, pH to 7.0 with NaOH.

Filter sterilised.

10 x YPC:

5% yeast extract (Difco), 1% peptone (Oxoid), 1% casamino acids, 17.6 mM KOH.

Not autoclaved, used immediately.

10 x Ca:

5% casamino acids, 17.6 mM KOH.

Not autoclaved, used immediately.

Hv-Ca salts:

362 mM CaCl₂, 8.3% v/v of trace elements, 615 µg/ml thiamine, 77 µg/ml biotin.

Chapter 2: Materials and Methods

KPO₄ Buffer:

308 mM K₂HPO₄, 192 mM KH₂PO₄ pH7.0.

Hv-YPC agar:

1.6% agar (Bacto), 18% SW, 1 x YPC, 3mM CaCl₂.

Microwaved without 10 x YPC to dissolve agar. 10 x YPC added, then autoclaved. CaCl₂ added prior to pouring, once cooled.

Hv-Ca agar:

1.6% agar (Bacto), 18% SW, 1 x Ca, 0.84% v/v of Hv-Ca salts, 0.002% v/v of KPO₄ buffer (pH 7.0).

Microwaved without 10 x Ca to dissolve agar. 10 x Ca added, then autoclaved. Hv-Ca salts and KPO₄ buffer added prior to pouring, once cooled.

Hv-Min agar:

1.6% Agar (Bacto), 18% SW, 30 mM Tris·HCl pH 7.5, 2.5% Hv- Min carbon source, 1.2% Hv-Min Salts, 0.002% v/v of KPO₄ buffer (pH 7.0), 444 nM biotin, 2.5 μM thiamine.

Microwaved to dissolve agar. Tris·HCl pH 7.5 added, then autoclaved. Hv-Min carbon source, Hv-Min Salts, KPO₄ buffer, biotin and thiamine added prior to pouring, once cooled.

Hv-YPC broth:

18% SW, 1 x YPC, 3 mM CaCl₂.

CaCl₂ added after autoclaving, when cool.

Hv-Ca+ broth:

18% SW, 30 mM Tris.HCl pH 7.0, 1 x Ca, 2.5% v/v of Hv-Min carbon source, 1.2% v/v of Hv-Min Salts, 0.002% v/v of KPO₄ buffer (pH 7.0), 444 nM biotin, 2.5 μM thiamine.

30% SW, dH₂O and Tris.HCl pH 7.0 autoclaved. All other components added when cool.

Hv-Min broth:

18% SW, 30 mM Tris·HCl pH 7.5, 2.5% Hv- Min carbon source, 1.2% Hv- Min Salts, 0.002% v/v of KPO₄ buffer (pH 7.0), 444 nM biotin, 2.5 μM thiamine.

30% SW, dH₂O and Tris.HCl pH 7.0 autoclaved. All other components added when cool.

Haloferax volcanii Media Supplements

All solutions are filter sterilised using a 0.2 μm syringe filter (Sartorius).

Table 2.6: Media supplements for use with *Haloferax volcanii*

Supplement	Abbreviation	Final concentration
Tryptophan	Trp	50 µg/ml
Thymidine	Thy	50 µg/ml (+ 50 µg/ml hypoxanthine in Hv-Ca and Hv-Min)
Uracil	Ura	50 µg/ml
5-Fluoroorotic acid	5-FOA	50 µg/ml (+10 µg/ml uracil)
Leucine	Leu	50 µg/ml
Mevinolin	Mev	6 µg/ml
Novobiocin	Nov	0.2 µg/ml

Table 2.7: Growth of auxotrophic mutants on different *Haloferax* media.

+ indicates growth and – indicates no growth (additional supplement required).

**ΔhdrB* strains require supplementation with hypoxanthine in Hv-Ca and hypoxanthine, methionine, glycine and pantothenic acid in Hv-Min.

Genotype	Hv-YPC	Hv-Ca	Hv-Min
<i>ΔpyrE2</i>	+	-	-
<i>ΔleuB</i>	+	+	-
<i>ΔtrpA</i>	+	-	-
<i>ΔhdrB</i>	-	-*	-*

Escherichia coli Media

Sterilised by autoclaving at 121°C for 15 minutes and stored at room temperature.

LB (Lysogeny Broth):

1% tryptone (Bacto), 0.5% yeast extract (Difco), 170 mM NaCl, 2 mM NaOH, pH 7.0.

LB agar:

300 ml of LB broth, 1.5% agar (Bacto).

Escherichia coli Media Supplements

Table 2.8: Media supplements for use with *Escherichia coli*

Supplement	Abbreviation	Final concentration
Ampicillin	Amp	50 µg/ml
Tetracycline	Tet	3.5 µg/ml
5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside	X-gal	40 µg/ml

Chapter 2: Materials and Methods

Other Solutions

TE: 10 mM Tris.HCl pH 8.0, 1 mM EDTA.

Sodium Acetate: 3 M NaAc pH 5.2

Filter sterilised.

2.2 Methods

2.2.1 General *Escherichia coli* Microbiology

Growth and Storage of *Escherichia coli*

Cultures of *E. coli* grown on solid media were incubated overnight in a static incubator (LEEC) at 37°C. Small-scale liquid cultures (1-10 ml) were grown overnight in the same static incubator with 8 rpm rotation. Large-scale cultures (300 ml) were incubated overnight in an Innova 4330 floor-standing shaking incubator (New Brunswick Scientific) at 37°C with 150 rpm shaking. For short-term storage all cultures were stored at 4°C. For long-term storage, 20% (v/v) glycerol was added to cultures (from 80% glycerol stock), mixed and flash frozen using dry ice. Frozen stocks were then stored at -80°C.

Preparation of Electrocompetent Cells

Two strains of *E. coli*, XL-1 Blue (*dam*⁺, tetracycline resistant) and N2338 (*dam*⁻), were used to prepare electrocompetent *E. coli* cells.

A 5 ml culture was grown overnight with appropriate antibiotic selection at 37°C with 8 rpm rotation. Cells were diluted 1/100 in LB broth supplemented with appropriate antibiotics. These were grown at 37°C to A₆₅₀ = 0.5-0.8. Cells were pelleted at 6000 x g for 12 minutes at 4°C. The supernatant was removed, and the pellet resuspended in an equal volume of ice-cold sterile 1 mM HEPES (pH 7.5). This process was repeated using two thirds volume 1 mM HEPES (pH 7.5), one third volume 1 mM HEPES (pH 7.5), 0.1 volume 1 mM HEPES (pH 7.5) + 10% glycerol and finally 0.001 volume 1 mM HEPES (pH 7.5) + 10% glycerol. Cells were aliquoted into 100 µl aliquots, snap frozen on dry ice and stored at -80°C.

Transformation of *Escherichia coli* by Electroporation

Buffers and Solutions:

SOC Broth: 2% tryptone (Bacto), 0.5% yeast extract (Difco), 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose.

1-2 µg of DNA in 4 µl of sterile dH₂O was added to 40 µl of electrocompetent cells on ice. The DNA and cells were gently mixed and transferred to a pre-chilled sterile electroporation cuvette (1 mm electrode gap, GENEFLOW). The cuvette was placed in an *E. coli* gene pulser (BioRad) and pulsed at 1.8 kV. 1 ml of SOC was immediately added and samples were incubated at 37°C with 8 rpm rotation for 1 hour, to allow for recovery of the cells. Cells were plated onto LB+2xAmpl plates and incubated at 37°C overnight.

2.2.2 General *Haloferax volcanii* Microbiology

Growth and Storage of *Haloferax volcanii*

Cultures of *Haloferax volcanii* grown on solid media were incubated for 4-7 days in a static incubator (LEEC) at 45°C in a plastic bag to prevent drying. Small-scale liquid cultures (1-10 ml) were grown overnight in the same static incubator with 8 rpm rotation. Large-scale cultures (>50 ml) were incubated overnight in an Innova 4330 floor-standing shaking incubator (New Brunswick Scientific) at 45°C with 120 rpm shaking. For short-term storage, plates and cultures were stored at room temperature. For long-term storage, 20% (v/v) glycerol was added to cultures (from 80% glycerol 6% salt water stock), mixed and flash frozen using dry ice. Frozen stocks were then stored at -80°C.

Transformation of *Haloferax volcanii* using PEG600

Haloferax volcanii can be efficiently transformed using PEG600 (REF Cline 1989). *Haloferax volcanii* encodes a restriction endonuclease, Mrr (encoded by HVO_0682), that targets CTAG-methylated DNA motifs for degradation as a defence mechanism. As such, plasmid DNA must be passaged through a *dam*- strain of *Escherichia coli* to remove methylation prior to transformation (Holmes *et al.*, 1991) or strains must be deleted for *mrr* to allow transformation with methylated (*dam*+) plasmid DNA.

Buffers and Solutions:

All solutions are filter sterilised using a 0.2 µm syringe filter (Sartorius). Unless stated otherwise, all centrifuge spins were at 3300 x g, 25°C in a swing-bucket rotor.

Buffered Spheroplasting Solution: 1 M NaCl, 27 mM KCl, 50 mM Tris.HCl pH 8.5, 15% sucrose.

Unbuffered Spheroplasting Solution: 1 M NaCl, 27 mM KCl, 15% sucrose, pH 7.5.

Transforming DNA: 5 μ l 0.5 M EDTA, pH 8.0, 15 μ l unbuffered spheroplasting solution, 10 μ l DNA (~1-2 μ g).

60% PEG 600: 150 μ l PEG 600 and 100 μ l unbuffered spheroplasting solution.

Spheroplast Dilution Solution: 23% SW, 15% sucrose, 37.5 mM CaCl₂.

Regeneration Solution: 18% SW, 1 \times YPC, 15% sucrose, 30 mM CaCl₂.

Transformation Dilution Solution: 18% SW, 15% sucrose, 30 mM CaCl₂

5-10 ml of YPC (+ thy if required) was inoculated with 1-4 colonies and incubated for ~16 hours at 45°C with 8 rpm rotation. When the A₆₅₀ = 0.6-0.8, cells were pelleted by centrifugation in a 15 ml round-bottomed tube. The supernatant was removed, and the pellet was gently resuspended in 2 ml buffered spheroplasting solution. Cells were transferred to 2 ml round-bottomed tube, pelleted again, and the supernatant was removed. Cells were resuspended in 400-800 μ l buffered spheroplasting solution. 200 μ l of this suspension was transferred to a fresh 2 ml tube per transformation. 20 μ l of 0.5 M EDTA (pH 8) was added to the side of the tube, gently inverted and incubated at room temperature for 10 minutes, facilitating removal of the cells' S-layer. DNA for transformation was added in the same manner as EDTA and incubated for a further 5 minutes at room temperature. 250 μ l of 60% polyethylene glycol 600 (PEG600) was added to the side of the tube and mixed by gentle rocking, before incubation at room temperature for 30 minutes. 1.5 ml of spheroplast dilution solution was added and mixed by gentle inversion. Following a two-minute incubation at room temperature, cells were pelleted by centrifugation. The pellet was then transferred whole to a sterile 4 ml tube containing 1 ml regeneration solution (+ 40 μ g/ml thy if required). To allow recovery, cells were incubated statically at 45°C for 90 minutes. Cells were then resuspended by tapping the tube and incubated at 45°C with 8 rpm rotation for 3-4 hours. Cells were transferred to a fresh 2 ml round-bottomed tube and pelleted by centrifugation. The cell pellet was gently resuspended in transformation dilution solution. Appropriate dilutions were made and 100 μ l of chosen dilutions were plated on suitable selective media. Plates were incubated for at least 5 days at 45°C.

2.2.3 DNA Extraction

Plasmid Extraction from *Escherichia coli*

Plasmid DNA extraction from *E. coli* was performed using Macherey-Nagel NucleoSpin Plasmid (Mini) and NucleoBond Xtra (Midi) kits. Protocol was followed as described in the manufacturer's guidelines. For minipreps 2 ml *E. coli* cell culture (LB broth +Amp) was used and eluted in 30 μ l elution buffer. For midipreps 300 ml *E. coli* cell culture (LB broth +Amp) was used.

Midipreps were eluted using isopropanol. The DNA was ethanol precipitated, resuspended in 200 μ l of TE, and stored at -20°C.

Genomic DNA Extraction from *Haloferax volcanii*

Buffers and Solutions:

ST Buffer: 1 M NaCl, 20 mM Tris.HCl pH 7.5.

Lysis Buffer: 100 mM EDTA pH 8.0, 0.2% SDS.

Genomic DNA Extraction for Southern Blotting

A 5 ml Hv-YPC (+thy if required) liquid culture of *Haloferax volcanii* was grown at 45°C to $A_{650} = 0.6-0.8$. 1 ml of culture was transferred into a 2 ml round-bottomed tube and pelleted at 3300 $\times g$ for 5 minutes at 25°C. The pellet was resuspended in 200 μ l of ST buffer followed by addition of 200 μ l of lysis solution. The tube was mixed by inversion and overlaid with 1 ml of 100% EtOH. DNA was spooled at the interface onto a capillary tip until the liquid was homogenous and clear. The spool of DNA was washed twice in 1 ml of 100% EtOH, and excess EtOH was allowed to drain from the DNA. The DNA was resuspended in 450 μ l of TE and precipitated with NaAc and 100% EtOH. Following this, the pellet was resuspended in 100 μ l of TE and stored at 4°C for up to 6 months.

High-Quality Genomic DNA Extraction for Genome Sequencing

A 5 ml Hv-YPC (+thy if required) liquid culture of *Haloferax volcanii* was grown at 45°C to $A_{650} \sim 1$. This culture was used to inoculate 1 L Hv-YPC and grown for approximately 16 h in a FerMac 360 controlled Bioreactor (Electrolab) at 45°C with agitation to an A_{650} of $\sim 0.05-0.1$. The cells were pelleted at 3300 $\times g$ for 8 minutes at 25°C. The cells were resuspended in 5 ml ST buffer and transferred to a 50 ml falcon tube. 5 ml lysis solution was added, inverted and incubated at RT for 5 minutes. This mix was then overlaid with 25 ml 100% EtOH and DNA was spooled at the interface using a glass pipette until the liquid was homogenous and clear. The spool of DNA was washed in 40 ml of 100% EtOH, and excess EtOH was allowed to drain from the DNA. The DNA was resuspended in 15 ml of TE and precipitated with NaAc and isopropanol. The pellet was then resuspended in 4 ml TE and was treated with 10 mg/ml RNase A (ThermoFisher) for 1 hour at 37°C. The solution was subsequently treated with 0.1 mg/ml proteinase K (Sigma) overnight at 45°C with agitation.

This solution was then overlaid with the same volume of phenol:chloroform:isoamyl alcohol (25:24:1 mix) (ThermoFisher) and was agitated for 10 minutes to allow complete mixing. The mix was then centrifuged for 5 minutes at 15000 $\times g$ at 4°C. The top interface was carefully

removed and transferred to a fresh 50 ml falcon tube. Both the phenol and the phenol:chloroform:isoamyl alcohol steps were repeated where samples appeared to be contaminated with lower phases. The DNA solution was then precipitated using NaAc and 100% EtOH, before being resuspended in 10 mM Tris.HCl pH 8.0.

2.2.4 Nucleic Acid Manipulation

PCR Amplification

Amplification of DNA was carried out by Q5 HotStart or OneTaq HotStart polymerases (NEB). These enzymes are suitable for use with templates containing a high percentage of GC in the template. Q5 HotStart was the enzyme of choice for high-fidelity reactions, while OneTaq HotStart was used for diagnostic amplifications. Reaction conditions are listed below. All reactions were carried out using a Techne TC-512 Thermocycler.

Table 2.9: PCR Components

Component	Q5 HotStart	OneTaq HotStart
dNTPs	200 μ M of each dNTP	200 μ M of each dNTP
Primers	0.5 μ M of each primer	0.5 μ M of each primer
Template DNA	1 ng – 1 μ g genomic DNA 1 pg – 1 ng plasmid DNA	10 ng template DNA
Buffer	1x Q5 Reaction Buffer	1x OneTaq GC Buffer
Enzyme	0.02 U/ μ l Q5 HotStart Polymerase	0.025 U/ μ l OneTaq HotStart Polymerase

Table 2.10: PCR Conditions

Step	Q5 HotStart	OneTaq HotStart	
Initial Denaturation	98°C, 30 seconds	94°C, 30 seconds	
Denaturation	98°C, 10 seconds	94°C, 30 seconds	30 cycles
Annealing	T _m °C, 20 seconds	T _m °C, 30 seconds	
Extension	72°C, 30 seconds/kb	68°C, 60 seconds/kb	
Final Extension	72°C, 5 minutes	68°C, 5 minutes	

Annealing temperatures for primers (T_m°C) were calculated using the equation below (**Equation 2.1**) (Howley *et al.*, 1979).

Equation 2.1: Calculating annealing temperature of primers. %GC = percentage guanine and cytosine in the primer, Homology = percentage homology shared between primer and template, Length = length of primer in bases.

$$81.5 + (16.6 \times \log_{10} [\text{Na}^+]) + (0.41 \times \%GC) - (100 - \%homology) - \frac{600}{\text{length}}$$

Touchdown PCR

Where primers do not have 100% homology to the template sequence (e.g. when introducing restriction sites or mutations), the annealing temperature is altered to allow for this. Two alternate annealing temperatures were calculated; T_{mS} was based on the original homology between the template and the primer, while T_{mE} was based on 100% homology of the primer to the template. The reaction started with T_{mS} and increased linearly over 10 cycles to T_{mE} . The remaining 20 cycles used T_{mE} as the annealing temperature.

Colony PCR

In order to screen large numbers of colonies for a desired plasmid or chromosomal gene, colony PCRs were used. *H. volcanii* colonies growing on solid media were touched gently with a sterile yellow tip, ensuring only a small number of cells were picked up and the colony was not disturbed. The yellow tip was used to pipette up and down in 100 μl of dH₂O. This was boiled at 100°C to lyse cells then cooled on ice. 1 μl of this was then used in a PCR reaction with OneTaq HotStart Polymerase. *E. coli* colonies were gently touched with a sterile yellow tip and those cells were used to directly inoculate the PCR reaction, without the prerequisite of boiling.

Restriction Digests

Restriction digests were carried out following manufacturer's instructions (NEB). All digests were supplemented with 200 ng/ μl BSA (NEB). For double digests NEB buffers were selected so that each enzyme had at least 75% activity. Plasmid DNA was digested for at least one hour. Genomic DNA and PCR products were digested for 16 hours.

Blunt-end filling with Klenow

Should overhangs generated by restriction digests have been required to be blunt-ended, the ends were filled in using Klenow (NEB). Samples were incubated with 1 unit of Klenow per μg of DNA, 1 mM dNTPs and 1x NEB buffer 4 for 30 minutes at 25°C. The reaction was stopped by heat inactivation at 75°C for 20 minutes.

Dephosphorylation of vector DNA

To prevent self-ligation of vector DNA, Shrimp alkaline phosphatase (NEB) was used to remove 5' phosphate groups. Samples were incubated with

Chapter 2: Materials and Methods

5 units of Shrimp alkaline phosphatase per μg of DNA and 1x Antarctic phosphatase buffer (or CutSmart Buffer, commonly used in restriction digests) for 30 minutes at 37°C. Phosphatase was heat inactivated at 65°C for 10 minutes.

Ligation of DNA

Ligations were performed using T4 DNA Ligase (NEB). For each μg of DNA, 5 units of ligase were used in a reaction with 1 \times T4 Ligase buffer. For vector:insert ligations, reactions contained a molar ratio of ~3:1 insert to vector DNA. Ligations were carried out at 15°C overnight or 4°C for 36 hours, followed by ethanol precipitation and resuspension in 5 μl dH₂O. This DNA was then used for transformation into *E. coli*.

Ethanol precipitation of DNA

To ethanol precipitate DNA, 2 volumes of 100% EtOH and 1/10 volume of 3 M sodium acetate (pH 5.2) were added to DNA and incubated at -20°C for at least 1 hour. Samples were centrifuged at 20,000 \times g for 30 minutes at 4°C and the supernatant removed. Pellets were washed in 400 μl 70% EtOH followed by centrifugation at 20,000 \times g for 10 minutes at 4°C. The supernatant was removed, and the pellets air-dried thoroughly before resuspension in sterile dH₂O.

Nucleic Acid Purification

PCR products, ligations, restriction digests and dephosphorylated DNA products were purified using Macherey-Nagel DNA purification kits. Protocol was followed according to manufacturer's instructions. In these kits, DNA is bound pH-dependently to a silica membrane and is separated from contaminants (such as small oligonucleotides or proteins) by washing with ethanol. DNA was eluted in 30 μl of the provided elution buffer.

Nucleic Acid Quantification

To determine the concentration and purity of plasmid preparations the absorbance at 260 nm and the 260:280 nm absorbance ratio, respectively, were measured by an Epoch 2 spectrometer (BioTek).

DNA Sequencing

All DNA sequencing reactions and analysis were performed by the Biopolymer Synthesis and Analysis Unit, University of Nottingham. Sequencing was carried out using the dideoxy chain termination method (Sanger *et al.*, 1977).

Oligonucleotide Synthesis

Oligonucleotides were synthesised by Eurofins MWG, Germany.

Phosphorylation of Oligonucleotides

Phosphorylation of oligonucleotides was carried out using T4 Polynucleotide Kinase (PNK; NEB). 200 pmol of primer was used in a reaction with 1x T4 PNK Reaction Buffer and 10 units of PNK. Phosphorylation was carried out at 37°C for 1 hour, following which PNK was heat inactivated for 20 minutes at 65°C. Phosphorylated oligonucleotides were stored at -20°C for future use.

Agarose Gel Electrophoresis

Buffers and Solutions:

TBE (Tris/Borate/EDTA): 89 mM Tris.HCl, 89 mM boric acid, 2 mM EDTA.

TAE (Tris/Acetic acid/EDTA): 40 mM Tris.HCl, 20mM acetic acid, 1 mM EDTA.

Gel Loading Dye (5×): 50 mM Tris·HCl, 100 mM EDTA, 15% Ficoll (w/v), 0.25% Bromophenol Blue (w/v), 0.25% Xylene Cyanol FF (w/v).

TBE buffer was used as standard practice for casting and running agarose gels. TAE buffer was used when high quality resolution and/or Southern blotting was required. Agarose gels were cast using agarose powder (SeaKem Lonza) and either TBE or TAE buffer. Gel loading dye was added to the DNA samples to give a final concentration of 1x. All samples and molecular markers, either 1 kb or 100 bp (both NEB), were loaded onto the gel. TBE gels (10 cm) were run at 110 V for ~1 hour. TAE gels (25 cm) were run overnight (16 hours) at 50 V with buffer circulation. For visualisation of bands, gels were stained with ethidium bromide at a final concentration of 0.5 µg/ml or SYBR Safe (Invitrogen) at a 0.5x final concentration. Gels used for DNA extraction were stained with SYBR Safe.

Agarose Gel Extraction and Purification of DNA

To purify DNA from agarose gels without UV exposure, gels were only exposed to SYBR Safe stain. DNA was visualised using a Dark Reader (Clare Chemical Research). DNA was purified using the Macherey-Nagel DNA purification kit following manufacturer's guidelines.

2.2.5 Genetic Manipulation of *Haloferax volcanii*

Plasmid Construction

Generating a Deletion Construct by PCR

Deletion constructs are generated by the insertion of the upstream and downstream regions of the gene of interest into cloning vector pTA131 (or a derivative). pTA131 is a derivative of standard *E. coli* cloning vector, pBluescript II SK+, whereby the *H. volcanii* uracil biosynthesis gene *pyrE2* (encoding orotate phosphoribosyl transferase) has been inserted (Allers *et al.*, 2004). The schematic for generating a deletion construct is shown in **Figure 2.1**.

Upstream and downstream regions were generated by PCR in two separate reactions: one to generate the upstream region (US) and one for the downstream region (DS). These PCRs were performed against the genomic clone of the gene of interest (GOI) to be deleted. External primers were designed to incorporate specified novel restriction sites within the product, specifically one of those compatible with the multiple cloning site (MCS) of vector pTA131 or derivative. Internal primers were designed with a *Bam*HI site, giving a *Bam*HI site at the site of gene deletion. This gives a product of the US and DS regions flanking a *Bam*HI site, which allows ease of downstream cloning when adding a marker flanked by *Bam*HI (e.g. *trpA* from pTA298). PCR products were cut with *Bam*HI and the newly introduced external restriction site, as appropriate, and ligated into compatible sites within pTA131 or derivative. Plasmids were transformed into *E. coli* XL-1 Blue cells and plated onto LB+Amp. Selected colonies were grown and screened by diagnostic restriction digest to check for the presence and correct orientation of the insert. Once confirmed by digest, DNA was sequenced to check for the absence of any point mutations.

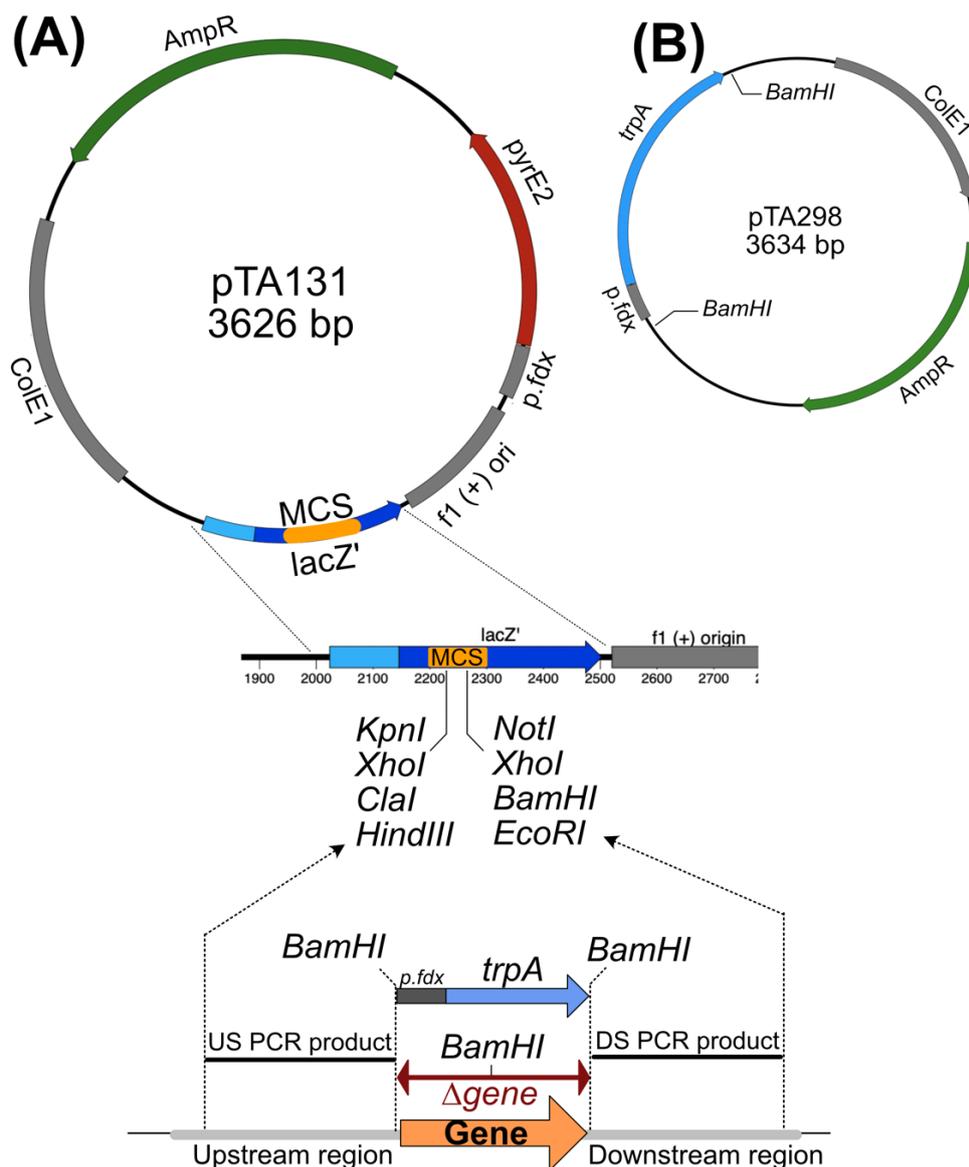


Figure 2.1: Schematic for generating a gene deletion construct by PCR. (A) Upstream (US) and downstream (DS) regions surrounding the gene for deletion are amplified by PCR. Primers will introduce novel restriction sites compatible with the multiple cloning site (MCS) in the *lacZ* gene of pTA131. *BamHI* sites within the internal primers will ligate the US and DS sequences, which will then be cloned into pTA131 (or derivative). (B) The *trpA* marker can be added to the deletion construct by removing *trpA* (and its promoter) from pTA298 using *BamHI* and inserting it into the internal *BamHI* site between US and DS sequences.

Generating a Gene Replacement Construct

Gene replacement constructs were made by inserting the gene of interest (usually with the addition of an inducible promoter or tags) along with its flanking regions into pTA131 or derivative. The exact protocol for doing

this was different for each gene replacement construct and is detailed in relevant chapters.

Strain Construction

Gene Deletion and Replacement in *Haloferax volcanii*

Gene deletion and replacement constructs were used to transform $\Delta pyrE2$ strains. Transformants are plated on Hv-Ca (+ additives as required) to select for the integration of the *pyrE2*-marked plasmid at the targeted gene locus (pop-in, *pyrE2*⁺). A pop-in colony was picked and restreaked onto the same selective media.

This pop-in was then used to set up the pop-out culture: a 5 ml Hv-YPC (+Thy if required) culture was inoculated with this single pop-in colony and the culture was grown overnight until $A_{650} \approx 1$. This culture was then diluted 1/500 into a fresh 5 ml Hv-YPC (+Thy) culture and the growth and dilution were repeated again. When the third culture reached $A_{650} \approx 1$, the culture was diluted in 18% salt-water and was plated on Hv-Ca +5-FOA (+ additives as required).

The relief in selection for uracil in the subsequent Hv-YPC overnights allows for the integrative plasmid and native gene to be lost by homologous recombination. This loss of *pyrE2* can then be selected for using 5-FOA to select for pop-outs (*pyrE2*⁻). These potential pop-out candidates were restreaked onto selective media and screened for the desired genotype. Depending on the orientation of the pop-out, resulting colonies will be either wild-type or deletion/replacement mutants. The pop-in/pop-out gene replacement method was developed by Bitan-Banin *et al.* (Bitan-Banin *et al.*, 2003) and a schematic can be seen in **Figure 2.2** (A-D). Addition of a selectable marker (e.g. *trpA*) allows for direct selection of deletion mutants (**Figure 2.2E**).

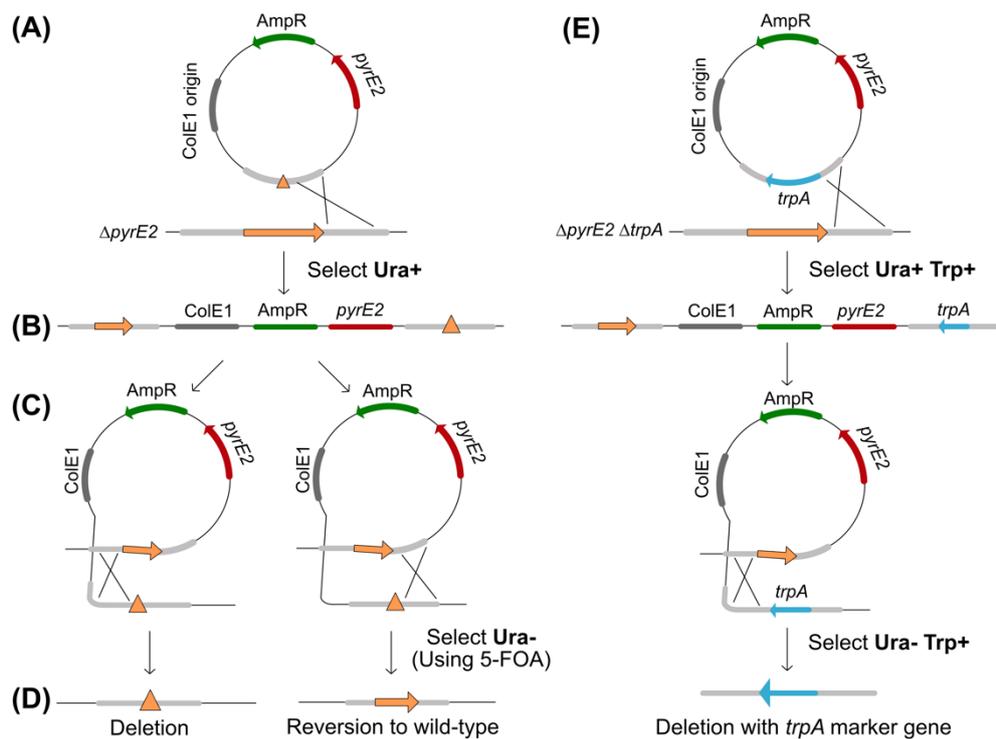


Figure 2.2: Gene deletion utilising the pop-in/pop-out method. (A) A $\Delta pyrE2$ strain is transformed with a *pyrE2*-marked deletion construct. (B) Pop-ins are selected by their ability to grow on media lacking uracil (*ura*⁺ phenotype). (C) Pop-out can be in the upstream (left) or downstream (right) orientation, resulting in the loss of the plasmid backbone (including *pyrE2*). The loss of *pyrE2* in pop-outs is selected for by plating on 5-FOA. (D) The gene is deleted (left) or reverts to wild-type (right). (E) A *trpA* marker can be used in deletion constructs to allow for direct selection of deletion pop-out candidates.

2.2.6 Genotype Screening

Various methods exist for the screening of genotypes in *Haloflex volcanii*. Firstly, if the genotype being screened has a selectable phenotype (e.g., $\Delta trpA$ strains cannot grow in the absence of tryptophan), then candidate colonies can be screened using selective media. However, the polyploid nature of *Haloflex volcanii* means genotypes can be merodiploid within a cell, where some chromosome copies may carry mutant alleles, while others retain wild-type alleles. Therefore, techniques need to ensure screening of all chromosome copies. For this reason, colony hybridisation and Southern blotting are used; these methods both require the denaturing of the total genomic DNA and transfer to a positively charged membrane by either colony lift or vacuum transfer.

Colony Lift

Buffers and solutions:

20×SSPE: 3 M NaCl, 230 mM NaH₂PO₄, 32 mM EDTA, pH 7.4.

Denaturing Solution: 1.5 M NaCl, 0.5 M NaOH.

Neutralising Buffer: 1.5 M NaCl, 0.5 M Tris·HCl, 1 mM EDTA.

Colony lifts allow for screening of large numbers of colonies with relative ease. Candidate colonies were patched onto Hv-YPC (+Thy if required) using sterile wooden toothpicks and incubated at 45°C until growth (~3 days). Patched colonies were lifted from the plate using circles of positively charged Nylon membrane (Amersham Hybond N+) following incubation on the plate for 1 minute. The membrane was transferred, colony side up, to Whatman paper soaked in 10% SDS for >5 minutes to lyse the cells. The membrane was then transferred to Whatman paper soaked in denaturing solution for >5 minutes to denature proteins and DNA. After this the membrane was transferred to Whatman paper soaked in neutralising solution for >5 minutes, which was then repeated. The membrane was then briefly washed for <30 seconds in 2 x SSPE before being thoroughly air-dried. DNA was crosslinked to the membrane with 120 mJ/cm² UV.

Southern Blot Vacuum Transfer

Buffers and Solutions:

20 x SSPE: 3 M NaCl, 230 mM NaH₂PO₄, 32 mM EDTA, pH 7.4.

Denaturing Solution: 1.5 M NaCl, 0.5 M NaOH.

Purified *H. volcanii* genomic DNA was digested with enzymes that cut either side of the region of interest. The resulting DNA fragments were separated on a 200 ml 0.75 % TAE for 16 hours at 50 V, with buffer circulation. The gel was post-stained with Ethidium Bromide (0.5 µg/ml) for 30 minutes and visualised. The gel-embedded DNA was acid-nicked for 20 minutes in 0.25 M HCl, followed by washing for 10 minutes in dH₂O. DNA was then denatured by washing in denaturing solution for >45 minutes. Membrane (BioRad Zeta-Probe GT or Amersham Hybond-XL) was soaked in dH₂O for 5 minutes before equilibrating in denaturing solution for a further 2 minutes. Vacuum transfer was carried out using a Vacugene XL gel blotter and Vacugene Pump (Pharmacia Biotech) for 1 hour 15 minutes at 50 mBar. Following transfer, the membrane was washed briefly in 2 x SSPE for <30 seconds and air-dried before DNA was cross-linked with 120 mJ/cm² UV.

Hybridisation

Buffers and Solutions:

100 x Denhardt's Solution: 2% Ficoll 400, 2% PVP (Polyvinyl Pyrrolidone) 360, 2% BSA (Bovine Serum Albumin, Fraction V).

20 x SSPE: 3 M NaCl, 230 mM NaH₂PO₄, 32 mM EDTA, pH 7.4.

Pre-hybridisation Solution: 6 x SSPE, 1% SDS. 5 x Denhardt's solution, 200 µg/ml salmon sperm DNA (Roche, boiled for 5 minutes at 100°C prior to addition). **Hybridisation Solution:** 6 x SSPE, 1% SDS, 5% dextran sulphate.

Low Stringency Wash Solution: 2 x SSPE, 0.5% SDS.

High Stringency Wash Solution: 0.2 x SSPE, 0.5% SDS.

Membranes from either colony lifts or Southern blots were pre-hybridised for >3 hours at 65°C in 40 ml pre-hybridisation solution. Radiolabelled DNA probes were made with 50 ng of DNA and 0.74 MBq of [α -³²P] dCTP (Perkin Elmer). DNA was denatured at 100°C for 5 minutes then incubated with the radioisotope and HiPrime random priming mix (Roche) for 15-20 minutes at 37°C. The radiolabelled probe was then purified on a BioRad P-30 column and mixed with 10 mg/ml salmon sperm DNA, followed by denaturing at 100°C for 5 minutes and quenching on ice. For Southern blots 3 µl of 1 µg/ml 1 kb ladder or 3 µl of 1 µg/ml bacteriophage lambda DNA for Pulsed-Field Gel Southern blots was also included in the radiolabelling reaction. The pre-hybridisation solution was replaced with 30 ml of hybridisation solution, the probe DNA was added and then the membranes were incubated overnight at 65°C. The membranes were washed twice with 50 ml low stringency wash solution, once for 10 minutes and once for 30 minutes. This was followed by another two washes with high stringency wash solution, both for 30 minutes. Membranes were air-dried before being wrapped in Saran wrap and exposed to a phosphorimager screen (Fujifilm BAS Cassette 2325) for >24 hours. The screen was scanned using a Molecular Dynamics STORM 840 scanner. Alternatively, membranes were visualized using Amersham Hyperfilm MP (GE Healthcare).

2.2.7 Phenotype Screening

Flow cytometry

DNA content and cell size of *H. volcanii* cells were determined via flow cytometry. Cultures were prepared in 5 ml Hv-YPC or Hv-Cas broth and grown at 45°C with 8 rpm rotation in two successive dilutions until an A₆₅₀ of ~0.4 was reached. Acridine orange solution was added to the cells at a final concentration of 1 µg/ml. Samples were analysed using an FC500 flow cytometer (Beckman Coulter; University of Nottingham Flow Cytometry

facility) equipped with 488nm laser and 528/28 emission filter to measure Acridine Orange fluorescence. Samples were run on the lowest speed setting and at least 20,000 cells were acquired for each sample. Data was analysed using Flow Jo v7.6 (Tree Star Inc.). Cells were gated based on forward and side scatter and doublets excluded by height/area analysis.

Tryptophan Gradient Plates

To generate a gradient of tryptophan across a plate, with the desired tryptophan concentration at one side fading to no tryptophan on the other side, plates were first poured with 17 ml Hv-Ca +Ura +Trp (of the desired concentration) on a 7° slant to form a tapered wedge (Figure 2.4). Once set, the plate was placed flat and the wedge was covered with 43 ml Hv-Ca + Ura agar, lacking tryptophan (**Figure 2.3**) (Bryson and Szybalski, 1952, Hawkins *et al.*, 2013a).

5 ml cultures of *H. volcanii* strains were grown with 8 rpm rotation in Hv-Cas (+Trp where required) at 45°C until an A650 of 0.6-0.8. These were then diluted into fresh Hv-Cas and incubated at 45°C until an A650 of 1.0. Serial dilutions of the cultures in 18% SW to 10⁻⁵ were prepared. Autoclaved paintbrushes (The Range) were first wetted in 18% SW. Using a fresh paintbrush for each strain, the paintbrush was dipped into the diluted culture and then painted in one direction across the gradient plate. Using the same paintbrush dipped again in the diluted culture, a second line was painted over the first in the opposite direction. Once dry, the plates were incubated at 45 °C for 5 days.

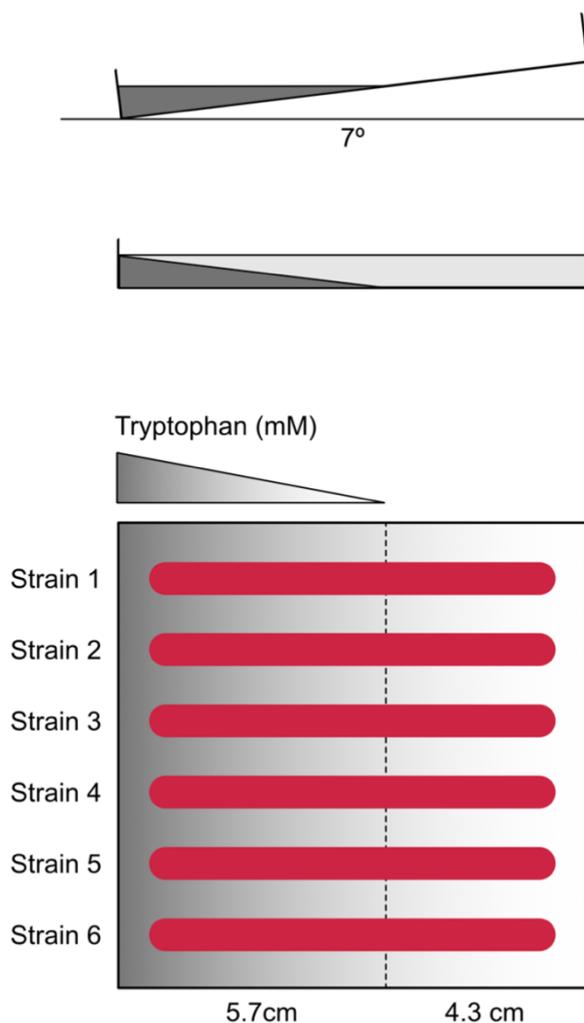


Figure 2.3: Tryptophan gradient plates. Plates were first poured with 17 ml Hv-Ca +Ura +Trp (of the desired concentration) on a 7° slant to form a tapered wedge. Once set, the plate was placed flat and the wedge was covered with 43 ml Hv-Ca + Ura agar, lacking tryptophan. Strains are painted across the tryptophan gradient.

Standard Growth Assay

Standard growth rate in liquid media was determined using an Epoch 2 Microplate Spectrophotometer (BioTek). Cultures were prepared in 5 ml Hv-YPC or Hv-Cas broth and grown to mid-late exponential phase, which corresponds to an A₆₅₀ of 0.4-0.8. These cultures were diluted and again grown to mid exponential phase. Serial dilutions of the cultures were made before loading 150 µl in duplicate, alongside appropriate blanks, to the wells of a 96 well microtiter plate (Corning). Where drug treatment was required, this was added to the media with which dilutions were made. For aphidicolin

(VWR), a 1 mg/ml stock was diluted in DMSO and untreated cells received the same volume of DMSO as a control. For H₂O₂ (Sigma Aldrich), a 10 M stock was diluted in dH₂O and untreated cells received the same volume of dH₂O as a control.

The plate was sealed around the edges with microporous tape (Boots UK Ltd) and incubated at 45°C with double orbital shaking at 1000 rpm for 72 hours in the Epoch. Readings at A₆₀₀ were taken every 15 minutes and converted to a 1 cm pathlength by dividing the raw A₆₀₀ value by 0.14. The generation time was calculated by plotting the growth on a log₂ scale and using the following equation, **Equation 2.2**.

Equation 2.2: Calculating doubling time of strains from growth curves.

$$G = \frac{t}{n}$$

G = generation time
t = time
n = number of generations

$$n = \frac{\log b - \log B}{\log 2}$$

b = end A₆₀₀
B = start A₆₀₀

Generation times varied between experiments therefore comparisons were only made between sets of strains within the same experiment i.e. strains that have been incubated on the same 96 well microtiter plate with the A₆₀₀ being measured simultaneously during a single run. Since generation times vary between experiments, the generation times stated are not absolute however, the relationships between the strains are consistent. Therefore, growth curves generated by this method are used to illustrate the relationship of generation times between a set of strains.

Microscopy

Depending on the aim of the study, cells were inoculated in either 5 ml of Hv-YPC or Hv-Cas (+ supplements) and grown for 16 hours at 45°C with 8 rpm rotation. Cultures were then diluted to reach an A₆₅₀ of ~0.4-0.6 the following day (depending on downstream applications).

Chapter 2: Materials and Methods

Where cells were to be treated with aphidicolin, 20 µg/ml aphidicolin (dissolved in DMSO) was added to cells and incubated for 3 hours. This dose was increased from that used for other aphidicolin experiments (e.g., growth curves) due to time constraints following COVID shift implementation and thus direct comparison between aphidicolin results at different dosages and time points are not applicable.

Following any treatments (if applicable), cells were then spun down at 3300 x g, 25°C, in a swing-bucket rotor and resuspended in 1 ml 18% salt water. DAPI stain was added (final concentration 2.5 µg/ml) and samples were inverted and incubated in the dark for 10 minutes. Cells were then spun down again, as previously, and resuspended in fresh 18% salt water.

10 µl samples were immobilized on 1% agarose pads containing 18% salt water and imaged using a Nikon Ti-E inverted fluorescence microscope equipped with a Plan Apo 100x/1.45 Ph3 objective and Andor Neo sCMOS camera. Acquisition was with the Nikon NIS software. Images were analysed using FIJI software (Schindelin *et al.*, 2012). For each cell preparation, five fields of view were taken, and a representative image was selected.

Recombination assay

Recombination frequency can be measured using an assay quantifying the amount of recombination occurring between the chromosome and a plasmid. This assay requires strains with an edited genetic background, whereby they carry a frame-shift mutation (*leuB-Ag1*) and are unable to grow on media lacking leucine. The plasmid utilised for this assay, pTA163, has been designed to carry a second frame-shift leucine-auxotrophic mutant (*leuB-Aa2*). Should a recombination event occur between this plasmid and the chromosome, the strain will become prototrophic for leucine and can grow in media lacking leucine (Lestini *et al.*, 2010) (**Figure 2.4**).

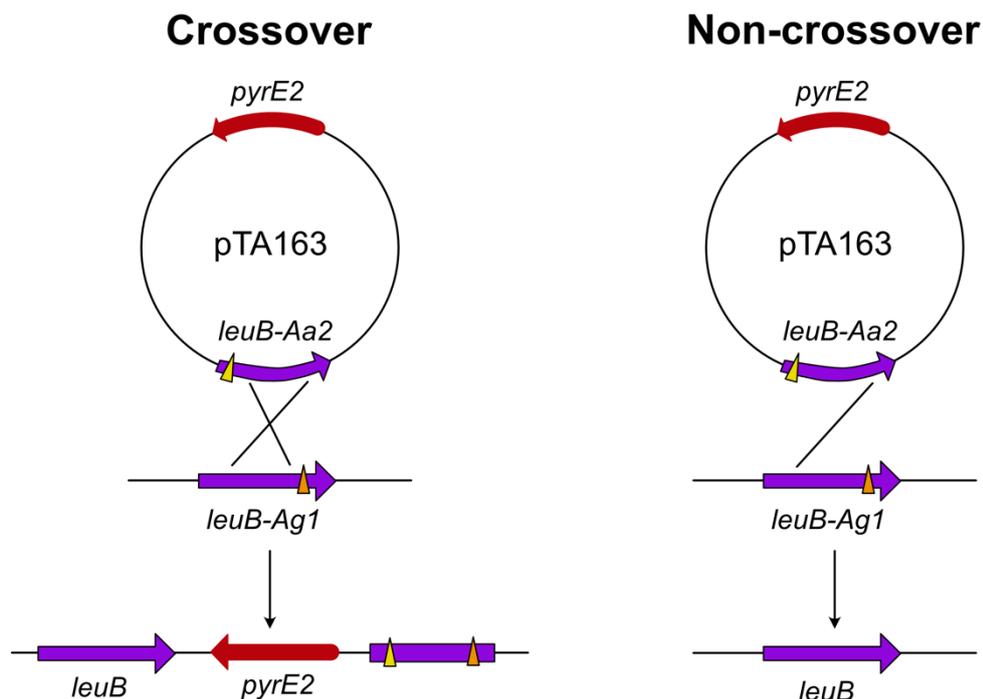


Figure 2.4: Recombination assay. $\Delta pyrE2$ strains with a chromosomal *leuB-Ag1* (*leu*⁻) allele are transformed with pTA163, containing *pyrE2* and *leuB-Aa2* (*leu*⁻) allele. A recombination event between the chromosome and the plasmid mutant *leuB* alleles can generate a wild type *leuB* allele and thus cells can grow in the absence of leucine. Depending on whether a crossover or non-crossover event occurred, strains will either retain (crossover) or lose (non-crossover) the *pyrE2* marker from pTA163.

To perform the assay, strains carrying the *leuB-Ag1* allele were transformed with 1 μg of pTA163 (*pyrE2*⁺ *leuB-Aa2*) according to the standard *Haloferax* transformation protocol. Transformants were plated on Hv-Min +trp +ura at dilutions between 10^0 - 10^{-5} to select for cells that had undergone the recombination event. Transformants were also plated on non-selective media (Hv-YPC) at 10^{-4} - 10^{-6} dilutions to determine the total viable cell count. Reversion of the chromosomal allele to wild type *leuB* is extremely rare and is considered to have no significant effect on the results of this assay (Haldenby, 2007).

Crossover (CO) recombination events (where the *pyrE2*-marked plasmid is integrated onto the chromosome) and non-crossover (NCO) events (conversion to the wild type *leuB* allele without integration of the *pyrE2* gene) were distinguished by patching *leu*⁺ transformants in duplicate on Hv-Min

+trp plates with and without uracil (ura). All colonies would be expected to grow on Hv-Min +trp +ura plates, while only those that have undergone a CO event will grow on Hv-Min +trp plates. The fraction of CO events (*leuB*+*pyrE2*+) is derived by comparison to the total recombination frequency and the remaining fraction are determined to be NCO events (*leuB*+*pyrE2*-).

Transformation efficiencies for all strains tested was also calculated; strains were transformed with 1 µg of episomal plasmid pTA354 (*pyrE2*+) as per the standard *Haloferax* transformation protocol. Transformants were plated on Hv-Ca +trp at 10⁻¹-10⁻³ dilutions to select for cells retaining the plasmid along with on Hv-YPC at 10⁻⁴-10⁻⁶ to determine the total viable cell count. The recombination efficiency for each strain was normalised to its transformation efficiency.

DNA Damage assays

Ultraviolet (UV) irradiation sensitivity

5 ml of Hv-YPC was inoculated with 1 colony and grown overnight at 45 °C with 8 rpm rotation. The culture was diluted into 5 ml of fresh Hv-YPC and grown to an A₆₅₀ of 0.4. A range of serial dilutions (10⁻¹ – 10⁻⁸) of the cells in 18% saltwater were made and duplicate 20 µl samples were spotted onto Hv-YPC agar and allowed to air-dry. Plates were exposed to UV light at 254 nm, 1 J/m²/sec for varying amounts of time and shielded from visible light to prevent photo- reactivation DNA repair. Plates were incubated at 45°C for 4-7 days. During this period, colonies were counted, and survival fractions were calculated relative to a non-irradiated control.

Mitomycin C (MMC) sensitivity

5 ml of Hv-YPC was inoculated with 1 colony and grown overnight at 45°C with 8 rpm rotation. The culture was diluted into 5 ml of fresh Hv-YPC and grown to an A₆₅₀ of 0.4. Cells were then serially diluted (10⁻¹ – 10⁻⁸) in 18% salt water and 20 µl duplicates were spotted onto Hv-YPC agar containing 0 – 0.02 µg/ml MMC. Plates were allowed to air-dry before incubation at 45 °C for 4-7 days. During this period, colonies were counted, and survival fractions were calculated relative to an untreated control.

Phleomycin sensitivity

5 ml of Hv-YPC was inoculated with 1 colony and grown overnight at 45°C with 8 rpm rotation. The culture was diluted into 5 ml of fresh Hv-YPC and grown to an A₆₅₀ of 0.4. Samples were split into 1 ml aliquots, and an

appropriate amount of phleomycin was added (0-2 mg/ml diluted in 18% salt water). Samples were vortexed gently to mix, then incubated for 1 hour at 45°C with 8 rpm rotation, followed by pelleting by centrifugation and resuspension in 1 ml Hv-YPC to remove the mutagen. Samples were serially diluted (10^{-1} – 10^{-8}) in 18% salt water and 20 μ l duplicates were spotted onto Hv-YPC agar. Plates were allowed to air-dry before incubation at 45°C for 4-7 days. During this period, colonies were counted, and survival fractions were calculated relative to an untreated control.

Protein interaction assay

Split-GFP

Split-GFP methodology was recently optimised for use in *H. volcanii* (Winter *et al.*, 2018). Constructs for its use were kindly provided by Felicitas Pfeifer.

Plasmids were generated carrying ^NGFP or ^CGFP fragments fused to the protein of interest. These plasmids carried mevinolin resistance (Mev^R) and novobiocin resistance (Nov^R) markers respectively. Pairs of ^N- and ^C-GFP expression plasmids were transformed into *H. volcanii* strain H26 simultaneously (as per the standard transformation methodology explained earlier). Transformants were plated on Hv-Cas +Ura +Mev +Nov plates and incubated at 45°C for 4-6 days. Candidates were restreaked on the same selective media before being screened for the presence of both episomes by colony PCR.

To screen for interaction (via GFP signal), a single restreaked colony was inoculated in 5 ml Hv-YPC +Mev +Nov media and was grown to $A_{650} \sim 1$ at 37°C for ~ 16 hours with 8 rpm rotation to obtain sufficient cell mass, followed by incubation of the same culture at 30°C with 8 rpm rotation overnight. 2 ml of culture was then spun down at 3300 x g, 25°C, in a swing-bucket rotor. The pellet was then washed in 18% salt water before being pelleted again, as previously. This pellet was then resuspended in 500 μ l of 18% salt water. Duplicate 150 μ l samples were loaded onto a 96-well plate and measured for fluorescence using a fluorescence imager (GE Healthcare Typhoon; excitation wavelength at 488 nm). As controls, untransformed H26 and 18% salt water were also plated and measured. The same plate was then subsequently measured for optical density (A_{600}) using an Epoch 2 spectrometer (BioTek).

The fluorescence was measured using Fiji software (Schindelin *et al.*, 2012) and normalised against optical density measurements and background readings for salt water alone were subtracted. The average of the two technical

replicate samples was calculated. The relative fluorescence (rf) was then calculated using the formula given below (**Equation 2.3**).

Equation 2.3: Calculation of relative fluorescence (rf) for split-GFP strains.

$$\text{rf} = \frac{\text{transformant} - \text{untransformed H26}}{\text{untransformed H26}}$$

For each strain tested, two independent trials were completed, each with two technical replicates. Statistical analysis was completed using GraphPad Prism, with *p*-values generated via a one-way ANOVA with Dunn's pairwise comparison against untransformed H26.

2.2.8 Protein Expression and Purification

Whole Cell Lysate preparation

Buffers and solutions:

10×DNase buffer: 100 mM Tris HCl (pH 7.5), 25 mM MgCl₂, and 5 mM CaCl₂

2×SDS-PAGE loading buffer: 100 mM Tris HCl, pH 6.8, 200 mM β-mercaptoethanol, 4% SDS, 0.2% bromophenol blue, 20% glycerol

To isolate whole cell lysates for analysis by SDS-PAGE, 5 ml of Hv-YPC (or selective media if required) was inoculated with 1 colony and grown overnight at 45°C with 8 rpm rotation. The culture was diluted into 5 ml of fresh Hv-YPC (or selective media) and grown to an A₆₅₀ of ~0.6. Cells were pelleted at 3300 x g for 8 minutes in a 14 ml round-bottomed tube (Sarstedt), and the supernatant was removed. The pellet was resuspended in 80 μl of dH₂O to lyse the cells. To this, 10 μl of 10×DNase buffer and 1 μl DNase I (ThermoFisher) was added and incubated at 37°C with 450 rpm shaking for 30 minutes. Following incubation, 100 μl of 2×SDS-PAGE loading buffer was added. Lysates were boiled for 10 minutes at 94°C and stored at -20°C. 15 μl was used for loading on SDS-PAGE gels.

Protein Purification using Nickel Affinity Chromatography

Buffers and Solutions:

Binding Buffer A: 20 mM HEPES pH 7.5, 2 M NaCl, 1 mM PMSF, imidazole to desired concentration

Chapter 2: Materials and Methods

Primarily, IMAC Sepharose 6 Fast Flow beads (GE Healthcare) were charged with Ni²⁺. Beads (0.5 ml per column) were washed twice with 10 column volumes (CV) of dH₂O, followed by equilibration of the column with 2 CV 0.2M NiSO₄ for 30 minutes at 4°C with rotation. 6 subsequent washes of the beads were performed, twice with 10 CV H₂O, once with 10 CV Buffer A + 500 mM imidazole, and finally 3 washes with 10 CV Buffer A + 20 mM imidazole. Beads were resuspended in 1 CV Buffer A + 20 mM imidazole. The filtered supernatant from the cell lysates was incubated with the Ni²⁺ charged beads for 1 hour at 4°C with rotation. Working at 4°C, the slurry was applied to a Poly-Prep column (Bio-Rad), the flow-through was collected and was subsequently reloaded onto the column. The column was washed with 10 CV Buffer A and then with 20 CV Buffer A, before eluting in 4 CV Buffer A +100 mM imidazole, 4 CV Buffer A +200 mM imidazole, and 4 CV Buffer A +500 mM imidazole.

Issues regarding contamination were very common when using histidine-based purification methods in *H. volcanii*. Therefore, purification using streptavidin was the preferred methodology for isolating proteins of interest in *H. volcanii*.

Protein Purification using Streptavidin

Buffers and Solutions:

Binding Buffer B: 20 mM HEPES pH 7.5, 2 M NaCl, 1 mM PMSF

Streptavidin Affinity Elution Buffer: 20 mM HEPES pH 7.5, 2 M NaCl, 1 mM PMSF, 5 mM D-desthiobiotin (IBA)

A 5 ml starter culture was grown overnight in Hv-YPC at 45°C to an A₆₅₀ of ~1.0, then diluted 10⁻² into fresh Hv-YPC and grown for a further 8 hours to an A₆₅₀ of ~1.0. The culture was diluted into 50 ml Hv-YPC to allow growth to A₆₅₀ of ~0.5 over a period of 24 hours. This culture was used to inoculate 2 L Hv-YPC and grown for approximately 16 h in a FerMac 360 controlled Bioreactor (Electrolab) at 45°C with agitation to an A₆₅₀ of ~0.5. The cells were pelleted at 3300 x g for 10 minutes at 4°C. Pellets were resuspended in 7 ml of buffer B and cells were lysed via sonication (4-8 x 20 seconds at ≤8 μm amplitude) on ice. Cell lysates were then pelleted at 20000 x g for 30 minutes at 4°C, and the supernatant was filtered sequentially through 0.8 μm, 0.45 μm, and 0.2 μm filters to remove DNA.

Working at 4°C, 1 ml of Strep-tactin Sepharose (IBA LifeSciences) was applied to a Poly-Prep gravity column (BioRad) and equilibrated twice with double the column volume (CV) of Buffer B. The protein sample was applied to the equilibrated column and reloaded twice. Following flow-through

of the sample, the column was washed five times with 1xCV Buffer B. The column-bound proteins were then sequentially eluted using 0.8, 1.4 and 0.8 CV of Streptavidin affinity elution buffer, giving samples E1, E2 and E3 respectively.

Elutions were stored at 4°C for the short term (<1 week). For long-term storage, 10% glycerol was added to samples before snap freezing on dry ice and storage at -80°C.

SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Buffers and Solutions:

10% SDS-PAGE gel (resolving): 10% acrylamide/bisacrylamide Protogel (National Diagnostics), 0.37 M Tris pH 8.8 with 0.5% 2,2,2-trichloroethanol (TCE; Sigma T54801), 0.1% SDS, 0.05% AMPS (ammonium persulfate), 0.05% TEMED (tetramethylethylenediamine)

3% SDS-PAGE gel (stacking): 3% acrylamide/bisacrylamide Protogel, 0.25 M Tris pH 6.8, 0.2% SDS, 0.125% AMPS, 0.125% TEMED

10×Tris-Glycine running buffer: 0.25 M Tris, 1.92 M glycine, 1% SDS

2× Laemmli buffer: 4% SDS, 20% glycerol, 10% β-mercaptoethanol, 0.004% bromophenol blue and 0.125 M Tris HCl pH 6.8

Protein samples were analysed using SDS-PAGE. Gels were cast in Novex 1 mm cassettes (Invitrogen). Laemmli buffer was added to protein samples at a 1x concentration and proteins were denatured by boiling at 94°C for 10 minutes and vortexed prior to loading onto the gel. Samples were run on the gel at 150V, 36 mA, for 120 minutes in 1xTris-Glycine running buffer, with either Blue Protein Standard Broad Range (NEB), Dual Colour or Dual Colour Kaleidoscope (BioRad) ladder running alongside samples. For visualisation of proteins, gels were either incubated with PageBlue Protein Staining solution (ThermoScientific) or imaged via exposure to UV to develop the 2,2,2-trichloroethanol (TCE; 1 minute) before being exposed and imaged for 5 seconds using a Fusion FX system (Vilber).

Western Blotting

Buffers and Solutions:

10×TBS: 198 mM Tris base, 936 mM NaCl₂, pH adjusted to 7.6 with 1 M HCl

TBST: 1 x TBS, 0.1% Tween20

Freshly run, unstained SDS-PAGE gels were removed from cassettes and the stacking gel and foot removed. The ladder was removed to prevent overexposure and the gel was imaged via exposure to UV to develop the TCE

(1 minute) before being exposed and imaged for 5 seconds using a Fusion FX system (Vilber). The gel was then reassembled, and proteins were transferred to a PVDF membrane by wet transfer using a BioRad Trans-Blot system at 25 V for 10 minutes. Following transfer, membranes were blocked in 5% milk in TBST (w/v) for 1 hour at room temperature, with agitation. The membrane was then incubated with an anti-6xHis mouse antibody conjugated to HRP (Proteintech HRP-66005) in 5% milk at a 1:10000 dilution at room temperature overnight. The membrane was then washed with TBST for 4x5 minutes, before being incubated with Amersham ECL Prime reagent (GE Healthcare) for 1 minute. Excess ECL reagent was removed and chemiluminescence was detected using a Fusion FX system (Vilber).

Mass Spectrometry

Mass spectrometry was performed by the Cambridge Centre for Proteomics, University of Cambridge. Proteins in gel bands were reduced (DTT), alkylated (iodoacetamide) and subjected to enzymatic digestion with Trypsin. The resulting peptides were loaded onto an autosampler for automated electrospray ionization-tandem mass spectrometry (MS/MS) analysis. All LC-MS/MS experiments were performed using a Dionex Ultimate 3000 RSLC nanoUPLC system and a QExactive Orbitrap mass spectrometer (Thermo Fisher Scientific Inc, Waltham, MA, USA). Peptides were separated on a PepMap C18 reversed-phase, 2 μ m particle size, 100A pore size, 75 μ m-inner-diameter, 50 cm column at a flow rate of 300 nL/min. Protein Discoverer software version 2.1 (Thermo Scientific) was used to generate a peak list file of uninterpreted fragment mass data, which was used to search against the UniProt Archaea database (ref: 20190528, 3149228 sequences; 899013487 residues) and common contaminant sequences (123 sequences; 40594 residues) using the MASCOT search engine. Only protein identifications with probability based MOWSE scores above a threshold of $p < 0.05$ were accepted.

2.2.9 Bioinformatic analysis

Protein domain analysis was performed using Pfam protein family database (El-Gebali *et al.*, 2019). Predicted protein structures were mapped using the intensive model setting for Phyre2 (Kelley *et al.*, 2015).

To carry out alignments of proteins, sequences were obtained from UniProt and alignment was carried out in MacVector using T-Coffee (Myers Miller; penalty for open gap = -50; extend gap = -50). Phylogenetic mapping was performed using MacVector and was calculated using the Neighbour Joining method (Bootstrap 1000 reps).

Chapter 2: Materials and Methods

Synteny analysis was performed using SyntTax Prokaryotic Synteny and Taxonomy Explorer using its best match settings with a minimal score of 10% (Oberto, 2013).

Chapter 3: Cellular requirement for replicative DNA polymerases PolB and PolD in *Haloferax volcanii*

3.1 Background

DNA polymerases

DNA polymerases (DNAPs) are key components of genome replication and are involved in diverse DNA repair mechanisms (Lujan *et al.*, 2016, Loeb and Monnat, 2008). Based on amino acid sequence alone, DNA polymerase enzymes were originally placed into six main families: A, B, C, D, X and Y (Ito and Braithwaite, 1991). Since then, PrimPol and reverse transcriptase (RT) enzymes have also been defined as DNA polymerases of separate families (Rudd *et al.*, 2014, Nakamura and Cech, 1998). **Table 3.1** shows the occurrence of different DNAP families across all domains of life.

Table 3.1: Comparison of the DNA polymerases in the three domains of life. Adapted from (Patel and Loeb, 2001) and (Ishino and Ishino, 2014).

		Bacteria	Eukaryotes	Archaea
Family	A	Pol-I	Pol γ , Pol θ	
	B	Pol-II	Pol α /primase, Pol δ , Pol ϵ , Pol ζ	PolB1, PolB2, PolB3
	C	Pol-III		
	D			PolD
	X		Pol β , Pol λ , Pol μ , Pol σ	PolX
	Y	Pol-IV, Pol-V	Pol η , Pol ι , Pol κ	PolY

All DNAP families share a core structure, shaped like a ‘right hand’ that holds the primer-template in place. This structure can be broken down into subdomains (thumb, fingers and palm domains) which each play their own roles in catalysis (Wu *et al.*, 2014a). DNAPs also carry accessory domains that specify their functions; for example, in the case of replicative DNA polymerases, they carry a 3'-5' exonuclease domain for proofreading to increase fidelity during DNA replication.

Replicative DNA polymerases

Replicative DNAPs are DNA-dependent DNA polymerases that use ssDNA as a template to copy the entire genome. They usually extend synthesis from a short RNA primer, with their 5' to 3' directionality allowing for continuous synthesis of the leading strand and discontinuous synthesis of the lagging strand. Replicative DNAPs are responsible for the correct replication of the genome once per cell cycle. The replicative polymerases used are

divided between Bacteria, Archaea and Eukaryotes, covering DNAPs from Families A, B, C and D (Leipe *et al.*, 1999) (**Table 3.2** and **Figure 3.1**).

Table 3.2: Comparison of DNA polymerases (DNAPs) involved in DNA replication across all domains of life. Adapted from Raia *et al.*, 2019.

Replicative DNAP	Family	Organisms found in	Functions	Exonuclease activity
DnaE	C	Bacteria	Leading and lagging strand synthesis	Proofreading, co-proofreading or inactivated
Pol III	C	Gram positive bacteria	Leading strand synthesis when associated with DnaE	Proofreading, co-proofreading
Pol I	A	Bacteria	Maturation of Okazaki fragments	Proofreading, flap endonuclease activity
PolB1	B	Archaea (<i>Crenarchaea</i> , <i>Thaumarchaeota</i> and <i>Korarchaeota</i>)	Leading and lagging strand synthesis	Proofreading
PolB2	B	Archaea (mainly <i>Euryarchaeota</i> and <i>Crenarchaeota</i>)	Regulation of DNA replication	Inactivated
PolB3	B	Archaea (except <i>Thaumarchaeota</i>)	Leading and/or lagging strand synthesis	Proofreading
PolD	D	Archaea (except <i>Crenarchaea</i>)	Leading and/or lagging strand synthesis	Proofreading
Pol α	B	Eukarya	Initiation of leading and lagging strand synthesis	Inactivated
Pol δ	B	Eukarya	Lagging strand synthesis	Proofreading
Pol ϵ	B	Eukarya	Leading strand synthesis	Proofreading

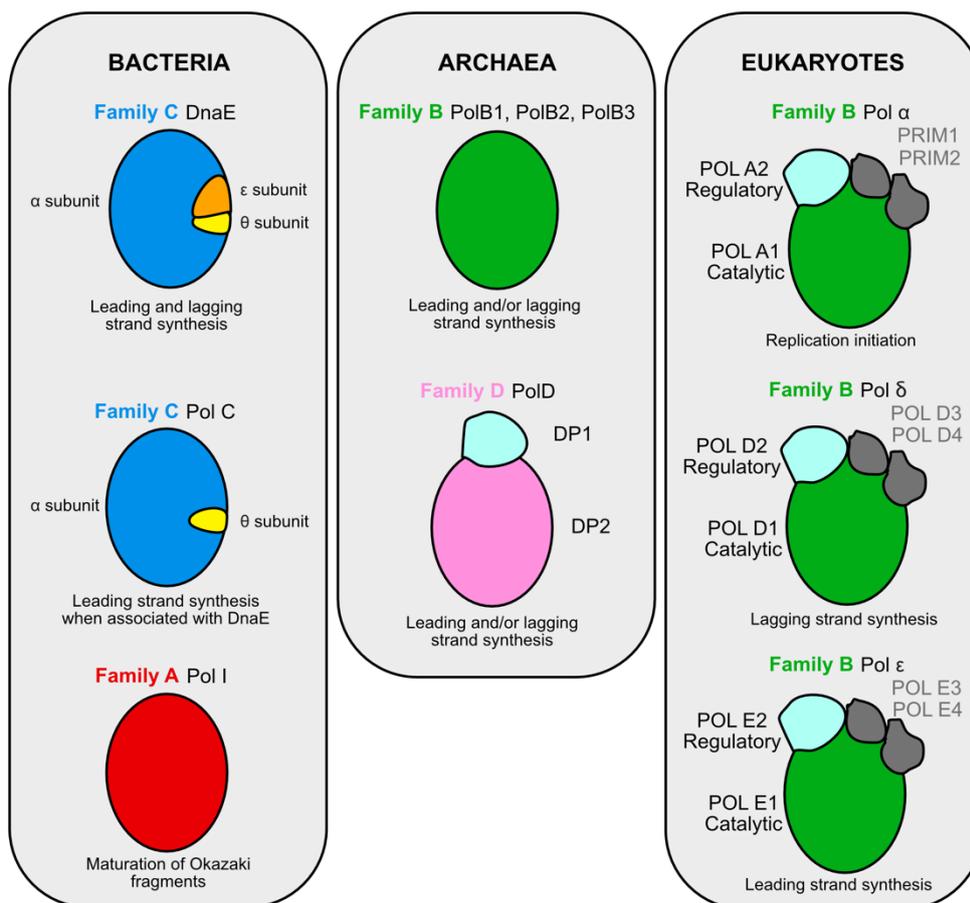


Figure 3.1: Comparison of replicative DNA polymerases across the three domains of life. Schematic representation of replicative DNA polymerases used by three domains of life. Eukaryotic subunits have been named according to the nomenclature of the corresponding human genes. Adapted from Raia *et al.*, 2019.

Bacterial replicative DNA polymerases

Genome replication in Bacteria is reliant on the Pol-III holoenzyme (HE); two copies of Pol-III are arranged asymmetrically at the replication fork. One copy will continuously synthesise the leading strand while the other discontinuously synthesises the Okazaki fragments of the lagging strand. Each Pol-III consists of an $\alpha\epsilon\theta$ core, where the α subunit is a Family C polymerase, ϵ is a separate 3'-5' exonuclease subunit from the DnaQ Family and θ acts to stabilise ϵ (Welch and McHenry, 1982). The Pol-III core associates with both the clamp protein β and the clamp loader γ to form the full HE (Jeruzalmi *et al.*, 2001).

In *Bacillus subtilis*, a prototypical Gram positive bacterium, two distinct copies of Pol-III, named Pol-C and DnaE, are utilised for genome duplication (Inoue *et al.*, 2001, Zhao *et al.*, 2006). DnaE and Pol-C differ in

that each carries a distinct α -subunit (the subunit required for catalysis of polymerising both DNA strands). Pol-C carries out the majority of the synthesis, while DnaE extends the RNA primers before handoff to Pol-C later (Sanders *et al.*, 2010). Additionally, rather than tracking along DNA, the replisome in *B. subtilis* remains relatively stationary, with template DNA being pulled into the replisome for duplication (Lemon and Grossman, 1998, Sawitzke and Austin, 2001).

Due to the discontinuous nature of lagging strand synthesis, the RNA primers attached to the Okazaki fragments must be removed to allow fragment ligation. For this function, bacteria utilise a Family A DNAP, Pol-I. This polymerase carries three active sites: a polymerase active site (referred to as Klenow fragment), a 3'-5' proofreading nuclease and a 5'-3' nuclease used to degrade the RNA primer.

Eukaryotic replicative DNA polymerases

Eukaryotes utilise various multimeric DNAPs for replication of the genome; namely Pol α , Pol δ and Pol ϵ (Burgers, 2009). All contain a catalytic core identifiable as a Family B polymerase, along with various accessory subunits fundamental to their roles in replication (Hubscher *et al.*, 2002). Pol α primarily extends the RNA primer for a limited time before passing the DNA to Pol δ and Pol ϵ on the lagging and leading strands respectively (Hubscher *et al.*, 2002). Because of this critical link between primer production and initial synthesis by Pol α , primase tightly associates with Pol α and its regulatory subunit (sometimes referred to as PrimPol or the primosome) (Garcia-Gomez *et al.*, 2013). Pol α lacks any proofreading activity, unlike both Pol δ and Pol ϵ , hence the early pass-over event to ensure high fidelity processive replication along the chromosome (Burgers and Kunkel, 2017).

Archaeal replicative DNA polymerases

All archaea encode Family B polymerases, which share homology with the catalytic subunits of eukaryotic Family B replicative polymerases (Raia *et al.*, 2019b, Kazlauskas *et al.*, 2020). They also, with the exception of Crenarchaea, encode an archaea-specific Family D polymerase, PolD (Makarova *et al.*, 2014).

Analysis of the euryarchaeal replication fork suggests a co-ordinated action of PolB and PolD with sliding clamp protein PCNA, similar to that seen in eukaryotes (Pan *et al.*, 2013). However, the intricacies of these replicative polymerases are less well defined within the archaea. While the replicative

polymerases are thought to have been identified, the mechanisms of which polymerase replicates the leading and/or the lagging strand remains under question. Generally, the level of confidence regarding replicative polymerase mechanisms in archaea is much lower than within the other two domains where much deeper interrogation has been carried out (Greenough *et al.*, 2015, Cubonova *et al.*, 2013, Henneke *et al.*, 2005).

PolB

Family B polymerases are found across all domains of life, as well as in viruses. At least one Family B polymerase is encoded by all archaeal species, however a single species may encode multiple PolB proteins (Sarmiento *et al.*, 2014). It is usually present as a monomer, with a single polypeptide encoding both the catalytic and proofreading activities (Barry and Bell, 2006, Makarova *et al.*, 2014), with the exception of *Methanothermobacter thermautotrophicus*, where it exists as two separate polypeptides (Kelman *et al.*, 1999b, Barry and Bell, 2006, Makarova *et al.*, 2014).

Archaeal Family B polymerases can be separated into three groups, historically named PolB1, PolB2 and PolB3 (Makarova *et al.*, 2014). PolB1 and PolB3 are active polymerases, while PolB2 proteins are generally inactive, carrying disrupted catalytic and exonuclease domains (although active PolB2 proteins are known) (Makarova *et al.*, 2014, Guy and Ettema, 2011, Rogozin *et al.*, 2008). The distribution of the PolB groups varies throughout the Archaea; PolB1 is absent from Euryarchaeota, PolB2 is scattered across the domain and PolB3 is missing from Thaumarchaeota (Makarova *et al.*, 2014, Cooper, 2018).

Several groups of archaeal PolB proteins contain multiple inteins; sometimes up to three per gene (Perler, 2002). Inteins are selfish genetic elements, often encoding endonucleases, that insert themselves into coding sequences and self-splice at the polypeptide level to propagate further intein insertions (Gogarten *et al.*, 2002, Liu, 2000). Within the PolB3 group, intein sites are generally conserved, however some are specific to lineage (Perler, 2002, MacNeill, 2009).

Generally, archaeal Family B polymerases are composed of a polymerase core (containing palm, fingers and thumb domains, an N-terminal 3'-5' exonuclease domain and a uracil-recognition domain (the latter being specific to archaea) (Makarova *et al.*, 2014, Wardle *et al.*, 2008). The uracil-recognition domain provides a damage sensing mechanism, whereby the

polymerase is able to scan ahead of the catalytic site and pause when meeting misincorporated uracil or hypoxanthine moieties that have escaped canonical repair by uracil-N-glycosylase (Greagg *et al.*, 1999, Connolly, 2009). Such a pocket has, as yet, not been found in any eukaryotic Family B polymerase (Wardle *et al.*, 2008) and hypothetically this mechanism may have developed due to the selection pressures of the harsh environments in which many archaea live.

Being the only processive polymerase in crenarchaea, it was hypothesised PolB was the main archaeal replicative polymerase, capable of both leading and lagging strand synthesis (Ishino *et al.*, 1998, Makarova *et al.*, 2014). However, crenarchaeal species will often encode more than one Family B polymerase. It is therefore possible that, within crenarchaea, each PolB has gained specialised roles at either the leading or lagging strands. For example, in *S. solfataricus*, PolB1 (Dpo1) has been implicated in replicating the leading strand, while PolB3 (Dpo3) is believed to replicate the lagging strand.

Most archaeal species, however, encode both Family B and archaeal-specific Family D polymerases. While PolB has been shown to extend DNA-primed templates efficiently, it struggles to extend RNA primers in an effective manner (Greenough *et al.*, 2015). This suggests extension must occur prior to handover to PolB; whether this is due to the inherent DNA polymerase activity of archaeal primases or PolD in non-crenarchaeal species remains to be shown.

Recent studies have revealed that PolB is not required for viability in all archaea: deletion of PolB is possible in three euryarchaeal species: *Thermococcus kodakarensis* (Cubonova *et al.*, 2013), *T. barophilus* (Birien *et al.*, 2018) and *Methanococcus maripaludis* (Sarmiento *et al.*, 2013). In these species, PolB is dispensable and PolD alone is essential and seemingly capable of both leading and lagging strand synthesis.

While it could be imagined that PolB is functionally redundant with PolD, a proofreading-deficient mutant of *T. kodakarensis* PolB did not show increased mutation rates overall, arguing against a role for PolB in canonical replication (Cubonova *et al.*, 2013). Additionally, *T. kodakarensis* $\Delta polB$ strains were shown to express similar levels of PolD protein as the wild type strain (Cubonova *et al.*, 2013). It would be expected if both PolB and PolD were acting simultaneously at the replication fork in the wild type strain, there would be an increase in PolD expression to compensate for the role of PolB in its absence; again, this suggests no role for PolB in canonical genome replication within *T. kodakarensis*.

Interestingly, the *T. kodakarensis* $\Delta polB$ strain showed increased sensitivity to gamma irradiation, suggesting PolB is involved in DNA synthesis during homologous recombination (i.e., during repair of DSBs) (Kushida *et al.*, 2019). While both PolB and PolD are capable of DNA synthesis, PolB has also been shown to be more efficient than PolD at extending a RadA recombinase-primed recombination intermediate (Hogrel *et al.*, 2020), again placing PolB at the heart of DNA repair.

However, PolB deletion is not possible in all euryarchaeal species. For example, in halophile *Halobacterium* NRC-1, both PolB and PolD are required for viability (Berquist *et al.*, 2007). It could be that PolB has gained an extra role in these species, potentially in DNA repair, or it could be reasoned that the high ploidy associated with halophilic species could increase the demand on replication proteins generally. Further work is needed to explain this differential requirement for DNA polymerases (specifically PolB) within euryarchaea.

PolD

Since its identification in 1998 (Ishino *et al.*, 1998), archaeal-specific Family D polymerase PolD has been shown to be distributed widely throughout archaea alongside PolB, being absent only from Crenarchaea. It is made up of two subunits encoded as two separate polypeptides: DP1 and DP2. DP1 is a small 3'-5' proofreading subunit, while DP2 is the large catalytic subunit (Raia *et al.*, 2019b, Natsuki *et al.*, 2019). It has been shown interaction between DP1 and DP2 is required for maximal exonuclease and DNA polymerase activities (Uemori *et al.*, 1997, Shen *et al.*, 2004).

Recently, the crystal structure of PolD has been resolved in the euryarchaeon *Pyrococcus abyssi* (Raia *et al.*, 2019a). While DP1 showed similarity to non-catalytic subunits of eukaryotic Family B polymerases (Jokela *et al.*, 2004, Makarova *et al.*, 2014), the catalytic DP2 subunit shows homology to the two-double-psi- β -barrel (DPBB) 'two-barrel' superfamily of polymerases (Raia *et al.*, 2019a). Well-known members of this superfamily include both DNA- and RNA-dependent transcriptases, homodimeric RNA silencing pathway RNAPs and atypical viral RNAPs (Fouqueau *et al.*, 2017, Sauguet, 2019, Ruprich-Robert and Thuriaux, 2010, Iyer, 2003). PolD is the first DNAP to be placed within this Family (Raia *et al.*, 2019b), therefore extending the repertoire of catalytic folds identified that are able to perform DNA replication. The evolutionary history of replication theorises that RNA was used as a genetic material prior to DNA (Leipe *et al.*, 1999). It can therefore be hypothesised that PolD may be the ancestral replicative DNA

polymerase of the last universal common ancestor (LUCA), having historically been utilised for RNA synthesis (Koonin *et al.*, 2020).

As with PolB, arguments have been made for PolD as either a leading or a lagging strand polymerase (Henneke *et al.*, 2005, Bauer *et al.*, 2012, Sarmiento *et al.*, 2014). However, evidence for its role as a lagging strand polymerase has gained momentum recently; PolD is able to efficiently extend synthesis from an RNA primer where PolB struggles (Greenough *et al.*, 2015, Uemori *et al.*, 1997) and PolD preferentially binds primed DNA, suggesting that PolD primes synthesis, even if there is a downstream handoff event to PolB (as seen between bacterial Pol I and Pol III). In support of this argument, it has been shown in *Pyrococcus abyssi* that PolB has the ability to displace PolD from DNA (Rouillon *et al.*, 2007).

Should PolD perform the lagging strand synthesis, it would require strand displacement activity to remove the primers associated with Okazaki fragments. Multiplex capillary electrophoresis with *Thermococcus* sp. 9°N proteins identified a failure of PolD to synthesise past downstream Okazaki fragments, halting 4 nucleotides prior to the fragment (Greenough *et al.*, 2015). Since PolB is capable of this activity, this would therefore indicate a requirement for PolB in Okazaki fragment maturation. However, it is worth noting this is a strain in which PolB can be deleted; therefore, either PolD can compensate *in vivo* and carry out this behaviour or an unknown player may be involved.

Studies in *P. abyssi* have shown that the RNA extension activity of PolD requires stimulation from PCNA (Henneke *et al.*, 2005). The binding of PolD to PCNA is also required for high polymerase processivity (Raia *et al.*, 2019a). This binding occurs through multiple sites in both DP1 and DP2 subunits, including a conserved PIP motif at the C-terminus of DP2. This interaction acts to ‘latch’ PolD onto the DNA duplex, preventing the enzyme from falling off prematurely. Recent work on *T. kodakarensis* has revealed that, prior to its interaction with PCNA, PolD interacts with primase (Oki *et al.*, 2021). This interaction utilises the PIP motif and primase dissociates upon binding of PCNA. This provides a direct link for PolD between initiation and elongation of DNA during replication and will warrant further study.

Further work is required on the archaeal replicative polymerases before such a definition of leading and lagging strand polymerase roles can be certain.

Modifications of archaeal DNA polymerases

Both bacterial (Sutton *et al.*, 2002) and eukaryotic (El-Andalousi *et al.*, 2006) polymerases have been shown to undergo modification(s) post-translationally. For example, eukaryotes are known to implement ubiquitin-mediated polymerase switching at the replication fork: where a DNA lesion would usually block replication, specialised polymerases can be ‘switched in’ to facilitate translesion synthesis (TLS) past the damaged DNA. Thus, it can be questioned whether similar modifications are at play in archaeal species.

Archaeal species have been shown to utilise an E1-E2-E2 ubiquitin-like small ubiquitin-like archaeal modifier protein (SAMP) system (Anjum *et al.*, 2015) and thus it could be hypothesised that such a system is involved in protein turnover or modifications at the replication fork. Other critical replication components have been seen to be modified in archaeal species, namely the helicase activity of *Sulfolobus* MCM is enhanced by lysine methylation (Xia *et al.*, 2015). It remains of interest as to the extent to which post-translational modifications impact DNA replication in archaea.

Haloferax volcanii DNA polymerases

Haloferax volcanii carries six genes encoding five DNA polymerases (Table 3.3) (Hartman *et al.*, 2010). These include members of DNA polymerase families B, D, X and Y.

Table 3.3: DNA polymerase genes encoded by *Haloferax volcanii*

Name	Gene locus	Location	% rare codon usage
PolD1	HVO_0003	Main chromosome	6.37
PolD2	HVO_0065	Main chromosome	3.82
PolB1	HVO_0858	Main chromosome	8.61
PolB2	HVO_A0065	pHV4	13.65
PolX	HVO_0741	Main chromosome	2.74
PolY	HVO_1302	Main chromosome	1.86

Family X polymerases are small, relatively inaccurate enzymes that are involved in numerous DNA repair pathways where they act to fill small gaps, including DSB repair and base excision repair (Moon *et al.*, 2007). Family Y polymerases act as translesion polymerases, where they synthesise short runs of error-prone sequence to bypass lesions within the DNA template (Sale *et al.*, 2012). Both PolX and PolY are known to be dispensable in *H. volcanii* (Thorsten Allers, unpublished data), suggesting their repair roles are not essential for canonical genome replication.

Chapter 3: Cellular requirement for replicative DNA polymerases PolB and PolD in *Haloflex volcanii*

As a euryarchaeon, *H. volcanii* encodes members of both Family B and Family D DNAPs. Two Family B polymerases are encoded: one from euryarchaeal active polymerase Family PolB3, gene *polB1* (HVO_0858) and one from the predicted-inactive Family PolB2, gene *polB2* (HVO_A0065). PolB1 contains an intein within its coding sequence. The deletion of this intein has been shown to not affect the growth of *H. volcanii*, indicating the intein sequence has no active role in replication of *H. volcanii* (Naor *et al.*, 2011).

When compared to PolB1, PolB2 has a high rare-codon usage: this is suggestive of gene transfer from another species (likely viral) and thus it is unlikely this polymerase is active within *H. volcanii* cells. It has been shown that deletion of PolB2 in *H. volcanii* is possible, fitting the theory it is non-native and likely inactive (Thorsten Allers, unpublished data).

PolD in *H. volcanii* is encoded by two subunits, DP1 (HVO_0003) and DP2 (HVO_0065). Small subunit DP1 is located in close proximity to replication origin *oriC1* and its associated Orc protein, Orc1, while the catalytic DP2 subunit is distal to the origin. Structurally, *H. volcanii* PolD is comparable to the known structure of *P. abyssi* PolD and has been shown to carry a comparable C-terminal PIP box motif (MacNeill, 2009).

3.2 Aims and Objectives

The specific requirements for PolB and PolD in archaea remain ill-defined. While both polymerases are thought to be involved in chromosomal replication, their roles remain largely undefined. Some species rely on both polymerases, while others may rely on PolB or PolD alone. The requirements for both PolB1 and PolD in *H. volcanii* remain unknown, and thus the objectives of this chapter are:

- Assess the requirement for PolB1 in *Haloferax volcanii* in the presence and absence of replication origins using an inhibitor screen
- Test the presumed essentiality of PolB1 and PolD by attempting gene deletions
- Where deletion is not possible, assess requirement for these polymerases using tryptophan-inducible promoters
- Develop strains and plasmids for CRISPR interference (CRISPRi) as an alternative method to inhibit *polB1* and *polD* expression

3.3 Results

3.3.1 Altered requirement for PolB in the absence of origins

Inhibition of PolB using Family B-specific inhibitor, Aphidicolin

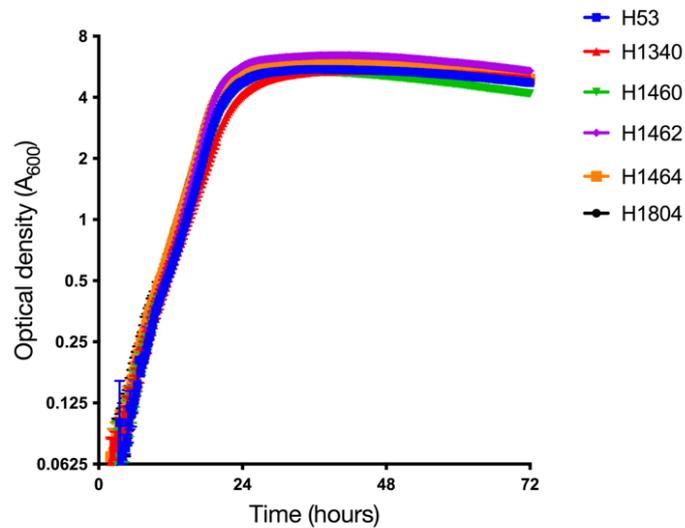
Growth rate

Aphidicolin is an antibiotic known to inhibit DNA replication. It was previously shown to block the polymerase activity of Family B polymerases in a Family-specific manner, while having no effect on the replicative ability of the distinct Family D polymerases (Ishino *et al.*, 1992, Cann, 1998, Ishino *et al.*, 1998).

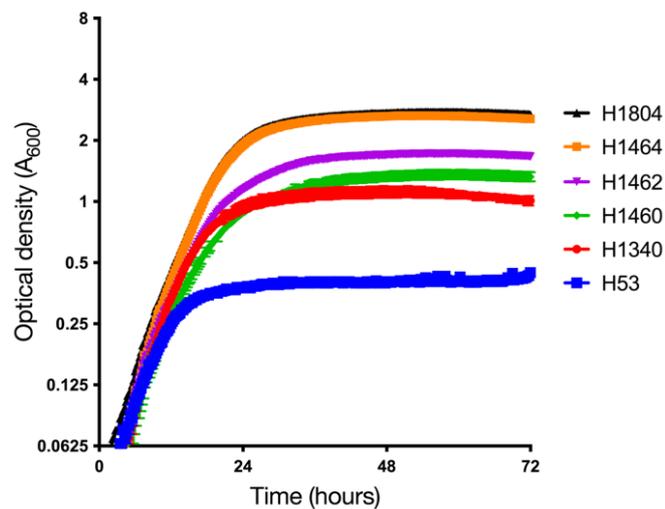
To elucidate whether the usage of B and D Family polymerases changes in the presence or absence of origins, strains with all origins present (H53; *ori+*) and with varying numbers and combinations of origin deletions (H1340, H1460, H1462, H1464 and H1804; *ori-*) were subjected to 72 hours chronic treatment with aphidicolin (7 µg/ml) or DMSO (empty vector control). Aphidicolin only inhibits Family B polymerases, and therefore any sensitivity to aphidicolin is directly related to requirement for PolB in *H. volcanii*.

Strains were grown for two consecutive overnights in Hv-YPC, ensuring on day three that actively dividing cells were used for the assay. Cells were diluted and treated with aphidicolin (or DMSO as empty vehicle control) in a 48 well plate. Optical density (OD; A₆₀₀) was continuously measured, allowing the plotting of growth curves for each strain (**Figure 3.2**).

(A) No aphidicolin



(B) 7 $\mu\text{g/ml}$ aphidicolin



Strain	Genotype
H53	$\Delta\text{pyrE2 } \Delta\text{trpA}$
H1340	$\Delta\text{pyrE2 } \Delta\text{trpA } \Delta\text{oriC1 } \Delta\text{oriC2}$
H1460	$\Delta\text{pyrE2 } \Delta\text{trpA } \Delta\text{oriC2 } \Delta\text{oriC3}$
H1462	$\Delta\text{pyrE2 } \Delta\text{trpA } \Delta\text{oriC1 } \Delta\text{oriC3}$
H1464	$\Delta\text{pyrE2 } \Delta\text{trpA } \Delta\text{oriC1 } \Delta\text{oriC2 } \Delta\text{oriC3}$
H1804	$\Delta\text{pyrE2 } \Delta\text{trpA } \Delta\text{oriC1 } \Delta\text{oriC2 } \Delta\text{oriC3 } \Delta\text{ori-pHV4-2}$

Figure 3.2: Effect of aphidicolin treatment on survival of wild type and origin-deleted strains. Wild type (H53) and strains with various combinations of chromosomal origin deletions (H1340, H1460, H1462, H1464 and H1804) were monitored for growth through measurement of optical density (A_{600}). All strains were treated with either DMSO (control) or aphidicolin (final concentration 7 $\mu\text{g/ml}$) chronically for 72 hours. The experiment was set up in a single 48-well plate and measured simultaneously using an Epoch Microplate Spectrophotometer (BioTek).

All strains, independent of origin number, grew equally well in the absence of aphidicolin (**Figure 3.2 A**). However, in the presence of aphidicolin (and therefore in cells with inhibited PolB activity), the number and combination of origins present proved important for cell growth. Wild-type cells (H53) are severely affected by aphidicolin treatment (**Figure 3.2 B**), with cell density plateauing by ~15 hours at an OD of ~0.4.

Contrastingly, all strains carrying at least one origin deletions have a growth advantage during aphidicolin treatment compared to wild type (**Figure 3.2 B**). This growth advantage can be judged by both i) the amount of time spent in exponential phase (where cells are actively dividing) and ii) by the final OD value reached when the curve plateaus. H1804 has all chromosomal origins deleted, along with that of mini-chromosome pHV4 ($\Delta oriC1 \Delta oriC2 \Delta oriC3$ and $\Delta ori-pHV4-2$). Both H1804 and H1464 (carrying deletions of the 3 chromosomal replication origins $\Delta oriC1 \Delta oriC2$ and $\Delta oriC3$), reach a final OD of ~2.7, significantly higher than wild type (OD ~0.4).

Cell morphology

To assess whether aphidicolin treatment disrupts cell division or altering morphology, strains H53 (*oriC+*) and H1804 ($\Delta oriC$) were grown for two subsequent overnights, reaching a final OD of 0.4, before being treated with either DMSO (control) or 20 $\mu\text{g/ml}$ aphidicolin for 3 hours. Cells were then spun down, resuspended in 18% salt water, stained with DAPI (final concentration 2.5 $\mu\text{g/ml}$) and incubated in the dark for 10 minutes. Cells were then washed of excess dye, resuspended in fresh 18% salt water and placed onto prepared agarose pads containing 18% salt water. Cells were imaged using a Nikon Ti-E inverted fluorescence microscope (**Figure 3.3**).

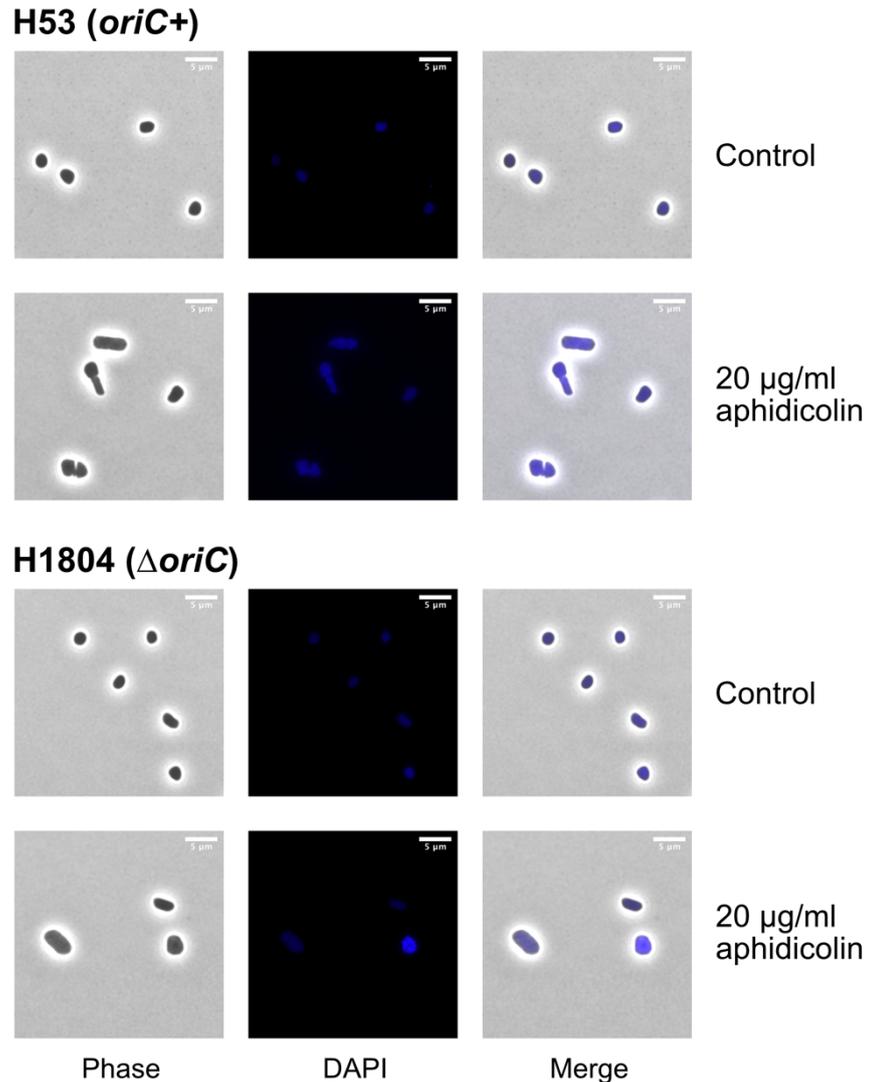


Figure 3.3: Microscopy showing cell morphology and DNA content for strains H53 (*oriC*+) and H1804 (Δ *oriC*) following treatment with aphidicolin. Both H53 and H1804 control samples show rounded cells with diffuse DNA (represented by DAPI stain). When treated with aphidicolin, H53 shows a stronger DAPI signal in all cells, and some cells show cell phenotypes differing from untreated. H1804 treated with aphidicolin shows an increase in DNA signal in some cells, but no change in morphology. Fluorescence was acquired with a one second exposure. Scale bars are 5 µm.

Control samples for H53 and H1804 are generally comparable; cells are rounded and show diffuse DNA staining throughout the cell. However, when treated with aphidicolin, H53 shows a range of cell morphologies, potentially suggesting issues with division (possibly caused by replication stress), and a strong increase in DAPI staining. Cell stress has previously been shown to induce nucleoid compaction in *H. volcanii* (Delmas *et al.*, 2013). However, compaction of the DNA was not observed here, instead treatment with aphidicolin led to an increased cell size.

3.3.2 Deletion of *polB1* (HVO_0858)

Both PolB and PolD are expected to be essential, based on gene essentiality in the close relative *Halobacterium* sp. NRC-1 (Berquist *et al.*, 2007). In wild type *H. volcanii*, *polB1* has previously been shown to be essential (Thorsten Allers, personal communication). However, with the reduction in requirement for PolB in the absence of origins, it poses the question as to whether PolB is an essential gene only in strains carrying out canonical origin-based replication. Therefore, a gene deletion using the existing pop-in/pop-out methodology was attempted for *polB1* in both a WT and Δori background.

Construction of pTA2294 (*polB1* deletion construct)

To create a deletion construct for *polB1* (HVO_0858), the genomic clone pTA193 (Figure 3.4; constructed by Thorsten Allers, unpublished data) acted as a template for PCR amplification of the upstream (US) and the downstream (DS) sequences flanking *polB1* in the genome.

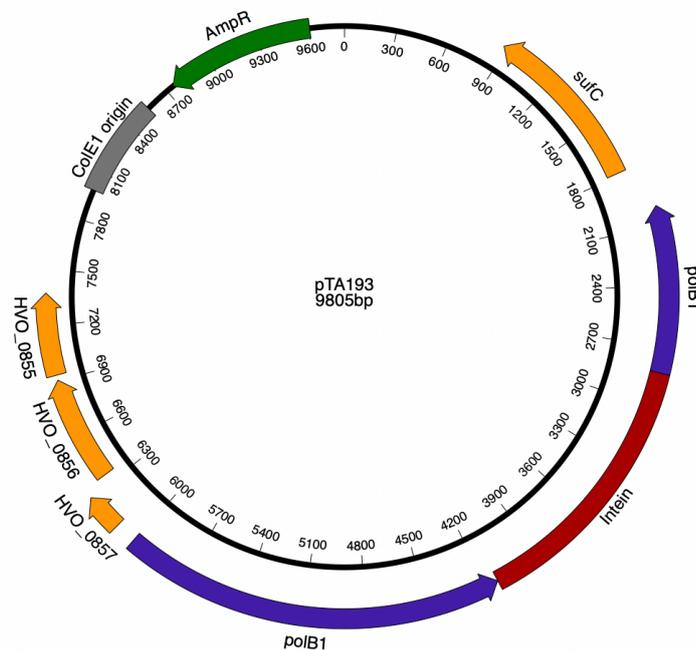


Figure 3.4: pTA193. Genomic clone of *polB1* (HVO_0858). Previously constructed by Thorsten Allers (unpublished data). PolB1 contains an intein (selfish genetic element; shown in red) within its coding sequence.

Primers *polBUSKpnF* and *polBUSClaR* amplified a 500 bp product containing the US sequence of *polB1*. This was digested with *KpnI* and *ClaI* and inserted into cloning vector pTA131 at *KpnI* and *ClaI* sites to give

intermediate plasmid pTA2291. Primers *polBDSBamF* and *polBDSXbaR* amplified a 520 bp product containing the DS sequence of *polB1*. This product was digested with *Bam*HI and *Xba*I and inserted into pTA2291 at *Bam*HI and *Xba*I sites to give the Δ *polB1* construct pTA2294 (Figure 3.5).

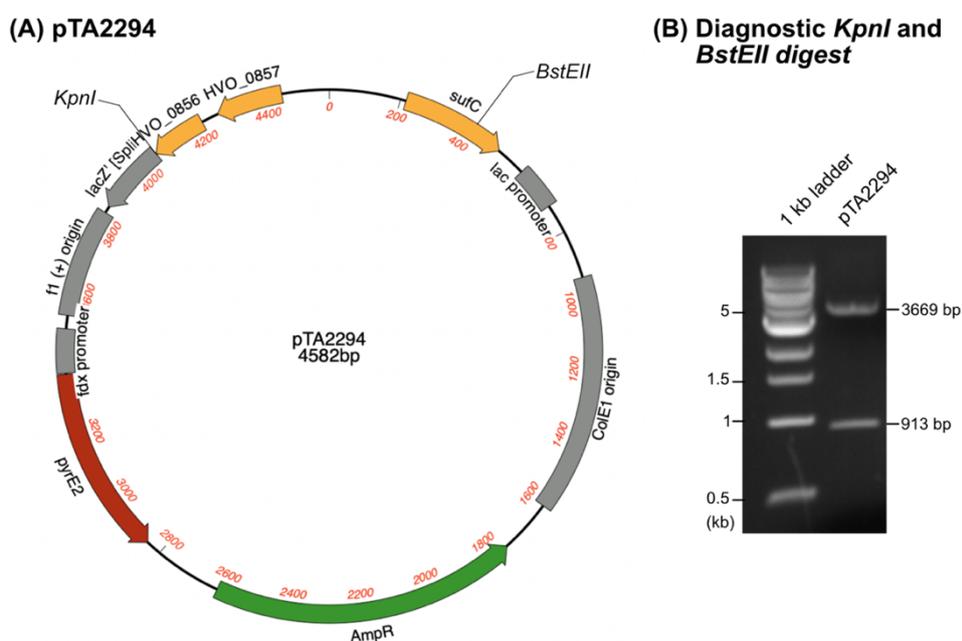


Figure 3.5: Generation of Δ *polB1* construct. (A) pTA2294 Δ *polB1* construct was generated by PCR against genomic clone pTA193 (previously constructed by Thorsten Allers, unpublished data). (B) Diagnostic digest of pTA2294 with *KpnI* and *BstEII* shows bands of 3669 bp and 913 bp, as predicted.

Adding *trpA* selection to pTA2294

To allow direct selection for the successful deletion candidates, the *trpA* marker (encoding tryptophan synthase) was added to the deletion construct. Since *polB1* is not within an operon, *trpA* was inserted into the construct along with synthetic promoter *p.fdx* to ensure adequate expression of tryptophan for selection. The *p.fdx::trpA* cassette (965 bp) was digested from pTA298 (Lestini *et al.*, 2010) with *Bam*HI and ligated into pTA2294 at the *Bam*HI site bordering the US and DS regions of the deletion construct. Both orientations of *trpA* were screened for, giving rise to two *trp*-marked constructs, pTA2300 and pTA2307 (Figure 3.6). pTA2300 carries *trpA* in the opposite orientation to *polB1* while pTA2307 carries *trpA* in the same orientation as *polB1*.

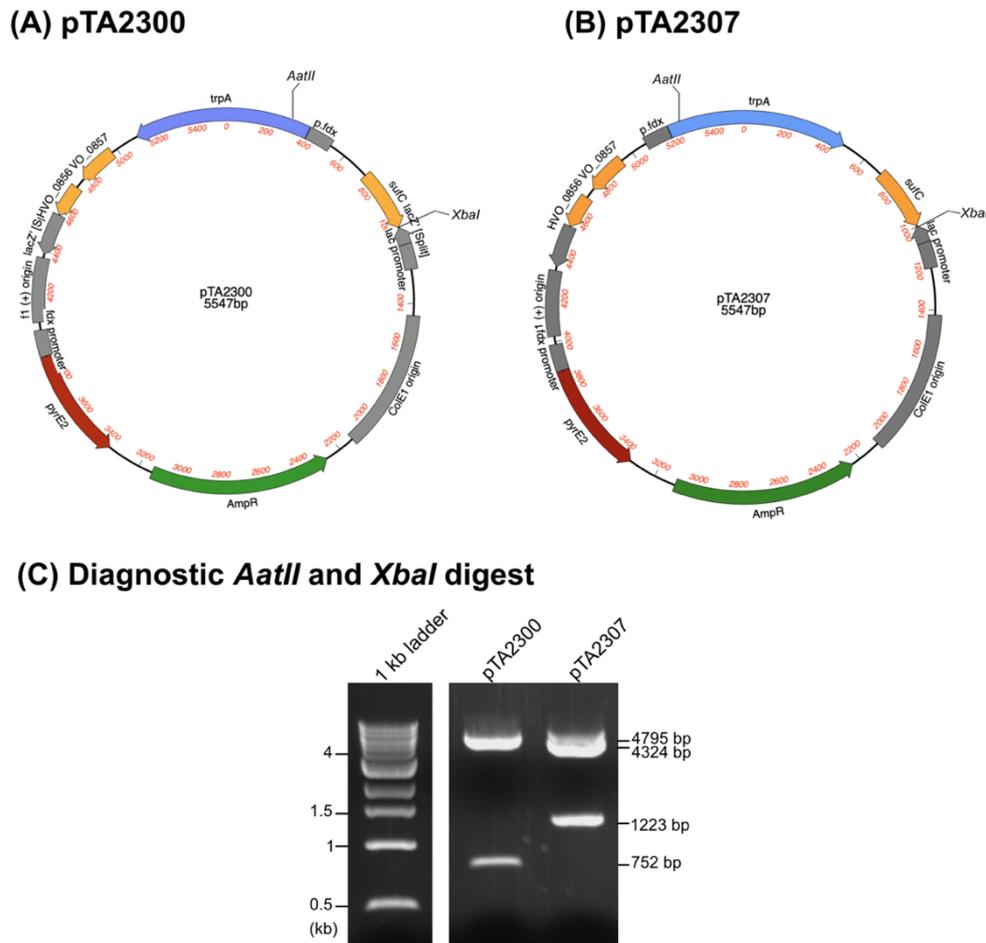


Figure 3.6: *trpA*-marked deletion constructs for *polB1*. (A) pTA2300 carries *trpA* in the opposite orientation as *polB1*. (B) pTA2307 carries *trpA* in the same orientation as *polB1*. (C) Diagnostic digest of clones with *AatII* and *XbaI* give bands of 4795 bp and 752 bp for pTA2300, and bands of 4324 bp and 1223 bp for pTA2307, as predicted.

Technical difficulties surrounding *dam* methylation of replication-associated plasmids

Haloferax volcanii contains a restriction endonuclease encoded by gene *mrr* (HVO_0682), which targets DNA methylated at *dam* sites (5' GATC 3') for degradation. Mrr acts as a defence mechanism against intruding foreign DNA carrying this type of methylation. For this reason, plasmids generated for transformation into *H. volcanii* are passed through a *dam*- *Escherichia coli* strain (N2338 (GM121)), which will leave the DNA unmethylated. Such a *dam*⁻ plasmid will then be able to persist within *H. volcanii* cells, whereas methylated *dam*⁺ DNA would usually be targeted for degradation.

However, the construction of *dam*⁻ plasmids encoding key components of *H. volcanii* DNA replication apparatus proved problematic; sequence data for *dam*⁻ plasmids was consistently weak, carrying mixed signals or mutant sequence compared to the correct clean sequencing of its *dam*⁺ partner. To test the quality of *dam*⁺ and *dam*⁻ plasmids, restriction digests were designed for plasmids containing either of the full length main replicative polymerases, PolB or PolD. **Figure 3.7** shows plasmid pairs (*dam*^{+/-}) generated as part of this project, with each pair being subject to the same diagnostic digest.

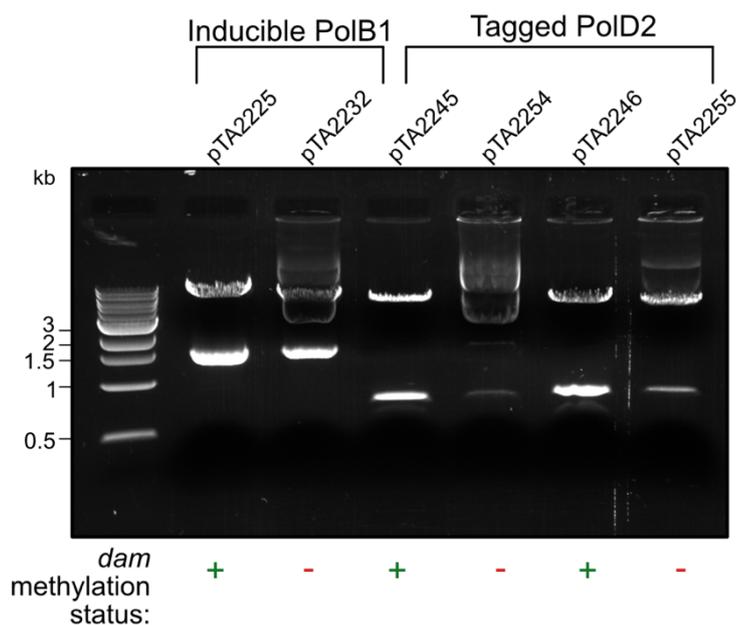


Figure 3.7: Comparison of *dam*⁺ and *dam*⁻ plasmid quality. *dam*⁺ and *dam*⁻ pairs were subjected to restriction digests to assess whether passaging through *dam*⁻ *Escherichia coli* affected DNA quality. Both PolB and PolD plasmids suffer DNA degradation following passaging, as seen by smears for the *dam*⁻ plasmids.

While the banding patterns for *dam*⁺ DNA are clear and defined, the *dam*⁻ DNA appears smeared, suggesting the DNA has undergone degradation within the *dam*⁻ cells. This could be due to errors in mismatch repair, which relies on strand methylation to identify the parent vs nascent strand for repair, however this was not proven in this study.

Deletion of mrr endonuclease (HVO_0682)

To overcome this issue and prevent degradation of *dam*⁺ DNA in *H. volcanii* upon transformation, the restriction endonuclease *mrr* (HVO_0682) gene was deleted from strains used for any subsequent construction involving PolB and/or PolD. As an example, the origin-deleted strain H1804 was

Chapter 3: Cellular requirement for replicative DNA polymerases PolB and PolD in *Haloferax volcanii*

subjected to deletion of *mrr* HVO_0682 using deletion construct pTA1150 (Allers *et al.*, 2010). This gave rise to Δmrr strain H4598. Deletion of the *mrr* gene was confirmed by colony hybridisation using a 520 bp probe generated by PCR against the wild-type (H53) genome, using primers *mrrF* and *mrrR* (**Figure 3.8 A**). Deletion of *mrr* was further confirmed by Southern blot (**Figure 3.8 B-D**).

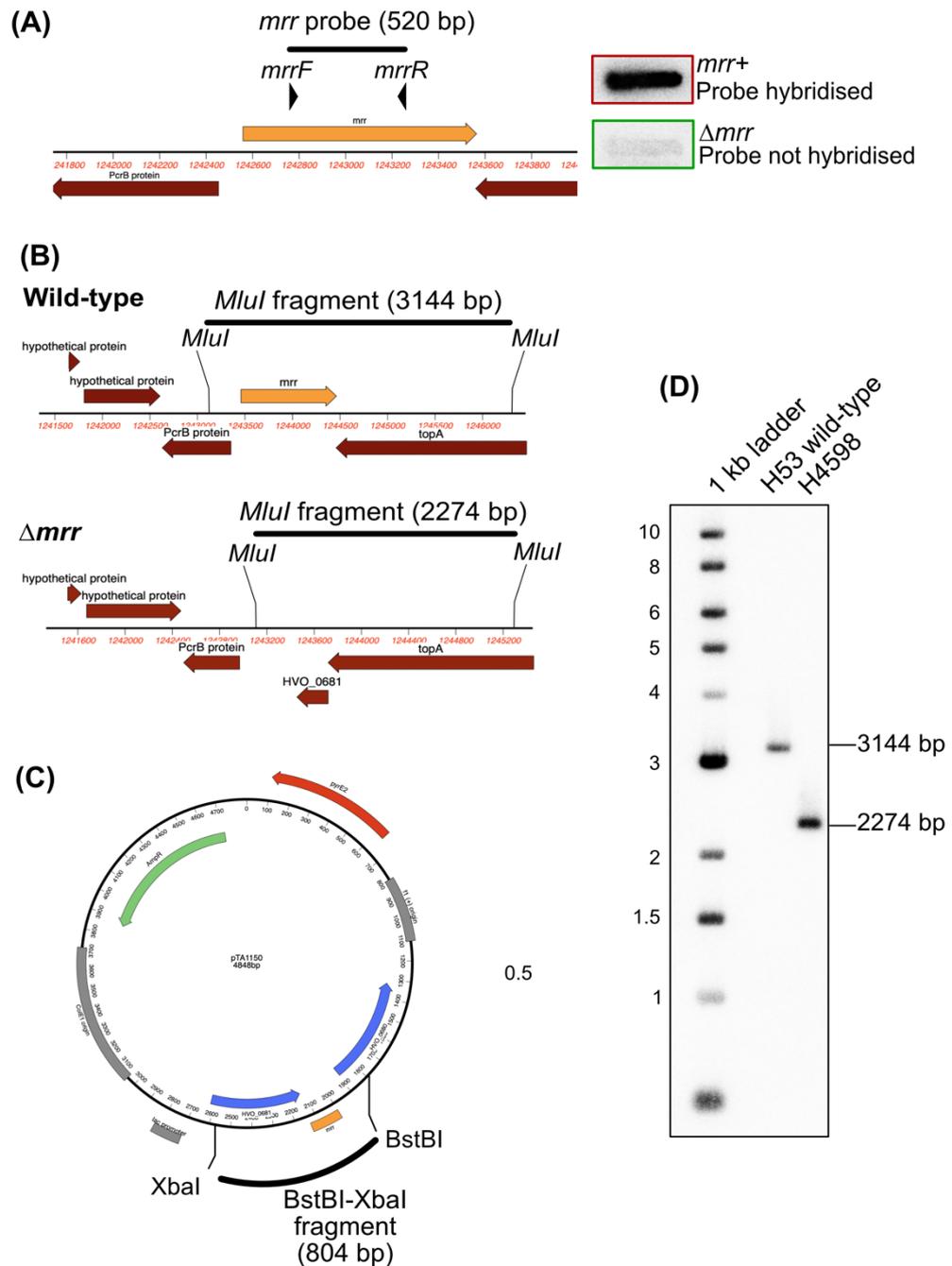


Figure 3.8: Deletion of *mrr* (HVO_0682). (A) Colonies were screened using a radioactive probe against the *mrr* gene, generated by PCR with primers *mrrF* and *mrrR*. Colonies where the probe did not hybridise were Δmrr candidates. (B) Expected Southern blot band sizes for *MluI*-digested genomic DNA (wild type 3144 bp and Δmrr 2274 bp). (C) 804 bp Δmrr Southern probe consisting of a *BstBI-XbaI* fragment of *pTA1150*. (D) Southern blot confirming strain H4598 as Δmrr .

Attempted deletion of PolB1 from originless strains

Strain H4598 ($\Delta pyrE2 \Delta trpA \Delta mrr \Delta oriC1 \Delta oriC2 \Delta oriC3 \Delta ori-pHV4-2$) was transformed with pTA2300 and pTA2307 to generate pop-in strains H4625 and H4627 respectively. Pop-outs of H4625 and H4627 were performed in the absence of tryptophan (Hv-Cas+Ura media) to increase selection for the pop-out event and gave rise to colonies on 5-FOA plates. These colonies were patched on Hv-Cas+Ura for screening by colony hybridisation.

When attempting to delete a gene that is predicted to be essential, it is important to ensure enough colonies are screened to conclude with confidence that the gene is in fact essential. It is well-documented that RadA is essential in *H. volcanii* strains lacking origins *oriC1, C2, C3* and *ori-pHV4*, and compared to wild type strains, *radA* generally becomes more difficult to delete in the absence of origins (**Table 3.4**) (Hawkins *et al.*, 2013a).

Table 3.4: Fraction of screened clones successfully deleted for *radA* in the presence or absence of replication origins. Fraction successfully $\Delta radA$ indicates the number of successful mutants with $\Delta radA$ phenotype over the total number of assayed colonies. Deleting the *radA* gene becomes increasingly difficult the more origins that are deleted and is impossible when all main chromosomal origins are deleted ($\Delta oriC1 \Delta oriC2 \Delta oriC3 \Delta ori-pHV4$). Data taken from (Hawkins *et al.*, 2013a).

Genotype	WT	$\Delta oriC1$	$\Delta oriC2$	$\Delta oriC3$	$\Delta oriC1 \Delta oriC2 \Delta oriC3$	$\Delta oriC1 \Delta oriC2 \Delta oriC3 \Delta ori-pHV4$
Fraction successfully $\Delta radA$	62/66	66/66	19/44	9/73	1/70	0/455

Using 1/70 as a measure of confidence for generating a ‘hard-to-delete’ mutant, where deletion of *radA* in $\Delta oriC1 \Delta oriC2 \Delta oriC3$ strains was difficult, the number of candidates requiring screening can be calculated to ensure 90% confidence of identifying a rare mutant (**Equation 3.1**).

Equation 3.1: Probability of identifying ‘hard-to-delete’ mutant based on rate of discovery of 1/70. Based on a discovery rate of 1/70, 152.35 candidates must be screened to ensure 90% confidence the gene cannot be deleted.

$$P(\Delta radA) = \frac{1}{70} = 0.0142$$

$$P(radA +) = \frac{69}{70} = 0.985$$

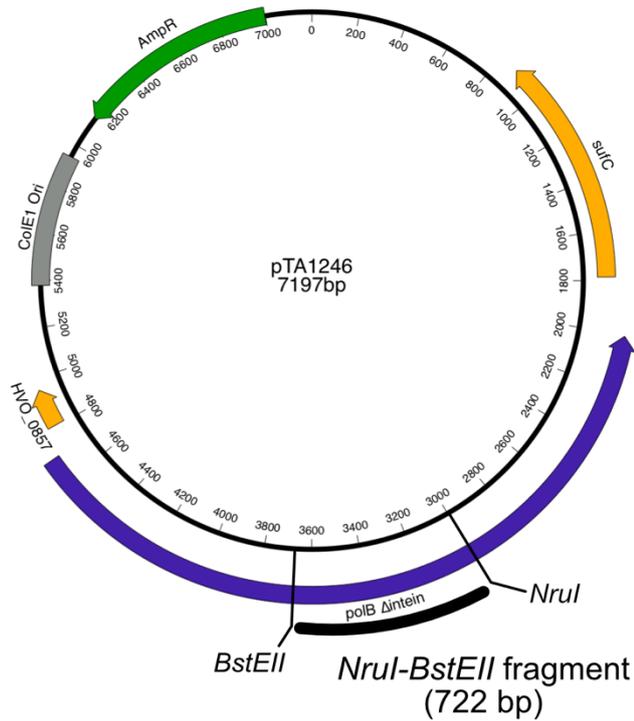
For 90% confidence:

$$\log_{10} 0.1 / \log_{10} P(radA+) = 152.35$$

Using this as a measure of confidence in finding ‘hard-to-delete’ mutants, screening of 160 colonies gives a 90% confidence level in finding successful *polB1* deletion candidates.

Hybridisation was carried out with a 722 bp probe, isolated by digesting pTA1246 (inteinless *polB1* genomic clone; constructed by Thorsten Allers, unpublished data) with *Bst*EI and *Nru*I, isolating only the coding sequence of *polB1* (**Figure 3.9 A**). All pop-out colonies for strains H4625 and H4627 were hybridised by the probe and thus carry at least one wild type *polB1* sequence (likely merodiploid candidates as able to grow in absence of tryptophan; **Figure 3.9 B**). Accordingly, it may be asserted with 90% confidence that *polB1* cannot be deleted from *H. volcanii* in the absence of replication origins.

(A) PolB1 probe



(B) Colony hybridisation



Figure 3.9: Colony hybridisation of $\Delta polB1::trpA+$ candidates. (A) Patches were probed with a *NruI*-*BstEII* fragment of pTA2325 (His6-*polB1* vector; construction data not shown). (B) Probed pop-outs all hybridised with the probe, suggesting all clones carry wild-type *polB1* sequence.

3.3.3 Attempted deletion of *polD2* (HVO_0065)

Similar to PolB1, essentiality of PolD in *H. volcanii* was predicted based on the requirement for both PolD1 and PolD2 in the closely related species *Halobacterium* sp. NRC-1 (Berquist *et al.*, 2007). To test the presumed essentiality of Family D polymerases in *H. volcanii*, a deletion attempt was made targeting *polD2* (HVO_0065; the catalytic subunit of PolD). Prior to this work, a deletion construct for *polD2*, pTA436, had been constructed by Sam Haldenby (unpublished data) (**Figure 3.10**). However, this deletion construct lacked additional selection for the deletion, for example the commonly-used marker *trpA*. As PolD would be presumed to play an important, potentially essential, role in *H. volcanii*, the deletion attempt should use additional selection (alongside 5-FOA to screen for pop-out events) to ensure a false result is not obtained.

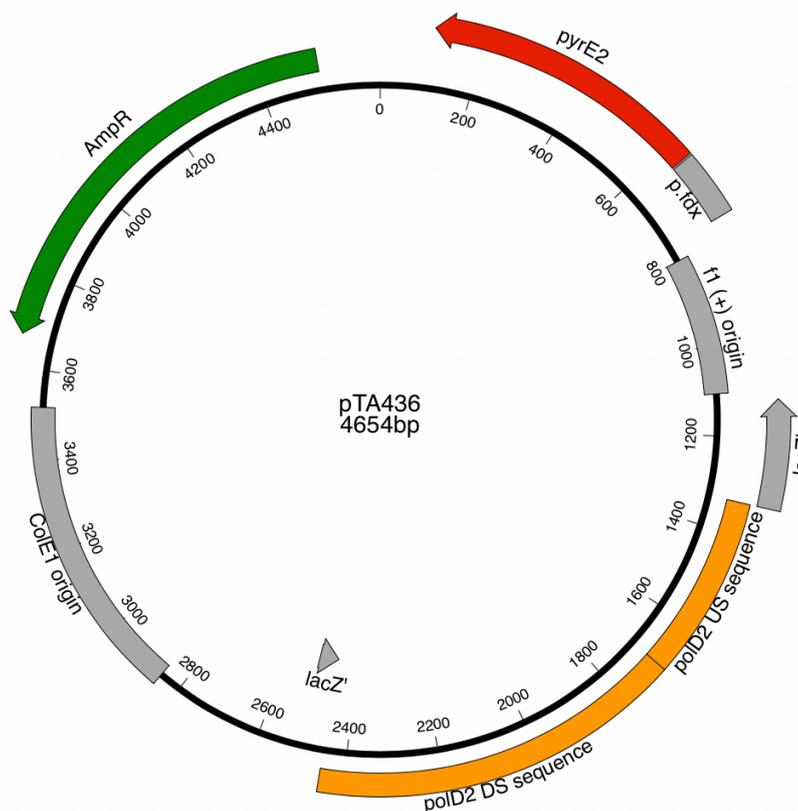


Figure 3.10: pTA436. Deletion construct targeting *polD2* (HVO_0065). Constructed by Sam Haldenby in 2005 (unpublished data).

Adding tryptophan selection to pTA436

To add extra selection for the successful deletion of *polD2*, the *trpA* marker was added to the deletion construct. Due to the method of construction, pTA436 lacks suitable restriction sites for addition of *trpA* at the US/DS

margin, harbouring only an *EagI* site. To overcome this issue, a pair of linker oligonucleotides were designed that when annealed to one another, harbour internal *Bam*HI and *Nde*I sites, and have *Mlu*I overhangs that are compatible with *Eag*I. Digestion of pTA436 with *Eag*I and insertion of the *Mlu*LinkF::*Mlu*LinkR annealed oligo linker gave rise to plasmid pTA2217 (Figure 3.11).

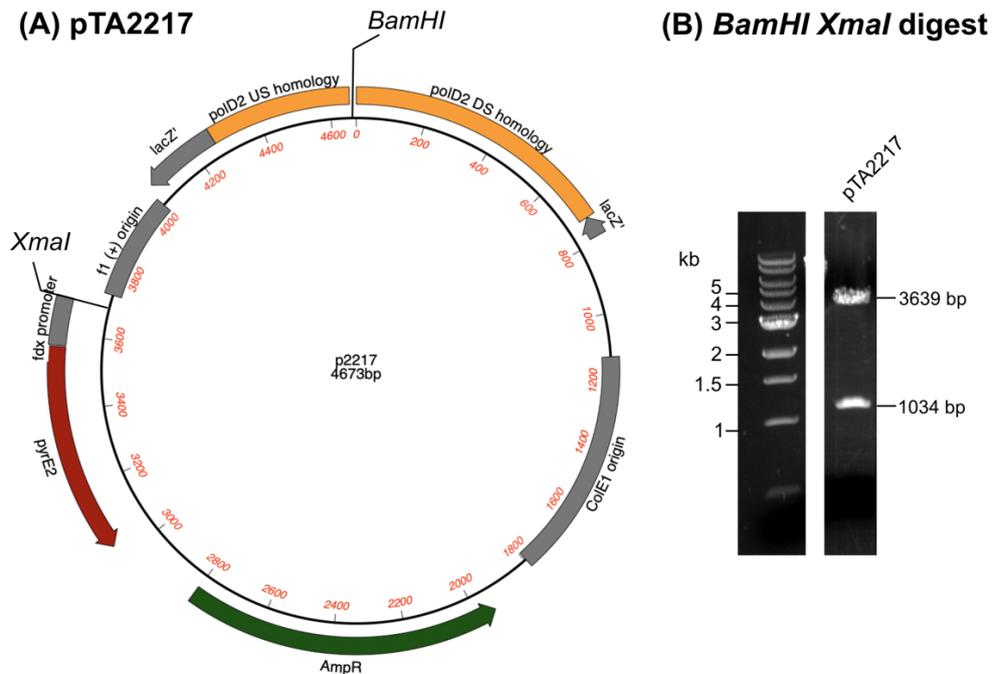


Figure 3.11: pTA2217. (A) To add further restriction sites to existing *polD2* deletion construct pTA436 (generated by Sam Haldenby, unpublished data), pTA436 was digested with *Eag*I and overlapping oligonucleotide linkers with an *Eag*I overhang were inserted and ligated to introduce unique cloning sites. (B) Digestion with *Bam*HI (cuts within linker) and *Xma*I showed bands at 3639 bp and 1034 bp, as predicted.

The introduction of an internal *Bam*HI allowed for digestion of *trpA* and its associated promoter *p.fdx* from plasmid pTA298 (Lestini *et al.*, 2010) and insertion at the US/DS sequence junction of pTA2217. Both orientations of *trpA* were selected for, giving rise to two *trpA*-marked constructs, pTA2367 and pTA2368 (Figure 3.12). pTA2367 carries *trpA* in the same orientation to *polD2* while pTA2368 carries *trpA* in the same orientation as *polD2*.

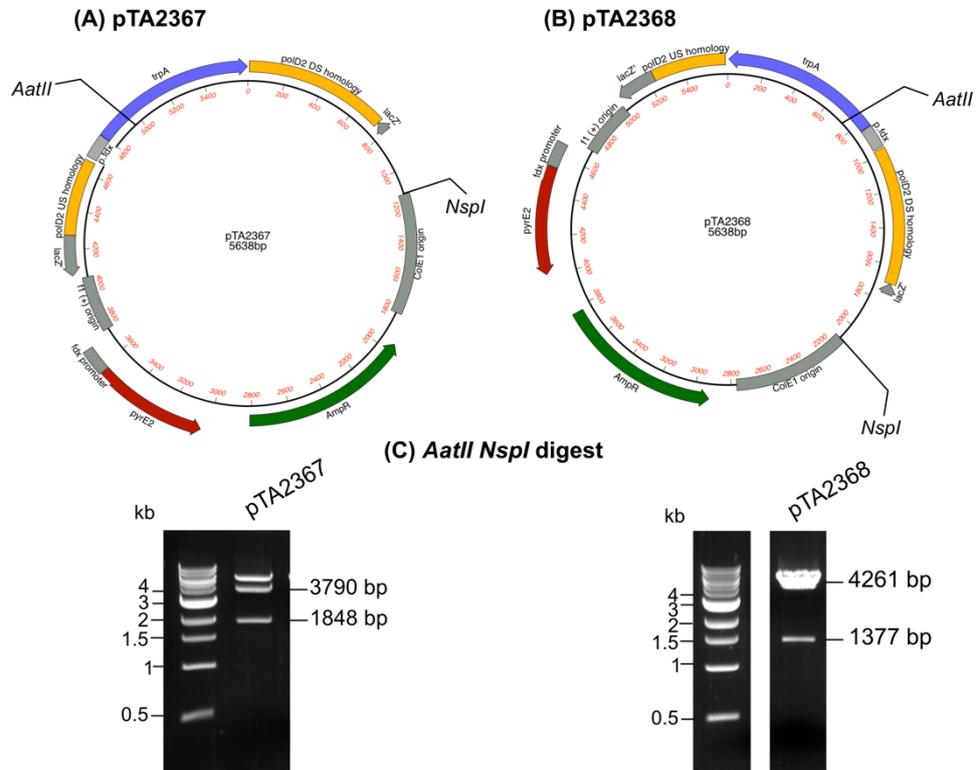


Figure 3.12: Construction of *trpA*-marked deletion constructs for *polD2*. (A) pTA2367 carries *trpA* in the same orientation as *polD2*. (B) pTA2368 carries *trpA* in the opposite orientation as *polD2*. (C) Diagnostic digest of clones with *AatII* and *NspI* give bands of 3790 bp and 1848 bp for pTA2367, and bands of 4261 bp and 1377 bp for pTA2368, as predicted. The high molecular weight band seen for pTA2367 is likely undigested DNA; both constructs were confirmed as correct via sequencing.

Attempted deletion of PolD2 from wild type and originless strains

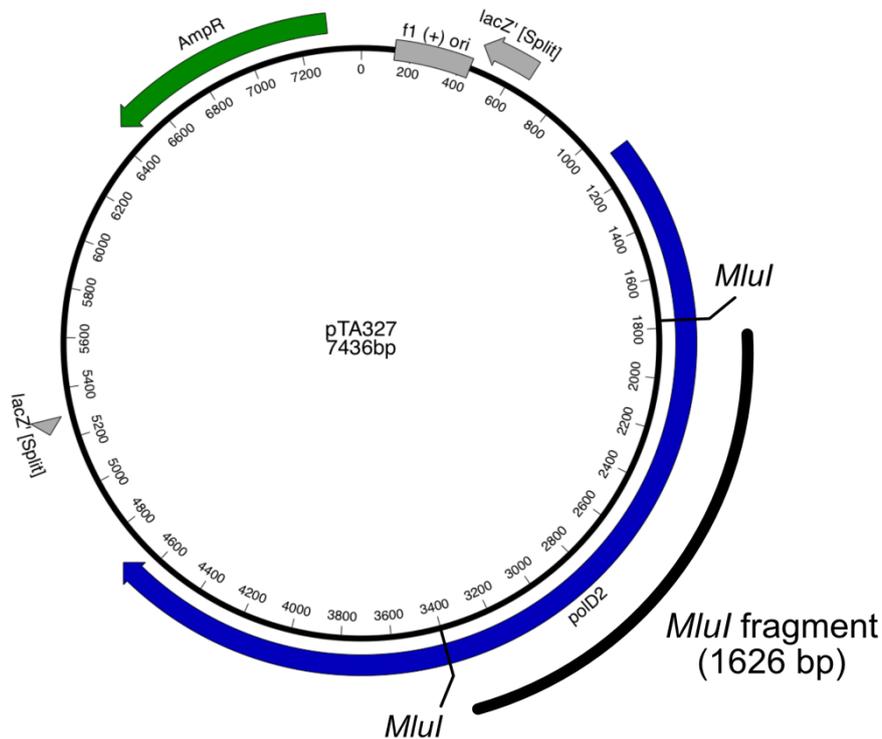
Strains H4045 (*oriC*⁺) and H4598 (Δ *oriC*) were transformed with pTA2367 to generate pop-in strains H4953 and H4955 respectively, and with plasmid pTA2368 to generate pop-in strains H4954 and H4956 respectively. Pop-outs were performed in Hv-Cas+Ura media to increase selection for the *trpA*-marked deletion construct. Pop-outs were patched on Hv-Cas+Ura and 160 clones per strain were screened for successful deletion using colony hybridisation.

Hybridisation was carried out with a 1626 bp probe, isolated by digesting pTA327 (Figure 3.13 A; *polD2* genomic clone, constructed by Sam Haldenby) with *MluI*, isolating only the coding sequence of *polD2*. All pop-out

Chapter 3: Cellular requirement for replicative DNA polymerases PolB and PolD in *Haloferax volcanii*

colonies for strains H4953-4956 hybridised by the probe and thus carry at least one WT *polD2* sequence (likely merodiploid candidates as able to grow in absence of tryptophan; **Figure 3.13 B**). According to the logic applied previously to *polB1* deletions, screening of 160 candidates yields a 90% confidence that *polD2* cannot be deleted in *H. volcanii*, both in strains encoding and lacking replication origins.

(A) PolD2 probe



(B) Colony hybridisation

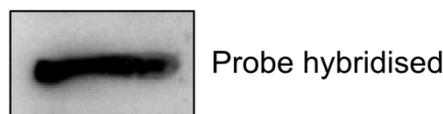


Figure 3.13: Colony hybridisation of $\Delta polD2::trpA^+$ candidates. (A) Patches were probed with a *MluI* fragment of pTA327 (*polD2* genomic clone). (B) Probed pop-outs for strains H4953-6 hybridised with the probe, suggesting all clones carry wild-type *polD2* sequence.

3.3.4 Assessing the altered requirement for PolB and PolD expression using inducible promoters

While both PolB1 and PolD2 appear to be essential in both WT and Δori strains, the differential response to aphidicolin in cells lacking replication origins suggests altered roles for the polymerases, depending on method of replication utilised (origin-dependent vs recombination-dependent replication). Placing the genes under inducible promoters will allow direct visualisation of their altered requirements in the presence/absence of origins.

Where a gene is essential, it is unlikely to allow integration of an inducible promoter (and therefore alteration of its physiological expression level), unless the event is forced to occur using selection pressure. Previously, a construct had successfully been used to place *radA* under an inducible promoter, including in the absence of origins where its expression is essential (Hawkins *et al.*, 2013a). The full-length *radA* gene was placed under the tryptophan-inducible promoter *p.tnaA* and was flanked at the 3' end by an *hdrB* marker (encoding thymidine synthesis). This cassette was then flanked by homologous sequences US and DS of *radA*, allowing for integration at the *radA* locus and subsequent removal of the WT *radA* gene and promoter. A similar methodology had also been used for *mcm* helicase, however due to its lower expression levels (8-fold lower expression compared to RadA; Thorsten Allers, personal communication), a lower-activity version of the tryptophanase promoter, *p.tnaM3*, had to be utilised for successful integration (Marriott, 2017). Due to the comparable expression levels of MCM and PolB1/PolD2 (RNAseq; Thorsten Allers, personal communication), the low-activity promoter *p.tnaM3* was utilised for inducible expression of both PolB1 and PolD2.

Cloning of an inducible *polB1* (HVO_0858) full-length construct

PolB1 contains an intein within its coding sequence, a selfish genetic element which is capable of splicing itself out at the polypeptide level. It has previously been shown the intein within *polB1* can be deleted with no effect on cell viability (Naor *et al.*, 2011). Due to the length of *polB1* (2.7 kb) and its intein (1.3 kb), working with the smaller inteinless version of *polB1* for cloning proved easier. Also, the mobile status of inteins meant that inteinless constructs inserted into strains with inteins would be at risk of invasion events, and thus strains used for these experiments were deleted for the PolB1 intein ($\Delta intein::polB1+$).

For cloning of inducible genes under control of the low-activity promoter *p.tnaM3* and with 3' *hdrB* selection, the existing cloning vector pTA1451 (Braun *et al.*, 2019) was utilised (**Figure 3.14**). Integration of a gene

at the *NdeI* site places it under *p.tnaM3*, while the *PciI* site allows for addition of an optional 6xHis protein affinity tag.

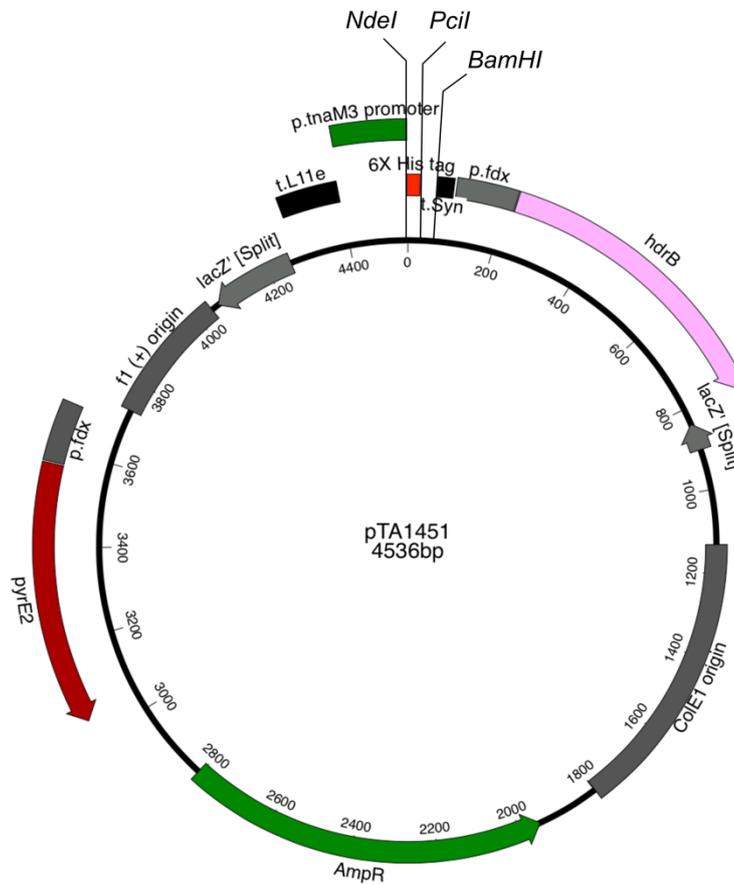


Figure 3.14: pTA1451. Plasmid for construction of 6xHis-tagged genes under the control of low-activity tryptophan inducible promoter *p.tnaM3* with additional *hdrB* selection. Constructed by Hannah Marriott, 2013. *p.tnaM3* promoter from pNPM-tnaM3-HfxMCM (Stuart MacNeill, unpublished) (Braun *et al.*, 2019).

The inteinless *polB1* was amplified by PCR from pTA1246, a genomic clone of *polB1* lacking the intein sequence (constructed by Thorsten Allers, unpublished data; **Figure 3.15**).

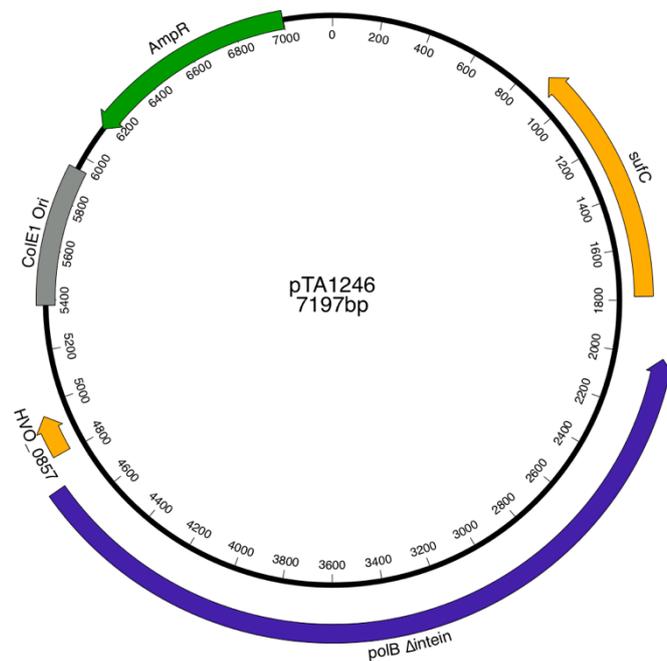


Figure 3.15: pTA1246. Genomic clone of *polB1* lacking internal intein sequence. Constructed by Thorsten Allers, unpublished data.

Primers *polBNdeF* and *polBBamR* were used to amplify the 2723 bp *polB1* coding sequence while integrating novel 5' *NdeI* and 3' *BamHI* sites. The product was digested with *NdeI* and *BamHI* and was inserted into pTA1451 at corresponding *NdeI* and *BamHI* sites, giving pTA2181 (**Figure 3.16**). pTA2181 contains inducible *polB1*, however, to facilitate integration of the construct as a gene replacement construct, the US and DS genomic sequences must be added. Thus, the *p.tnaM3-polB1-hdrB* cassette was digested from pTA2181 with *BglII* and inserted into the unique *BamHI* site of Δ *polB1* vector pTA2294, giving rise to the gene-replacement inducible-*polB1* construct, pTA2225 (**Figure 3.17**).

(A) pTA2225

(B) *NdeI* *KpnI* digest

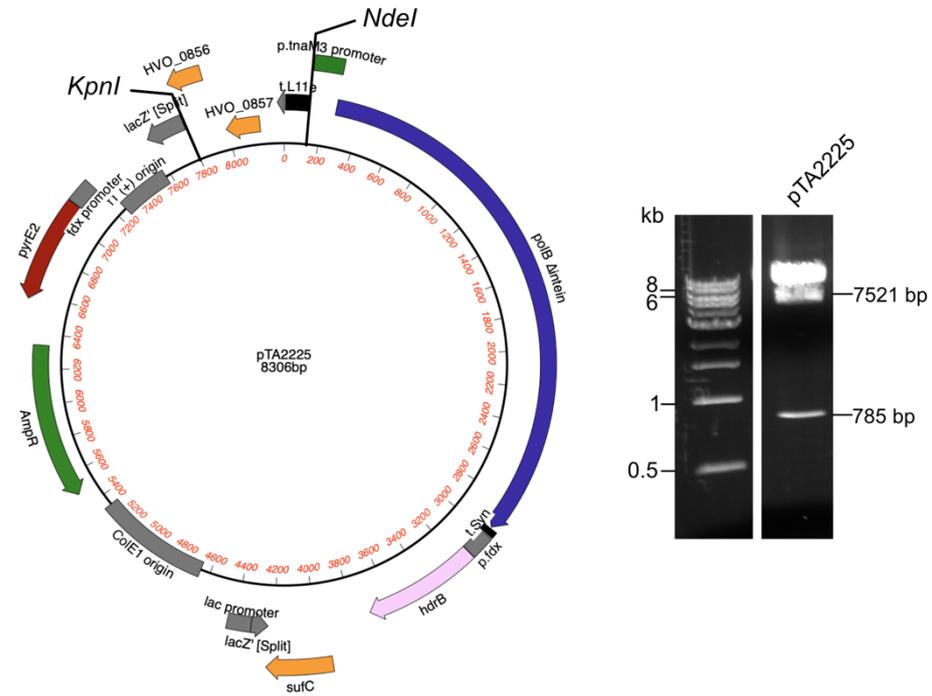


Figure 3.17: pTA2225. (A) pTA2225 was constructed by placing the *p.tnaM3-polB1-hdrB* cassette from pTA2181 into *polB1* deletion construct pTA2294. (B) Digestion of pTA2225 with *KpnI* and *NdeI* gave rise to bands at 7521 bp and 785 bp, as expected. The band at ~8 kb likely represents undigested DNA, as pTA2225 was further confirmed as correct by sequencing.

Cloning of an inducible *polD2* (HVO_0065) full-length construct

The genomic sequence of *polD2* was amplified by PCR from the genomic clone pTA327. Primers *dp2fwdNde* and *dp2revBam* were used to isolate the 3655 bp coding sequence of *polD2*. The product was digested with *NdeI* and *BamHI* before being inserted into pTA1451 at *NdeI* and *BamHI* sites. This gave generated the intermediate plasmid pTA2182 carrying the *p.tnaM3-polD2-hdrB* cassette (**Figure 3.18**).

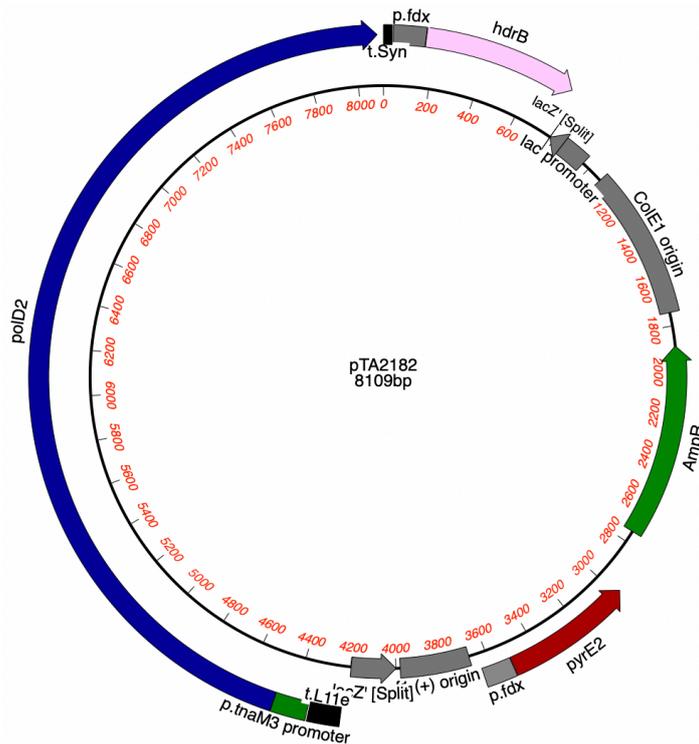
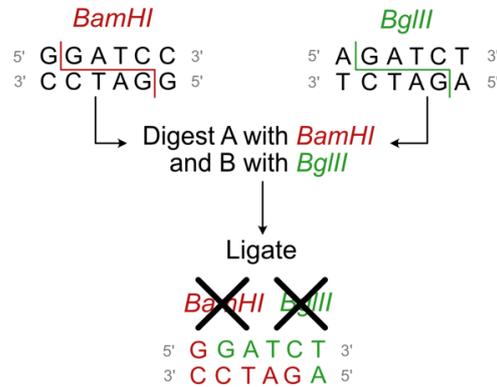


Figure 3.18: pTA2182. Coding sequence of *polD2* amplified from genomic clone pTA327 using primers *dp2fwdNde* and *dp2revBam* (3655 bp). The PCR product was digested with *Nde*I and *Bam*HI and inserted into *Nde*I and *Bam*HI sites in cloning vector pTA1451, giving intermediate plasmid pTA2182.

Initially, cloning of the *p.tnaM3-polD2-hdrB* cassette in the $\Delta polD2$ deletion construct, pTA2217, was attempted as above. *Bgl*II was used to digest the inducible gene cassette from pTA2182, allowing insertion into the deletion construct pTA2217 at its unique *Bam*HI site. However, repeated attempts failed and thus a new cloning strategy was designed. When *Bgl*II and *Bam*HI sites ligate with each other, their 5'-GATC-3' overhangs are compatible, but their external bases pairs differ and thus both sites are inactivated upon ligation (**Figure 3.19 A**). pTA2182 contains *Bgl*II sites surrounding the cassette, but also carries a *Bam*HI site at the 3' end of *polD2*. Removal of this *Bam*HI site would mean upon digestion of pTA2182 with *Bgl*II, digestion of pTA2217 with *Bam*HI and subsequent ligation, the product would be immune to digestion with *Bam*HI or *Bgl*II as sites are inactivated at ligation (**Figure 3.19 B**). This allows for a further level of selection for the correct product, where the ligation reaction can be digested with *Bam*HI to remove any vector-only candidates prior to transformation.

(A) BamHI BglII cloning



(B) Cloning for inducible cassette insertion

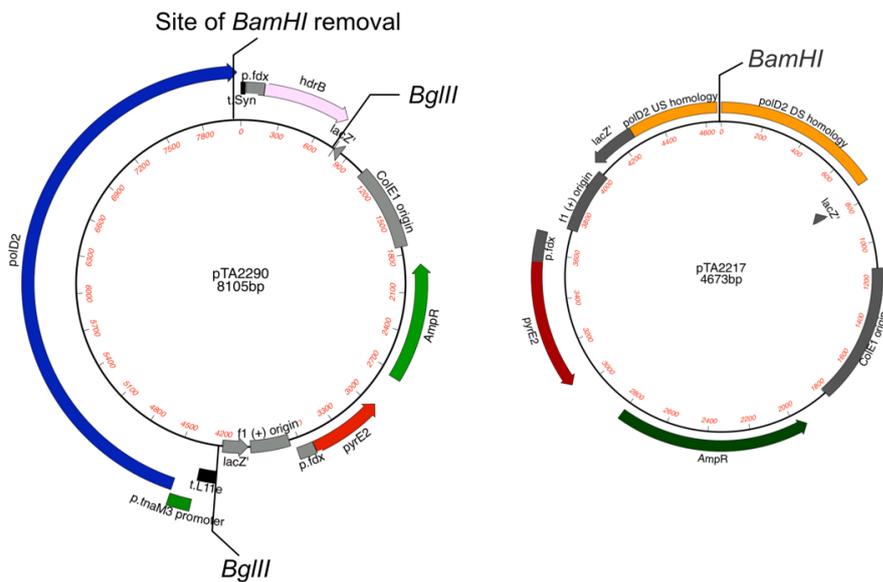


Figure 3.19: Cloning strategy for placing inducible *polD2* cassette into genomic vector. (A) While the overhang of 5'-GATC-3' is shared between *Bam*HI and *Bgl*II, when the two overhangs are ligated together, upstream bases are altered meaning both sites are inactivated. (B) The unwanted *Bam*HI site in pTA2182 was removed using Klenow, giving rise to pTA2290. The *p.tnaM3-polD2-hdrB* cassette was digested from pTA2290 using *Bgl*II and inserted into pTA2217 at its *Bam*HI site, inactivating all *Bgl*II and *Bam*HI sites upon ligation.

The *Bam*HI site was removed from pTA2182 by digesting the plasmid with *Bam*HI before being blunt-ended using Klenow fragment and self-ligated using T4 DNA Ligase. The resulting product, confirmed to be lacking the targeted *Bam*HI site by sequencing, was named pTA2290. pTA2290 was digested with *Bgl*II to remove the *p.tnaM3-polD2-hdrB* cassette and was inserted into pTA2217 at its *Bam*HI site. The ligation was digested with

*Bam*HI prior to transformation and the resulting successful plasmid named pTA2295 (**Figure 3.20**).

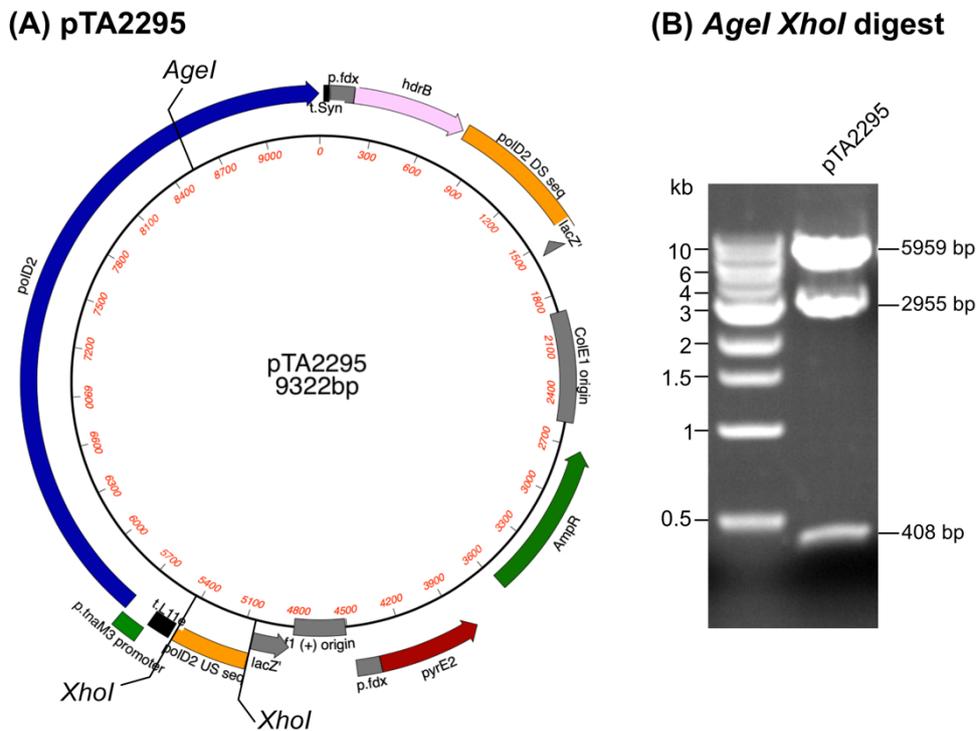


Figure 3.20: pTA2295. (A) Map of *p.tnaM3-polD2-hdrB* construct with genomic sequences for integration on the main chromosome. (B) Digestion of pTA2295 with *AgeI* and *XhoI* show bands at 5959 bp, 2955 bp and 408 bp, as expected.

Integration of full-length inducible polymerase constructs

Auxotrophic strain generation

To allow for selection of the inducible polymerase constructs, host strains needed to carry deletions for the selectable markers used. The pop-in/pop-out methodology relies on uracil selection and thus the strain must be Δ *pyrE2*. The inducible cassettes contain a *hdrB* marker for additional selection via thymidine auxotrophy and thus the strains must also be Δ *hdrB*. Due to the inducer of the promoters being tryptophan, the strains must also be *trpA*⁺ to ensure any difference seen in the absence of induction is due to the change in expression, not the auxotrophy of the strain (i.e., if the strain does not grow in the absence of tryptophan, it is due to the lack of polymerase expression as opposed to the absence of tryptophan synthesis).

Chapter 3: Cellular requirement for replicative DNA polymerases PolB and PolD in *Haloferax volcanii*

A wild type Δmrr strain, H4045, was previously generated in the H53 background strain prior to this project (Laura Mitchell, unpublished data). Strains H4045 and H4598 both carry the genotype $\Delta pyrE2 \Delta trpA \Delta mrr$, with H4045 being *oriC*⁺ and H4598 being $\Delta oriC$. The *hdrB* (HVO_2919) gene was deleted from strains H4045 and H4598 using deletion construct pTA155 (Allers *et al.*, 2004). This generated the $\Delta hdrB$ strains H4691 and H4695 respectively. Candidates were initially screened for thymidine auxotrophy on selective media, and genotypes were confirmed by Southern blot (**Figure 3.21**).

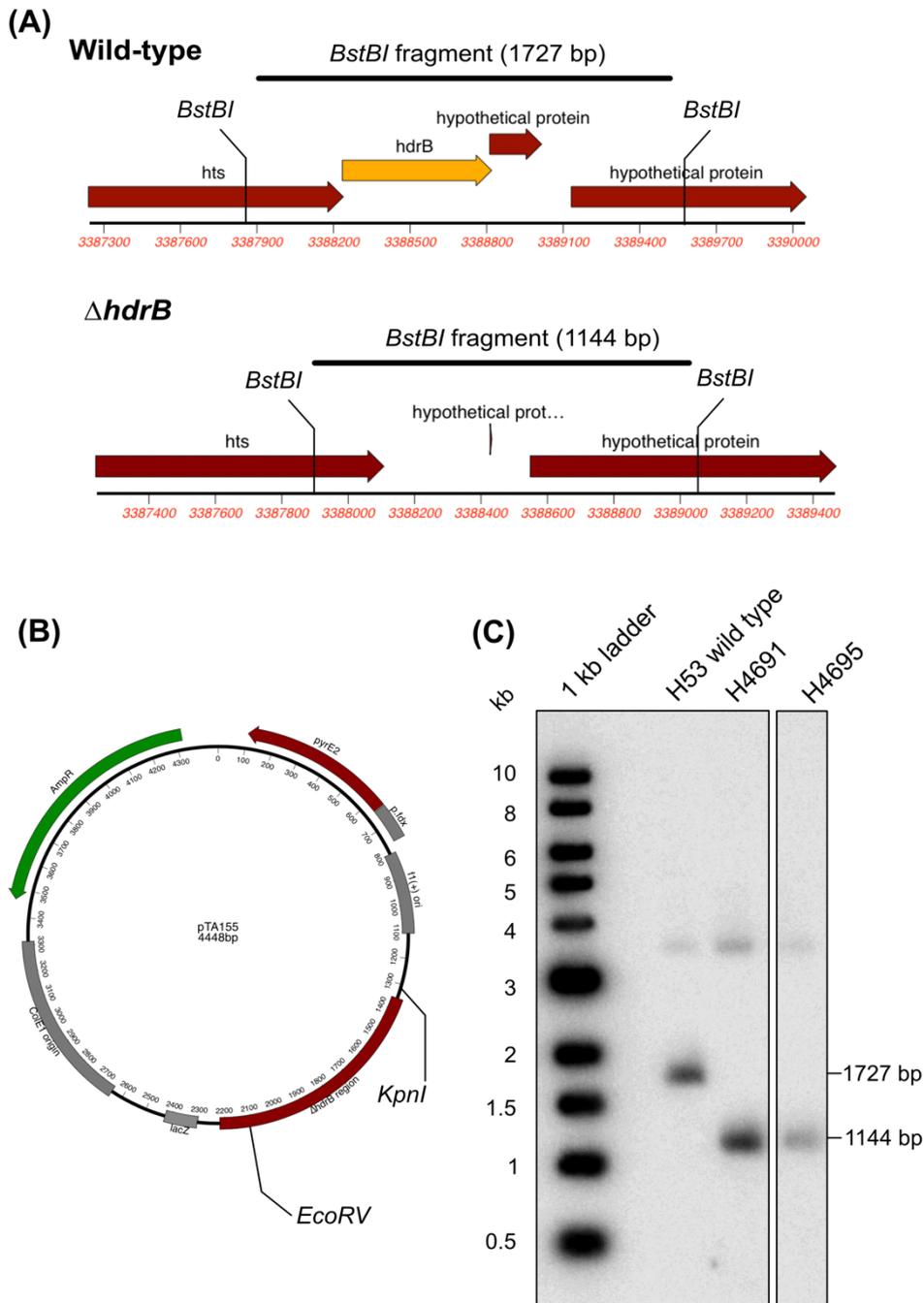


Figure 3.21: Deletion of *hdrB* (HVO_2919). (A) Expected Southern blot band sizes for *BstBI*-digested genomic DNA (wild type 1727 bp and Δ *hdrB* 1144 bp). (B) 764 bp Δ *hdrB* Southern probe consisting of a *EcoRV*-*KpnI* fragment of pTA155. (C) Southern blot confirming strain H4691 and H4695 as Δ *hdrB*. Faint band at \sim 3.5 kb likely due to non-specific binding of the probe.

Restoration of *trpA* was carried out in strains H4691 and H4695 via linear DNA transformation. The 2049 bp sequence encoding *trpA* and its

surrounding genomic sequence was digested from pTA49 (Lam *et al.*, 1990, Allers *et al.*, 2004) using *Bst*XI and *Bam*HI. Candidates were screened on selectable media lacking tryptophan, where colonies should only grow if they have successfully integrated the *trpA* gene. This gave rise to *trpA*⁺ strains H4829 (*oriC*⁺) and H4832 (Δ *oriC*).

Due to the use of an inteinless version of PolB1 during cloning and the inherent mobility associated with inteins, the inducible PolB1 construct would ideally be transformed into a strain deleted for *polB1-intein*. This had been shown to be possible (Naor *et al.*, 2011), but required screening at both the pop-in and pop-out stage as a consequence of the mobile nature of inteins. Due to the laborious nature of this process, deletion of Δ *polB1-intein* was carried out alongside trials of generating the inducible PolD2.

Full-length inducible polymerase integration

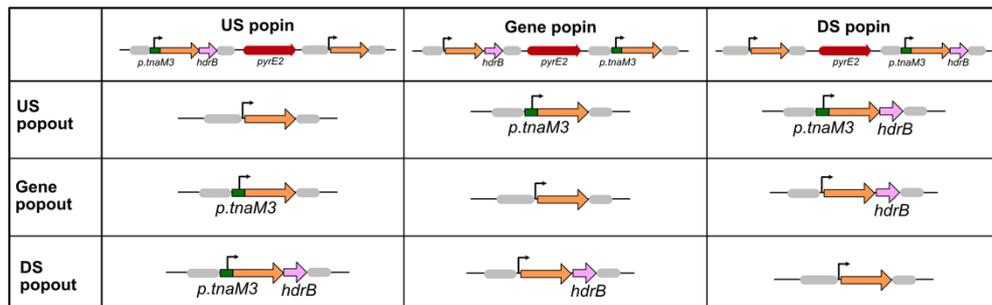
The aim is to utilise both *pyrE2* (uracil selection) and *hdrB* (thymidine selection) to ensure the correct integration of the inducible promoter at the target gene. Gene *pyrE2* allows for selection of pop-in events using media lacking uracil, and selection of pop-out events using 5-FOA. Gene *hdrB* will provide selection at both stages where colonies will only be capable of growth on media lacking thymidine.

Due to the multiple sections within the generated constructs that share homology with the chromosome, there are numerous pop-in and pop-out orientations that can occur (namely between the upstream [US] sequence, the gene itself, or the downstream [DS] sequence). The orientation of the pop-in event can be screened using PCR.

Due to the possible outcomes (summarised in **Table 3.5**), an US event will be required; once this US pop-in is confirmed, only a DS pop-out event would result in correct integration of the inducible promoter with the associated *hdrB* marker. This latter event can be selected for using media lacking thymidine and containing 5-FOA.

A DS pop-in and subsequent US pop-out can also result in the integration of the promoter and *hdrB* marker. However, a gene pop-out of this DS pop-in would result in the gene and *hdrB* marker being integrated into the chromosome, in the absence of the promoter. This, therefore, reduces the selection power of the pop-out from a DS pop-in as the ability to synthesise thymidine is not specific to the required product (unlike the US pop-in). A summary of the differing orientations and resulting pop-out products are summarised in **Table 3.5**.

Table 3.5: Possible outcomes of pop-in and pop-out events of different orientations for an inducible gene with added *hdrB* selection. Only an upstream (US) pop-in followed by a downstream (DS) pop-out or a DS pop-in followed by US pop-out give the inducible gene and *hdrB* cassette. However, the DS pop-in gives rise to two products with *hdrB* selection and thus only selection of *thy*-5-FOA^R candidates of an US pop-in would give the correct product.



Strains H4829 and H4832 were transformed with full-length inducible PolD2 plasmid pTA2295. Pop-ins were patched on Hv-Cas media and screened for pop-in orientation using colony PCR with primers *dp2USF* and *dp2intR* (**Figure 3.22**). While *dp2intR* will bind within the *polD2* sequence of both the inducible and wild type copies, *dp2USF* binds sequence upstream of the genomic sequence included in the plasmid and thus will only bind once within the integrated pop-in strain, upstream of the integration event. An US pop-in will give rise to a product of 1291 bp, while both DS and gene pop-in events give a product of 1007 bp.

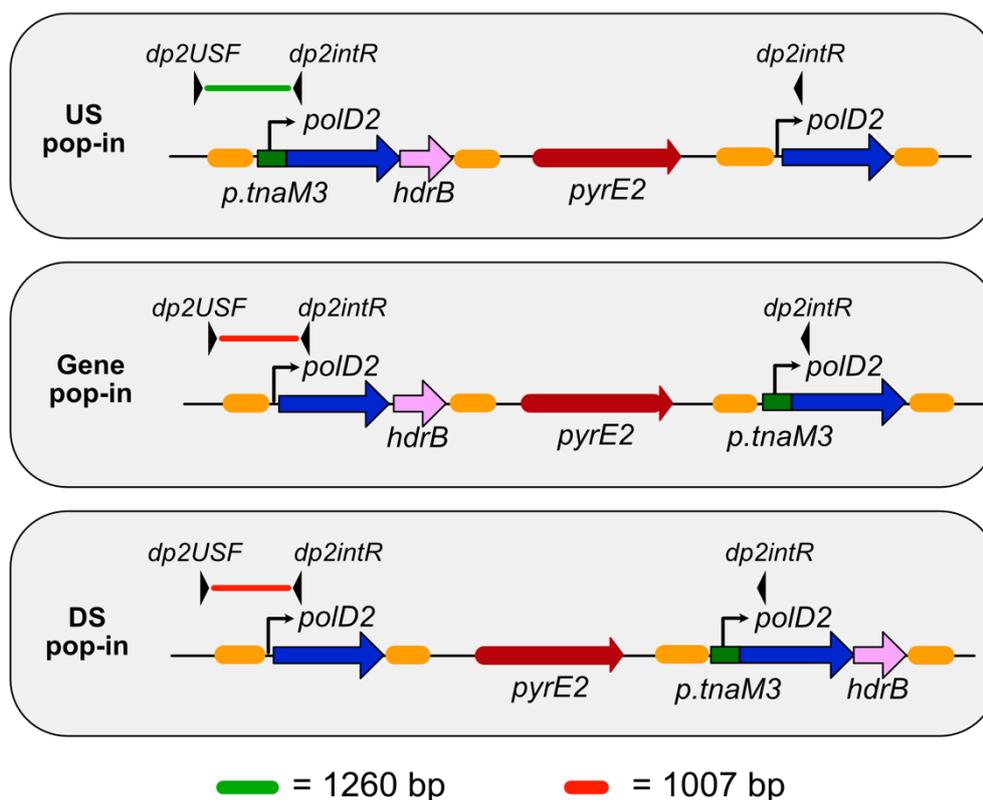


Figure 3.22: Colony PCR to screen for orientation of pop-in event for inducible *p.tnaM3::polD2::hdrB*. Colony PCR was performed on pop-in candidates using primers *dp2USF* and *dp2intR*. Elements coloured as in plasmid maps used previously; orange represents US/DS genomic sequence included within the plasmid. Where an US pop-in event occurs, the primers will give a product of 1260 bp, while a gene or DS pop-in event give a product of 1007 bp.

Of 146 pop-in candidates screened for both H4829 and H4832 transformants, no US pop-ins were isolated. This is likely due to the large size of the *polD2* gene; pTA2295 contains 391 bp of US sequence for the pop-in event to occur within, compared with 3609 bp of gene and 729 bp of DS sequence. It is likely that the same problem would apply to inducible *polB1* pTA2225 (486 bp US vs 2697 bp gene and 509 bp DS sequence).

To overcome this bias, cloning was undertaken to increase the size of the US sequence, and therefore increase the probability of an US pop-in event occurring. The US sequence present in both pTA2295 and pTA2225 was extended, giving rise to pTA2394 and pTA2382 respectively, now carrying 1213 bp and 1111 bp of US sequence respectively (**Figure 3.23**).

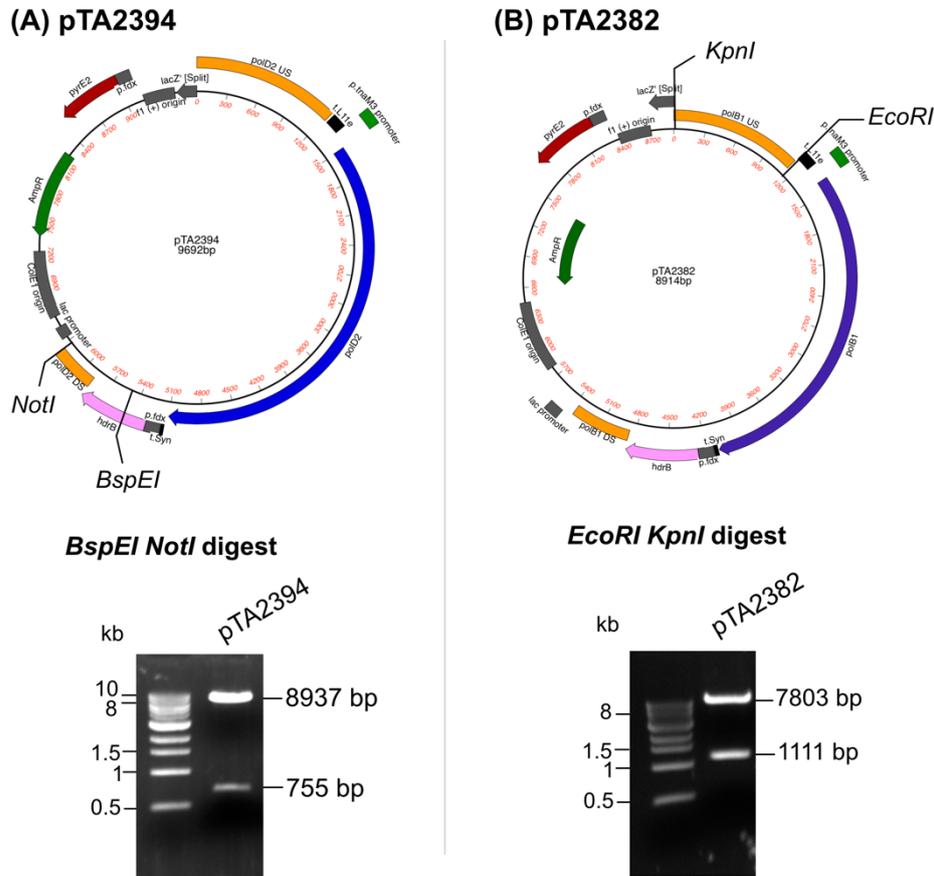


Figure 3.23: Addition of further upstream genomic sequence to inducible plasmids to give constructs (A) pTA2392 (*polD2*) and (B) pTA2382 (*polB1*). Digestion of pTA2394 with *NotI* and *BspEI* gives bands of 8937 bp and 755 bp, as predicted. Digestion of pTA2382 with *EcoRI* and *KpnI* gives bands of 7803 bp and 1111 bp, as predicted.

Using the newly enlarged inducible *polD2* construct, US pop-in candidates were able to be generated more readily (9 clones of 160 screened). However, a bias still existed, this time for the size of the DS portion of the plasmid; pop-out within the 729 bp of DS sequence was statistically not as likely as pop-out within the larger 3609 bp of *polD2* polymerase sequence. Alongside the relatively short DS sequence length decreasing the chance of a DS pop-out, the DS pop-out event would leave the polymerase under the control of *p.tnaM3*; a less favourable outcome for the cell than maintaining WT expression. Of the 9 US pop-in clones, pop-outs were attempted and patched +/- *trp* to assess whether under the control of *p.tnaM3* (assuming the gene is essential, pop-outs should not be able to grow in the absence of induction/tryptophan). Over 200 patches were screened, but no successful candidates were isolated.

Overcoming issues regarding integration of inducible promoters at large genes

To overcome the problems associated with such large inducible constructs containing the full-length gene, a new cloning strategy was devised. The addition of *hdrB* selection is known to work well to force genes of a shorter length under the inducible promoter (e.g., *radA*, ~1 kb) (Hawkins *et al.*, 2013a). Thus, the new strategy should implement *hdrB* selection while overcoming the bias associated with gene size and pop-in probabilities. The aim of this renewed strategy is to ensure larger genes are not reliant on performing US followed by DS recombination events, as the probabilities of both events occurring are highly skewed by the size of the gene between these sequences.

Instead, a truncated allele of the polymerase would be used, removing the majority of the gene sequence and all DS sequence. Thus, the gene would be placed under *p.tnaM3* directly at the pop-in stage, should a gene pop-in event occur (**Figure 3.24**). The *hdrB* marker can be placed in between US sequence and the gene, ensuring the same level of selection can be applied as with the full-length construct.

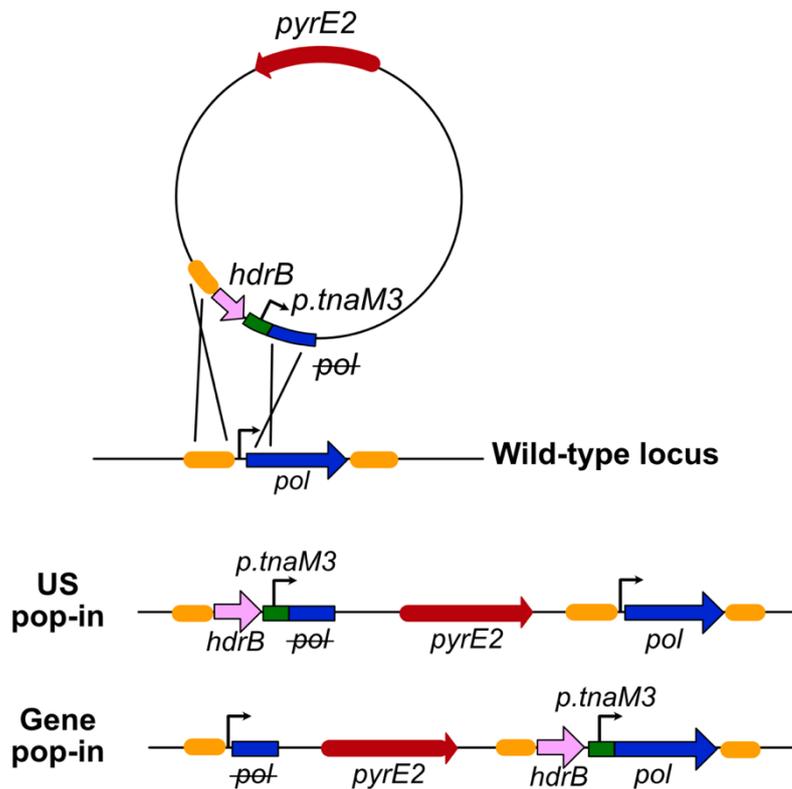


Figure 3.24: New strategy for placing large genes under the inducible promoter *p.tnaM3* using a truncated allele with *hdrB* selection. To force a gene under the inducible promoter at the pop-in stage, a truncated inducible allele must be cloned into a plasmid, with *hdrB* selection between the upstream sequence and the truncated gene (here marked with a strikethrough). A pop-in event between the gene sequences will lead to integration of the full-length gene under the inducible promoter, while an upstream (US) pop-in event will not.

When cloning the construct, by altering the lengths of the gene fragment vs US sequence, a bias can be introduced towards a gene pop-in, which would immediately place the wild type full-length gene under the control of *p.tnaM3*. A subsequent pop-out event with the US genomic sequence would leave the inducible promoter in place of the WT, removing the truncated gene and WT promoter. This strategy would allow for phenotypic screening at the level of the pop-in, and while the pop-out event could still revert to WT, it would not rely on altered expression of the polymerase compared to the pop-in and therefore is more favourable than the previous methodology.

Alongside the advantage of being able to screen for induction at the pop-in stage, this strategy also removes the requirement for a strain deleted for $\Delta polB1-intein$, as the truncated *polB1* sequence cloned into the new vector can be designed to only include sequence prior to that of the intein.

Previous work utilising varying lengths of DNA homology in an *H. volcanii* recombination assay allowed the frequency of recombination to be correlated with homology length (Jones, 2019) (**Figure 3.25**). The data displayed an exponential relationship, and the derived equation can be utilised to predict frequencies of recombination using DNA of a given length (**Equation 3.2**).

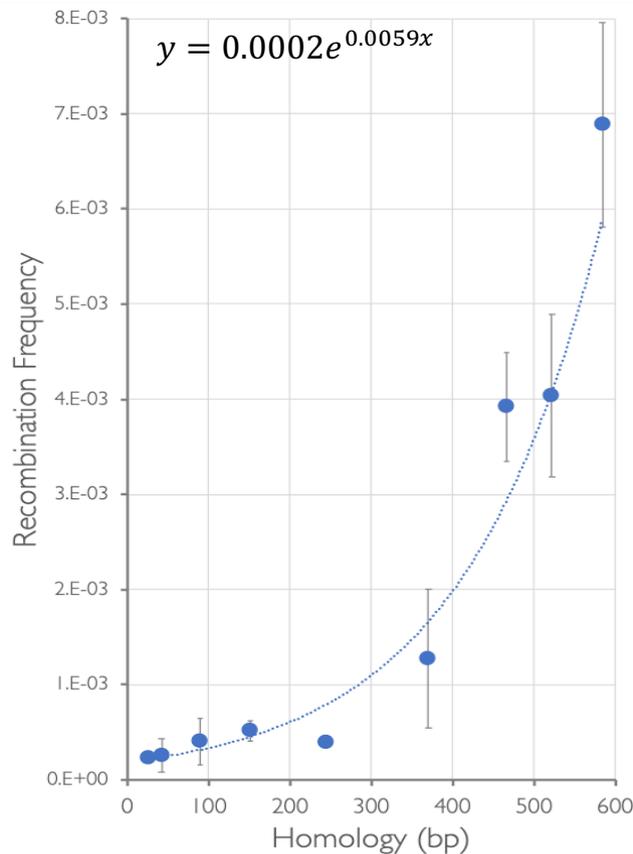


Figure 3.25: Recombination frequency of truncated *leuB* alleles in a wild-type background. The line of best fit is exponential with an equation of $y=0.0002e^{0.006x}$ and an R^2 value of 0.9514. Data from (Jones, 2019).

Equation 3.2: Calculating the frequency of recombination. Rearrangement of the exponential line equation will allow input of x bp to calculate frequency of recombination (f).

$$y = 0.0002e^{0.0059x}$$

$$f = \text{pow}(e, 0.0059 \times x) \times 0.0002$$

Cloning of a truncated polB1 (HVO_0858) inducible promoter construct

This new truncated allele strategy was primarily trialled for *polB1*. PCR amplification of the full-length inducible construct pTA2225 using primers *polB1USKpnF* and *polBintBamR* generated the 225 bp US genomic sequence, the *p.tnaM3* cassette and the first 1383 bp of *polB1* sequence. Utilising the previous equation, this would give rise to recombination frequencies of $f.US = 0.00075$ and $f.gene = 0.699$, providing a 932-fold bias to a gene pop-in occurring over an US pop-in.

Primers *polB1USKpnF* and *polBintBamR* integrated novel *KpnI* and *BamHI* sites at the 5' and 3' ends of the product respectively. The PCR product was digested with *KpnI* and *BamHI* and inserted into vector pTA2422 at its respective sites (**Figure 3.26 A**). This generated the intermediate plasmid pTA2515 for truncated inducible *polB1* (**Figure 3.26 B**).

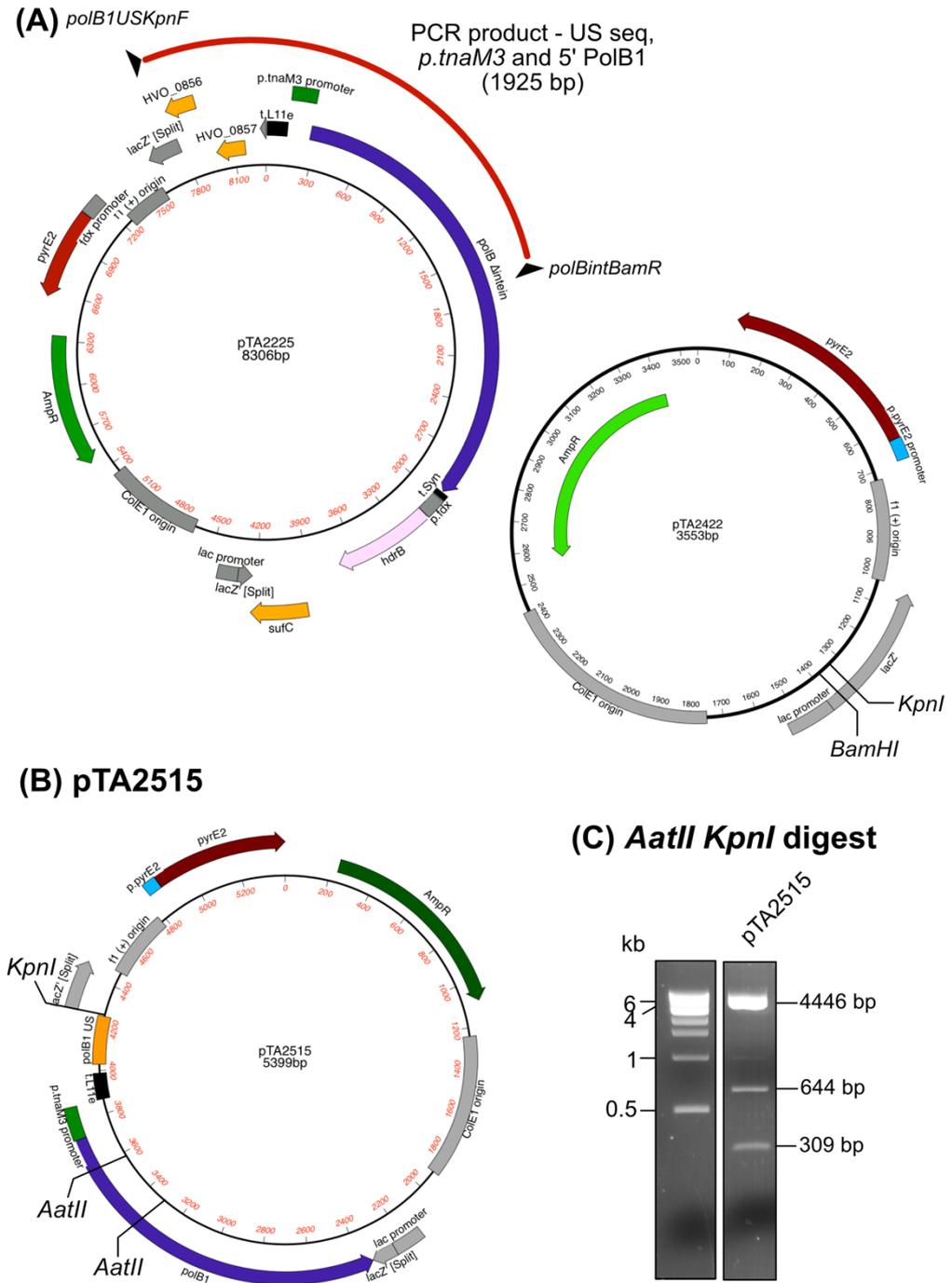


Figure 3.26: Construction of intermediate plasmid pTA2518 containing a truncated inducible *polB1* allele. (A) PCR of pTA2225 using primers *polB1USKpnF* and *polBintBamR* amplified a 1925 bp product containing *polB1* upstream (US) sequence, *p.tnaM3* and the 5' end of the *polB1* gene. Primers integrated *KpnI* and *BamHI* sites, which were utilised for digestion and ligation into vector pTA2422 at compatible *KpnI/BamHI* sites. (B) This generated plasmid pTA2515, containing a truncated inducible allele of *polB1*. (C) Digestion of pTA2515 with *AatII* and *KpnI* shows bands at 4446 bp, 644 bp and 309 bp as predicted.

Chapter 3: Cellular requirement for replicative DNA polymerases PolB and PolD in *Haloferax volcanii*

To add *hdrB* selection, the *hdrB* gene was amplified by PCR from the inducible-*mcm* intermediate plasmid pTA1407 (Marriott, 2017). Primers *pfdxhdrBHindF* and *pfdxhdrBEcoR* amplified *hdrB* and its associated promoter *p.fdx*, while inserting novel *HindIII/EcoRI* sites into the product. The product was digested with *HindIII* and *EcoRI* and inserted into pTA2515 at its compatible *HindIII/EcoRI* sites, located between the US genomic sequence and the *p.tnaM3* promoter. This gave rise to truncated inducible *polB1* construct pTA2518 (**Figure 3.27**).

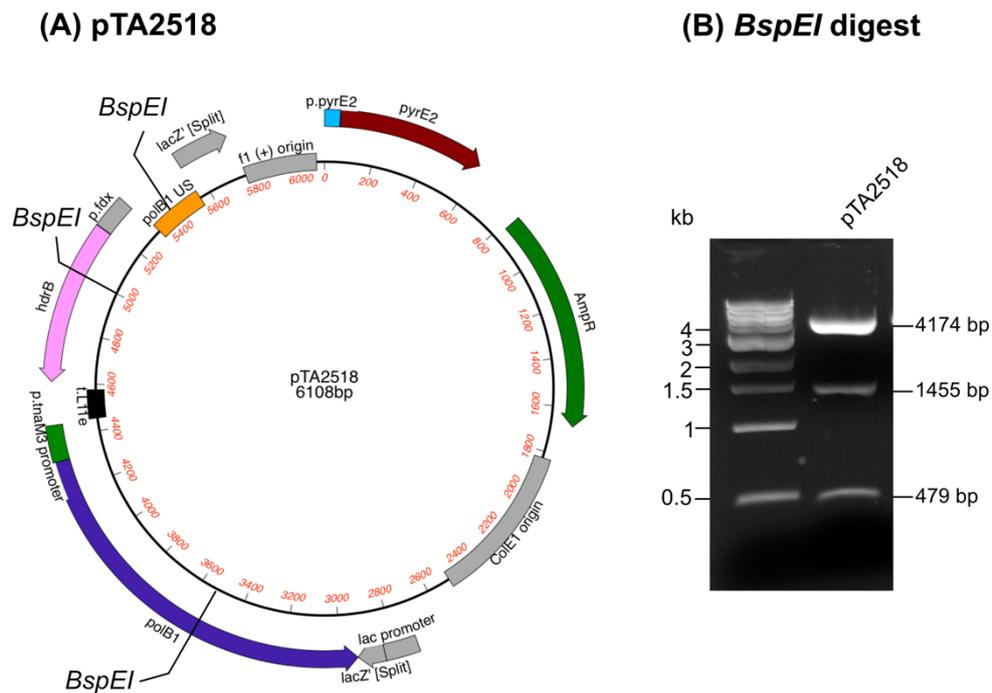


Figure 3.27: pTA2518. (A) Map of truncated inducible *p.tnaM3::polB1* allele with added *hdrB* selection. (B) Digestion of pTA2518 with *BspEI* shows bands at 4174 bp, 1455 bp and 479 bp, as expected.

Preliminary screen of pTA2518 pop-ins

Prior to cloning of a truncated *polD2* construct, a preliminary screen was undertaken of pTA2518 pop-ins to ensure that the gene pop-ins integrated in both US and gene orientations as expected. H4829 was transformed with pTA2518 and pop-in candidates were screened for sensitivity to tryptophan using selective media (**Figure 3.28 A**), in addition to colony PCR to confirm pop-in orientation (**Figure 3.28 B**). Primers *pBSF2* and *polR* amplify differentially sized products depending on the orientation of the pop-in event;

Chapter 3: Cellular requirement for replicative DNA polymerases PolB and PolD in *Haloferax volcanii*

US pop-ins will give a product of 1,111 bp while gene pop-ins will give a product of 2,112 bp. Of 80 clones screened, several showed some level of tryptophan auxotrophy (**Figure 3.28 A**); colony PCR identified six gene pop-in candidates (**Figure 3.28 C**). This result validated the truncated method of generating inducible mutants and thus cloning was carried out for *polD2*. It is worth noting that primer *pBSF2* used for colony PCR binds within the plasmid backbone and thus this PCR screen alone is not sufficient to ensure all genome copies have integrated a copy of the plasmid; steps to ensure correct strain generation using truncated alleles are discussed in detail following the construction of a truncated inducible *polD2* construct.

Chapter 3: Cellular requirement for replicative DNA polymerases PolB and PolD in *Haloferax volcanii*

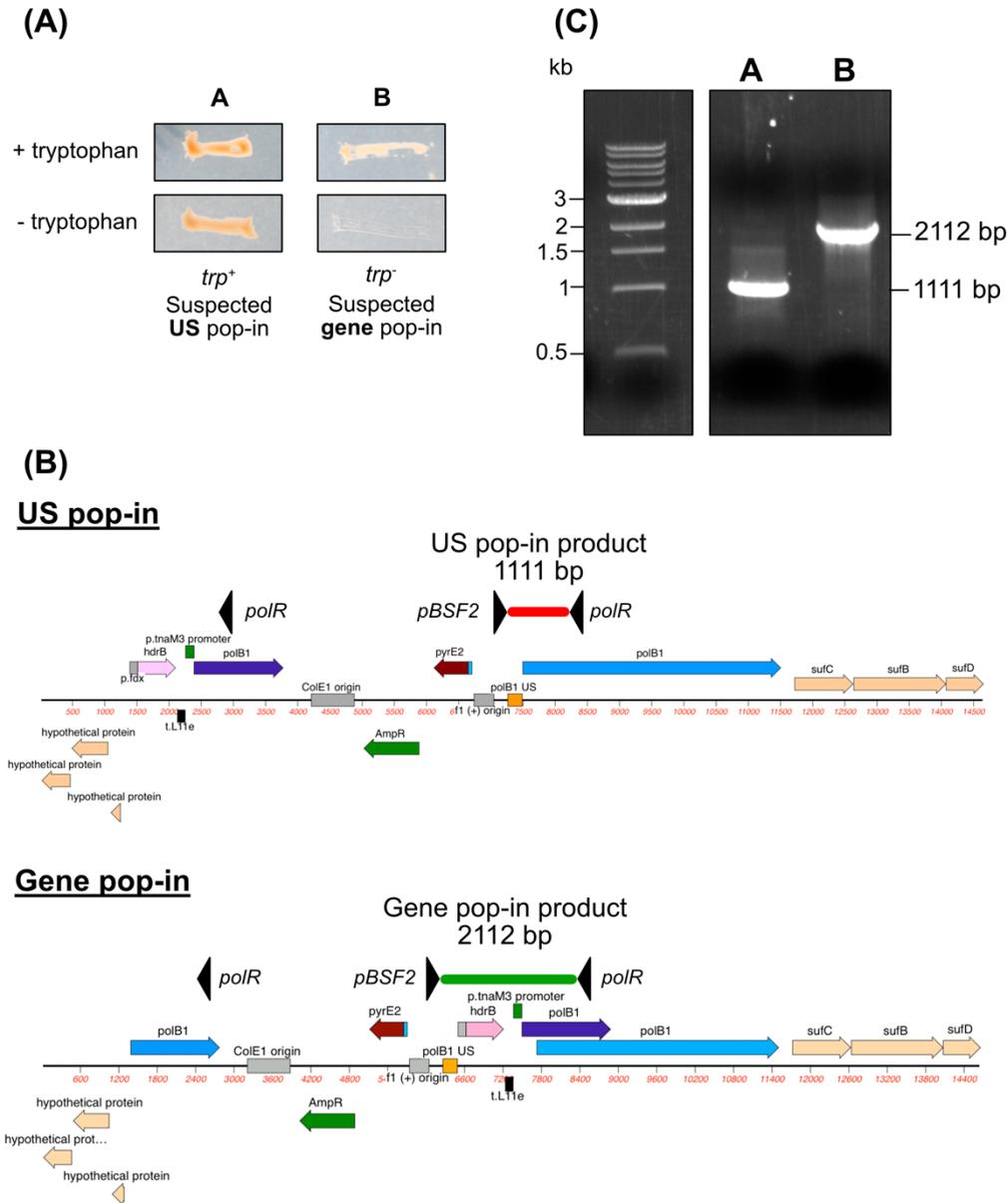


Figure 3.28: Screening for pop-in orientation of truncated inducible *p.tnaM3::polB1* plasmid pTA2518. (A) Candidates were patched on media containing (*trp*⁺) and lacking (*trp*⁻) tryptophan. Clones that cannot grow in the absence of tryptophan (e.g. clone B) are likely to have undergone a gene pop-in event, which places essential gene *polB1* under tryptophan-inducible promoter *p.tnaM3*, while upstream (US) pop-in candidates maintain wild type expression (e.g. clone A). (B) Pop-in orientation may be confirmed using colony PCR. Primers *pBSF2* and *polR* amplify differentially sized products depending on whether an US pop-in event has occurred (1111 bp) or whether a gene pop-in has occurred (2112 bp). (C) Colony PCR products showing clone A is an US pop-in while clone B is a gene pop-in.

Cloning of a truncated inducible promoter vector

The apparent success of the truncated inducible *polB1* trial suggested this technique could be applied not only to polymerase genes, but to all large genes where it can be imagined that gene size would be problematic with respect to selecting the correct pop-in and pop-out orientation. Therefore, prior to cloning truncated inducible *polD2*, an empty cassette with 5' *hdrB* and *p.tnaM3* was created as a future resource for integration of *p.tnaM3* at other such large genes.

The 1283 bp *hdrB::p.tnaM3* cassette from truncated inducible *polB1* plasmid pTA2518 was isolated by PCR using primers *RBDX1* and *ptnaM3rev_Not*; the latter primer introduced a novel *NotI* site into the product. The product was digested with *NotI* and *HindIII* (a site present within the template and therefore the product) and inserted into pTA2422 at compatible *NotI/HindIII* sites, generating the empty 5' *hdrB::p.tnaM3* vector pTA2554 (Figure 3.29).

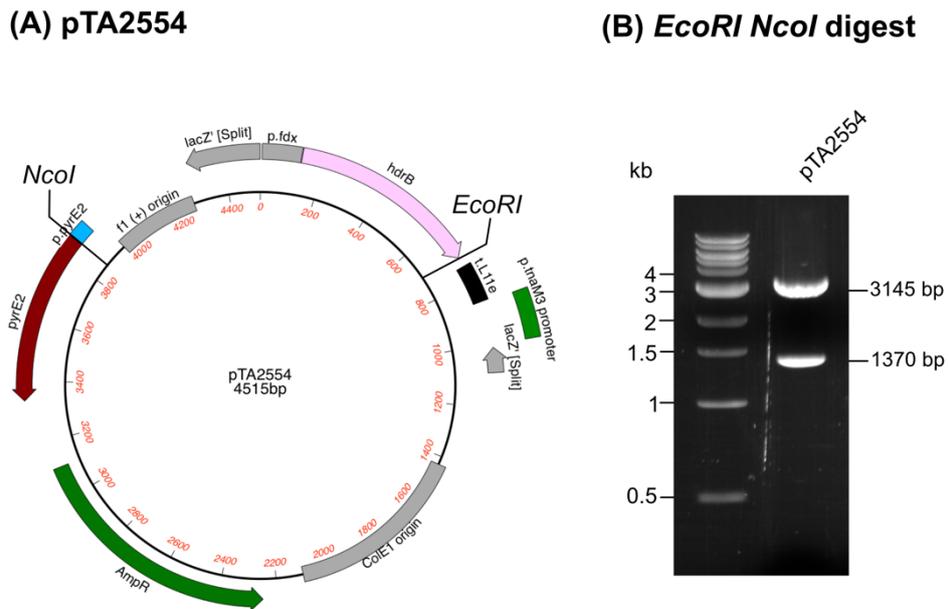


Figure 3.29: pTA2554. (A) Map of pTA2554, empty 5' *hdrB::p.tnaM3* vector for creation of truncated inducible alleles. (B) Digestion of pTA2554 with *EcoRI* and *NcoI* shows bands at 3145 bp and 1370 bp, as expected.

Due to the disruption of the multiple cloning site of pTA2422, *KpnI* and *HindIII* are available for integration of the US sequence, while insertion of the truncated gene of choice can be carried out using *NdeI* and *NotI* sites.

Cloning of a truncated polD2 (HVO_0065) inducible promoter construct

Cloning was designed to introduce 509 bp of US sequence and 1766 bp of the 5' end of *polD2*. Utilising the previous equation, this would give rise to recombination frequencies of $f_{US} = 0.004$ and $f_{gene} = 6.7$, providing a 1765-fold bias to a gene pop-in occurring over an US pop-in.

PCR using full-length inducible *polD2* construct pTA2295 was performed using primers *polD2USF* and *dp2intNotR* to amplify a 1766 bp fragment of *polD2*. The product was digested using an internal *NdeI* site present in the template/product and newly incorporated *NotI* at the 3' end of the gene fragment and inserted into inducible vector pTA2554. This gave rise to intermediate truncated *polD2* plasmid pTA2555 (**Figure 3.30**).

(A) pTA2555

(B) *MluI XmnI* digest

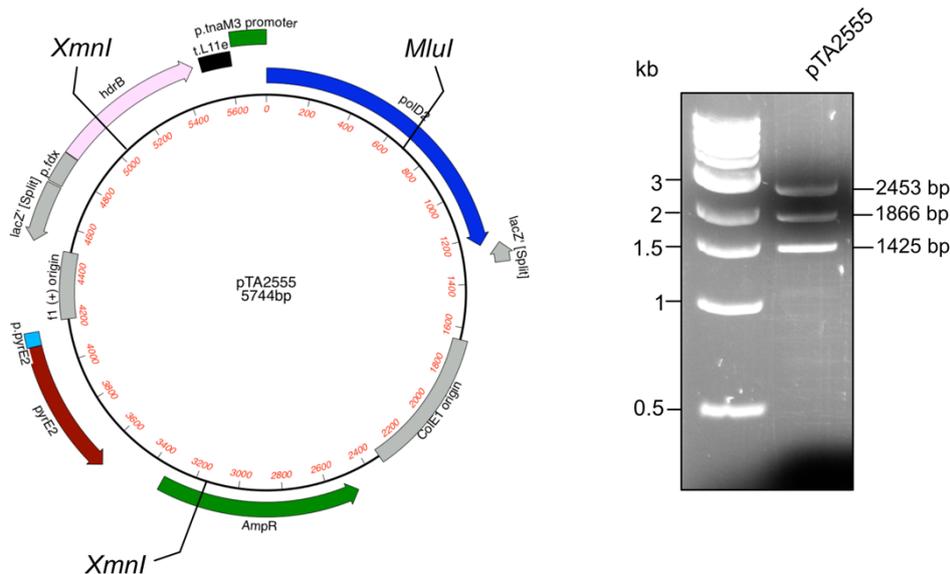


Figure 3.30: pTA2555. (A) Map of pTA2555, intermediate in creation of a truncated *polD2* inducible plasmid. (B) Digestion of pTA2555 with *MluI* and *XmnI* shows bands at 2453 bp, 1866 bp and 1425 bp, as expected.

Sequence US of *polD2* was amplified from pTA2295 by PCR using primers *pBSF2* and *dp2USHindR*, with the latter primer introducing a novel *HindIII* site. This 509 bp product was digested at its internal *KpnI* site and novel *HindIII* site and inserted into intermediate plasmid pTA2555 at compatible *KpnI/HindIII* sites. This gave rise to truncated inducible *polD2* plasmid pTA2560 (**Figure 3.31**).

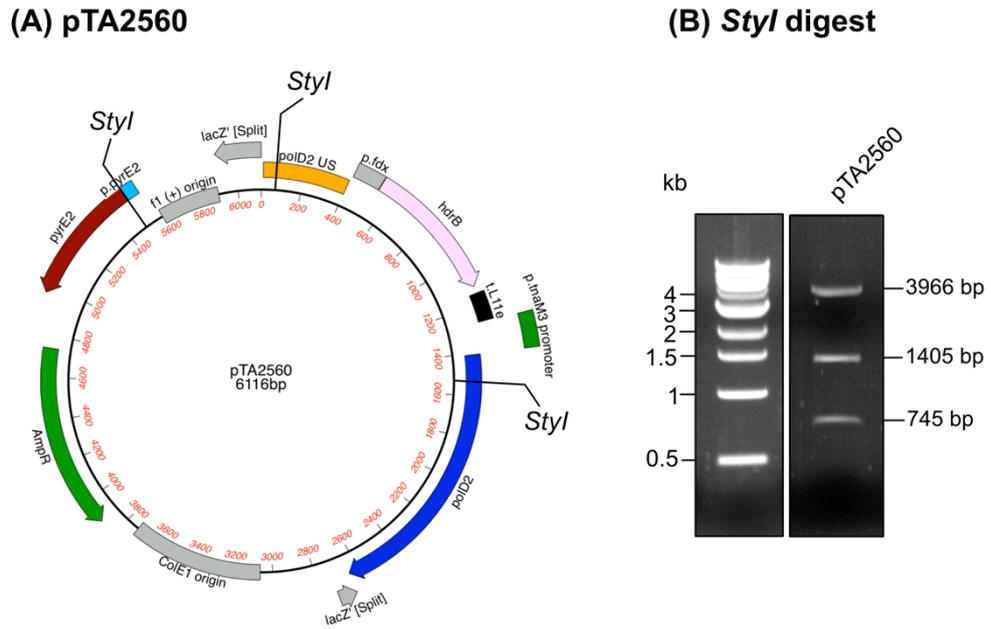


Figure 3.31: pTA2560. (A) Map of pTA2560, truncated *polD2* inducible plasmid with 5' *hdrB* selection. (B) Digestion of pTA2560 with *StyI* shows bands at 3966 bp, 1405 bp and 745 bp, as expected.

Truncated inducible polymerase strain construction

Strains H4829 (*oriC*⁺) and H4832 (Δ *oriC*) were transformed with truncated inducible *polB1* plasmid pTA2518 and inducible *polD2* plasmid pTA2560. Pop-in candidates were patched on Hv-Cas +Ura agar plates with and without tryptophan, to screen for gene pop-ins (which will directly be under the control of *p.tnaM3*). While numerous candidates showed some level of tryptophan auxotrophy when first patched, suggesting correct integration of the promoter, the observed tryptophan auxotrophy did not align with the results of the PCR (where the promoter was not integrated). The PCR screen for PolD2 pop-in candidates is shown in **Figure 3.32** (PolB1 pop-in screening details shown in **Figure 3.28**).

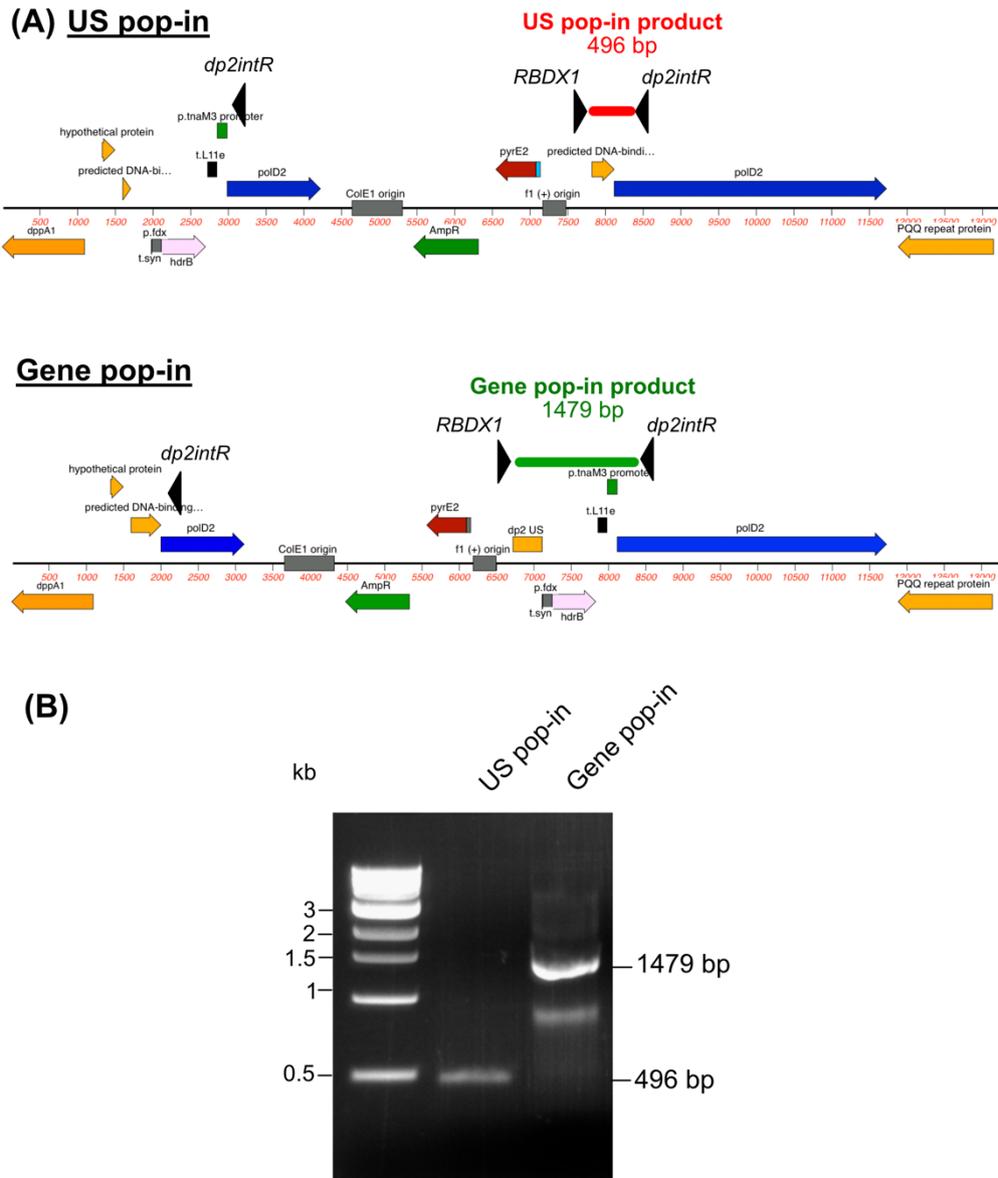


Figure 3.32: Screening for pop-in orientation of truncated inducible *p.tnaM3::polD2* plasmid pTA2560. (A) Pop-in orientation can be confirmed using colony PCR. Primers *RBDX1* and *dp2intR* give rise to differentially sized products depending on whether an US pop-in event has occurred (496 bp) or whether a gene pop-in has occurred (1479 bp). (B) Colony PCR products showing US pop-in and gene pop-in products.

Any gene pop-ins identified by PCR were then restreaked on both Hv-Cas +Ura with and without tryptophan. Of 17 *polB1* and 12 *polD2* gene pop-ins restreaked, all re-gained the ability to grow in the absence of tryptophan. PCR on these now *trp*⁺ clones revealed they still had a product associated with a gene pop-in and thus it is likely that the pop-in was merodiploid and some genome copies retained wild type *polB1*. This flexibility in the pop-in status

Chapter 3: Cellular requirement for replicative DNA polymerases PolB and PolD in *Haloferax volcanii*

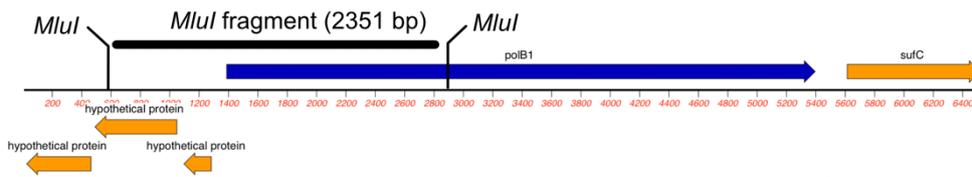
suggested that screening at the pop-in stage alone may not be inherently reliable, likely a result of the high polyploidy of *H. volcanii*.

To maximise the probability of generating an inducible clone for each polymerase, one gene pop-in and one US pop-in (as confirmed by PCR) were subjected to pop-out out-growth, with an overnight culture being diluted into fresh Hv-YPC daily for five subsequent overnights. This extended out-growth allows increased time for the pop-out event to occur fully, following which cultures were plated on Hv-Cas +Trp +5-FOA. For each strain, 240 pop-out candidates were patched in duplicate on Hv-Cas +Ura +/-Trp. Of the total of 960 colonies patched, only one had a *trp*- phenotype; this was a product of H4829 being transformed with inducible *polB1* plasmid pTA2518 (the product of an US pop-in).

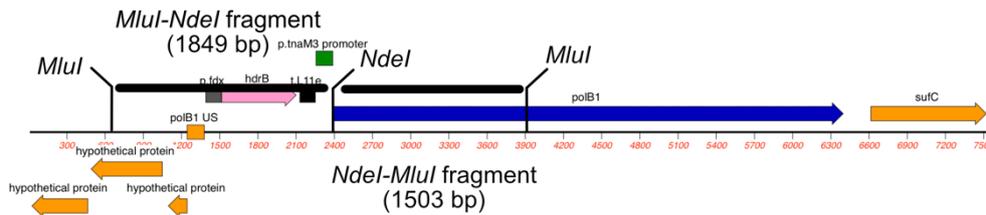
When restreaked, this single candidate was consistently *trp*- and therefore was further screened for *polB1* status by Southern blot (**Figure 3.33**).

Chapter 3: Cellular requirement for replicative DNA polymerases PolB and PolD in *Haloferax volcanii*

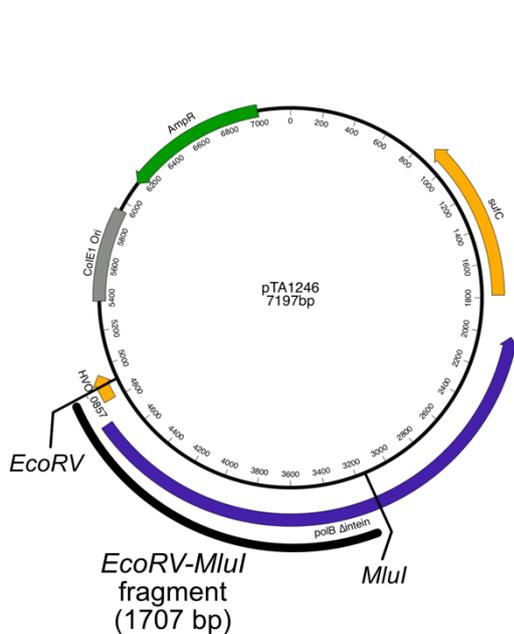
(A) Wild-type



hdrB+::p.tnaM3-polB1



(B) Southern probe



(C) Southern blot

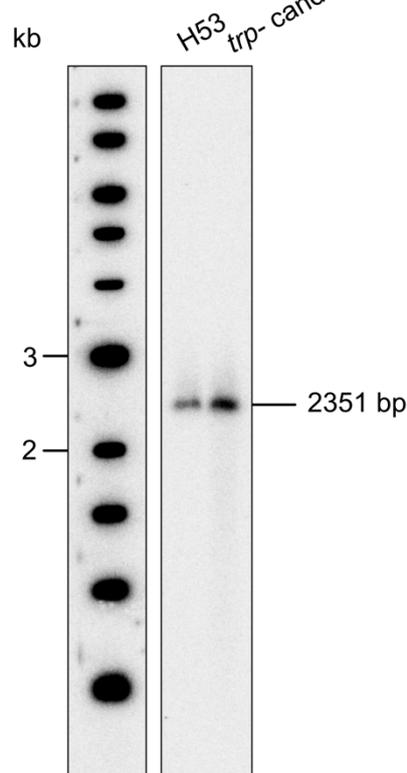


Figure 3.33: Screening of *trp-* candidate for pop-out of pTA2518 (*hdrB+::p.tnaM3-polB1*). (A) Expected Southern blot band sizes for *NdeI* and *MluI* digested genomic DNA (wild type 2351 bp and *hdrB+::p.tnaM3-polB1* 1849 bp and 1503 bp). (B) 1707 bp *polB1* Southern probe consisting of a *EcoRV-MluI* fragment of pTA1246. (C) Southern blot confirming *trp-* candidate is wild type at the *polB1* locus.

Chapter 3: Cellular requirement for replicative DNA polymerases PolB and PolD in *Haloferax volcanii*

Southern blotting showed that the *trp*- candidate was in fact wild type at the *polB1* locus. This is likely due to contamination with a $\Delta trpA$ strain, or a mutation at the *trpA* locus explaining the inability of this candidate to synthesise tryptophan. The high number of candidates screened failing to generate an inducible strain suggests inherent problems with placing the replicative polymerases under *p.tnaM3*. Thus, it was not possible to generate these strains during the timeframe of this project.

3.3.5 CRISPR interference (CRISPRi) against replicative polymerases

Alongside cloning and integration of the inducible promoters, an alternative strategy was designed to investigate the altered requirement for DNA polymerases in *H. volcanii* in the absence of origins. As opposed to controlling expression levels, as with the inducible promoters, CRISPR interference (CRISPRi) blocks the WT promoter of genes, assessing the impact of transcription inhibition for both PolB1 (HVO_0858) and PolD1 (HVO_003). Whilst PolD2 is the major catalytic subunit of the PolD complex, its promoter overlaps with a gene of unknown function, meaning targeting the promoter of PolD2 may lead to unknown off-target effects. PolD1, the exonuclease domain of PolD, has been shown to be required for full functionality of PolD (Cann *et al.*, 1998, Shen *et al.*, 2004), therefore direct targeting of PolD1 should still present the phenotype expected for the effect of PolD2 (and more generally PolD) knockdown.

CRISPR-derived RNA (crRNA) design

Spacer sequences targeting the promoters of both PolB1 (HVO_0858) and PolD1 (HVO_0003) (listed in **Table 3.4**) were designed by members of the group of Professor Anita Marchfelder (University of Ulm, Germany). For each gene, PAM sequences within the promoter were identified to ensure correct binding of the crRNA at the targeted spacer sequence. *H. volcanii* CRISPR-Cas9 is activated by six different PAM sequences: TTC, ACT, TAA, TAT, TAG, and CAC (Maier *et al.*, 2019). One of these sequences was integrated into each spacer sequence designed, along with a 5' handle and the homologous seed sequence. Stachler & Marchfelder (2016) previously assessed if spacer position within the promoter affected the efficiency of transcription inhibition, however it was shown that inhibition was not determined by position within the promoter.

For each gene targeted, three separate spacer sequences were designed, to increase likelihood of a successful knockdown, as different crRNAs have been shown to have different levels of interference activity (Maier *et al.*, 2013). All sequences target the template strand in the region of the promoter and transcriptional start site, as inhibition of transcription initiation leads to the most efficient knockdown (Anita Marchfelder, personal communication). The spacer sequences designed were 36 base pairs in length, fitting within the normal range for *H. volcanii* CRISPR spacer sequences (34-39 nucleotides) (Maier *et al.*, 2015a).

Table 3.4: Spacer sequences for CRISPRi targeting replicative polymerases PolB1 and PolD1. PAM sequences are highlighted within the spacer sequence in green.

Gene targeted		Spacer sequence
PolB1 (HVO_0858)	Spacer #1	cgggtgcggttcgcggaacgccggggttttagcc
	Spacer #2	gcggaacgccggggttttagccgcgccgccgaag
	Spacer #3	ccgcgccgccgaagcgattgccatgacgcagacggg
PolD1 (HVO_0003)	Spacer #1	aacgttcgggttcggcatccttttgccgcgtg
	Spacer #2	ggcatccttttgccgcgtgctgcgcactccgggtg
	Spacer #3	ccgggtgtgccactggagacgccggcgccgcgcgtg

Creating shuttle vectors for crRNA expression

A system for expression of crRNAs in a Cas6-independent manner had already been previously established for *H. volcanii* (Maier *et al.*, 2015b). This system utilises episomal plasmid expression, whereby the plasmid contains a *Haloferax* replication origin and, as such, can maintain itself in the cell without integrating onto the chromosome. The spacer sequence will be engineered into a spacer cassette, driven by the high expression promoter *p.syn* and will be flanked by an 8 nucleotide 5' handle (same length as would normally be generated by Cas6) and a 5' and 3' t-element; t-elements are tRNA-like structures that are key for CRISPR function (Stachler *et al.*, 2017). These t-elements will allow recognition by RNase P and tRNase Z, which will process the crRNA into its mature form. This processing ensures the ends of the crRNA match those of the 'natural' crRNA and thus will not be targeted for degradation.

Plasmids for expression of crRNA were generated in a two-step process: 1) Inverse PCR of vector pMA-telecRNA19 (Maier *et al.*, 2015b) integrating designed spacer sequences within the cassette and 2) Transfer of the spacer cassette to the episomal vector pTA232 (Allers *et al.*, 2004) (**Figure 3.34**). This two-step process allowed the accurate inverse amplification of the relatively small template plasmid (pMA-telecRNA19, 2601 bp) before being placed into the much larger vector pTA232 (7818 bp).

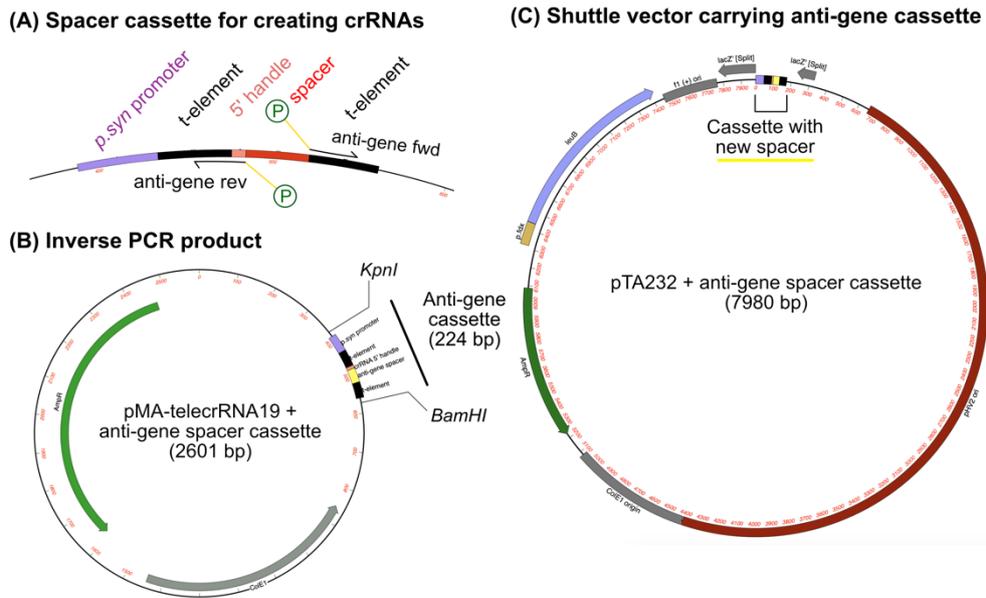


Figure 3.34: Construction of shuttle vectors for production of crRNAs. (A) pMA-telecRNA19 contains a spacer cassette, consisting of the *p.syn* promoter, a crRNA 5' handle, the anti-gene spacer sequence and 5' and 3' flanking t-elements (Maier, Stachler *et al.*, 2015). To introduce a new spacer sequence, primers are engineered with homology to the template (black) as well as the new spacer sequence (yellow). (B) The product of the inverse PCR will contain the new spacer sequence within the anti-gene cassette. Restriction digest with *KpnI* and *BamHI* will remove the spacer cassette. (C) *KpnI/BamHI* sites in pTA232 allow insertion of the cassette into *ori*-pHV2 shuttle vector pTA232 for expression of the crRNA in the cell.

Inverse PCR was used, with each primer containing 18 nt of the new spacer sequence while omitting the spacer sequence present in the template. The primers used for each inverse PCR are listed in **Table 3.7**; sequence homologous to the template is in black, while the spacer sequence for integration is in red. Primers were phosphorylated using PNK, allowing the product to self-ligate in the presence of T4 DNA Ligase. The annealing temperature was calculated using only the primer sequence homologous to the template. Following the self-ligation of the PCR product, the initial template DNA was selectively digested using *DpnI* and candidates were screened by sequencing. Anti-polB1 spacer sequences #1, #2 and #3 gave rise to intermediate plasmids pTA2227, pTA2228 and pTA2293 respectively. Anti-polD1 spacer #2 gave rise to intermediate plasmid pTA2334, while inverse PCRs to introduce spacers #1 and #3 repeatedly failed to yield a product.

Table 3.7: Primers used for integration of spacer sequences in pMA-telecRNA19. Sequence homologous to the template is in black, spacer sequence for integration is in red.

Gene target	Primer	Spacer sequence
PolB1 (HVO_0858)	polB1#1antirev	TTCCGCGAACCGCACCCGGCTTC AACTACCGATCA
	polB1#1antifwd	ACGCCGGGGTTTTAGCCACCGA TATTGGTATGGC
	polB1#2antirev	AAACCCCGGCGTTTCCGCGCTTC AACTACCGATCA
	polB1#2antifwd	TTAGCCGCGCCGCCGAAGCCGAT ATTGGTATGGC
	polB1#3antirev	ATCGCTTCGGCGGGCGGGCTTC AACTACCGATCA
	polB1#3antifwd	TGCCATGACGCAGACGGGACCGA TATTGGTATGGC
PolD1 (HVO_0003)	polD1#1antirev	GCCGAAACCCAGAACGTTGCTTC AACTACCGATCA
	polD1#1antifwd	ATCCTTTTGGCCGCGTGTACCGAT ATTGGTATGGC
	polD1#2antirev	CGCGGCAAAAAGGATGCCGCTTC AACTACCGATCA
	polD1#2antifwd	TGTCGCGCACTCCGGGTGACCGA TATTGGTATGGC
	polD1#3antirev	CTCCAGTGGCACACCCGGGCTTC AACTACCGATCA
	polD1#3antifwd	ACGCCGGCGCGCATCGTCACCGA TATTGGTATGGC

The spacer cassette was then digested from pTA2227, pTA2228, pTA2293 and pTA2334 using *KpnI* and *BamHI*. This 224 bp fragment was ligated into the compatible *BamHI* and *KpnI* sites of vector pTA232. This gave the final shuttle cassette for expression of crRNAs against polB1: #1 pTA2251, #2 pTA2252 and #3 pTA2305 and polD1: #2 pTA2342. pTA2251 is shown in **Figure 3.35** as an example.

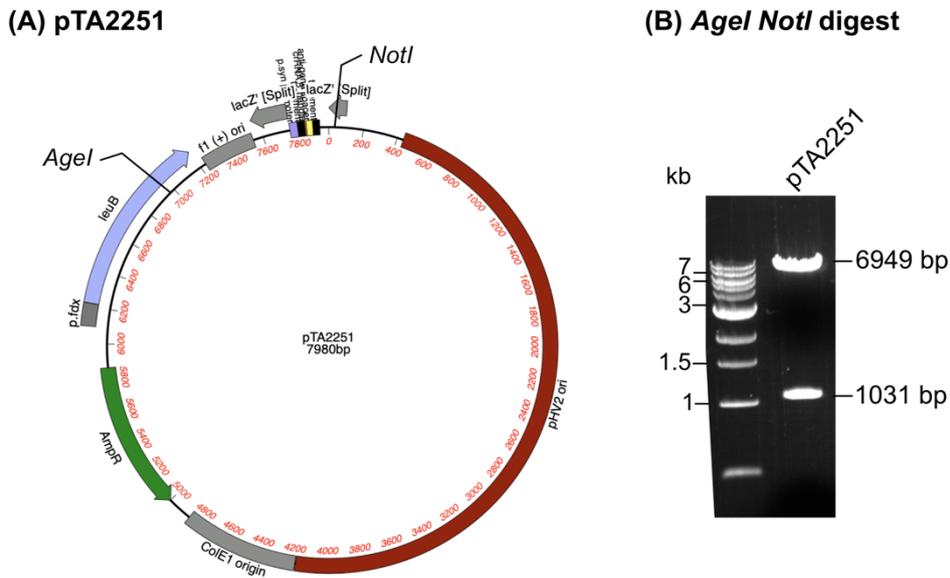


Figure 3.35: pTA2251. (A) The anti-gene cassette was cut from the pMA-telecRNA19 inverse PCR product using *KpnI* and *BamHI* and placed into pTA232 at *KpnI* and *BamHI* sites. The same methodology was used for all anti-gene cloning; pTA2251 is shown as an example. (B) Digestion with *AgeI* and *NotI* showed bands at 6949 bp and 1031 bp, as predicted.

CRISPRi strain construction

Deletion of cas3 (HVO_A0209)

Cas3 is usually recruited by the Cascade complex to degrade foreign DNA. Degradation of the gene targeted using CRISPRi was not the goal, instead merely inhibiting its transcription through binding of Cascade to the promoter. Therefore, strains for use in CRISPRi were generated that were deleted for *cas3* (HVO_A0209) in a wild type (*ori+*; H53) and originless (Δ *ori*; H1804) background. The *cas3* gene was deleted from strains H53 and H1804 using deletion construct p131-updo(*cas3*) (Stachler and Marchfelder, 2016). This gave rise to Δ *cas3* strains H4385 and H4387 respectively. These strains were constructed by Rebecca Lever and confirmed by colony hybridisation and Southern blotting (data not shown).

Deletion of cas6 (HVO_A0205)

Cas6 is involved in maturing crRNAs for use in CRISPR. CRISPRi relies on the artificial crRNA added being the only mature crRNA present; in the absence of Cas6, only the plasmid-derived crRNA will be mature. Therefore, *cas6* was deleted in strains used for CRISPRi. The *cas6* gene

Chapter 3: Cellular requirement for replicative DNA polymerases PolB and PolD in *Haloferax volcanii*

(HVO_A0205) was deleted from $\Delta cas3$ strains H4385 and H4387 using deletion construct p131-updo(*cas6*) (Brendel *et al.*, 2014). This gave rise to $\Delta cas6$ strains H4606 and H4608 respectively. Deletion of the *cas6* gene was confirmed by colony hybridisation using an 813 bp probe generated by PCR against the genome, using primers *cas6F* and *cas6R* (**Figure 3.36 A**). Deletion of *cas6* was further confirmed by screening using Southern blot (**Figure 3.36 B-D**).

Deletion of mrr restriction endonuclease (HVO_0682)

Deletion of *mrr* will remove the need to passage the large CRISPRi plasmids through *dam*- *E. coli*. The *mrr* gene (HVO_0682) was deleted from $\Delta cas3 \Delta cas6$ strains H4606 and H4608 using deletion construct pTA1150 (Allers *et al.*, 2010). This gave rise to Δmrr strains H4739 and H4741 respectively. Deletion of the *mrr* gene was confirmed by colony hybridisation using a 520 bp probe generated by PCR against the wild-type (H53) genome, using primers *mrrF* and *mrrR*. Deletion of *mrr* was further confirmed by screening using Southern blot (as previously shown in **Figure 3.8**).

Deletion of leuB (HVO_1502)

The *leuB* marker within the pTA232 backbone will select for propagation of the episome, thus strains used for CRISPRi must be deleted for gene *leuB* (HVO_1502). The *leuB* gene was deleted from strains H4739 and H4741 using deletion construct pTA73 (Allers *et al.*, 2004). This gave rise to $\Delta leuB$ strains H4886 and H4888 respectively. Deletion of the *leuB* gene was primarily confirmed by screening for leucine auxotrophy and the genotype was further confirmed by Southern blot.

CRISPRi trial

Where targeting an essential gene, it would be expected the cell would attempt any mechanism available to prevent its loss of viability. When using CRISPRi, the knockdown of the targeted gene is reliant on the CRISPR-Cas system. To this end, cells are capable of incapacitating their CRISPR-Cas system or the synthetic crRNA in place of decreasing the expression level of a key gene (escapers). The most commonly reported way for this to occur is deletion of the *cas* gene cassette by homologous recombination (HR) via the repeats found within CRISPR loci P1 and P2 (Fischer *et al.*, 2012). P1 and P2 flank the *cas* genes and their sequences differ by only one nucleotide, meaning that such a HR event is probable. Such recombination could also occur between flanking t-elements on the plasmid, resulting in deletion of the crRNA. Therefore, when targeting an essential gene, the majority of clones isolated would be expected to revert to WT-like colonies. The CRISPRi knockdown of PolB1 and PolD2 is predicted to hinder growth, as both are essential genes. To allow for this predicted phenotype, the transformation was incubated for two weeks to allow for the appearance of any slow-growing colonies.

CRISPRi strains H4886 and H4888 were transformed with anti-polB1 episomal plasmids #1 pTA2251, #2 pTA2252 and #3 pTA2305, anti-pold1 plasmid #2 pTA2342 and empty vector control pTA232, and left to grow at

45°C for 2 weeks. All plates had colonies at both 10^{-1} and 10^{-3} dilutions, and of ~300 colonies per plate, 5 small candidate colonies were identified (**Figure 3.37**). The rest were comparable in size to wild type colonies and thus were likely escapers.



Figure 3.37: Example of CRISPRi knockdown of essential gene candidate. Following introduction of the synthetic crRNA targeting essential gene PolB1, plates were incubated for two weeks. The majority of colonies formed were wild type-like and predicted to carry inactivated CRISPR systems. Where knockdown was presumed to have occurred, growth was impacted and thus colonies were much smaller. These small colonies were selected as candidates for successful CRISPRi.

When such candidates were restreaked on fresh selective medium, they lost the associated slow-growing phenotype and showed no growth defect compared to wild type. This was likely due to loss of the Cas cassette. The screen was repeated three times, generating a total of 54 small colonies, however once restreaked all colonies were no longer slow-growing. The inherent instability of the Cas cassette meant it was not possible to isolate a stable knockdown candidate; this methodology had previously not been used against essential genes and it suggests that, without further strain modification (e.g., to remove the P1/P2 loci), it is not a suitable technique for targeting essential genes due to its inherent instability.

3.4 Discussion

Family B polymerases are predicted to be key players in genome replication. The finding that the aphidicolin resistance *H. volcanii* increases with the successive deletion of replication origins is somewhat unexpected. The combination of origin deletions seems to have an additive effect, where increased numbers of origin deletions (or of increased strength) leads to an increased resistance to aphidicolin treatment. Previous replication profiles have been mapped using frequency marker analysis for *H. volcanii*, whereby peak height will allow the definition of the most and least active replication origins (**Figure 3.38**). Screening of multiple origin deletion mutants for aphidicolin response suggests loss of the most active origins (*oriC1*>*ori2*>*oriC3*>*ori-pHV4-2*) corresponds with reduced requirement for PolB activity. This places PolB (and the requirement for its activity) at the heart of origin-based DNA replication.

Chapter 3: Cellular requirement for replicative DNA polymerases PolB and PolD in *Haloferax volcanii*

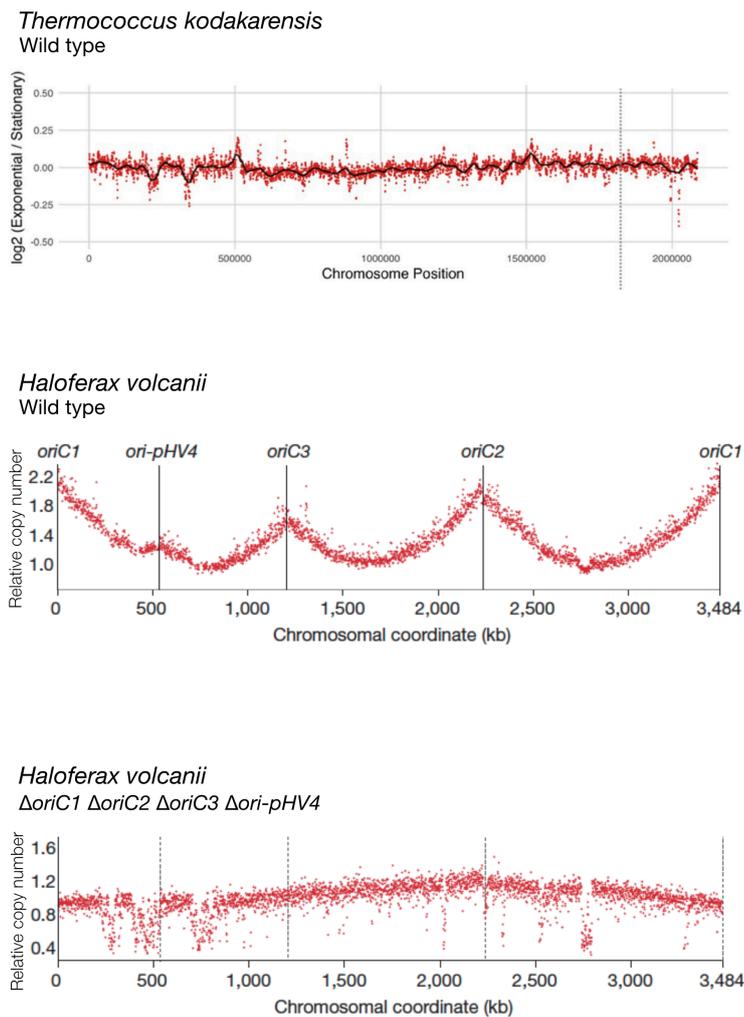


Figure 3.38: Replication profiles of euryarchaeal species *Thermococcus kodakarensis* and *Haloferax volcanii* and essentiality of PolB. (A) Wild type *T. kodakarensis* has a flat replication profile, signifying it is unlikely to utilise origins for genome duplication, instead priming replication along the length of the chromosome. (B) Wild type *H. volcanii* utilises the four origins on the main chromosome for its replication; this is seen as peaks on the replication profile. (C) Originless *H. volcanii* has a flat replication profile, comparable to *T. kodakarensis*, where replication is primed along the length of the chromosome. Figure adapted from (Hawkins *et al.*, 2013a, Gehring *et al.*, 2017).

Deletion of *polB1* and *polD2* using the standard pop-in/pop-out methodology was not possible. PolD has yet to be deleted in any archaeal species, and as such its essentiality in *H. volcanii* was unsurprising. Regarding PolB, while originless strains showed an increased resistance, suggesting a reduced requirement for PolB in these strains, the aphidicolin-treated sample

never matched the optical density associated with untreated cells. This difference suggests that there is still some level of requirement for PolB in the absence of origins.

PolB is non-essential in the euryarchaeon *T. kodakarensis* (Cubonova *et al.*, 2013). While the same may be predicted for *H. volcanii*, mapping of the replication profile of wild type *T. kodakarensis* reveals that, even when origins are present, they are not utilised for canonical replication under laboratory conditions (**Figure 3.38**).

The *T. kodakarensis* $\Delta polB$ strain has an increased sensitivity to gamma irradiation (Kushida *et al.*, 2019). Analogously, PolB in *H. volcanii* may have adopted a more central role in DNA repair, whereby it is now essential. The haloarchaeon *Halobacterium* NRC-1 also requires both PolB and PolD for viability (Berquist *et al.*, 2007) and the environment for halophilic archaea may select for the use of PolB to replicate their highly polyploid genomes. Should an inducible promoter be generated for PolB, this could be tested by exposing strains to DNA damaging agents at varying levels of PolB induction.

To further confirm the essentiality of these genes, it may be necessary to generate episomes carrying either PolB or PolD, and delete the gene from the chromosome in the presence of *in trans* expression from such an episome. Selection for the episome could then be subsequently removed and, should the strain lose viability, it would confirm that the organism is reliant on expression of the associated polymerase. However, the level of expression from the episome may itself be detrimental to the cell, whether there is too little or too much expression to allow deletion of the wild type locus.

Numerous attempts were undertaken to place *polB1* and *polD2* under tryptophan-inducible promoters. An early problem proved to be the cloning of full-length polymerase genes, specifically passaging through a *dam* methylation defective strain of *E. coli* (required for transformation into *mrr*⁺ *H. volcanii*). Overexpression of foreign replication components could cause issues within *dam*⁻ *E. coli*, whereby the strain is both deficient in mismatch repair (which relies on strand methylation to identify the parent vs nascent strand for repair) and retains the recombinase protein RecA. Being reliant on homologous recombination (and therefore RecA) for repair, introduction of an additional polymerase could alter this process and over-replication or misreplication of the DNA could cause the degradation of DNA as seen during this project. This issue was overcome through deletion of the *mrr* gene, which usually targets *dam*-methylated DNA for degradation. While a simple gene deletion, this additional requirement for strains added to the time taken to begin trials for generating inducible polymerases. It would be of interest to see

whether the truncated versions cloned later would be stable in a *dam*⁻ strain of *E. coli*, which could therefore reduce the need for Δmrr strains going forward.

The large genes proved problematic when using the standard protocol for integration of inducible promoters; where the pop-out event causes a detrimental phenotype, there will always be a bias to carry out the alternative pop-out and revert to wild type. While screening of numerous candidates would allow isolation of US pop-ins, it remains unlikely a subsequent DS pop-out would occur for all genome copies when there is >2.5 kb of gene sequence that would allow reversion to wild type.

Generation of the truncated constructs overcame this issue, however it still proved impossible to isolate a candidate under *p.tnaM3*. Pop-out attempts were made at varying levels of induction and for varying amounts of time, however all resulting candidates were able to grow in the absence of tryptophan. The failure to control expression of the polymerases may be due, in part, to induction of *p.tnaM3* being at a constant level; if PolB and PolD levels vary during cycles of replication, introduction of promoters that require constant level of expression may not be tolerated. It has previously been suggested that wild type (*oriC*⁺) *H. volcanii* may switch between origin usage and RDR, depending on growth phase (Thorsten Allers, personal communication). If the former requires PolB and the latter requires PolD, the levels of expression required would be varied over time and expression at a constant level from *p.tnaM3* may not meet the demand for each polymerase at any one time.

Instead of relying on negative selection (inability to grow without tryptophan or thymidine), to force integration of the inducible promoters it may be beneficial to include a positive selectable marker alongside, for example, an antibiotic resistance marker. This could go in place of the *hdrB* selection and allows for positive selection using plates containing the appropriate antibiotic at both the pop-in and pop-out stage.

CRISPRi was an alternative method employed to assess if requirements for PolB and PolD differ in the absence of replication origins. While the host strains and plasmids were able to be generated successfully, no stable knockdown candidates were isolated. The >99% identity between P1 and P2 meant that recombination between these loci would lead to loss of the Cas cassette, and thus reversion to wild type was a common occurrence. Should a knockdown candidate have retained its slow-growing phenotype upon restreaking, there would be no way to ensure the deletion event would not occur when the cells are used for downstream experiments. Therefore, these would all require regular screening by Southern blotting to ensure the Cas

Chapter 3: Cellular requirement for replicative DNA polymerases PolB and PolD in *Haloferax volcanii*

cassette has not been lost since, and this would be very time-consuming. This suggests that CRISPRi, while a useful technique, is not an ideal technique for targeting essential genes.

Alongside failed attempts to generate stable knockdowns, a method for screening knockdown candidates proved challenging. Trials were carried out to detect either *polB1* or *polD1* in wild type strains on a Northern blot, but this was unsuccessful, likely due to the low levels of expression of both *polB1* and *polD1* (RNAseq data; Thorsten Allers, personal communication). Alongside this, trials of quantitative reverse transcriptase PCR (qRT-PCR) were attempted but repeatedly failed. Antibodies against PolB and PolD in *H. volcanii* are not readily available and therefore to assess knockdown at the protein level, these antibodies would either need to be generated against the protein, or epitope tags would need to be engineered onto the polymerase in question.

The differing requirements for the two polymerases still remains an interesting observation and merits additional work. It would be interesting to assess the levels of expression of both proteins over time, either via integration of tags or production of antibodies against the polymerases. Since the wild type strain appears to be capable of switching from origin-dependent replication to RDR, it would be pertinent to assess protein expression of both PolB and PolD over time, and to correlate these results directly with replication profiles; does origin use directly correlate with an increase in protein expression?

It could also be interesting to alter the levels of polymerases through overexpression. If PolB were overexpressed from an episome, would the cell now be more resistant to aphidicolin? Would the cell use origins consistently or still switching to RDR sporadically? Does it have an altered response to DNA damaging agents? It would also be interesting to see if an increase in either PolB or PolD usage in the wild type or originless strain can alter replication efficiency; can the originless strain still grow faster when PolB is overexpressed?

3.5 Conclusion

The requirement for replicative DNA polymerases PolB and PolD is altered in the absence of origins, as seen by altered aphidicolin response. However, both genes proved essential in both wild type and originless backgrounds.

Chapter 4: Genetic analysis of *recJ* genes in *Haloferax volcanii*

4.1 Background

RecJ proteins are part of the DHH phosphoesterase protein superfamily, found throughout all domains of life (Aravind and Koonin, 1998). They have primarily been studied in bacterial species, where their enzymatic role has been defined as being a ssDNA-specific 5'-3' exonuclease. Subsequently, prominent members of the DHH superfamily have been identified in archaeal and eukaryotic species, including RecJ exonuclease, exopolyphosphatase PPX1, Prune phosphodiesterase, and cell cycle and replication protein Cdc45 (Cdc45 biology is covered in more detail in *Chapter 5*). The ubiquity of RecJ-like proteins suggests an ancient origin and an important function within cells, whether catalytically active or inactive.

The DHH superfamily is one of several known families of phosphoesterases with a broad spectrum of substrates. The domains associated with the family (DHH and subdomains DHHA1 and DHHA2) can be characterised both structurally and by assessment of active sites dispersed within the defining motifs of the domain (Aravind and Koonin, 1998). The DHH domain is thought to be important for the catalytic activity of these enzymes, while the presence of either subdomain plays a part in defining the substrate specificity (Aravind and Koonin, 1998). Subfamily DHHA1 is only found in archaeal and bacterial proteins, while subfamily DHHA2 is also found in eukaryotic proteins.

Bacterial RecJ proteins

RecJ was primarily identified in *Escherichia coli*; cells lacking both RecBCD and RecJ showed an extreme deficiency in homologous recombination when compared to either single mutant (Lovett and Clark, 1984, Lovett and Kolodner, 1989). It has since been identified in almost all bacterial species, with most RecJ proteins carrying an N-terminal catalytic core consisting of DHH and DHHA1 domains, and a C-terminal oligonucleotide binding (OB) fold (**Figure 4.1**). The OB fold is thought to aid in binding ssDNA, where ssDNA passes through the β -barrel structure. Some bacterial species, for example radiation-tolerant *Deinococcus radiodurans*, have RecJ proteins carrying an additional C-terminal domain, in this case named domain IV. Here, domain IV of *DrRecJ* promotes substrate binding and allows for interaction with HerA helicase (Cheng *et al.*, 2015).

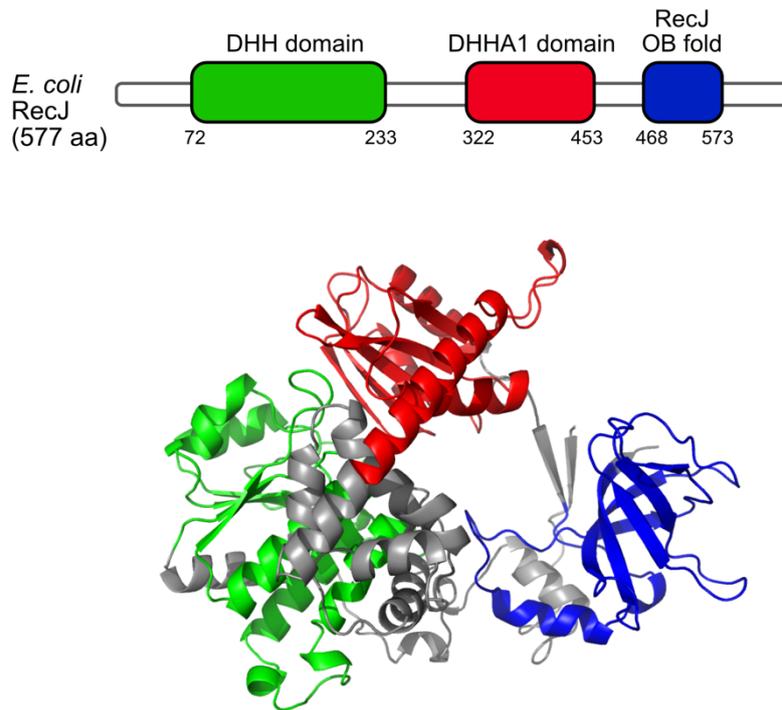


Figure 4.1: Domain and structural representation of *Escherichia coli* RecJ. Domains are coloured on both the gene and protein models. Domain diagram is not to scale. Protein structure was predicted from protein sequence using Phyre2 prediction software.

Studies have shown that within *E. coli*, RecJ has a clear role in the RecF pathway of homologous recombination, whereby RecJ and RecQ work in parallel to process DNA breaks, and is essential in the process of recombination-dependent replication (Kowalczykowski, 2000, Courcelle and Hanawalt, 1999). Of the numerous genes within the RecF recombination pathway (including *recF*, *recJ*, *recO*, *recN*, *recQ* and *ruv*), mutations in *recJ* have the most extreme effects on recombination efficiency (Kolodner *et al.*, 1985, Lloyd *et al.*, 1988). RecJ has also been implicated in a multitude of different DNA repair pathways, including ssDNA gap repair, base excision repair (BER), and mismatch repair (MMR) (Cheng *et al.*, 2016, Dianov *et al.*, 1994).

Alongside its role in DNA repair pathways, RecJ has also been implicated in genome replication. When the replication fork undergoes damage, RecJ can aid in the rescue of stalled replication forks (Chow and Courcelle, 2007, Courcelle and Hanawalt, 1999). RecJ acts alongside RecQ to degrade nascent lagging strands, allowing repair to occur at the site of damage, before RecF acts prior to the replication-restart primases PriA and PriC to

resume replication. It has been shown this role is specific to RecJ; no other exonucleases are able to compensate for this role as an ‘interrupter’ of replication (Courcelle and Hanawalt, 1999).

Bacterial RecJ proteins have been shown to degrade ssDNA directionally, moving in a 5' to 3' direction. While RecJ alone is capable of binding and performing nucleolytic resection of 5' ssDNA overhangs, it has been observed that RecJ can interact with ssDNA binding protein SSB, which stimulates its DNA binding and nuclease activities (Han *et al.*, 2006, Sharma and Rao, 2009). RecJ also requires this interaction with SSB to perform exonucleolytic degradation of ssDNA (Morimatsu and Kowalczykowski, 2014).

Archaeal RecJ proteins

The DHH superfamily of phosphoesterases has undergone an expansion event within the archaea (namely the euryarchaea), where multiple species now encode numerous RecJ-like proteins; some retain sequence identity while others evolved quickly and developed specialised functions (Makarova *et al.*, 2012). Archaeal RecJ proteins remain members of the DHH superfamily but differ from their bacterial counterparts in terms of sequence, domain organisation and substrate specificity (Li *et al.*, 2017, Oyama *et al.*, 2016).

All archaeal species encode at least one RecJ protein. Most species carrying a RecJ protein have maintained an intact DHH nuclease domain, suggesting they remain active (Makarova *et al.*, 2012). For example, in species *Methanocaldococcus jannaschii*, two *recJ* genes are encoded. Interestingly, expression of either RecJ protein encoded by *M. jannaschii* in $\Delta recJ$ *E. coli* cells has the ability to rescue the defect in recombination and repair, suggesting active archaeal RecJ proteins can play a role similar to that of bacterial RecJ proteins (Rajman and Lovett, 2000). However, it is worth noting inactive RecJ proteins have also been identified in species encoding more than one RecJ homologue.

Makarova *et al.* (2012) showed that when bacterial, archaeal and eukaryotic DHH superfamily proteins are mapped phylogenetically via alignment of the catalytic DHH domain, three distinct clades are identified (**Figure 4.2**): namely the Ppx1 clade, the COG2404 clade, and the RecJ clade. Alignment was performed using Clusters of Orthologous Genes for bacterial (COGs) and archaeal (arCOGs) species, with protein families being used for eukaryotes.

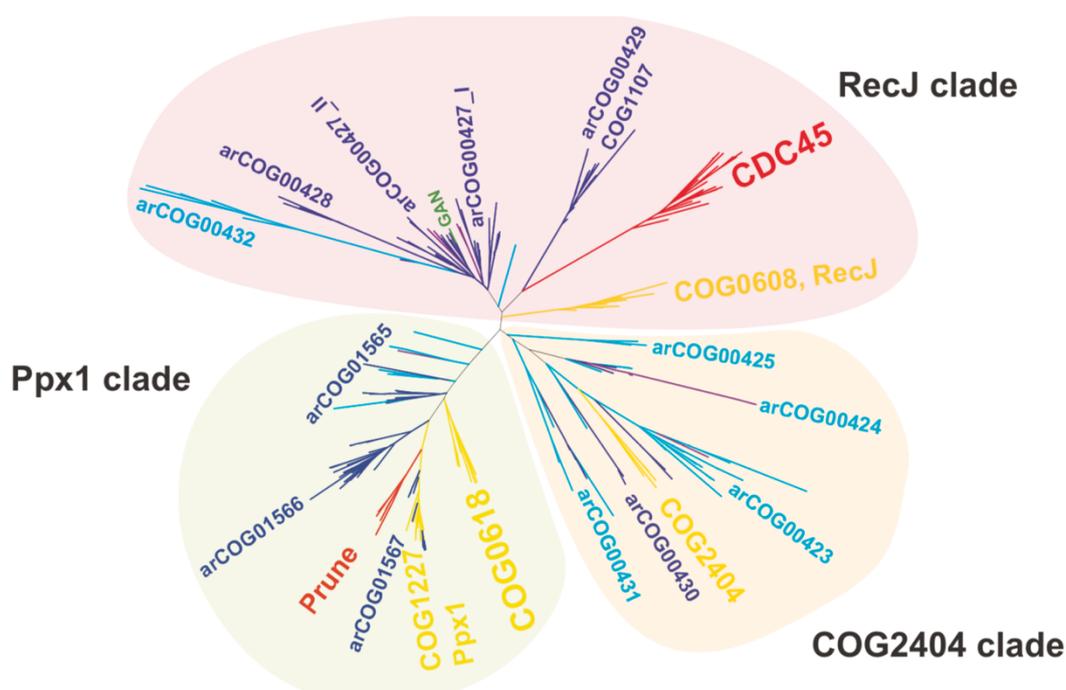


Figure 4.2: Phylogeny of the DHH superfamily showing distribution of RecJ homologues in Archaea. Eukaryotes - red, bacteria - yellow, euryarchaea – dark blue, crenarchaea – light blue, deeply branched archaeal lineages – purple. The tree was aligned using the DHH catalytic domain. Branches are labelled with COG (Clusters of Orthologous Genes) and arCOG (archaeal COG) numbers or family names (for eukaryotes). Three major clades are shaded: RecJ clade (pink), Ppx1 clade (green) and COG2404 clade (orange). *Thermococcus kodakarensis* GAN protein is highlighted in green within the RecJ clade. Adapted from Makarova *et al.* (2012).

The Ppx1 clade contains bacterial inorganic pyrophosphatases/exopolyphosphatases, members of arCOG1567 and members of the eukaryotic Prune family (Makarova *et al.*, 2012). In bacteria and yeast, these proteins were shown to hydrolyse inorganic polyphosphates as critical players in phosphate metabolism (Kornberg *et al.*, 1999). All members of the clade carry an active DHH domain and C-terminal DHHA2 (subfamily II) domain. Their similar domain structures and grouping within the clade suggests the archaeal counterparts are likely to share functions with those of bacteria and eukaryotes, however this is yet to be confirmed biochemically.

The COG2404 clade are predicted to carry active DHH domains and C-terminal DHHA1 (subfamily I) domains. These proteins are well distributed throughout archaea but are only present in only a subset of bacterial species. It

can therefore be postulated that bacteria have gained these proteins over time by horizontal gene transfer (HGT) from archaea (Makarova *et al.*, 2012). In *Bacillus subtilis*, its COG2404 member NrnB has been shown to act as a nanoRNase (degrades oligoribonucleotides of ≤ 5 residues) (Fang *et al.*, 2009). Again, similarity in domain architecture and activity would suggest shared protein functions, but the biochemical evidence for this is lacking.

The RecJ clade contains bacterial RecJ proteins, eukaryotic replication protein Cdc45 and archaeal homologues spanning various arCOGs. Previous studies have defined the roles of bacterial RecJ and eukaryotic Cdc45 well: bacterial RecJ is active as a nuclease with an integral role in DNA repair, while eukaryotic Cdc45 has lost key residues required for nuclease activity, and instead plays a structural role in the activation of the CMG replicative helicase at the replication fork. Within the RecJ clade, the positioning of the bacterial RecJ (yellow in **Figure 4.2**) compared with that of eukaryotic Cdc45 (red) and the remainder being archaeal RecJ proteins, Makarova *et al.* (2012) make the argument that Cdc45 clusters with archaeal RecJ proteins, away from the bacterial RecJ family. This provides some evidence for archaeal RecJ proteins having gained a different function to their bacterial counterparts, acting similar to Cdc45 in replication as opposed to DNA damage repair and recombination. However, the argument is not without fault: the predicted archaeal Cdc45, GAN, is not the closest-mapping arCOG to Cdc45, suggesting that phylogenetic mapping alone does not answer this question. It is also worth noting that while some arCOGs within the RecJ clade are predicted to carry inactivated DHH domains (akin to Cdc45), the main arCOG represented (arCOG00427) carries an intact DHH domain.

TkoGAN (gene TK1252) is one of two RecJ-like proteins encoded by *T. kodakarensis*, the other being Hef-associated nuclease (*TkoHAN*; gene TK0155) (Nagata *et al.*, 2017a). *TkoGAN* is a processive 5'-3' exonuclease active only on ssDNA. Its structure and domain layout bears similarity to both bacterial RecJ and eukaryotic Cdc45 (Oyama *et al.*, 2016). *TkoGAN* shares similar biochemical properties with bacterial RecJ, including substrate and metal co-factor requirement, and enzyme directionality (Makarova *et al.*, 2012, Nagata *et al.*, 2017a, Li *et al.*, 2011). Its interaction with central replication component GINS implicates *TkoGAN* in genome replication, suggesting that it may act as the Cdc45 homologue in *T. kodakarensis* (Li *et al.*, 2011). However, the maintenance of catalytic activity of *TkoGAN* is in contrast to eukaryotic Cdc45, which has lost key residues required for nuclease activity. It could be that the GAN has a function specific to archaea that has been lost in Cdc45. It was also shown the *TkoGAN*:GINS complex was able to stimulate the activity of MCM helicase *in vitro* (akin to Cdc45 in eukaryotes) (Nagata *et al.*, 2017a). However, it is worth noting that deletion of *TkoGAN* has no effect on cell viability (Burkhart, 2017, Nagata *et al.*, 2017b); in eukaryotes, Cdc45 is

essential and thus it would be predicted the Cdc45 homologue would be essential for replication as a key component of the replicative helicase complex (CMG complex).

The second *T. kodakarensis* RecJ protein, *TkoHAN*, was primarily identified through its interaction with Hef (helicase-associated endonuclease for fork-structured DNA). Hef was discovered in the hyperthermophile *P. furiosus*, where it was shown to have both helicase and nuclease activity on a range of branched DNA structures, including replication forks (Komori *et al.*, 2002, Komori *et al.*, 2004). Genetic analysis of Δ *hef* strains showed an involvement in DNA repair, with a high sensitivity to mitomycin C (MMC) implicating Hef in repair of interstrand crosslinks (ICLs) (Fujikane *et al.*, 2010). Further study of Hef in *H. volcanii* also implicated Hef in stalled replication fork repair, where it acts in an alternative pathway to Hjc resolvase (Lestini *et al.*, 2010, Lestini *et al.*, 2013).

TkoHAN has 3'-5' exonuclease activity on ssDNA and RNA substrates, which is stimulated in the presence of Hef (Nagata *et al.*, 2017b). This difference in both directionality and substrate specificity when compared to *TkoGAN* provides evidence for the differentiation of RecJ proteins within euryarchaeal species encoding more than one RecJ protein. *TkoHAN* is also non-essential, with Δ *han* strains showing few defects. Interestingly, Δ *gan* Δ *han* strains are viable but show a major growth defect compared to either single mutant (Nagata *et al.*, 2017b). It can be imagined in the absence of GAN, the replicative helicase complex is destabilised and the replication fork will be prone to more stoppages; in the presence of HAN this situation can be resolved through its interaction with Hef to overcome stalled forks, however in the absence of HAN this deficiency would be increased. The interactions between GINS and GAN, and Hef and HAN, have been shown *in vitro* to be specific, suggesting there is no complementarity occurring between these two proteins (Nagata *et al.*, 2017b). The HAN protein family has been shown to be present only within Euryarchaea (as seen with distribution of Hef), supporting the co-operative functions of Hef and HAN in euryarchaeal-specific DNA and stalled replication fork repair.

P. furiosus encodes two RecJ-like proteins, PF2055 and PF0399. Of these, PF2055 is predicted to be a 'GAN-like' protein and its protein structure has been solved (now referred to as *PfuRecJ*). Akin to *TkoGAN*, *PfuRecJ* has 5'-3' exonuclease activity on ssDNA and shows direct interaction with GINS (Li *et al.*, 2017), but is also capable of 3'-5' exonuclease activity on RNA where it is predicted to function in proofreading 3'-mismatched ribonucleotides (Yuan *et al.*, 2013). At the structural level, *PfuRecJ* is very similar to *TkoGAN*, and is therefore comparable to eukaryotic Cdc45. The similarity in structure, interactors and sequence (73% sequence identity by Clustal Omega)

suggests *TkoGAN* and *PfuRecJ* perform the same function in their respective organisms. The second RecJ-like protein in *P. furiosus*, PF0399, has not been studied in any detail: it maps to the same arCOG as HAN and thus could be predicted to play a role in stalled replication fork repair. However, this proposal lacks evidence, and it cannot be ruled out that PF0399 does not play a different species-dependent role in *P. furiosus*.

In addition to the euryarchaeal *TkoGAN* and *PfuRecJ* interactions with GINS, a RecJ-GINS interaction has also been seen within Crenarchaeota. In species *Sulfolobus solfataricus*, its RecJ/Cdc45-like protein, RecJdbd/Cdc45, is seen to interact with GINS and MCM (Xu *et al.*, 2016). The association of RecJdbd/Cdc45 with GINS stimulates the helicase activity of MCM, however GINS or RecJdbd/Cdc45 alone did not stimulate MCM activity (as seen for other species) (Marinsek *et al.*, 2006).

The crystal structure of two archaeal RecJ proteins have now been solved: *Thermococcus kodakarensis* GAN (*TkoGAN*) (Oyama *et al.*, 2016) and *Pyrococcus furiosus* RecJ (*PfuRecJ*; predicted GAN) (**Figure 4.3**) (Li *et al.*, 2017).

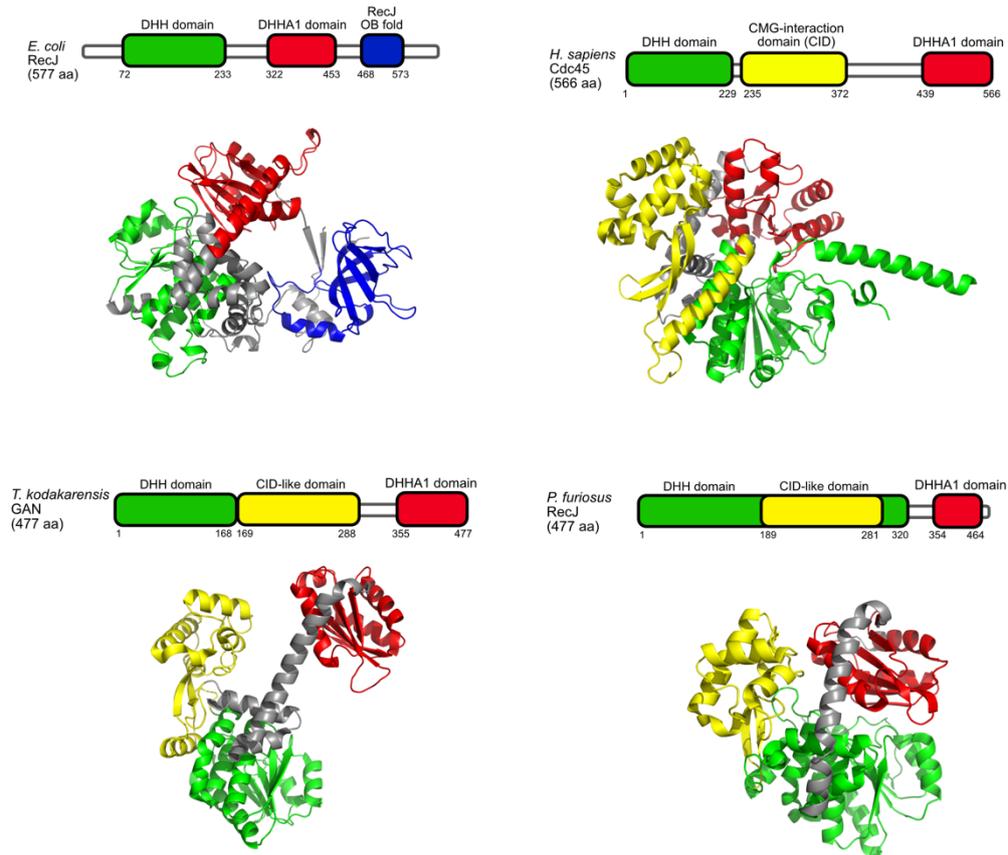


Figure 4.3: Comparison of *E. coli* RecJ, *H. sapiens* Cdc45 and solved archaeal RecJ protein structures for *T. kodakarensis* GAN and *P. furiosus* RecJ. Structure and domain information for *Escherichia coli* RecJ (mapped using Phyre2) was annotated according to Pfam domain analysis. Structure and domain information for *Homo sapiens* Cdc45 (PDB ID: 5DGO), *Thermococcus kodakarensis* GAN (PDB ID: 5GHT) and *Pyrococcus furiosus* RecJ (PDB ID: 5X4H) were annotated according to papers defining their crystal structures (Simon *et al.*, 2016; Oyama *et al.*, 2016; Li *et al.*, 2017 respectively).

At the structural level, clear similarities can be seen between the archaeal RecJ/GAN proteins and both *E. coli* RecJ and eukaryotic Cdc45. With regards to the structure of human Cdc45 (hCdc45), shown in **Figure 4.3**, the CMG-interaction domain (CID) was identified due to its prominent role mediating MCM interactions and some of the GINS interactions within the CMG complex (Simon *et al.*, 2016). The same CID domain was defined structurally in the crystal structures of both *TkoGAN* and *PfuRecJ*. However, this may be somewhat misleading: while both GAN-like proteins have been seen to interact with GINS, this interaction does not take place not through the CID domain but through the DHH domain (Li *et al.*, 2017, Oyama *et al.*, 2016). This difference in eukaryotic and archaeal proteins suggests the mechanism of action of the predicted Cdc45 archaeal homologues could be

different to what is known for the eukaryotic counterpart, and thus knowledge regarding eukaryotic Cdc45 mechanisms should not be assumed for the archaeal RecJ proteins.

Haloferax volcanii RecJ proteins

Within the RecJ clade defined by Makarova *et al.* (2012), *Haloferax volcanii* encodes four RecJ homologues, namely RecJ1 (HVO_0073), RecJ2 (HVO_1147), RecJ3 (HVO_1018), and RecJ4 (HVO_2889). It also encodes a single protein within the COG2404 clade (HVO_1824) and three proteins which fall within the Ppx1 clade (HVO_0756, HVO_0990 and HVO_1690). As this study primarily focusses on the biology of DNA replication, this work will be concerned only with proteins within the RecJ clade (RecJ1-RecJ4).

Previous work has shown that *recJ1*, *recJ3* and *recJ4* are dispensable in *H. volcanii*, and their deletion (either alone or in combination) has little effect on viability as a measure of cell growth (Lever, 2019). However, previous attempts to delete *recJ2* in our lab proved it is not possible using the pop-in/pop-out method, and thus it was determined that RecJ2 is essential for viability. It is worth noting that it is stated that *recJ2* is nonessential in the publication Giroux and MacNeill (Giroux and MacNeill, 2015), however, any scientific evidence to support the published claim is absent. Phenotypic analysis of *recJ* mutants has not been carried out and thus the roles of these proteins in *H. volcanii* remains elusive.

4.2 Aims and Objectives

While it is known that *recJ1*, *recJ3* and *recJ4* can be deleted with little effect on growth rate, the specific role of these proteins within *H. volcanii* remains largely unknown. The objectives of this chapter are to:

- Utilise phylogenetic analysis of known archaeal RecJ proteins to predict the roles of the RecJ proteins in *H. volcanii*;
- Analyse the phenotypes of strains deleted for *recJ1*, *recJ3* and/or *recJ4* after treatment with DNA damaging agents;
- Analyse the recombination frequency, and fractions of crossover and non- crossover products formed, by strains deleted for *recJ1*, *recJ3* and/or *recJ4*;
- Generate strains deleted for *recJ1*, *recJ3* and/or *recJ4* in combination with genes encoding key players in DNA replication, repair, and stalled fork restart;
- Utilise the well-studied tryptophan-inducible gene regulation system to confirm the supposed essentiality of *recJ2*.

4.3 Results

4.3.1 In silico analysis of RecJ proteins in Archaea

Domain analysis of RecJ proteins

Studies of bacterial RecJ family proteins have defined a consensus domain architecture carried by most bacterial RecJ proteins: an N-terminal DHH domain (Pfam Protein Family entry PF01368; (El-Gebali *et al.*, 2019)) followed by a DHHA1 subdomain (DHH-associated domain 1; PF02272), with a C-terminal RecJ OB-fold (oligonucleotide binding; PF17768). Domain analysis using Pfam database searches recognises no defined DHH domains or canonical OB folds in any of the four *H. volcanii* RecJ proteins (**Figure 4.4**). Only RecJ1 carries a DHHA1 domain, while RecJ3 and RecJ4 have diverged from the bacterial domain structure, instead carrying S1 RNA binding (PF00575) and tRNA OB-fold (PF01336) domains.

Prediction of three-dimensional structures for *H. volcanii* RecJ proteins was carried out using Phyre2 (Kelley *et al.*, 2015) and showed a distinct structural difference between RecJ1 and RecJ2 in comparison to RecJ3 and RecJ4 (**Figure 4.4**). Protein structures were annotated according to their Pfam-identified domains allowing easy comparison of defined protein structures (El-Gebali *et al.*, 2019). RecJ1 shows strong structural similarity to the bacterial RecJ proteins (shown here: *Escherichia coli* and *Thermus thermophilus*), with the exception of the C-terminal OB-fold β -barrel being absent from RecJ1. Visually, RecJ2 appears to share these structural similarities, however, lacks the defined DHHA1 domain of RecJ1. The structures of the RecJ proteins, specifically RecJ1 and RecJ2, are directly comparable, suggesting the absence of the defined DHHA1 domain in RecJ2 is likely due to sequence degradation over time.

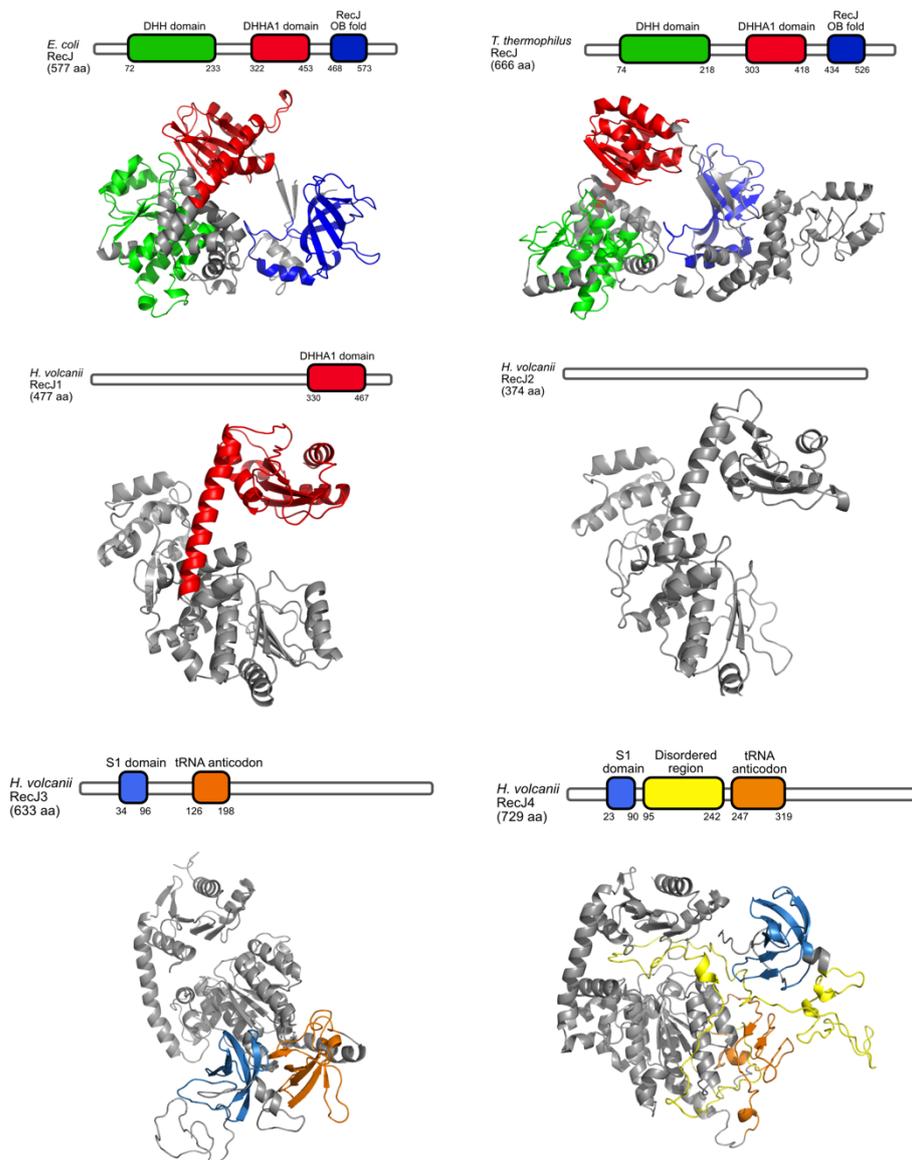


Figure 4.4: Structural comparison of RecJ proteins from *Escherichia coli* (*Eco*), *Thermus thermophilus* (*Tth*) and *Haloferax volcanii* (*Hvo*). Domain analyses of protein sequences for *Eco*RecJ, *Tth*RecJ and *Hvo*RecJ proteins (obtained from UniProt) were carried out using Pfam, and domains were coloured (as above). Protein structures for the *E. coli* and *H. volcanii* RecJ proteins were mapped using Phyre2 protein modelling software (Kelley *et al.*, 2015). The crystal structure for *Tth*RecJ was readily available (PDB ID: 2ZXO). Domain boxes are not to scale. Crystal structures for *Eco*RecJ and *Tth*RecJ (PDB ID: 1IR6) share structural and domain similarity, which is expected as they are both documented to function in similar pathways in their bacterial hosts. Structural similarity is clear between *Hvo*RecJ1 and *Hvo*RecJ2, and both may share some structural identity

Structurally, RecJ3 and RecJ4 differ from both bacterial RecJ proteins and the other *H. volcanii* RecJ proteins. While RecJ1 and RecJ2 lack the β -barrel RecJ OB-fold of bacterial RecJ, RecJ3 and RecJ4 carry two predicted DNA-binding domains: the S1 RNA-binding and tRNA OB-fold nucleic acid-binding domains. The S1 domain, originally identified in S1 ribosomal protein, is found in numerous RNA-associated proteins and is predicted to have RNA-binding capabilities (Bycroft *et al.*, 1997). Fusions of DHHA1 domains with S1 RNA-binding domains have previously been observed (Aravind and Koonin, 1998). It could be hypothesised such a fusion has occurred in this subset of RecJ proteins during evolution, and therefore RecJ3 and RecJ4, if active, are able to bind RNA.

With regard to the tRNA OB-fold domain, its family contains various OB-fold domains that are capable of nucleic acid binding (Koonin *et al.*, 2000, Theobald *et al.*, 2003). In particular, this family includes aminoacyl-tRNA synthetases that catalyse the addition of an amino acid to the appropriate tRNA molecule, RecG helicase involved in DNA repair by overcoming stalled replication forks, replication factor A (RPA) involved in binding ssDNA to prevent degradation, and the C-terminus of bacterial Pol III- α subunit of Pol III polymerase (Koonin *et al.*, 2000, Bochkarev *et al.*, 1997). This would suggest, depending on maintenance of structure and key residues, the ability of RecJ3 and RecJ4 to bind DNA. RecJ4 has a large intrinsically disordered region (IDR; yellow on **Figure 4.4**) and previous studies have shown disordered regions may be responsible for interactions with other proteins, for example between Hef and PCNA in *T. kodakarensis* (Ishino *et al.*, 2014). The IDR in RecJ4 contains an acidic patch (residues 185-227) and studies in yeast have shown a conserved acidic patch within nucleosomes is critical for protein-protein interactions (Cucinotta *et al.*, 2019); it could be predicted that the same applies for RecJ4 and thus it may act as a central hub for protein-protein interaction networks in *H. volcanii*.

The apparent similarity in models of RecJ1 and RecJ2 when compared to bacterial RecJ is surprising, given the lack of defined domains in *Haloferax* RecJ proteins. To further interrogate the extent to which the DHH domains of predicted DHH superfamily proteins RecJ1-4 have deteriorated over time, an alignment was carried out and key DHH superfamily motifs were mapped according to known consensus sequences, defined in Aravind & Koonin (1998). Alignment of *E. coli* RecJ, *H. volcanii* RecJ1-4 and *Thermococcus kodakarensis* GAN was carried out using T-Coffee in MacVector (Myers-Miller; penalty for open gap = -50, extend gap = -50) and key residues and motifs were annotated in bold and red (**Figure 4.5**).

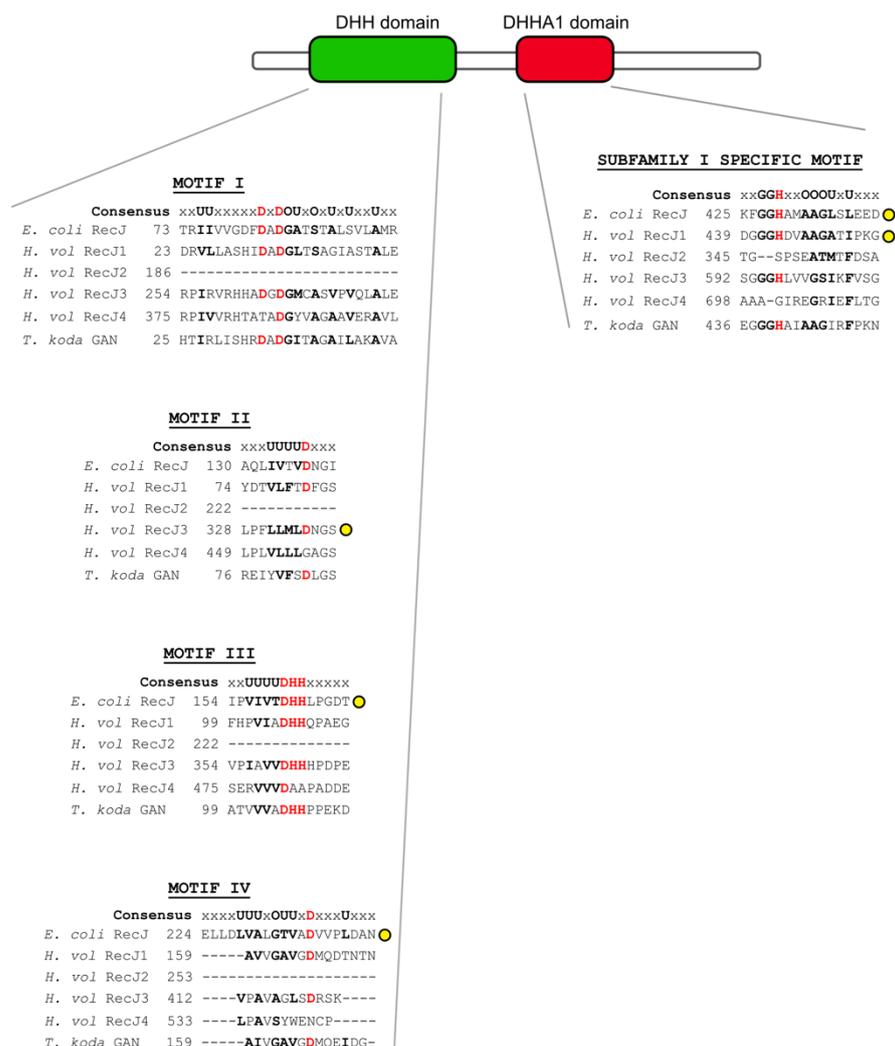


Figure 4.5: Motif analysis for DHH and DHHA1 domain activity for *Escherichia coli*, *Haloferax volcanii* and *Thermococcus kodakarensis* RecJ-like proteins. Consensus sequences for each domain motif were taken from Aravind & Koonin *et al.* (1998). Consensus sequences are highlighted in bold. Perfectly conserved aspartic acid (D) and histidine (H) residues are highlighted in bold and red. U indicates bulky hydrophobic residues (LIFVMA) and O indicates small residues (SCAGT). Yellow circles mark where the protein sequence perfectly matches the consensus. Protein sequences were obtained from UniProt and alignment was carried out in MacVector using T-Coffee (Myers Miller; penalty for open gap = -50; extend gap = -50).

Mutation analysis of *E. coli* RecJ has previously shown all DHH domain motifs (I-IV) are necessary for exonuclease activity (Sutera *et al.*, 1999). The failure of RecJ2 to align to any sequence at the positioning of the motifs suggests the sequence has undergone significant divergence since

historically carrying a DHH domain. The DHH domain consensus for RecJ4 is not strong: all motifs have lost catalytic residues (highlighted in bold and red on **Figure 4.5**). Conversely RecJ1 and RecJ3 retain all the catalytic residues required of a DHH domain but have lost some surrounding consensus residues. The same, however, can be applied to *T. kodakarensis* (*Tko*) GAN: this is of importance, as it has previously been shown that *Tko*GAN has catalytic activity without a Pfam-annotated DHH catalytic domain (DHH domain of *Tko*GAN has only been annotated following solving of crystal structure) (Li *et al.*, 2011, Oyama *et al.*, 2016). Therefore, the lack of a ‘defined’ DHH domain in RecJ1 and RecJ3 should perhaps be taken with caution; these proteins may in fact possess an intact DHH domain, which carries a consensus differing from that defined by Aravind & Koonin (Aravind and Koonin, 1998), potentially archaeal-specific.

With regards to the DHHA1 domain, only RecJ1 is identified using Pfam as carrying this domain, which is explicable as RecJ1 is the only *Haloferax* RecJ to carry the defined consensus sequence. RecJ3, as with the DHH domain, carries the catalytic residues but lacks some surrounding consensus. This may be due to the fusion event with the S1 domain, or RecJ3 may also carry an as-yet undefined archaeal-specific DHHA consensus.

Phylogenetic mapping of RecJ proteins

The euryarchaeal species *T. kodakarensis* encodes two RecJ genes, GINS-associated nuclease (GAN) and Hef-associated nuclease (HAN), which have been relatively well characterised (Li *et al.*, 2011, Nagata *et al.*, 2017b).

For archaeal genomes, databases containing archaeal Clusters of Orthologous genes (arCOGs) are readily available (Makarova *et al.*, 2015). The arCOG database clusters groups of genes predicted to function in similar ways and thus allows for the functional annotation of hypothetical or as yet unidentified genes.

Using *T. kodakarensis* GAN and HAN sequences, along with *H. volcanii* RecJ1-4, species *P. abyssi* and *P. furiosus* were screened for RecJ-like proteins using an arCOG search on the eggNOG database (Huerta-Cepas *et al.*, 2019) (**Table 4.1**). These *Pyrococcus* species were included in the screen as they are relatively well studied for RecJ proteins.

Table 4.1: Distribution of RecJ-like proteins in *Haloferax volcanii*, *Thermococcus kodakarensis*, *Pyrococcus furiosus* and *Pyrococcus abyssi*. For each RecJ-like protein, the protein name and gene locus are given along with the archaeal Clusters of Orthologous genes (arCOG) number associated with the gene according to eggNOG.

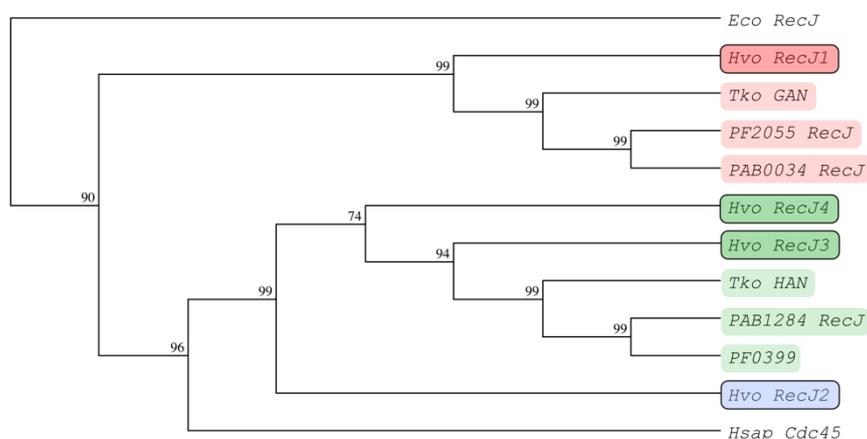
Organism	Protein name	Gene locus	arCOG number
<i>Haloferax volcanii</i>	RecJ1	HVO_0073	00427
	RecJ2	HVO_1147	00428
	RecJ3	HVO_1018	00429
	RecJ4	HVO_2889	00429
<i>Thermococcus kodakarensis</i>	GAN	TK1252	00427
	HAN	TK0155	00429
<i>Pyrococcus furiosus</i>	RecJ	PF2055	00427
	HAN	PF0399	00429
<i>Pyrococcus abyssi</i>	RecJ-like	PAB0034	00427
	RecJ	PAB1284	00429

The arCOG grouping suggests RecJ1 is akin to GAN-like proteins (arCOG00427), RecJ2 falls into its own grouping (arCOG00428), and RecJ3 and RecJ4 are HAN-like (arCOG00429). Further investigation into arCOG00428 (where RecJ2 alone resides) reveals that this arCOG is specific to Halobacteriales encoding more than one RecJ protein.

P. furiosus encodes two RecJ-like proteins with gene loci PF2055 and PF0399 respectively. *P. abyssi* encodes two RecJ-like proteins with gene loci PAB0034 and PAB1284 respectively. These proteins, along with human Cdc45, archetypal *E. coli* RecJ, *T. kodakarensis* GAN and HAN proteins and *H. volcanii* RecJ1-RecJ4 were aligned using T-Coffee in MacVector (Myers-Miller; penalty for open gap = -50, extend gap = -50). This alignment allowed for the plotting of a phylogenetic tree, rooted on either bacterial RecJ or eukaryotic Cdc45 (Figure 4.6).

(A) Outrooted to *E. coli* RecJ

Method: Neighbor Joining; Bootstrap (1000 reps); tie breaking = Systematic
 Distance: Uncorrected ("p")
 Gaps distributed proportionally



(B) Outrooted to *H. sapiens* Cdc45

Method: Neighbor Joining; Bootstrap (1000 reps); tie breaking = Systematic
 Distance: Uncorrected ("p")
 Gaps distributed proportionally

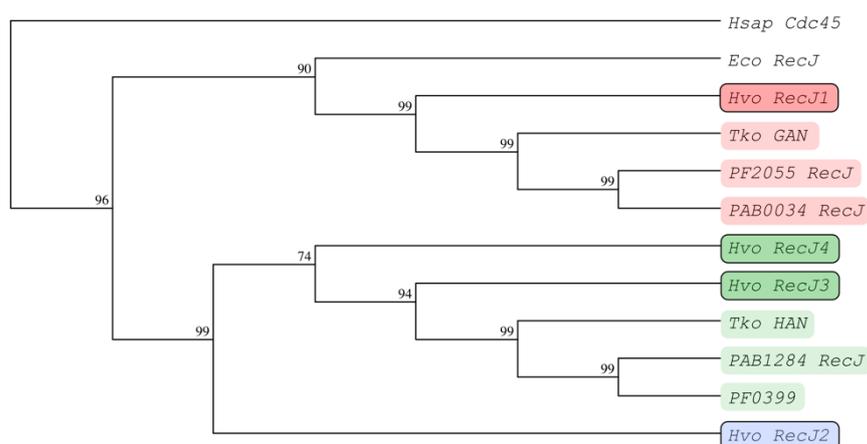


Figure 4.6: Phylogenetic mapping of RecJ/Cdc45 DHH superfamily proteins. Phylogenetic tree of RecJ/Cdc45 DHH superfamily proteins from eukaryotes: *Homo sapiens* (*Hsap*) Cdc45 protein, bacteria: *Escherichia coli* (*Eco*) RecJ protein and archaea: *Pyrococcus abyssi* (*Pab*) RecJ-like proteins, *Pyrococcus furiosus* (*Pfu*) RecJ-like proteins, *Thermococcus kodakarensis* (*Tko*) GAN and HAN proteins and *Haloferax volcanii* (*Hvo*) proteins RecJ1-RecJ4. (A) Tree is rooted to *Eco*RecJ. (B) Tree is rooted to *Hsap*Cdc45. Genes are coloured according to arCOG grouping; arCOG00427 is shown in red, arCOG00428 in blue and arCOG00429 in green. *Hvo* RecJ genes are circled in black. For both (A) and (B), of the *Hvo*RecJ genes, RecJ1 groups closest to both *Eco*RecJ and *Hsap*Cdc45. *Hvo*RecJ3 and RecJ4 group together, while generally RecJ2 is an outlier. Calculated using Neighbour Joining method (Bootstrap 1000 reps). Numbers above branches indicate the percentage likelihood supporting the nodes. Generated using MacVector.

The phylogeny in **Figure 4.6** suggests two distinct groupings of archaeal RecJs: the upper grouping containing *TkoGAN*, PF2055, PAB0034 and *HvoRecJ1*, and the lower group containing *TkoHAN*, PF0399, PAB1284 and *HvoRecJ3/RecJ4*. *HvoRecJ2* also falls within the latter group, but is an outlier compared to the other members of this grouping. When cross-referenced with **Table 4.1**, this phylogeny corresponds with the grouping of genes according to arCOG; the first group fitting within arCOG00427 and the latter falling within arCOG00429. *HvoRecJ2* groups alone within arCOG00428, fitting its positioning as an outlier. Analysis of arCOGs 00427, 00428 and 00499 on the eggNOG database provides some basic functional information. Further information specific to each arCOG is described below:

arCOG00427

This arCOG is defined as containing proteins with a phosphoesterase RecJ domain. It is found throughout various archaeal phyla, including *Euryarchaeota*, *Crenarchaeota*, *Thaumarchaeota* and *Korarchaeota*. Of the proteins mapped here, those that have been studied in detail have either been proven or have been predicted to have maintained nuclease activity. These would likely be predicted to be the ‘active’ RecJ gene within a species.

arCOG00428

Proteins encoded within arCOG00428 are predicted to have transferase activity, specifically the ability to transfer glycosyl groups. The distribution of arCOG00428 members is very narrow: these genes are only found within the *Halobacteriales* order. These genes are likely to have gained a specified function away from that of typical RecJ proteins.

arCOG00429

Genes within arCOG00429 are part of wider cluster of orthologous genes (COG) found in both archaea and bacteria; COG1107. Such genes are predicted to be DnaJ-type Zn finger domain containing proteins, with COG1107 being linked to ribosomal small subunit biogenesis. With regards to the archaeal proteins mapped here, all are predicted to contain S1 RNA-binding domains and tRNA oligonucleotide-binding domains and thus this may encompass the predicted domain of COG1107. Genes within arCOG00429 are found only within the phylum *Euryarchaeota*.

Large-scale alignment of DHH domains of predicted RecJ-like archaeal proteins and mapping of the RecJ clade by Makarova *et al.* (2012) specify the

‘main’ archaeal RecJ clade as arCOG00427 (Makarova *et al.*, 2012). From the analysis here, this would denote RecJ1 the ‘main’ RecJ protein in *H. volcanii*, grouping with the GAN proteins of *T. kodakarensis*, *P. furiosus* and *P. abyssi*. When compared structurally to these GAN proteins, there is structural similarity with RecJ1, again providing evidence to confirm this prediction (Figure 4.7).

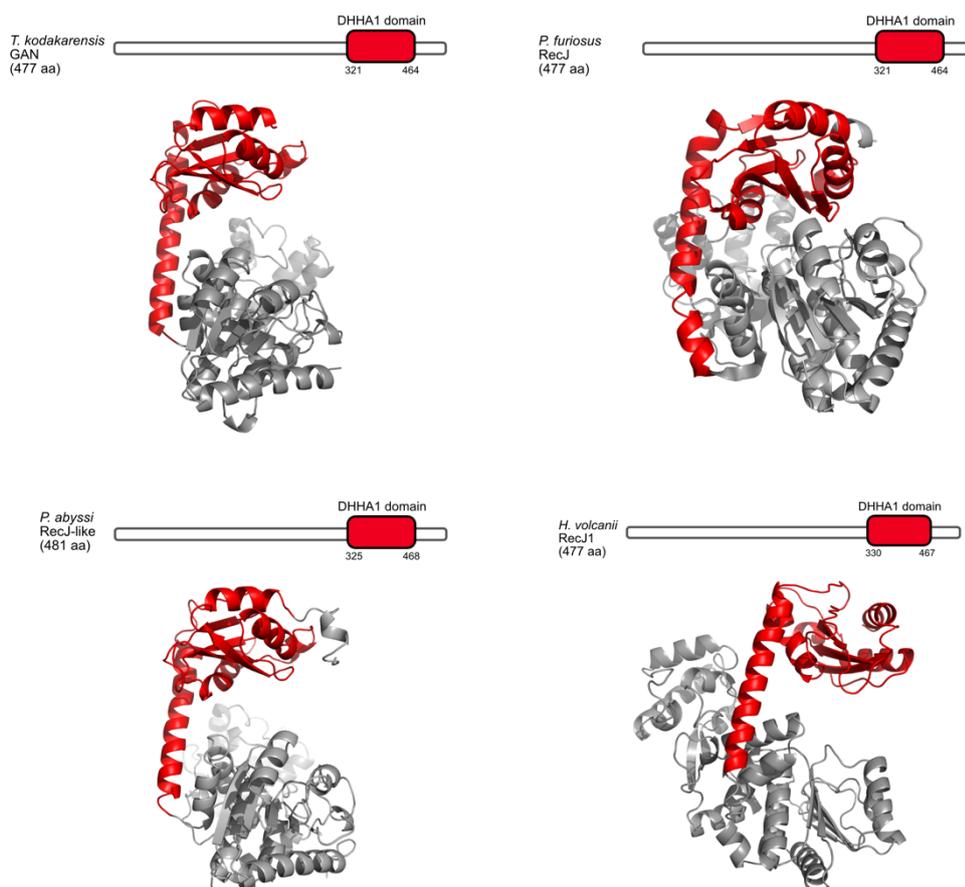


Figure 4.7: Comparison of structures of arCOG00427 GAN-like proteins in *Thermococcus kodakarensis* (TK1252), *Pyrococcus furiosus* (PF2055), *P. abyssi* (PAB0034) and *Haloferax volcanii* (HVO_0073). Domain analysis of protein sequences for proteins were carried out using Pfam, and domains were coloured (as above). Protein structures were mapped using Phyre2 protein modelling software (Kelley *et al.*, 2015). Domain boxes are not to scale. Similarities in domain layout and protein structure are seen between all arCOG00427 proteins shown.

The presence of arCOG00428 members only within *Halobacteriales* suggests this is a result of duplication and horizontal gene transfer, common of halophilic archaea where polyploid members regularly share genes between species (Rhodes *et al.*, 2011). The sequence degradation of RecJ2 compared to others (Figure 4.5) and outlying the two main arCOGs carried by other species

suggests RecJ2 is unlikely to play the role of the ‘active’ RecJ in *H. volcanii* and would be predicted to have instead gained a novel role. With regards to the arCOG00429 clade, it appears to have undergone a fast evolution following a duplication event within *Euryarchaea* and have gained a specialised function (Makarova *et al.*, 2012).

Due to its grouping with both *Tko*HAN and *Pfu*HAN, it can be predicted that *Hvo*RecJ3 and *Hvo*RecJ4 carry out similar roles in DNA replication and/or repair (Nagata *et al.*, 2017b, Feng *et al.*, 2018). However, the loss of catalytic residues seen for RecJ4 would argue against this; it cannot be ruled out that RecJ4 has gained a non-catalytic role elsewhere (potentially as a scaffold mediating protein interaction(s) through its IDR). At the structural level, all arCOG00429 proteins show similarities, providing good evidence for RecJ3 and RecJ4 being HAN-like proteins (**Figure 4.8**).

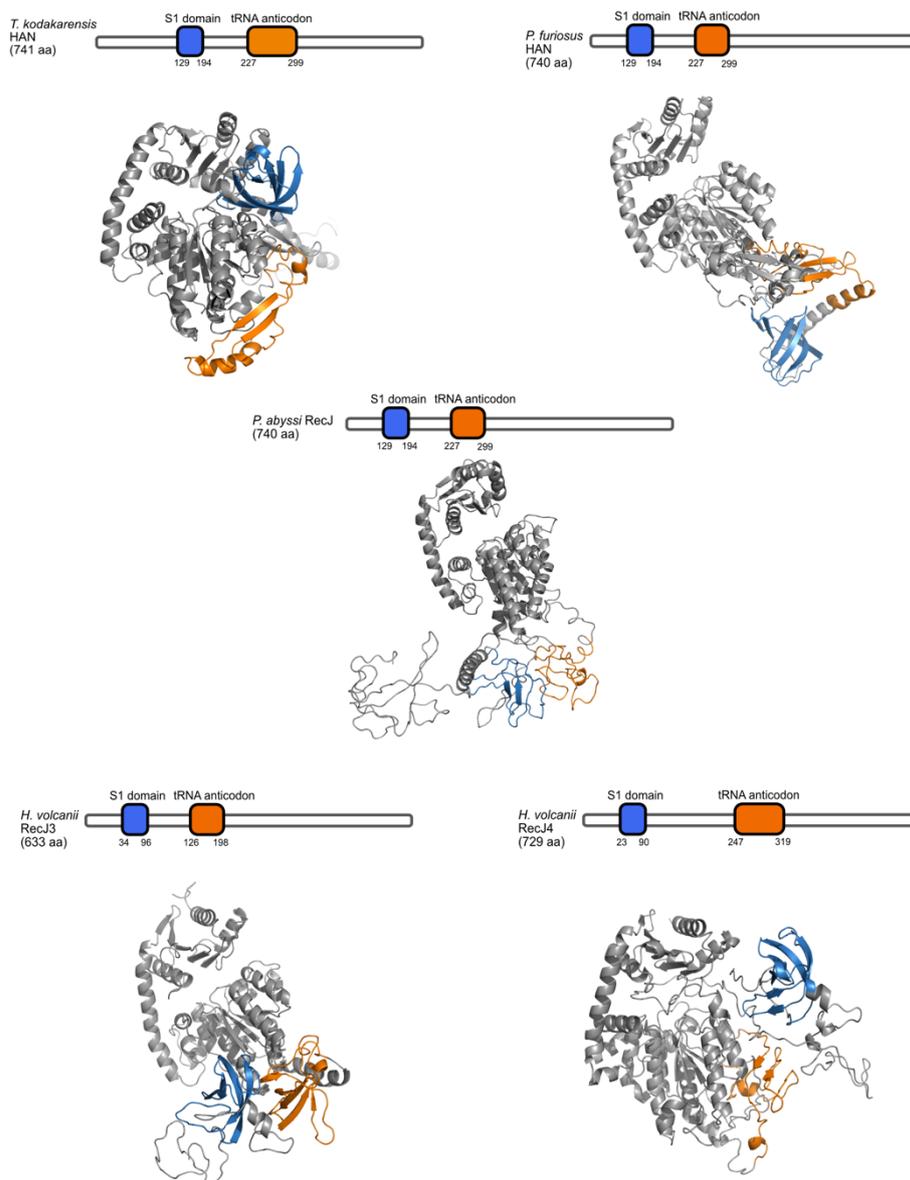


Figure 4.8: Comparison of structures of arCOG00429 HAN-like proteins in *Thermococcus kodakarensis* (TK0155), *Pyrococcus furiosus* (PF0399), *P. abyssi* (PAB1284) and *Haloferax volcanii* (HVO_1018 and HVO_2889). Domain analysis of protein sequences for proteins were carried out using Pfam, and domains were coloured (as above). Protein structures were mapped using Phyre2 protein modelling software (Kelley *et al.*, 2015). Domain boxes are not to scale. Similarities in domain layout and protein structure are seen between all arCOG00429 proteins shown; in particular, the structures of *Pfu*HAN/*Pab*RecJ and *Hvo*RecJ3, and *Tko*HAN and *Hvo*RecJ4 both show very strong structural similarities.

The ability to delete RecJ1, RecJ3 and RecJ4 in *H. volcanii* is surprising given that they are predicted to play active roles in DNA replication and/or repair. However, in fellow euryarchaeon *T. kodakarensis*, Δ gan Δ han double deletions are viable suggesting some redundancy in replication and repair mechanisms (Nagata *et al.*, 2017b). This would represent deletion of all members of arCOG00427 and arCOG00429 for this strain. Deletion of members of arCOG00427 and arCOG00429 in *H. volcanii* would cover the deletion of RecJ1, RecJ3 and RecJ4. RecJ2 is a member of a different arCOG (arCOG00428), and as yet it is currently unknown whether this can be deleted in species carrying arCOG00427 and arCOG00429 members.

Genomic context of RecJ genes

Genome neighbourhood analysis of *T. kodakarensis* GAN shows its inclusion within an operon containing ribosomal proteins S15 and S3, and a subunit of the tRNA-modifying KEOPS complex. This conserved neighbourhood is seen in numerous genomes, both where arCOG00427 representatives are present and absent. For example, in some *Halobacteriales* species, where no arCOG00427 candidates are encoded, instead arCOG00428 or arCOG00432 members are found in the same genomic context. Therefore, the argument is made for these alternative but syntenically conserved genes to perform similar functions to the missing arCOG00427 proteins (Makarova *et al.*, 2012).

Since *H. volcanii* contains members of both arCOG00427 (RecJ1) and arCOG00428 (RecJ2) it was of interest to see if either is found within this conserved operonic layout. Synteny analysis was performed using SyntTax (Oberto, 2013) against both *T. kodakarensis* and species within the order *Halobacteriales* using *H. volcanii* RecJ gene sequences to examine genomic neighbourhood conservation. Halophiles are often polyploid and are known to undergo a large amount of lateral gene transfer between species (both halophiles and non-halophilic species) (Rhodes *et al.*, 2011). As a result, tight gene linkages within halophilic species are even more likely to carry functional relevance, where the operon of genes has been selected for and maintained as a unit (Gabaldon and Huynen, 2004, Korbel *et al.*, 2004, Wolf *et al.*, 2001).

While analysis of RecJ1, RecJ3 and RecJ4 revealed no genome neighbourhood conservation, RecJ2 was shown to be part of a strongly conserved operonic layout maintained throughout all *Halobacteriales* species screened (**Figure 4.9**). However, synteny was not identified between *H. volcanii* RecJ2 and *T. kodakarensis* GAN; while the operonic layout is comparable, the difference in gene sequence (and arCOG grouping) means there is no definable syntenic link between *Tko*GAN and RecJ2. When

searching using the protein sequence for *TkoGAN*, the only syntenic link within *H. volcanii* was to that of *RecJ1* (BLAST score 23.25; both within the same arCOG and share structural similarities).

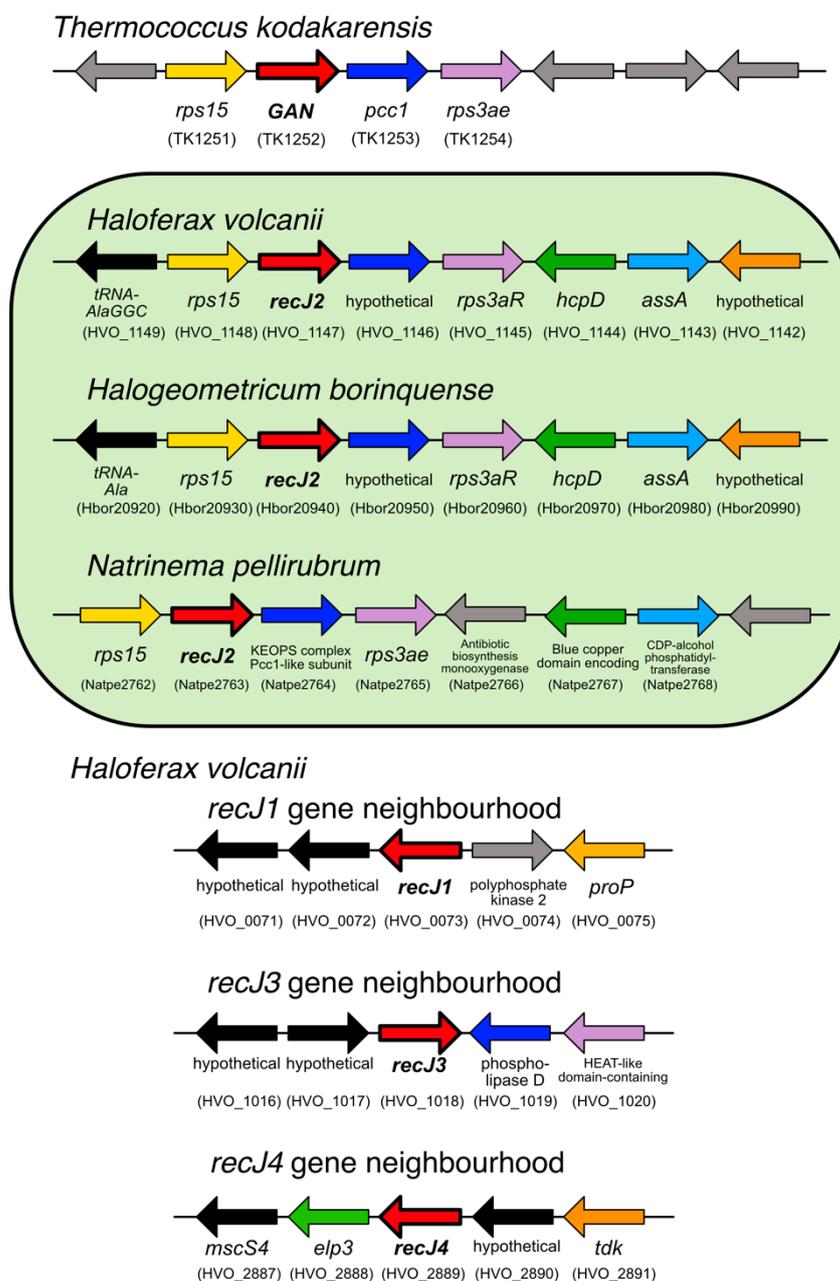


Figure 4.9: Synteny analysis of *Haloferax volcanii* RecJ gene neighbourhoods in halophilic species and comparison to gene neighbourhood of *Thermococcus kodakarensis* GAN. Genes located two upstream and five downstream of *recJ2/GAN* (red) labelled and coloured depending on similarity of predicted function. *T. kodakarensis* GAN shows similarity in neighbourhood to RecJ2 neighbourhoods of that of the RecJ2 region within halophilic species. However, only halophilic species (green box) show synteny for the RecJ2 neighbourhood. There is no synteny with the remaining *H. volcanii* *recJ* genes (*recJ1*, *recJ3* or *recJ4*). Locus tags for each gene are indicated in brackets.

Genes tightly linked to *recJ2* include those encoding ribosomal proteins S15 and S3Ae and a gene predicted to encode a member of the KEOPS complex (Kinase, Endopeptidase and Other Proteins of small Size; HVO_1146, *pcc1* in *T. kodakarensis*). The KEOPS complex is involved in tRNA modification and is conserved in eukaryotes with some bacterial and archaeal species carrying homologues (Srinivasan *et al.*, 2011, Wan *et al.*, 2016, Naor *et al.*, 2012). These genes are predicted to be essential, due to having key roles in transcription and translation processes. It is important to note that *H. volcanii* RecJ2 is positioned within a co-transcribed operon; ribosomal proteins are known to be highly expressed and thus the same can be expected of RecJ2 expression levels (Anita Marchfelder, personal communication).

arCOG00428 (i.e., RecJ2) are documented to act as a *bona fide* RecJ in species lacking the canonical arCOG00427 (Makarova *et al.*, 2012). However, in the case of *H. volcanii* RecJ2 its predicted lack of catalytic activity and the presence of arCOG00427 member RecJ1 may argue against its function as a canonical RecJ. It is unlikely the positioning of *H. volcanii* RecJ2 within this operon is the reasoning for its inability to be deleted: *TkoGAN* is found in the same operonic layout and Δ *gan* strains were able to be constructed and surrounding gene expression levels were shown to be unaltered (Burkhart, 2017). Further work is needed to elucidate whether RecJ2 has indeed gained a new role within *H. volcanii*.

4.3.2 Characterisation of strains deleted for *recJ1*, *recJ3* and/or *recJ4*

Deletions for *recJ1*, *recJ3* and *recJ4* (single, double and triple mutants) had previously been generated in a background in which no genes known or suspected to function in DNA replication or repair are mutated (H164; Δ *pyrE2 bgaHa-Bb leuB-Ag1 Δ trpA*), they showed no difference to WT in a standard growth rate assay (Lever, 2019). Based on studies of RecJ family proteins in other euryarchaea (specifically, *T. kodakarensis*), it would be predicted that in *H. volcanii*, one (or more) RecJ protein should act in replication as the GAN (GINS-associated nuclease) homologue, with other RecJ(s) acting as the HAN (Hef-associated nuclease) homologue to overcome stalled replication forks. A study of the sensitivity of the *recJ* deletion strains to DNA damaging agents and replication inhibitors was carried out, to clarify the roles of RecJ1, RecJ3 and RecJ4.

The intention was to screen that strains carrying single deletions for *recJ1*, *recJ3* and *recJ4* alongside the strain reported to carry deletions of all three *recJ* genes (H4376; Lever, 2019). However, further screening of H4376 by colony PCR and colony hybridisation revealed that this strain was only deleted for *recJ1*. Screening of alternative strains revealed that H4273 carries deletions of *recJ1*, *recJ3* and *recJ4* in a H164 background (where Δ *recJ4* is marked by *trpA*⁺); this strain was subsequently used for phenotypic analysis alongside the single *recJ* mutants.

Growth rate

Due to the presumptive triple mutant strain H4376 in fact being wild type for *recJ3* and *recJ4*, all single mutants H3929, H3931, H3932 were instead screened for growth rate in comparison to the genuine triple *recJ* mutant, H4273, and control strain H164.

Strains were grown for two consecutive overnights in Hv-YPC broth, ensuring on day three that actively dividing cells were used for the assay. Cells were diluted in fresh Hv-YPC broth and arrayed in a 96-well plate. Optical density (OD; A₆₀₀) was semi-continuously measured, allowing the plotting of growth curves for each strain (**Figure 4.10**).

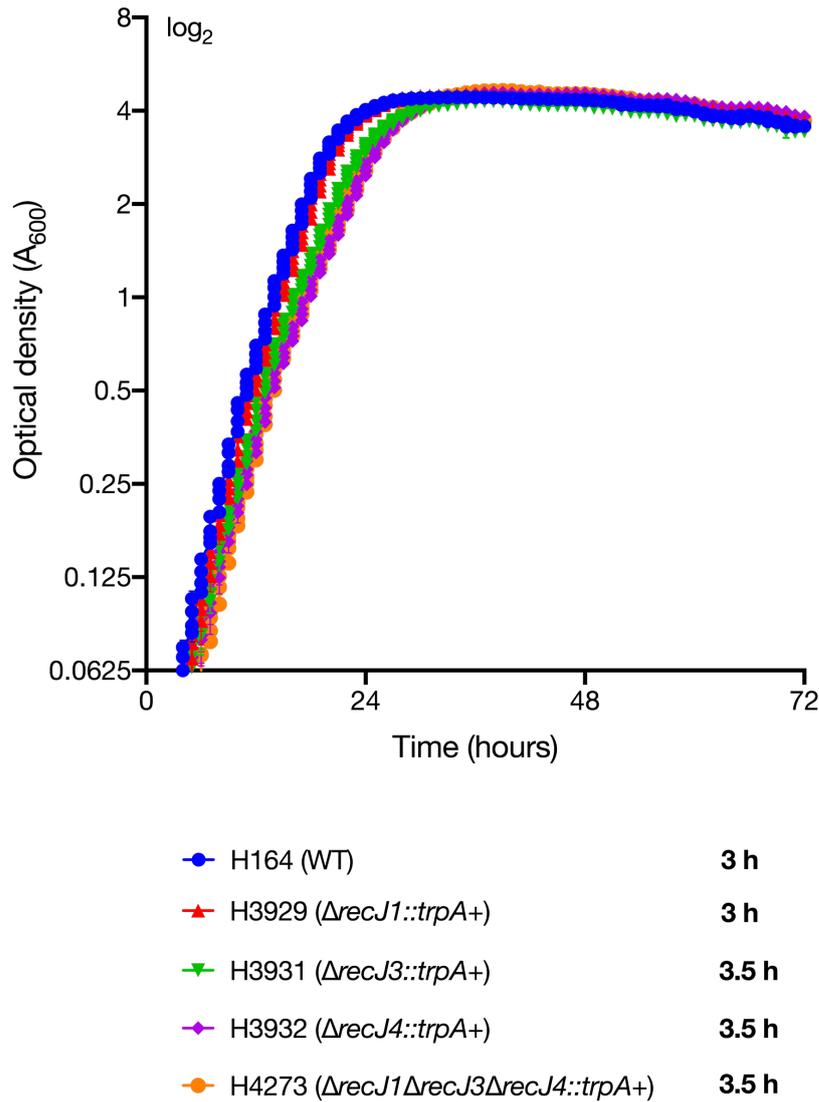


Figure 4.10: Exponential growth rate of strains deleted for *recJ1*, *recJ3* and/or *recJ4*. Generation time in hours (h) is indicated in bold in the legend. Strains were grown in Hv-YPC broth for two consecutive overnights before being diluted and plated. All strains (n=2) were incubated on the same 96-well plate and measured simultaneously for optical density (A_{600}) using an Epoch2 Microplate Spectrophotometer (BioTek). Error bars represent standard error of the mean (SEM).

The $\Delta recJ$ mutant strains have comparable generation times, with H3931 ($\Delta recJ3::trpA+$), H3932 ($\Delta recJ4::trpA+$) and H4273 ($\Delta recJ1 \Delta recJ3 \Delta recJ4::trpA+$) having a slight delay in doubling time (3.5 hours) compared to WT (3 hours).

DNA content and cell size

In order to determine the effect of *recJ1*, *recJ3* and/or *recJ4* deletions on DNA content and cell size, strains were analysed using flow cytometry (Figure 4.11).

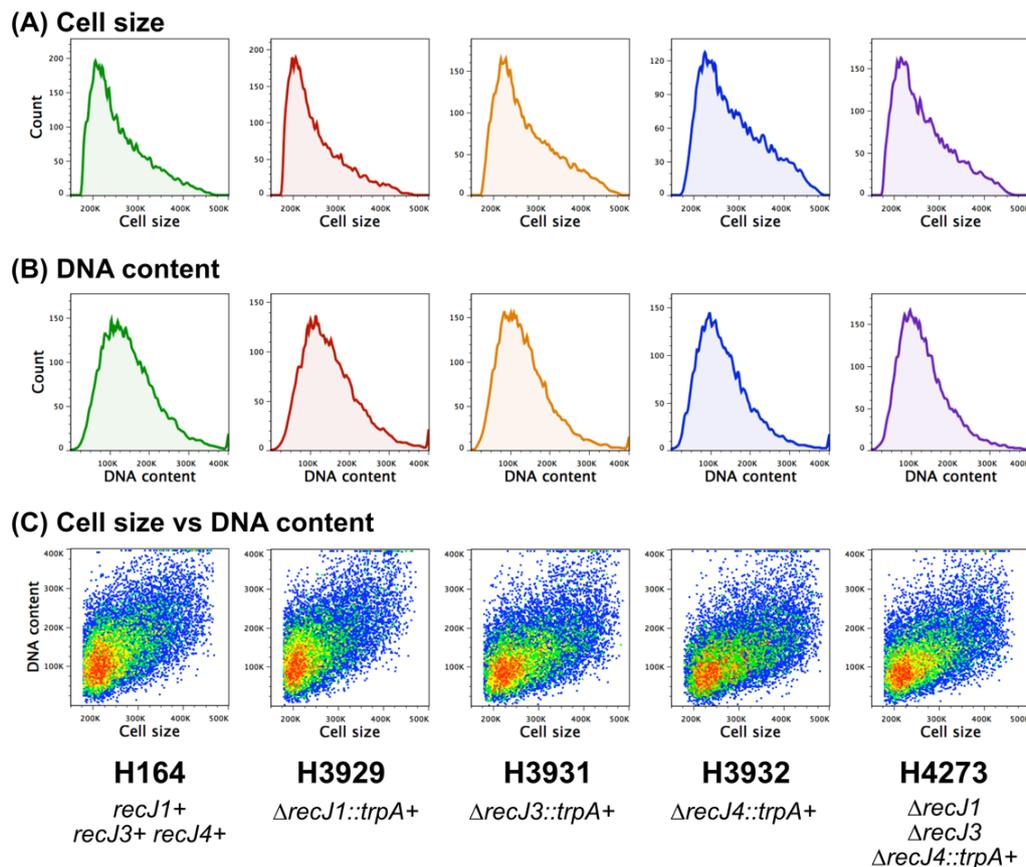


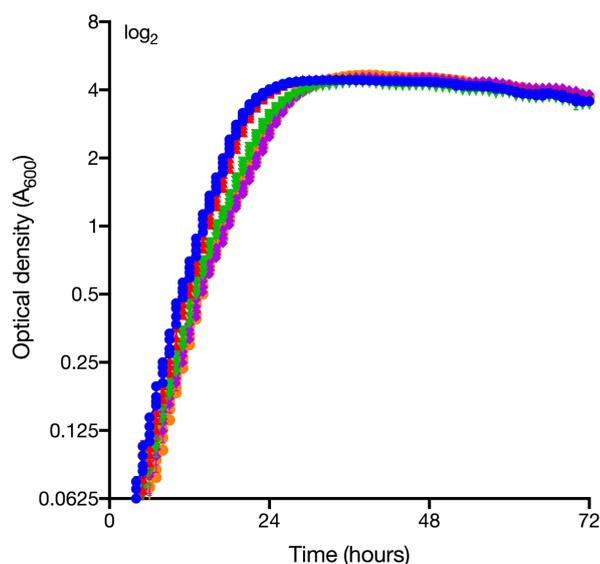
Figure 4.11: Flow cytometry analysis of strains deleted for *recJ1*, *recJ3* and/or *recJ4*. (A) Determination of cell size. H3931 ($\Delta recJ3::trpA+$), H3932 ($\Delta recJ4::trpA+$) and H4273 ($\Delta recJ1 \Delta recJ3 \Delta recJ4::trpA+$) have a broader spread of cell sizes than either H164 (wild type) or H3929 ($\Delta recJ1::trpA+$). (B) Determination of DNA content. All cells show similar profiles. (C) Density dot plots displaying cell size vs DNA content. H3929 is very similar to H164, while H3931, H3932 and H4273 show a larger distribution of cell sizes.

While the DNA content profiles of all strains were comparable, the range of cell sizes differed between mutants. H3929 ($\Delta recJ1::trpA+$) was generally comparable to the control strain (H164) in all plots, while H3931 ($\Delta recJ3::trpA+$), H3932 ($\Delta recJ4::trpA+$) and H4273 ($\Delta recJ1 \Delta recJ3 \Delta recJ4::trpA+$) showed a broader range of cell sizes, having a higher proportion of larger cells than H164 and H3929.

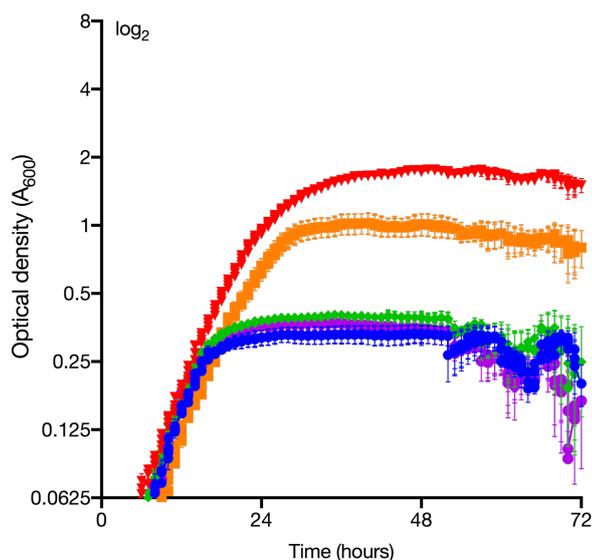
Survival following inhibition of replicative polymerase PolB

Aphidicolin is a specific inhibitor of Family B replicative polymerases, which play a critical role in genome replication (Forterre *et al.*, 1984). If one of the RecJ proteins is acting in replication as a key member of the CMG complex (akin to Cdc45 in eukaryotes), it could be argued that the deletion of this RecJ would impact the efficiency of DNA replication and treatment with an inhibitor of replication would show a more adverse response in the absence of the RecJ. Strains deleted for *recJ1* (H3929), *recJ3* (H3931), *recJ4* (H3932), and *recJ1*, *recJ3* and *recJ4* (H4273) were tested for response to aphidicolin treatment compared to the control strain (H164). Strains were grown for two consecutive overnights in Hv-YPC broth, ensuring on day three actively dividing cells were used for the assay. Cells were diluted in fresh Hv-YPC broth and treated with 10 µg/ml aphidicolin (or DMSO as control) in a 96-well plate. Optical density (OD; A₆₀₀) was semi-continuously measured, allowing the plotting of growth curves for each strain (**Figure 4.12**).

(A) 0 $\mu\text{g/ml}$ aphidicolin



(B) 10 $\mu\text{g/ml}$ aphidicolin



- H164 (WT)
- H3929 ($\Delta recJ1::trpA+$)
- H3931 ($\Delta recJ3::trpA+$)
- H3932 ($\Delta recJ4::trpA+$)
- H4273 ($\Delta recJ1 \Delta recJ3 \Delta recJ4::trpA+$)

Figure 4.12: Effect of aphidicolin treatment on survival of wild type and $\Delta recJ$ strains. Wild type (H164) and strains with *recJ1* (H3929), *recJ3* (H3931), *recJ4* (H3932) or *recJ1*, *recJ3* and *recJ4* (H4273) deletions were monitored for growth through measurement of optical density (A_{600}). All strains were treated with either DMSO (A; control) or aphidicolin (B; final concentration 7 $\mu\text{g/ml}$) chronically for 72 hours. All strains (n=2) were incubated in a single 96-well plate and measured simultaneously using an Epoch Microplate Spectrophotometer (BioTek). Error bars represent standard error of the mean (SEM).

In the presence of aphidicolin, there is an altered response in the strain deleted for *recJ1* (H3929) and in the triple deletion of *recJ1*, *recJ3* and *recJ4* (H4273) (**Figure 4.12 B**). Of the two, H3929 ($\Delta recJ1$) appears the most resistant to aphidicolin, with H4273 showing an intermediate phenotype between a control strain-like response and the increased resistance of H3929. However, the strain is still affected to some degree, with its final OD falling short of that of untreated cells (OD ~ 2 vs ~ 4). Strains H3931 ($\Delta recJ3::trpA+$) and H3932 ($\Delta recJ4::trpA+$) have a response to aphidicolin treatment comparable to H164.

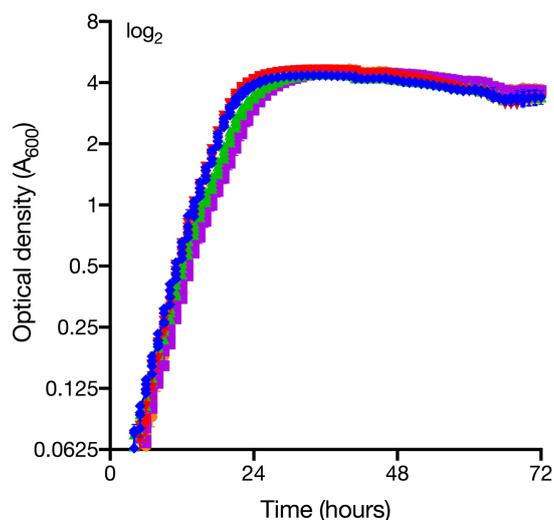
Survival following treatment with DNA damaging agents

Hydrogen peroxide treatment

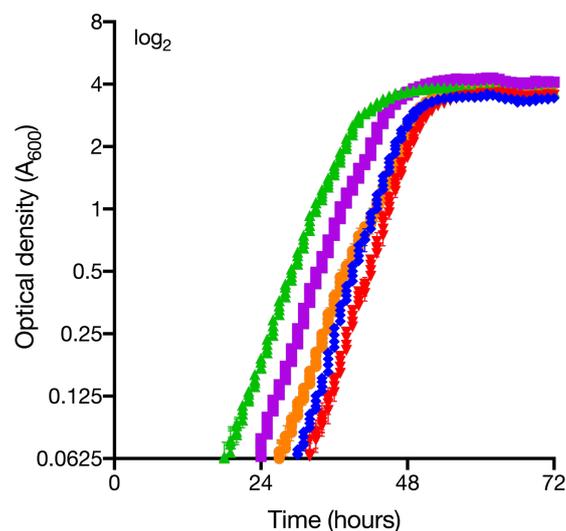
Hydrogen peroxide (H_2O_2) is an oxidising agent that leads to production of highly toxic hydroxyl radicals, which causes oxidative base damage and strand breaks in DNA (Goldstein *et al.*, 1993). Due to the many types of base lesions that can be caused by oxidative damage, there are a plethora of pathways involved in its repair. Base excision repair (BER), transcription-coupled repair (TCR), global genome repair (GGR), mismatch repair (MMR), translesion synthesis (TLS), homologous recombination (HR) and non-homologous end-joining (NHEJ) have all been identified as contributors to repair of oxidative lesions, however BER is the major repair pathway for oxidative DNA damage (Slupphaug *et al.*, 2003).

To test for the involvement of RecJ proteins in oxidative damage repair, strains deleted for *recJ1* (H3929), *recJ3* (H3931), *recJ4* (H3932) and *recJ1*, *recJ3* and *recJ4* (H4273) were tested for response to H_2O_2 treatment compared to the control strain (H164). Strains were grown for two consecutive overnights in Hv-YPC broth, ensuring on day three actively dividing cells were used for the assay. Cells were diluted in Hv-YPC broth and treated with H_2O_2 (or sterile distilled water as control) chronically for 72 hours in a 96-well plate. OD (A_{600}) was semi-continuously measured, allowing the plotting of growth curves for each strain (**Figure 4.13**).

(A) 0 mM H₂O₂



(B) 0.4 mM H₂O₂



- ◆ H164 (WT)
- ◆ H3929 ($\Delta recJ1::trpA+$)
- ◆ H3931 ($\Delta recJ3::trpA+$)
- ◆ H3932 ($\Delta recJ4::trpA+$)
- ◆ H4273 ($\Delta recJ1\Delta recJ3\Delta recJ4::trpA+$)

Figure 4.13: Effect of H₂O₂ treatment on survival of wild type and $\Delta recJ$ strains. Wild type (H164) and strains with *recJ1* (H3929), *recJ3* (H3931), *recJ4* (H3932) or *recJ1*, *recJ3* and *recJ4* (H4273) deletions were monitored for growth through measurement of optical density (A₆₀₀). All strains were treated with either sterile distilled water (A; control) or H₂O₂ (B; final concentration 0.4 mM) chronically for 72 hours. All strains (n=2) were incubated in a single 96-well plate and measured simultaneously using an Epoch Microplate Spectrophotometer (BioTek). Error bars represent standard error of the mean (SEM).

In the presence of H₂O₂, H3931 ($\Delta recJ3::trpA+$) and H3932 ($\Delta recJ4::trpA+$) show a slight resistance to treatment compared to H164 (WT), where these strains are able to begin growing exponentially prior to others (including H164). H3929 ($\Delta recJ1::trpA+$) and H4273 ($\Delta recJ1 \Delta recJ3 \Delta recJ4::trpA+$) show a response comparable to H164.

Mitomycin C treatment

Mitomycin C (MMC) is a naturally synthesised chemical mutagen produced by *Streptomyces caespitosus* and is commonly used as a chemotherapeutic anti-tumour agent (Tomasz, 1995). MMC acts to crosslink DNA, either within the same DNA strand (intra-strand) or between different strands of the duplex (inter-strand); MMC has a bias for interstrand crosslinks. Replication forks are unable to bypass DNA interstrand crosslinks, thus it is vital for these to be repaired in order to replicate DNA successfully. NER, homologous recombination, and translesion DNA synthesis are involved in the repair of DNA interstrand crosslinks.

To test for the involvement of RecJ proteins in DNA crosslink repair, strains deleted for *recJ1* (H3929), *recJ3* (H3931), *recJ4* (H3932) and *recJ1*, *recJ3* and *recJ4* (H4273) were tested for response to MMC treatment compared to the control strain (H164). Cultures were grown to mid-exponential phase, and replica-spotted onto Hv-YPC agar containing either water (control) or MMC. Plates were incubated for 4-7 days at 45°C, colonies were counted, and survival fractions were calculated (**Figure 4.14**).

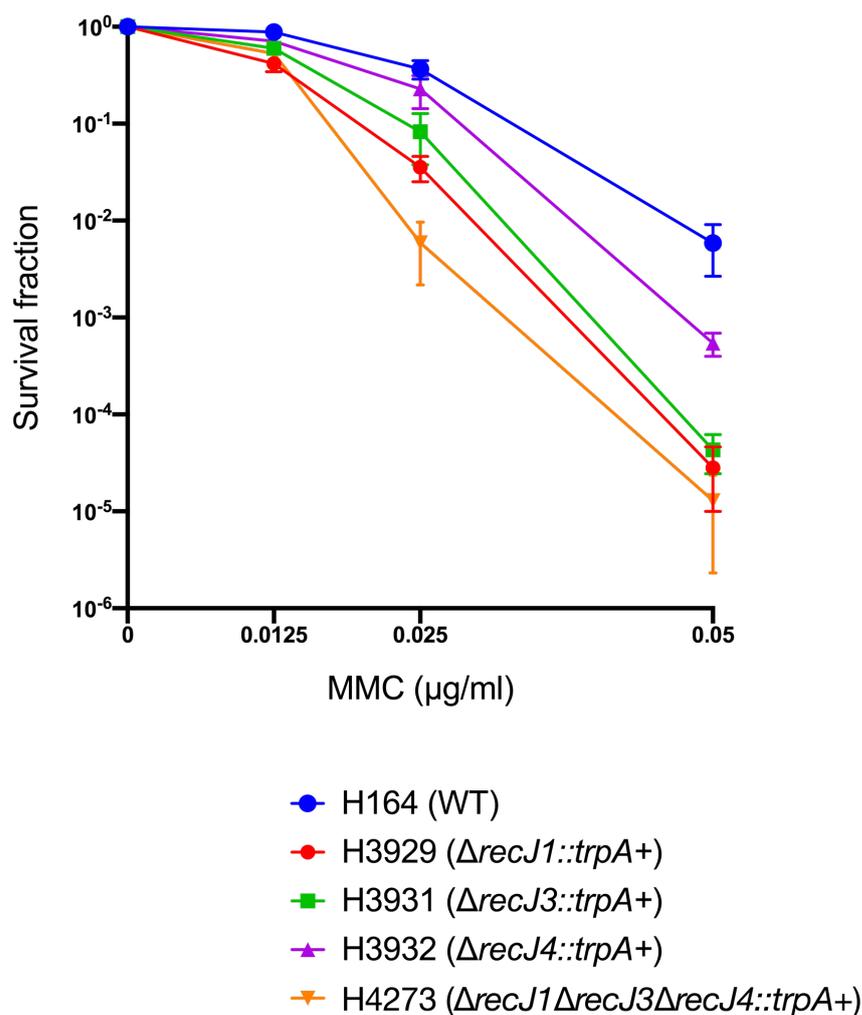


Figure 4.14: Survival frequency of $\Delta recJ$ strains following treatment with mitomycin C (MMC). Wild type (WT; H164) and strains with *recJ1* (H3929), *recJ3* (H3931), *recJ4* (H3932) or *recJ1*, *recJ3* and *recJ4* (H4273) deletions were chronically exposed to MMC within YPC plates. All strains carrying *recJ* deletions are more sensitive to MMC than wild type (H164), with H3929, H3931 and H4273 having the largest response. Survival fraction is calculated relative to untreated control. Each data point is generated as an average of at least 3 independent trials. Error bars represent standard error of the mean (SEM).

At the highest dose of MMC tested (0.05 µg/ml), all strains harbouring at least one *recJ* deletion are more sensitive to MMC than the control strain (H164). RecJ1 and RecJ3 mutant strains show an increased sensitivity to MMC compared to the RecJ4 mutant strain. The triple mutant H4273 has a response generally comparable to that of the single mutants.

UV treatment

Ultraviolet (UV) radiation is a source of major DNA damage, causing a variety of DNA lesions, including pyrimidine dimers, 6-4 photoproducts and single/double-stranded DNA breaks (Sinha and Hader, 2002). Pyrimidine dimers and 6-4 photoproducts are ‘bulky’ lesions and act to distort the sugar-phosphate backbone of the DNA duplex. These lesions, if left unrepaired, can result in mismatch errors during replication (Helleday *et al.*, 2014). With the exception of placental mammals, all organisms have the innate ability to repair these photolesions using photolyases, in a visible light-dependent process named photoreactivation (Sancar, 2000). NER is also used to remove bulky lesions from DNA (Sinha and Hader, 2002); this is the primary method for repairing bulky lesions in humans. Reactive oxygen species may also arise as a result of UV irradiation and the free radicals formed are then capable of causing single strand DNA (ssDNA) breaks. If two ssDNA breaks occur in close proximity, or if the replication fork encounters a ssDNA break on the leading strand, this can lead to the production of a double strand DNA break (DSB). DSBs can be repaired by homologous recombination (HR), non-homologous end joining (NHEJ) or microhomology-mediated end joining (MMEJ).

To test for the involvement of RecJ proteins in UV damage repair, strains deleted for *recJ1* (H3929), *recJ3* (H3931), *recJ4* (H3932) or all three (H4273) were tested for response to UV treatment compared to the control strain (H164). Cultures were grown to mid-exponential phase, spotted onto Hv-YPC agar and treated with varying doses of UV. Plates were incubated in the dark to prevent the innate visible light-dependent action of photolyases from repairing the bulky lesions in the treated strains. After 4-7 days of incubation at 45°C, colonies were counted, and survival fractions were calculated accordingly (**Figure 4.15**).

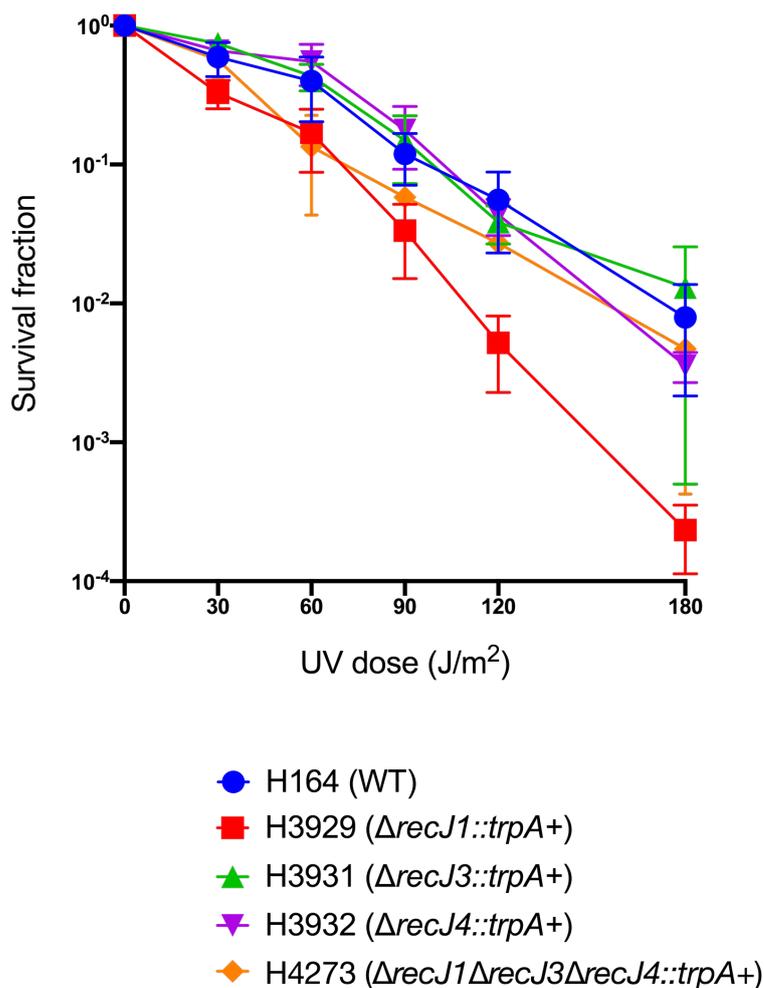


Figure 4.15: Survival frequency of $\Delta recJ$ strains following treatment with ultraviolet (UV) radiation. Wild type (H164) and strains with *recJ3* (H3931), *recJ4* (H3932) or *recJ1*, *recJ3* and *recJ4* (H4273) show no significant difference in survival fractions. Strain H3929 ($\Delta recJ1::trpA+$) shows an increased sensitivity to UV treatment at compared to other strains at higher doses (>100 J/m²). Survival fraction is calculated relative to untreated control. Each data point is generated as an average of at least 3 independent trials. Error bars represent standard error of the mean (SEM).

Strains H3931($\Delta recJ3::trpA+$), H3932 ($\Delta recJ4::trpA+$) and H4273 ($\Delta recJ1\Delta recJ3\Delta recJ4::trpA+$) show no difference in UV response to H164 (control). H3929 ($\Delta recJ1::trpA+$) shows a slight sensitivity to UV compared to H164 at high doses (>90 J/m²).

Phleomycin treatment

Phleomycin is a member of the bleomycin antibiotic family, isolated from a mutant strain of *Streptomyces verticillus*. Phleomycin, a common chemotherapeutic, binds and intercalates DNA. This may lead to both single- and double-strand DNA breaks (Sleigh, 1976). Phleomycin is also capable of producing free radicals in the presence of Fe^{2+} and O_2 , which can lead to DNA breakage. Breaks induced by phleomycin may be repaired by HR, NHEJ or MMEJ.

To test for the involvement of RecJ proteins in phleomycin-induced DNA break repair, strains deleted for *recJ1* (H3929), *recJ3* (H3931), *recJ4* (H3932) or all three (H4273) were tested for response to phleomycin treatment compared to the control strain (H164). Cultures were grown to mid-exponential phase and incubated with phleomycin in Hv-YPC broth for 1 hour. Following this, cells were washed to remove phleomycin and spotted onto Hv-YPC agar. Plates were incubated at 45°C for 4-7 days, colonies were counted, and survival fractions were calculated accordingly (**Figure 4.16**).

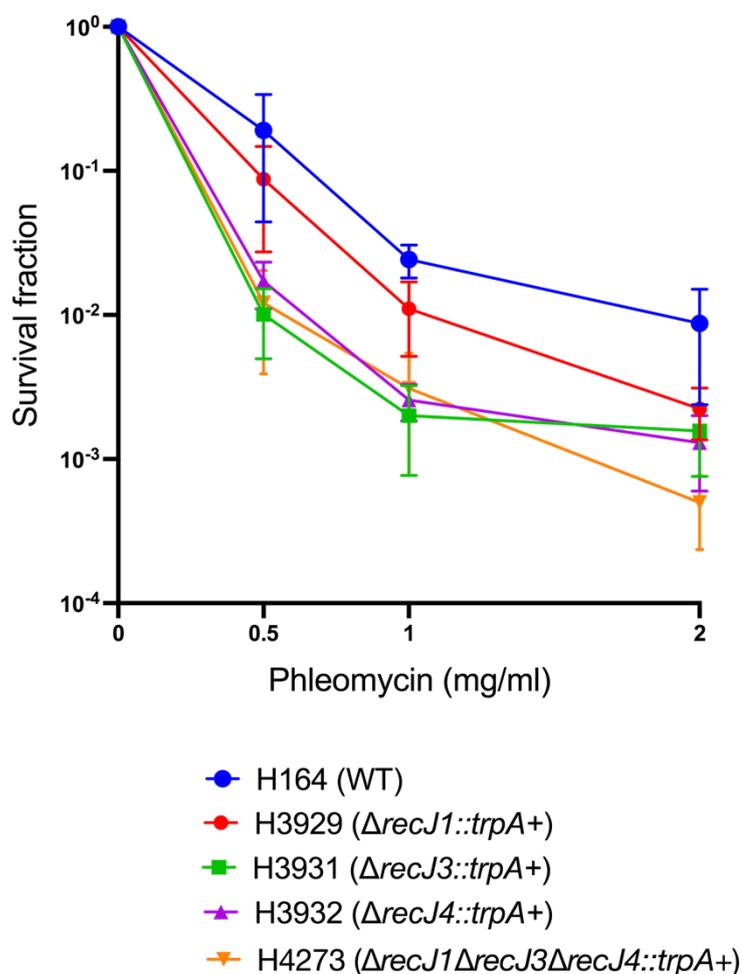


Figure 4.16: Survival frequency of $\Delta recJ$ strains following treatment with phleomycin. H3931 ($\Delta recJ3$), H3932 ($\Delta recJ4$) and H4273 ($\Delta recJ1\Delta recJ3\Delta recJ4$) show increased sensitivity compared to H164 at 0.5-1 mg/ml phleomycin. However, at the highest dose (2 mg/ml), only H4273 shows an increased sensitivity when compared to H164. Survival fraction is calculated relative to untreated control. Each data point is generated as an average of at least 3 independent trials. Error bars represent standard error of the mean (SEM).

At the lower doses of phleomycin tested (0.5-1 mg/ml), H3931 ($\Delta recJ3::trpA+$), H3932 ($\Delta recJ4::trpA+$) and H4273 ($\Delta recJ1\Delta recJ3\Delta recJ4::trpA+$) show an increased sensitivity to phleomycin compared to WT (H164). However, at the highest dose, only the survival fraction of H4273 differs from that of H164.

Recombination frequency

Recombination assays allow measurement of the frequency of homologous recombination between the chromosome and a closed circular plasmid through use of two mutant *leuB* alleles (reported in (Lestini *et al.*, 2010); a schematic is shown in **Figure 4.17** and detailed information on the assay is found in *Materials and Methods*). Screening of transformants for ability to synthesise uracil will also allow a definition of whether a crossover

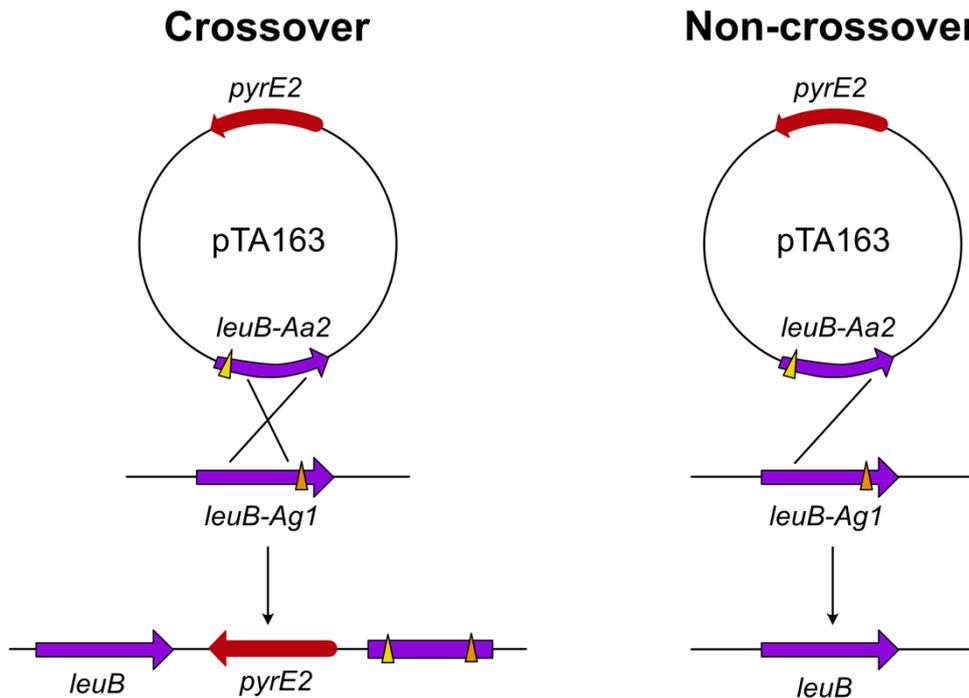


Figure 4.17: Recombination assay. $\Delta pyrE2$ strains with a chromosomal *leuB-Ag1* (*leu*-) allele are transformed with pTA163, containing *pyrE2* and *leuB-Aa2* (*leu*-) alleles. A recombination event between the chromosome and the plasmid mutant *leuB* alleles can generate a wild type *leuB* allele and thus cells having undergone a recombination event can now grow in the absence of leucine. Depending on whether a crossover (CO) or non-crossover (NCO) event occurred, strains will either retain (CO) or lose (NCO) the *pyrE2* marker of pTA163 on the chromosome.

(CO) or non-crossover (NCO) recombination event has taken place.

In order to assess whether deletion of *recJ1*, *recJ3* or *recJ4* has an effect on the frequency of recombination in *H. volcanii*, a recombination assay was performed (**Table 4.2**).

Table 4.2: Recombination frequencies of strains deleted for *recJ1*, *recJ3*, *recJ4* alone or in combination.

Strain	H164	H3929	H3931	H3932	H4273
	<i>recJ+</i>	$\Delta recJ1$	$\Delta recJ3$	$\Delta recJ4$	$\Delta recJ1\Delta recJ3$ $\Delta recJ4$
Recombination frequency (RF)	7.23×10^{-5} (+/- 1.96×10^{-5})	9.20×10^{-5} (+/- 2.28×10^{-5})	1.57×10^{-4} (+/- 2.44×10^{-5})	1.26×10^{-4} (+/- 1.73×10^{-6})	1.07×10^{-4} (+/- 2.34×10^{-5})
Transformation efficiency (TE)	5.09×10^{-2}	1.37×10^{-1}	1.77×10^{-2}	1.64×10^{-2}	5.20×10^{-2}
Relative recombination frequency (normalised by TE)	1.42×10^{-3}	6.71×10^{-4}	8.92×10^{-3}	7.68×10^{-3}	2.05×10^{-3}
	1×	0.5×	6.3×	5.4×	1.4×
Crossover fraction	17.5%	41.25%	25%	21.25%	26.25%
Non-crossover fraction	82.5%	58.75%	75%	78.75%	73.75%

Values in bold indicate the amount of recombination compared to wild type H164 (*recJ+*). Values are generated as an average of at least 3 independent trials, +/- standard error is shown in brackets. Cells are shaded blue to indicate a recombination defect and red to indicate hyper-recombination. The fraction of crossover and non-crossover events are represented as a percentage, with cells being shaded differently where values differ significantly from the wild type ($p < 0.05$ with one degree of freedom with a chi-squared test); blue indicates a decrease, red indicates an increase compared to *recJ+* strain H164. 80 colonies per strain were assayed for quantification of crossover vs. non-crossover events.

Both H3931 ($\Delta recJ3::trpA+$) and H3932 ($\Delta recJ4::trpA+$) show a hyper-recombinant phenotype with recombination levels of 6.3x and 5.4x respectively higher than that of H164 (*recJ+*). This could suggest a role for RecJ3 and/or RecJ4 as anti-recombinases. A decrease in the level of recombination was observed for the $\Delta recJ1$ strain H3929 (0.5x recombination level compared to H164), alongside an increase in CO events and decrease in NCO events compared to H164.

4.3.2.1 Dependence of $\Delta recJI$ mutant on *radA* expression

Following the observation that the $\Delta recJI$ mutant H3929 showed a resistance to aphidicolin comparable to that of strains lacking the main chromosomal origins ($\Delta oriC1\Delta oriC2\Delta oriC3\Delta ori-pHV4$: e.g., H1804), the question arose as to whether the mode of replication is altered in a *recJI* mutant, and whether this altered mode of replication shares any similarities with the predicted mode of replication utilised by originless strains (namely RDR).

Originless replication has been shown to be reliant on recombinase protein RadA (Hawkins *et al.*, 2013a). When placed under tryptophan-inducible promoter *p.tnaA*, induction of *radA* is essential in a background deleted for the main chromosomal origins of replication (**Figure 4.18**).

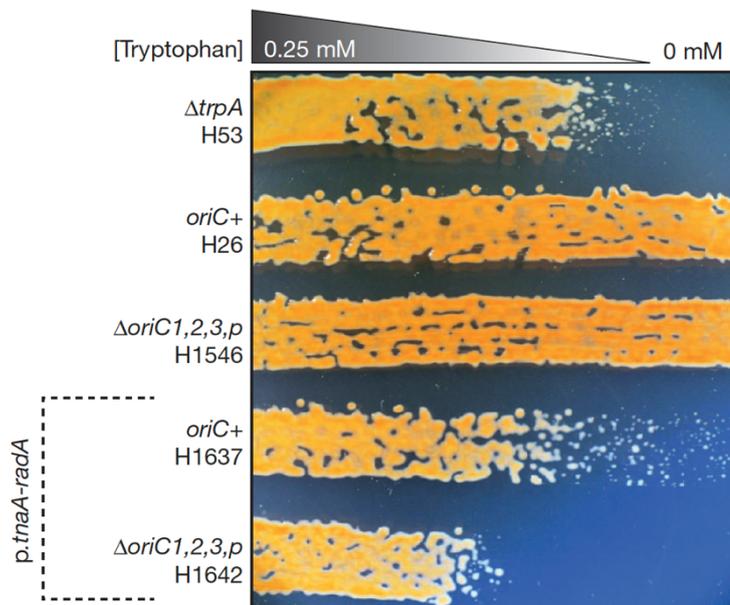


Figure 4.18: RadA recombinase is essential in a strain lacking origins *oriC1*, *oriC2*, *oriC3* and *ori-pHV4*. Where *radA* is under tryptophan-inducible promoter *p.tnaA*, its expression is essential for growth in the strain deleted for the chromosomal origins (H1642), while the *oriC+* strain remains viable in the absence of induction (H1637). Taken from (Hawkins *et al.*, 2013a).

Should deletion of *recJI* lead to a similar mode of replication to that of strains deleted for the main chromosomal origins, it would be predicted that a *recJI* mutant in an *oriC+* background should also now be dependent upon RadA expression. To test this hypothesis, a tryptophan inducible RadA construct was integrated into $\Delta recJI$ mutant H3929. Plasmid pTA1837 allows

for integration of full-strength tryptophan-inducible promoter, *p.tnaA*, at the start of the reading frame of *radA* (**Figure 4.19**) (Ausiannikava *et al.*, 2018).

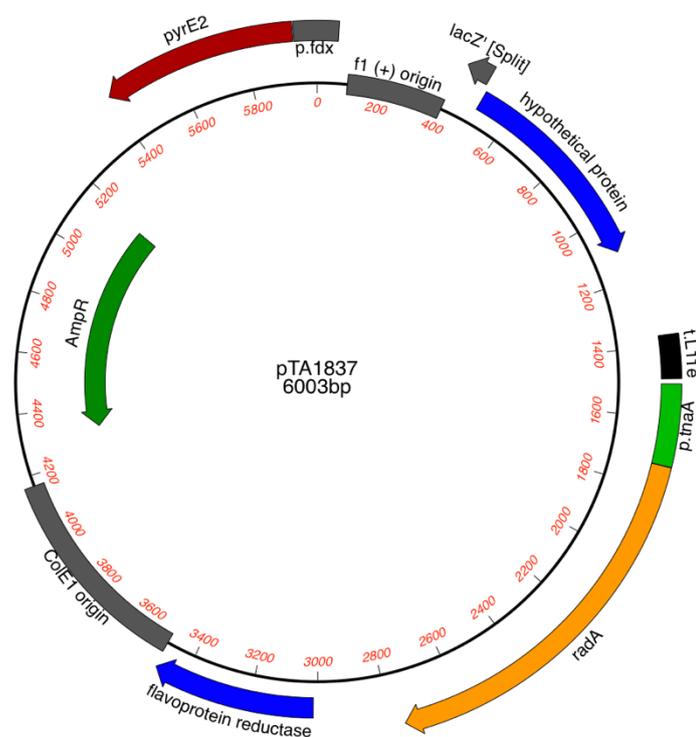
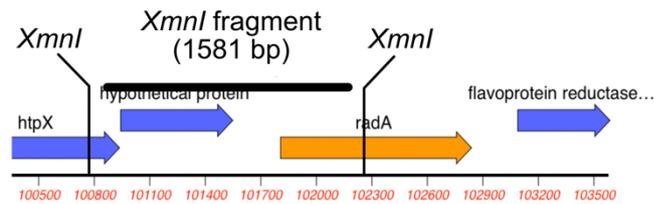


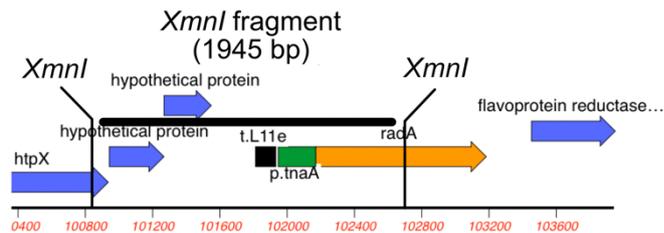
Figure 4.19: pTA1837. Construct to integrate tryptophan-inducible promoter, *p.tnaA* in frame with recombinase *radA*. (Ausiannikava *et al.*, 2018).

Plasmid pTA1837 was used to transform strain H3929 to give pop-in strain H5274. Pop-outs were plated on 5-FOA and screened for differential growth in the presence or absence of tryptophan (where $\Delta radA$ mutants are slow-growing compared to wild type). Southern blot further confirmed integration of *p.tnaA::radA*, giving rise to strain H5320 ($\Delta pyrE2$ bgaHa-Bb leuB-Ag1 $\Delta trpA$ $\Delta recJ1::trpA+$ *p.tnaA::radA*) (**Figure 4.20**).

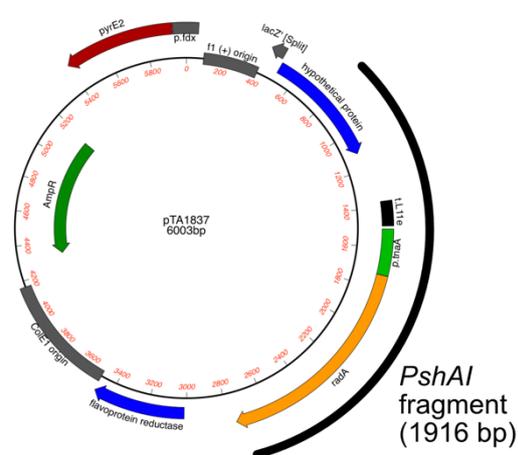
(A) Wild type *radA* region



(B) *p.tnaA::radA* region



(C) pTA1837 probe



(D) Southern blot

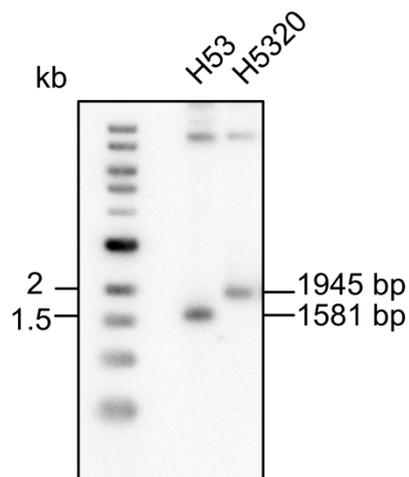
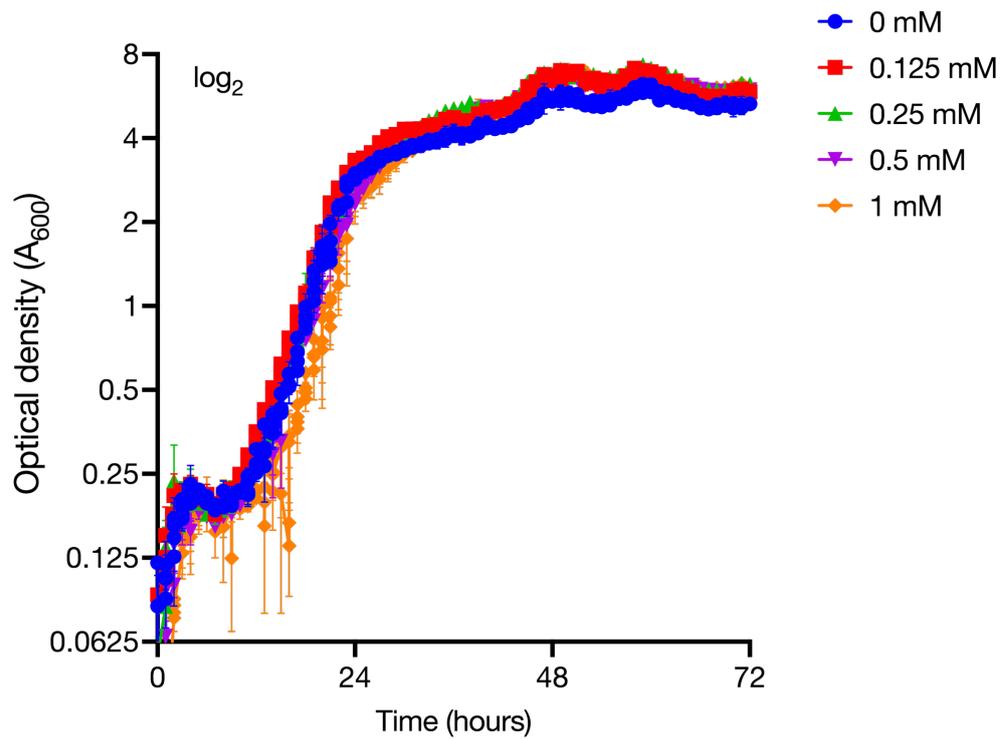


Figure 4.20: Confirmation of integration of *p.tnaA* at the *radA* locus. (A) Expected Southern blot band of 1581 bp for *XmnI* digested wild type genomic DNA. (B) Expected Southern blot band of 1945 bp for *XmnI* digested genomic DNA where *radA* is under control of *p.tnaA*. (C) 1916 bp *p.tnaA::radA* Southern probe consisting of an *PshAI* fragment of pTA1837. (D) Southern blot confirming strain H5320 has correctly integrated *p.tnaA* at the *radA* locus (H53 is a wild type control).

Since the essentiality for *radA* expression has previously been shown using strains H1637 (*oriC*+ *p.tnaA::radA*) and H1642 (Δ *oriC p.tnaA::radA*), these strains were used as controls alongside H5320 to test for the essentiality of *radA* in a strain deleted for *recJ1*. Tryptophan gradient plates were tested with these strains, however the results were repeatedly inconsistent. Instead, growth of these strains was measured in liquid culture. Strains were grown for two consecutive overnights in Hv-YPC broth, ensuring on day three that actively dividing cells were used for the assay. Once the A_{650} reached ~ 0.4 , cells were diluted to 10^{-2} in Hv-Cas+Ura media containing a range of different tryptophan concentrations (0.125 mM, 0.25 mM, 0.5 mM and 1 mM). While Hv-YPC does contain tryptophan, the carryover following 16+ hours of growth and subsequent 100-fold dilution is unlikely to affect the result; where H1642 is known to be inviable at 0 mM tryptophan, this was observed in the liquid media following this methodology. These dilutions were then arrayed in a 96-well plate and A_{600} was semi-continuously measured, allowing the plotting of growth curves for each strain (**Figure 4.21**).

(A) H1637 (*oriC*+ *p.tnaA::radA*)



(B) H1642 (Δ *oriC* *p.tnaA::radA*)

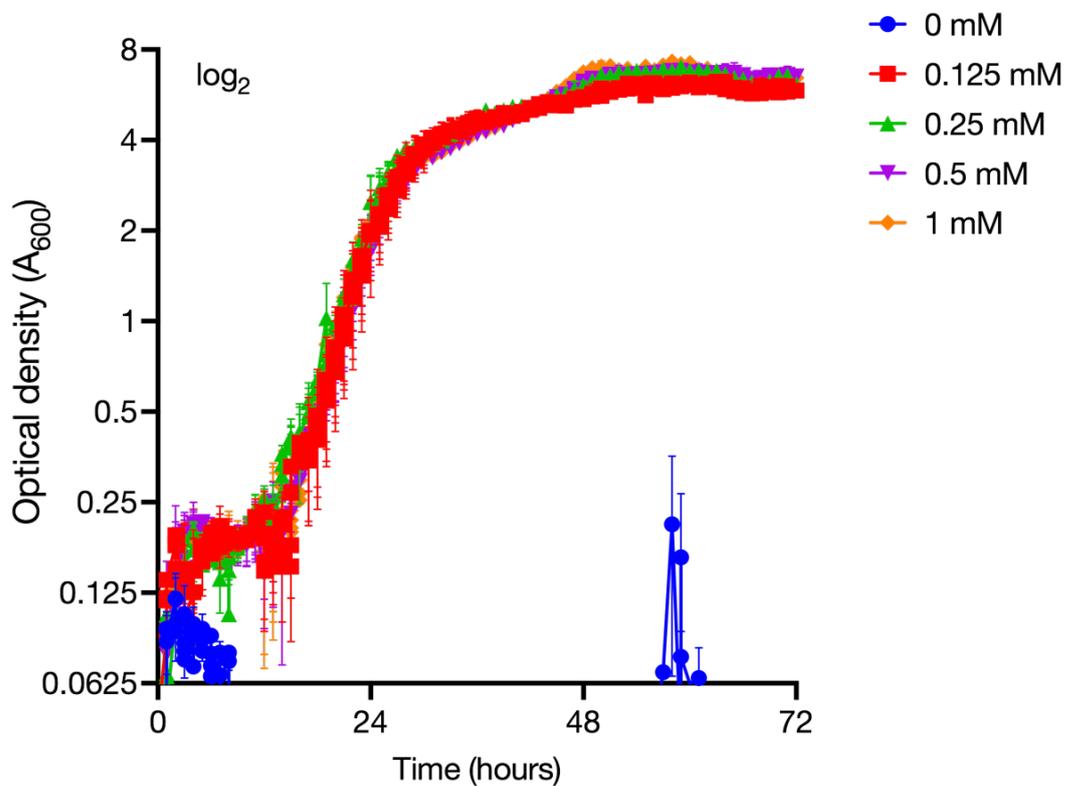


Figure 4.21 continued

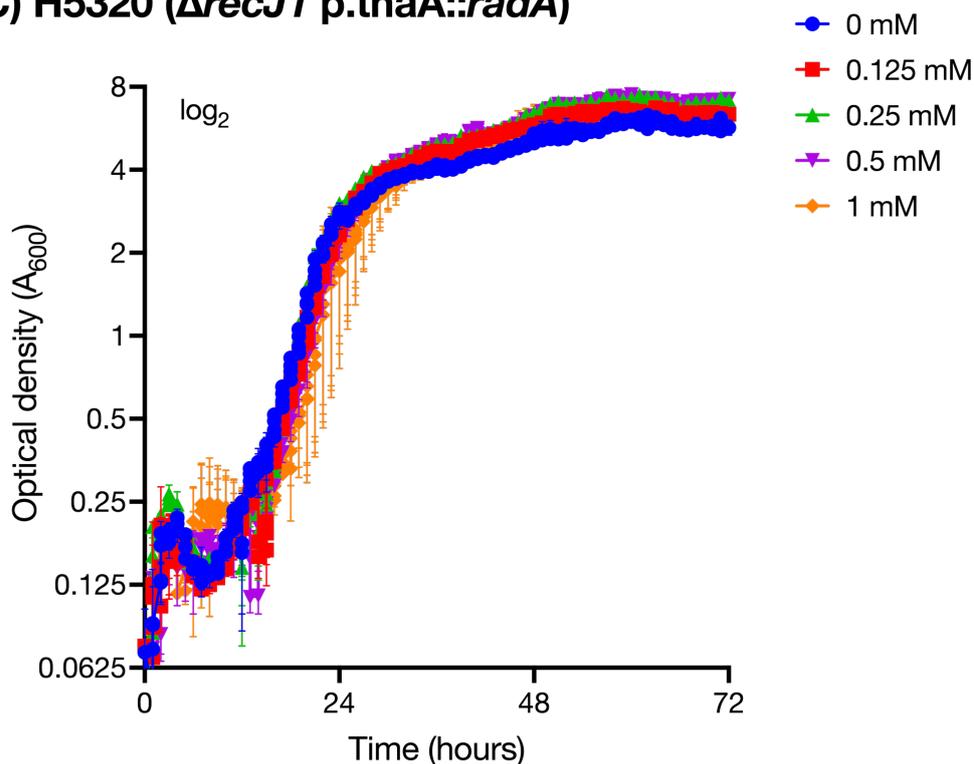
(C) H5320 ($\Delta recJ1$ p.tnaA::radA)

Figure 4.21: Exponential growth rate of *p.tnaA::radA* strains at varying tryptophan concentrations. Strains were grown in Hv-YPC broth for two consecutive overnights before being diluted to 10⁻² in Hv-Cas+Ura with varying concentrations of tryptophan (0/0.125/0.25/0.5/1 mM) and plated. (A) H1637 (*oriC+* *p.tnaA::radA*) is viable in the absence of induction and are therefore not reliant on RadA recombinase. (B) H1642 ($\Delta oriC$ *p.tnaA::radA*) is inviable in the absence of induction, showing originless strains are reliant on RadA recombinase. (C) H5320 ($\Delta recJ1$ *p.tnaA::radA*) are viable in the absence of induction and are therefore not reliant on RadA recombinase. All strains (n=2) were incubated on the same 96-well plate and measured simultaneously for optical density (A_{600}) using an Epoch2 Microplate Spectrophotometer (BioTek). Error bars represent standard error of the mean (SEM).

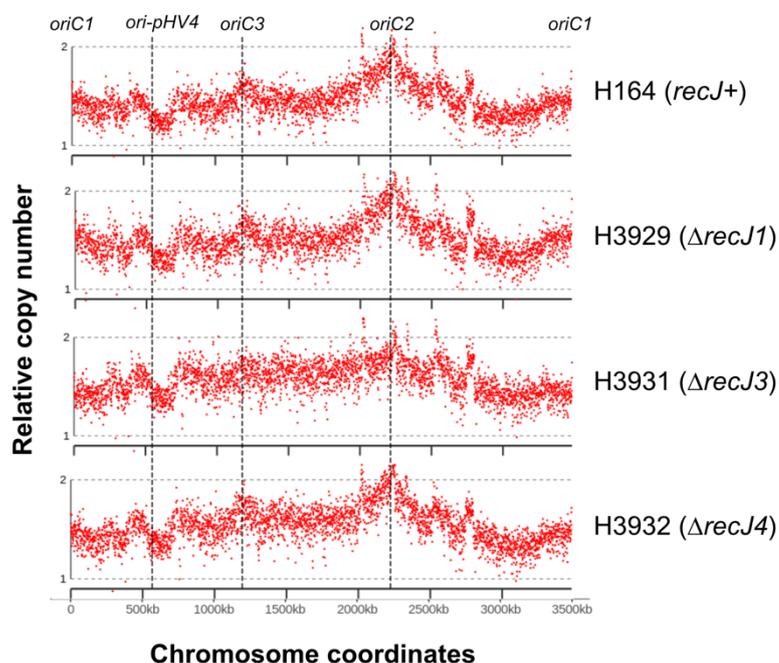
Analysis of liquid cultures showed the same result as seen for the tryptophan gradient plates, whereby the response of H5320 matches that of *oriC+* strain H1637, providing strong evidence that $\Delta recJ1$ mutants are not reliant on expression of *radA* unlike H1642 ($\Delta oriC$).

4.3.2.2 Replication profiling of *recJ* mutants

Since $\Delta recJ1$ mutants are not reliant on RadA, this suggests they utilise their origins of replication. However, the aphidicolin resistance seen for this strain is similar to that of the $\Delta oriC$ strain H1804. If the $\Delta recJ1$ mutant is requiring lower levels of PolB (as predicted by the observed resistance to aphidicolin), it could be predicted such a change in polymerase usage could impact the replication profile of this strain.

Replication profiling was carried out for strains H164, H3929, H3931 and H3932 to assess if origin utilisation differs in the absence of RecJ1, RecJ3 or RecJ4. A previously generated non-replicating control was utilised (H26, DNA purified at stationary phase), alongside a pseudo-control generated during bioinformatics. Both are shown in **Figure 4.22**; it is important to note that the previously generated H26 control utilised a different sequencing methodology and this can impact the output.

(A) Normalised to H26 stationary phase



(B) Pseudo-normalised

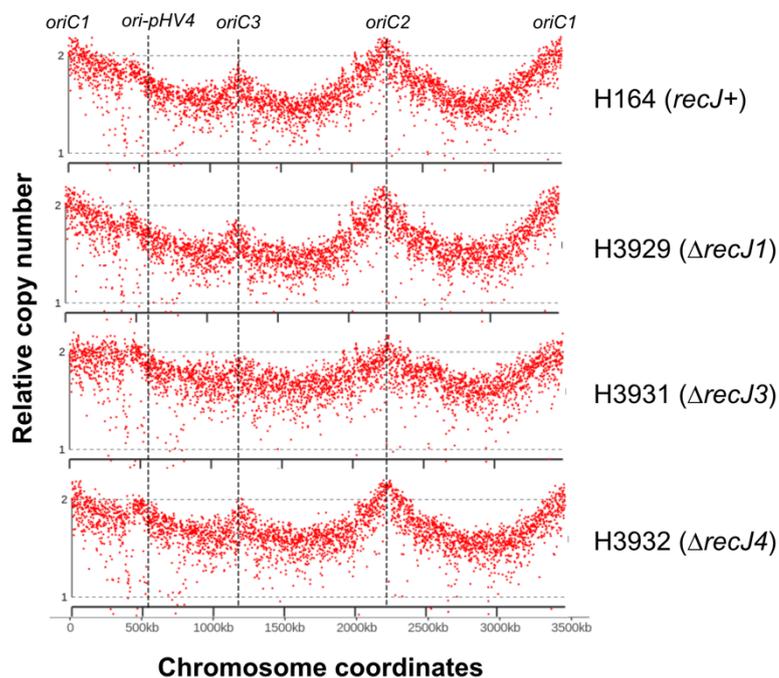


Figure 4.22: Replication profiles for wild type (H164), $\Delta recJ1$ (H3929), $\Delta recJ3$ (H3931) and $\Delta recJ4$ (H3932) strains. Relative copy number plotted against chromosomal coordinate for the main chromosome. DNA copy number from actively replicating samples were normalised against either (A) H26 DNA purified during stationary phase (non-replicating), or (B) a pseudo-value representing equal coverage across the genome. Dashed lines mark the location of deleted origins.

There are marked differences when comparing the profiles generated when normalising against the existing H26 data (**Figure 4.22 A**) and the pseudo-normalised data (**Figure 4.22 B**). Generally, the fit for the wild type model is more appropriate with expectations using the pseudo-controlled data, where clear peaks can be observed for each origin. This difference is likely due to differences in sequencing between the samples generated during this project (Illumina) and the previously generated samples (NextSeq500). While the pseudo-controlled data looks generally comparable to profiles generated previously (Hawkins *et al.*, 2013a), it is important to note that this technique does not account for pseudogenes or differences in copy number during stationary phase.

Generally, in both cases of normalisation, the replication profiles of the $\Delta recJ$ mutants are comparable to the wild type. In both cases, the peaks for the $\Delta recJ3$ strain (H3931) are shallower, which could be suggestive of RecJ3 having a role at the replication fork. The profile for the $\Delta recJ1$ strain (H3929) is indistinguishable from the wild type; this suggests the reason for the observed aphidicolin resistance in H3929 is not due to altered use of origins.

4.3.3 Genetic interactions with *recJ* genes in *H. volcanii*

In order to further elucidate the roles of RecJ proteins in *H. volcanii*, *recJ1* (HVO_0073), *recJ3* (HVO_1018) and *recJ4* (HVO_2889) were deleted in combination with genes predicted to be linked to RecJ/GAN/HAN function in archaea. RecJ2 (HVO_1147) was not utilised for this study due to its apparent essentiality. Plasmids for *trpA*-marked deletions of *recJ1*, *recJ3* and *recJ4* were readily available (Figure 4.23; pTA1958, pTA1960 and pTA1997 respectively; Lever, 2019).

(A) pTA1958: $\Delta recJ1::trpA+$

(B) pTA1960: $\Delta recJ3::trpA+$

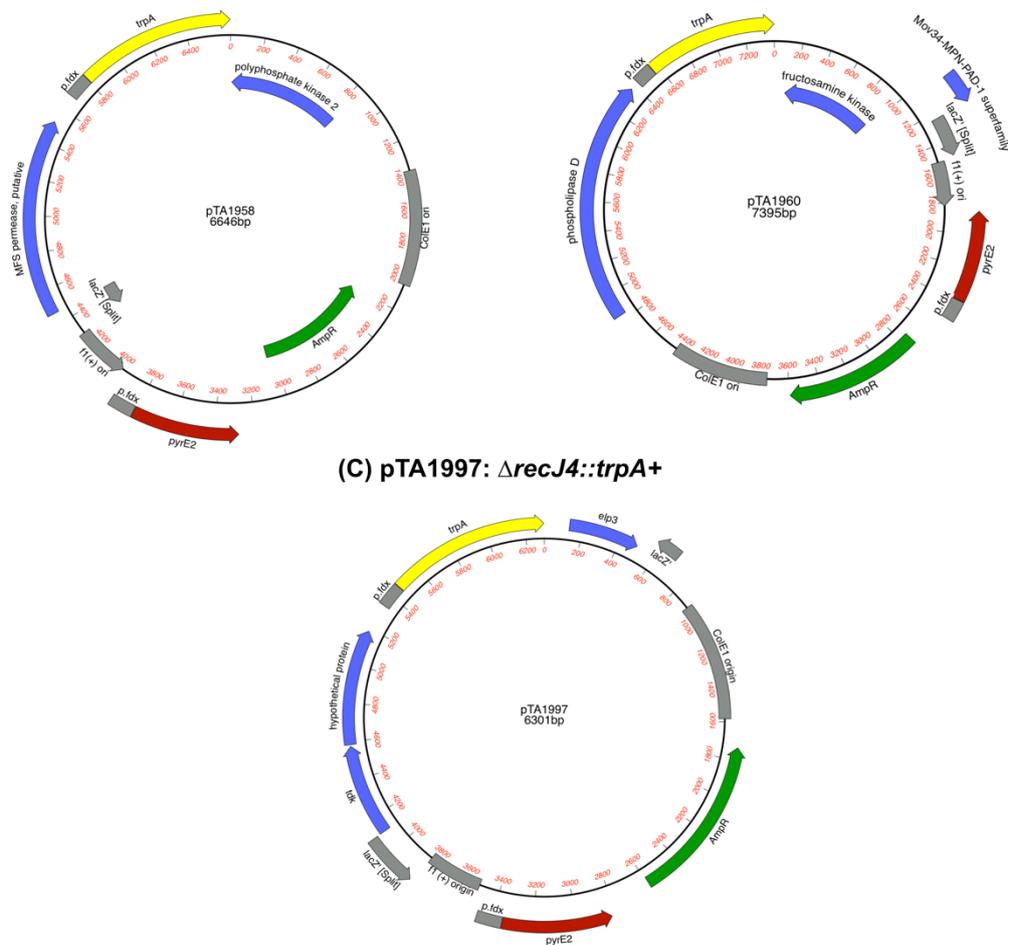


Figure 4.23: $\Delta recJ$ deletion constructs (A) pTA1958, (B) pTA1960 and (C) pTA1997. For deletion of *recJ1* (HVO_0073), *recJ3* (HVO_1018) and *recJ4* (HVO_2889) respectively. Constructed by Rebecca Lever (Lever, 2019).

Following identification of deletion candidates by colony hybridisation, *recJ* deletions were verified by restriction digest and Southern blot (see **Figure 4.24** for details; restriction digest shown here applies to all blots in this section).

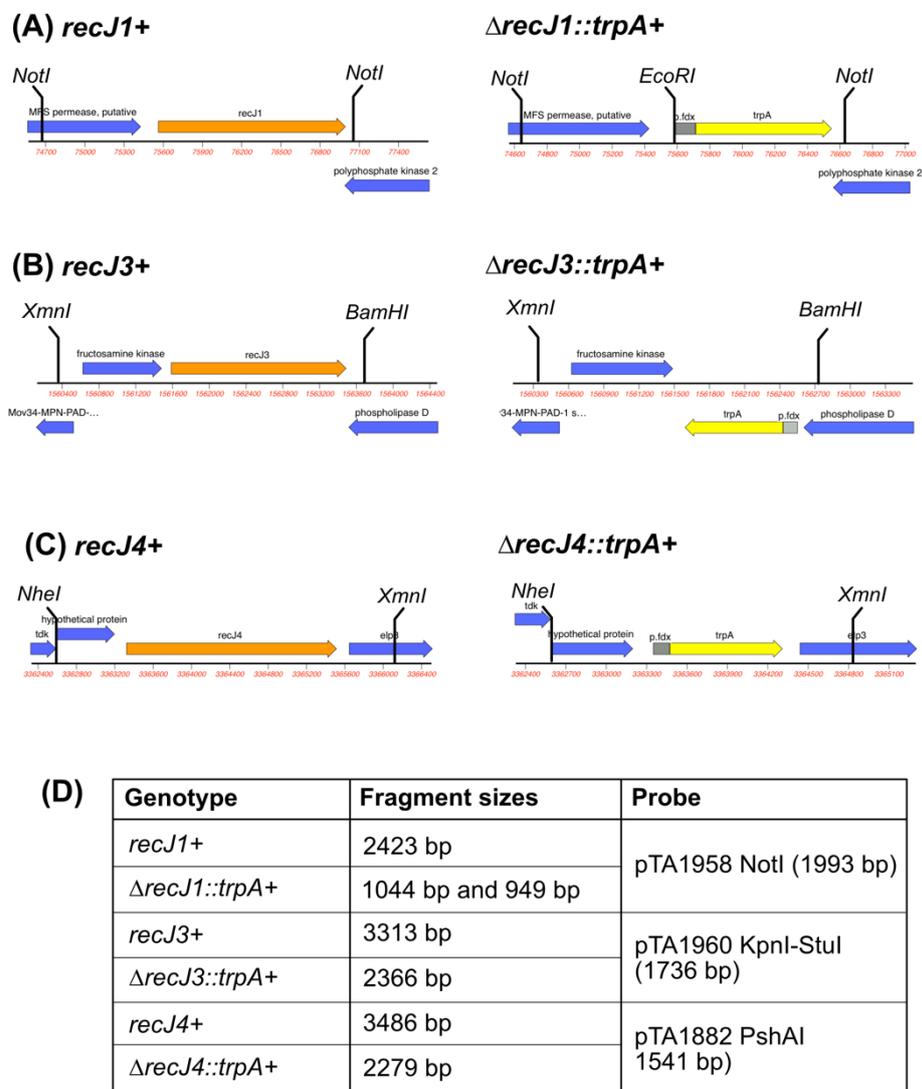


Figure 4.24: Overview of restriction digest and Southern blotting to confirm deletions of *recJ1*, *recJ3* or *recJ4* in *H. volcanii*. (A) gDNA of *recJ1* candidates was digested with *NotI* and *EcoRI*. (B) gDNA of *recJ3* candidates was digested with *XmnI* and *BamHI*. (C) gDNA of *recJ4* candidates was digested with *NheI* and *XmnI*. (D) Table showing predicted restriction fragment sizes for digested DNA when using probe DNA listed (Δ constructs for *recJ1* and *recJ3*, genomic clone for *recJ4*).

4.3.6.1 Deletion in combination with Hef and Hjc

Hjc and Hef are endonucleases proposed to act in the restart of stalled replication forks in *H. volcanii* (Lestini *et al.*, 2010) (**Figure 4.25**). Hjc is a Holliday junction (HJ) resolvase that binds HJs and cleaves the opposing strands symmetrically, generating two recombinant duplexes. Hef comprises two distinct domains: an N-terminal helicase domain (DEAH family) and a C-terminal XPF family endonuclease domain and is the archaeal homologue of eukaryotic FANCM. Hef acts on flapped, nicked or forked DNA and can convert a HJ into a forked structure by incision near the branch point (Komori *et al.*, 2002, Komori *et al.*, 2004). In *H. volcanii*, deletion of *hef* results in sensitivity to the crosslinking agent MMC, suggesting an involvement in interstrand crosslink repair, while deletion of *hjc* has little effect on survival fractions (Lestini *et al.*, 2010). However, $\Delta hjc \Delta hef$ strains are inviable and the double mutant is synthetically lethal. Deletion of $\Delta radA$ in Δhef is highly deleterious, suggesting a HR-independent function of Hef, while $\Delta radA \Delta hjc$ mutants are comparable to $\Delta radA$ strains, suggesting its involvement in HR exclusively (Lestini *et al.*, 2010).

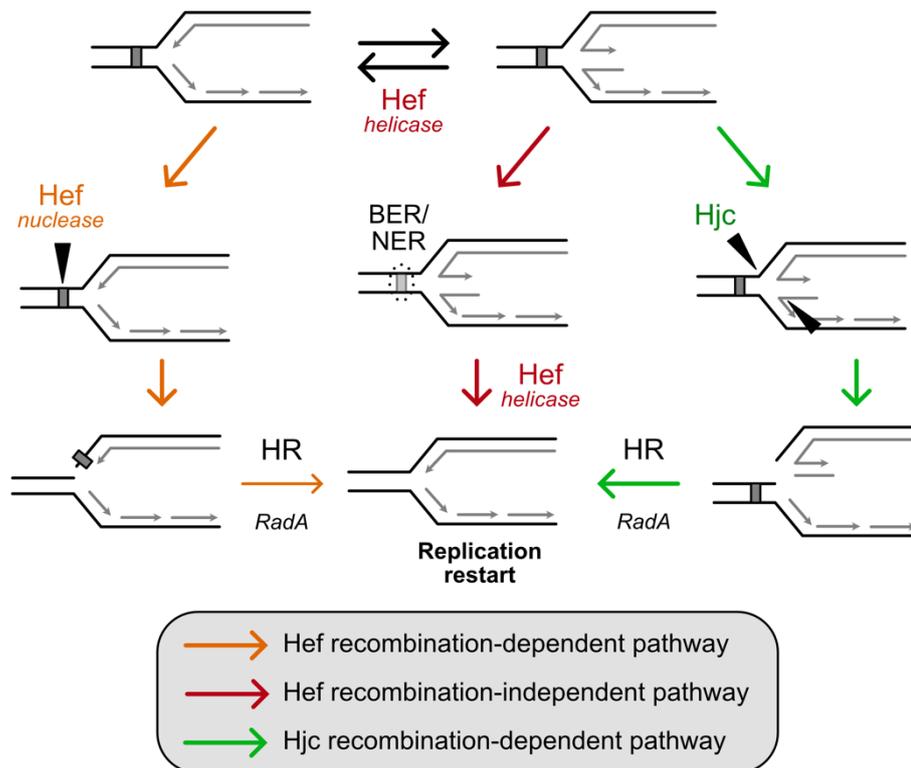


Figure 4.25: Involvement of Hef and Hjc in repair of stalled replication forks. The stalled fork may be remodelled by Hef helicase and subsequently cleaved by Hef nuclease (orange arrows). The broken arm remaining as a cleavage product can be utilised for replication restart by homologous recombination (HR; dependent on RadA). Hef can also act in a RadA-independent pathway (red arrows), where reversal of the fork exposes the DNA lesion for repair by base excision repair (BER) or nucleotide excision repair (NER). Once the lesion is removed, Hef helicase can reset the replication fork. If the stalled replication fork is reversed to form a Holliday junction, Hjc can cleave this symmetrically to generate a broken chromosomal arm (green arrows). This double-strand break can then be repaired using HR. Adapted from Lestini *et al* (2010).

In *H. volcanii*, the identity of HAN (Hef-associated nuclease) remains unknown. Deletion of HAN would be predicted to render the Hef pathway(s) inactive, therefore $\Delta hjc \Delta han$ strains should be inviable (as for $\Delta hjc \Delta hef$). In contrast, $\Delta hef \Delta han$ strains would retain a functional RadA/Hjc-dependent pathway and thus would be predicted to be viable.

The Δhef deletion strain H364 was transformed with plasmids pTA1958 ($\Delta recJ1::trpA^+$), pTA1960 ($\Delta recJ3::trpA^+$) and pTA1997 ($\Delta recJ4::trpA^+$) to give rise to pop-in strains H5244, H5245 and H5246 respectively. Pop-outs gave rise to 5-FOA^R candidates that were patched on

YPC and primarily screened via colony hybridisation. Strains H5286 ($\Delta hef \Delta recJ1::trpA+$), H5297 ($\Delta hef \Delta recJ3::trpA+$) and H5306 ($\Delta hef \Delta recJ4::trpA+$) were further verified by restriction digest and Southern blot (**Figure 4.26**).

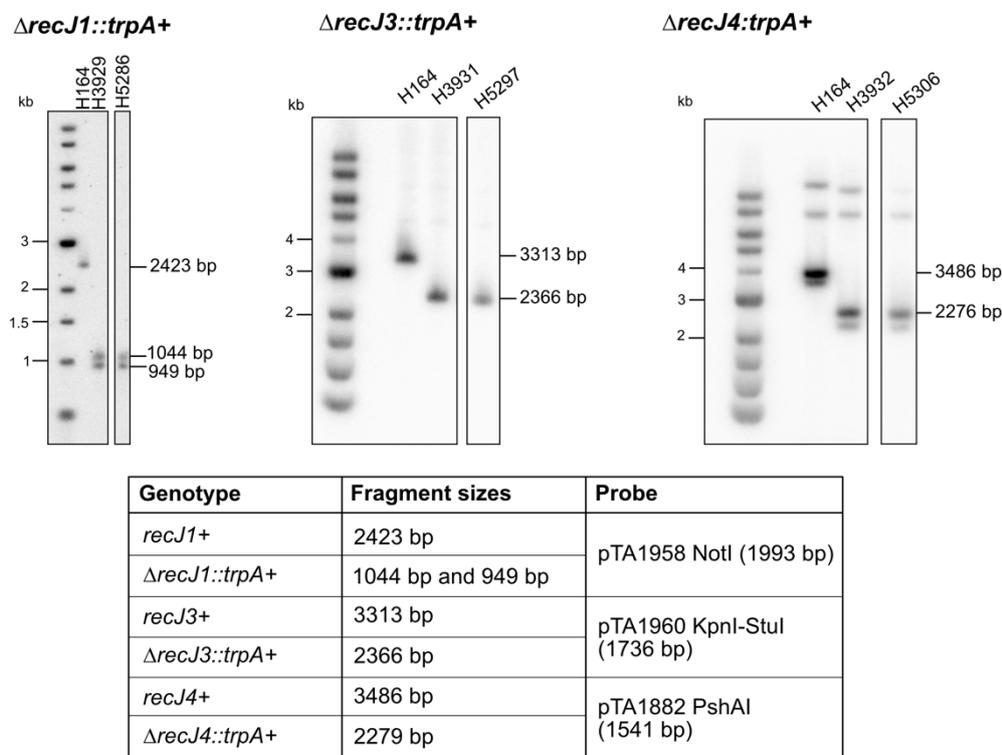
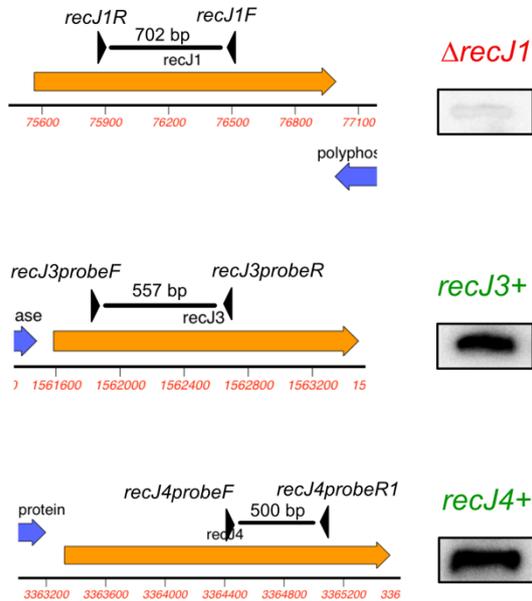


Figure 4.26: Restriction digest and Southern blotting to confirm deletion of RecJ1, RecJ3 and RecJ4 in Δhef strain H364. Strain H5286 is confirmed as $\Delta recJ1::trpA+$, H5297 is confirmed as $\Delta recJ3::trpA+$ and H5306 is confirmed as $\Delta recJ4::trpA+$. H164 is wild type, while H3929, H3931 and H3932 were used as positive controls for $\Delta recJ1::trpA+$, $\Delta recJ3::trpA+$ and $\Delta recJ4::trpA+$ respectively. Bands seen >6 kb on the *recJ4::trpA+* blot likely represent non-specific binding of the probe.

The Δhjc deletion strain H282 was transformed with plasmids pTA1958 ($\Delta recJ1::trpA+$), pTA1960 ($\Delta recJ3::trpA+$) and pTA1997 ($\Delta recJ4::trpA+$) to give rise to pop-in strains H5268, H5269 and H5270 respectively. Pop-outs gave rise to 5-FOA^R candidates that were patched on YPC and primarily screened via colony hybridisation (**Figure 4.27 A**). Strain H5291 ($\Delta hjc \Delta recJ1::trpA+$) was further verified by restriction digest and Southern blot (**Figure 4.27 B**). No candidates were isolated for deletion of *recJ3* or *recJ4* in the Δhjc background. Pop-outs of H5269 and H5270 were repeated and a total of 160 clones per strain were screened by colony

hybridisation to ensure the strain could not be generated. All clones hybridised with the probe, suggesting both RecJ3 and RecJ4 are essential in a Δhjc strain.

(A) Colony hybridisation



(B) Southern blot

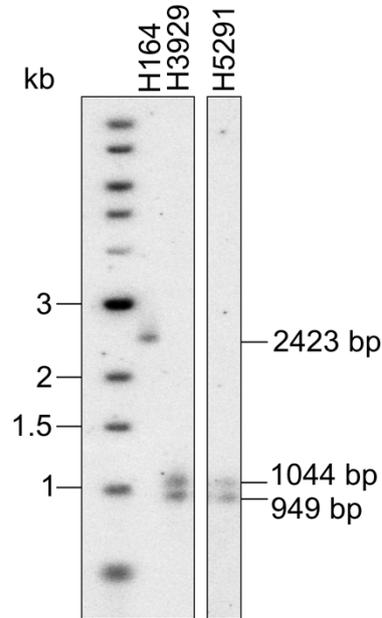


Figure 4.27: Generation of $\Delta recJ$ mutants in Δhjc strain H282. (A) Hybridisation with probes consisting of PCR products of *recJ1* sequence (702 bp), *recJ3* sequence (557 bp) and *recJ4* sequence (500 bp) gave rise to candidates for $\Delta hjc \Delta recJ1::trpA^+$, but all candidates probed for *recJ3* and *recJ4* were wild type. (B) Southern blotting of $\Delta hjc \Delta recJ1::trpA^+$ candidate H5291 showed it was successfully deleted.

Growth rate

In order to compare growth rates of the double mutants generated, strains were assayed for growth rate alongside the single mutants for each strain; H364 (Δhef), H282 (Δhjc), H3929 ($\Delta recJ1$), H3931 ($\Delta recJ3$) and H3932 ($\Delta recJ4$). Strains were grown for two consecutive overnights in Hv-YPC broth, ensuring on day three actively dividing cells were used for the assay. Cells were diluted in fresh Hv-YPC broth and arrayed in a 96-well plate. OD (A_{600}) was semi-continuously measured, allowing the plotting of growth curves for each strain (**Figure 4.28**).

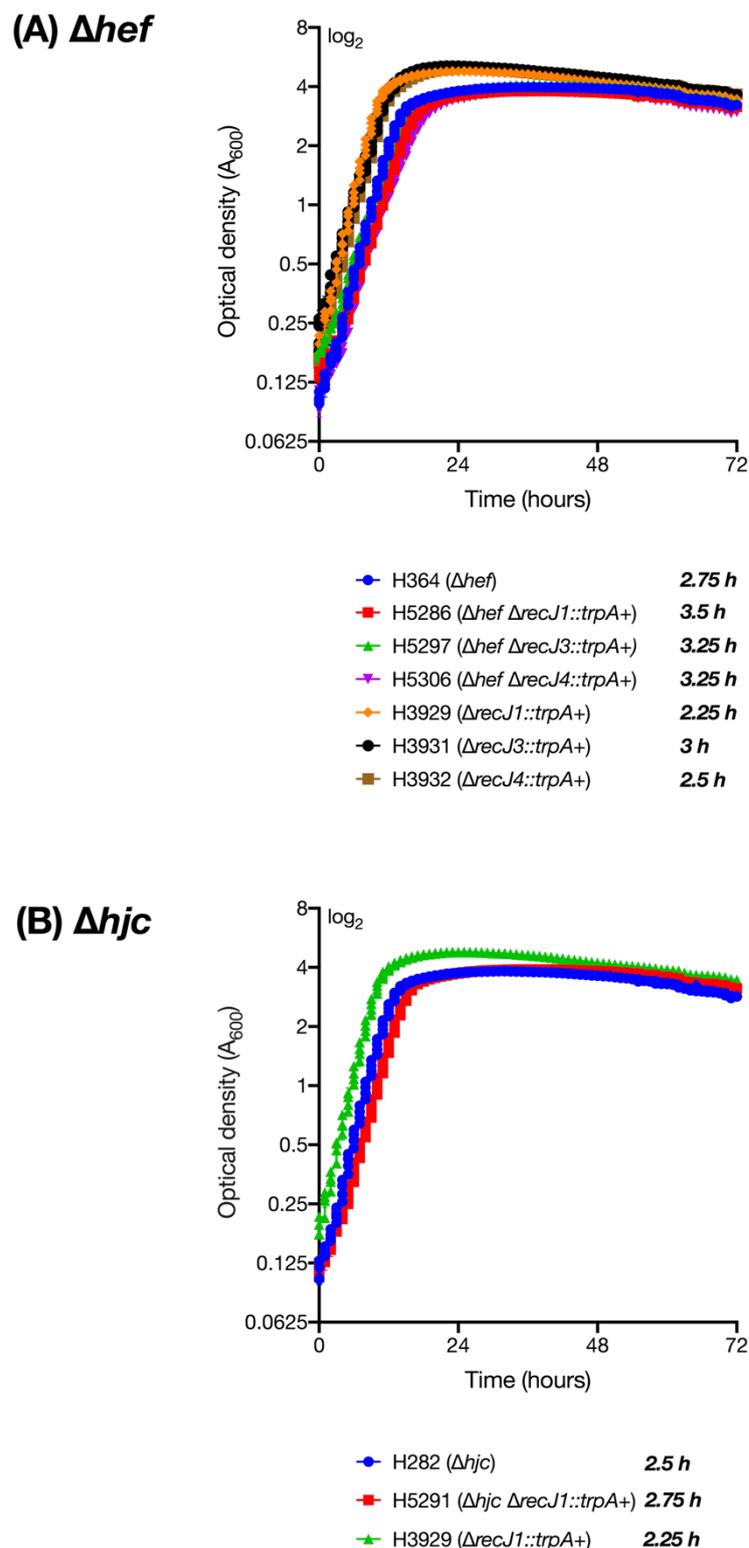


Figure 4.28: Exponential growth rate of strains deleted for (A) *hef* or (B) *hjc* in combination with *recJ1*, *recJ3* or *recJ4*. Generation time in hours (h) is indicated in bold beside each strain legend. Strains were grown in Hv-YPC broth for two consecutive overnights before being diluted in fresh Hv-YPC broth. All strains (n=2) were incubated on the same 96-well plate and measured simultaneously for optical density (A_{600}) using an Epoch2 Microplate Spectrophotometer (BioTek).

Figure 4.28 A shows all strains carrying Δhef have a longer generation time compared to $\Delta recJ$ mutants in a H164 background (H3929, H3931 and H3932). The additional deletion of *recJ1*, *recJ3* or *recJ4* in Δhef strain H364 does lead to a further (minor) increase in generation time, but the strains do not appear synthetically lethal.

Figure 4.28 B shows all strains carrying Δhjc have a longer generation time compared to the $\Delta recJ1$ mutant in a H164 background (H3929). The additional deletion of *recJ1* in Δhjc strain H282 does lead to a further (minor) increase in generation time, but the strains do not appear synthetically lethal.

DNA content and cell size

In order to determine the DNA content and cell size of cells deleted for *recJ1*, *recJ3* or *recJ4* in combination with Δhef , strains were analysed using flow cytometry (**Figure 4.29**).

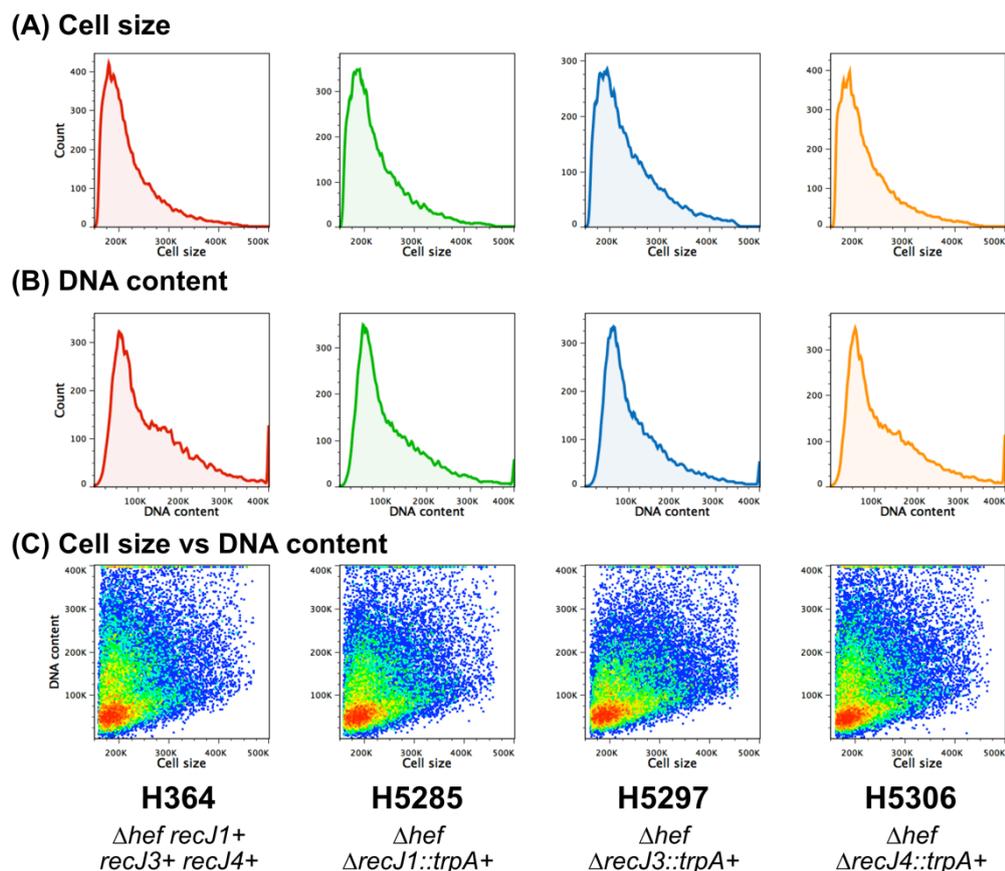


Figure 4.29: Flow cytometry analysis of strains deleted for *hef* and *recJ1*, *recJ3* or *recJ4*. (A) Determination of cell size. All cells show similar cell size profiles. (B) Determination of DNA content. H364 (Δhef) and H5306 ($\Delta hef \Delta recJ4::trpA+$) have a higher frequency of cells with high DNA content compared to H5285 ($\Delta hef \Delta recJ1::trpA+$) and H5297 ($\Delta hef \Delta recJ3::trpA+$). (C) Density dot plots displaying cell size vs DNA content. Both H364 and H5306 show a larger distribution of DNA content within cells compared to H5285 and H5297.

While the cell size profile for all Δhef +/- $\Delta recJ$ mutants were comparable, H364 (Δhef) and H5306 ($\Delta hef \Delta recJ4::trpA+$) showed a broader range of DNA content within cells, where there was a larger proportion carrying increased amounts of DNA.

In order to determine the DNA content and cell size of cells deleted for *recJ1*, *recJ3* or *recJ4* in combination with Δhjc , strains were analysed using flow cytometry (Figure 4.30).

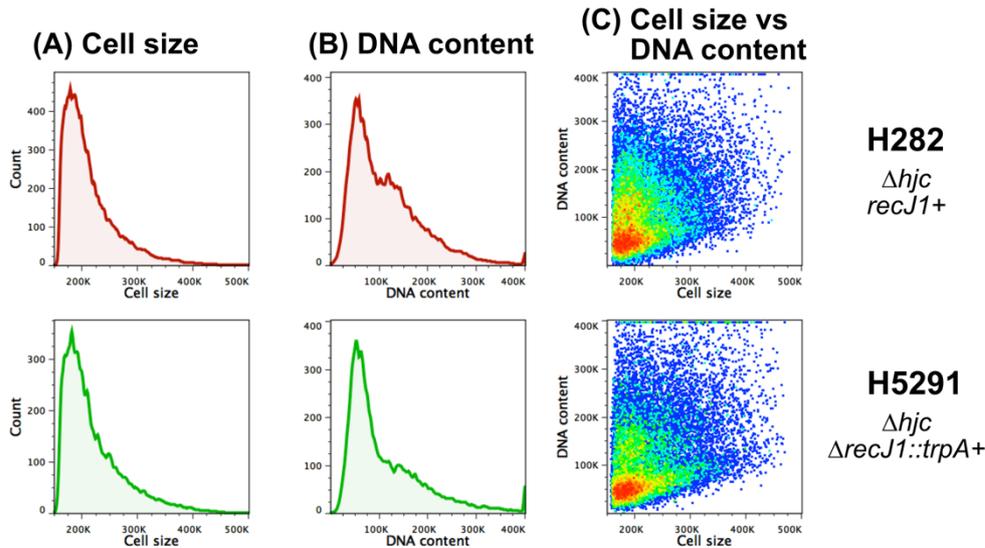


Figure 4.30: Flow cytometry analysis of strains deleted for *hjc* alone or *hjc* in combination with *recJ1*. (A) Determination of cell size. Both strains show similar cell size profiles. (B) Determination of DNA content. H282 (Δhjc) shows a larger proportion of cells with high DNA content than H5291 ($\Delta hjc \Delta recJ1::trpA+$). (C) Density dot plots displaying cell size vs DNA content. H282 shows a larger distribution of DNA content within cells compared to H5291.

The cell size profiles for H282 (Δhjc) and H5291 ($\Delta hjc \Delta recJ1::trpA+$) are comparable. However, H282 shows a broader range of DNA content within cells, where there was a larger proportion carrying increased amounts of DNA compared to H5291.

Survival following treatment with mitomycin C

Strains deleted for *hjc* or *hef* are not sensitive to UV irradiation but show a sensitivity to the crosslinking agent MMC (Lestini *et al.*, 2010). In order to test whether there was a synthetic defect between *recJ1/recJ3/recJ4* and *hef* or *hjc*, Δhef strains H364 (Δhef), H5286 ($\Delta recJ1 \Delta hef$), H5297 ($\Delta recJ3 \Delta hef$) and H5306 ($\Delta recJ4 \Delta hef$) and Δhjc strains H282 (Δhjc) and H5291 ($\Delta recJ1 \Delta hjc$) were tested for response to MMC treatment compared to control strain H164 (*hef+ hjc+ recJ+*). Cultures were grown to mid-exponential phase, and replica-spotted onto Hv-YPC agar containing either water (control) or MMC. Plates were incubated for 4-7 days at 45°C, colonies were counted, and survival fractions were calculated (**Figure 4.31**).

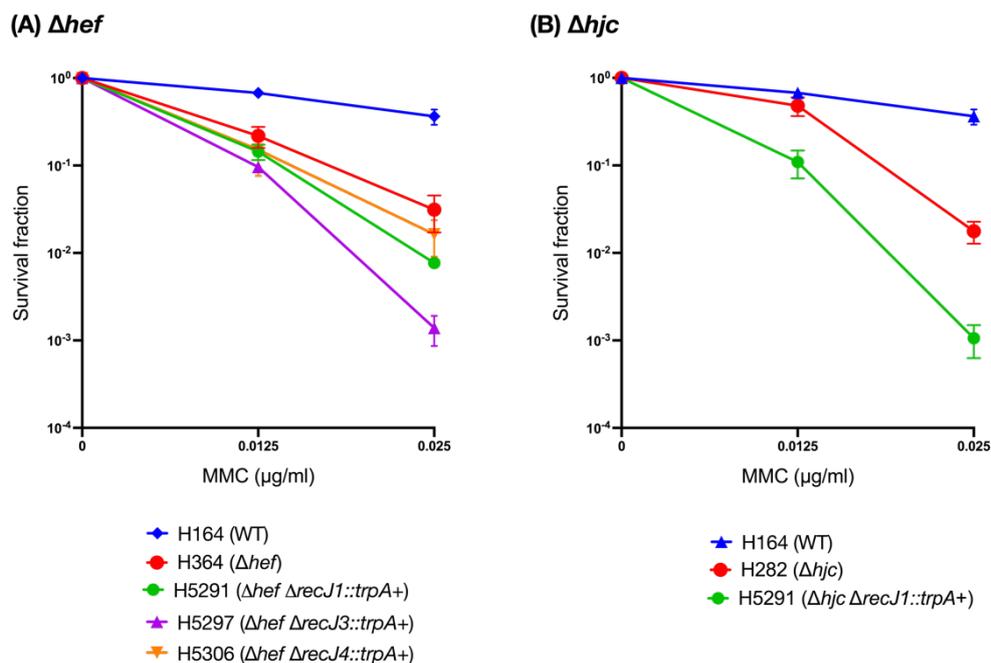


Figure 4.31: Survival frequency of (A) Δhef and (B) Δhjc strains following treatment with mitomycin C (MMC). Control strain H164, Δhef strains H364 (Δhef), H5286 ($\Delta recJ1 \Delta hef$), H5297 ($\Delta recJ3 \Delta hef$) and H5306 ($\Delta recJ4 \Delta hef$) and Δhjc strains H282 (Δhjc) and H5291 ($\Delta recJ1 \Delta hjc$) were chronically exposed to MMC within YPC plates. Both Δhef (H364) and Δhjc (H282) strains show an increased sensitivity to MMC compared to control strain H164. Strain H5297 ($\Delta hef \Delta recJ3::trpA+$) shows an increased sensitivity to MMC compared to parent strain H364. Strain H5291 ($\Delta hjc \Delta recJ1::trpA+$) shows an increased sensitivity to MMC compared to parent strain H282. Survival fraction is calculated relative to untreated control. Each data point is generated as an average of at least 3 independent trials. Error bars represent standard error of the mean (SEM).

As previously observed, strains H364 (Δhef) and H282 (Δhjc) show an increased sensitivity to MMC. Following treatment with MMC, both H5291 ($\Delta hef \Delta recJ1::trpA+$) and H5306 ($\Delta hef \Delta recJ4::trpA+$) show survival similar to the parent strain H364 (Δhef), while strain H5297 ($\Delta hef \Delta recJ3::trpA+$) shows an increased sensitivity at the highest dose (0.025 $\mu\text{g/ml}$). Analysis of this data by two-tailed *t*-test showed no significant difference ($p > 0.05$). Strain H5291 ($\Delta hjc \Delta recJ1::trpA+$) shows an increased sensitivity to MMC compared to parent strain H282 (Δhjc). Analysis of this data by two-tailed *t*-test showed at 0.025 $\mu\text{g/ml}$, there was a significant difference between response to MMC of H282 and H5291 ($p < 0.05$).

4.3.6.2 Deletion in combination with origins of replication

H. volcanii strains carrying deletions of all chromosomal origins of DNA replication have been shown to replicate in a manner different to origin-dependent strains, namely recombination-dependent replication (RDR). This alternate method of replication has been shown to result in an altered response to Family B replicative polymerase inhibitor, aphidicolin (see *Chapter 3* for details). The CMG complex is known to be critical for genome replication in the presence of origins; the differential response to aphidicolin treatment previously shown for H3929 ($\Delta recJ1::trpA+$) suggests its involvement in replication (likely as GAN within the CMG complex).

It remains unknown if the CMG complex is involved in RDR in originless strains. If the CMG complex plays an important role in originless replication, it could be predicted that the aphidicolin response of the $\Delta oriC \Delta gan$ strain in *H. volcanii* would differ from that of $\Delta oriC$ alone. However, if the CMG complex is not utilised for originless replication, the additional deletion of GAN should not affect the aphidicolin response of these strains.

Originless strain H1804 ($\Delta oriC1, C2, C3, pHV4$) was transformed with plasmids pTA1958 ($\Delta recJ1::trpA+$), pTA1960 ($\Delta recJ3::trpA+$) and pTA1997 ($\Delta recJ4::trpA+$) to give rise to pop-in strains H5175, H5174 and H5176 respectively. Pop-outs gave rise to 5-FOA^R candidates that were patched on YPC agar and primarily screened via colony hybridisation. Strains H5282 ($\Delta oriC1, C2, C3, pHV4 \Delta recJ1::trpA+$), H5294 ($\Delta oriC1, C2, C3, pHV4 \Delta recJ3::trpA+$) and H5303 ($\Delta oriC1, C2, C3, pHV4 \Delta recJ4::trpA+$) were further verified by restriction digest and Southern blot (**Figure 4.32**).

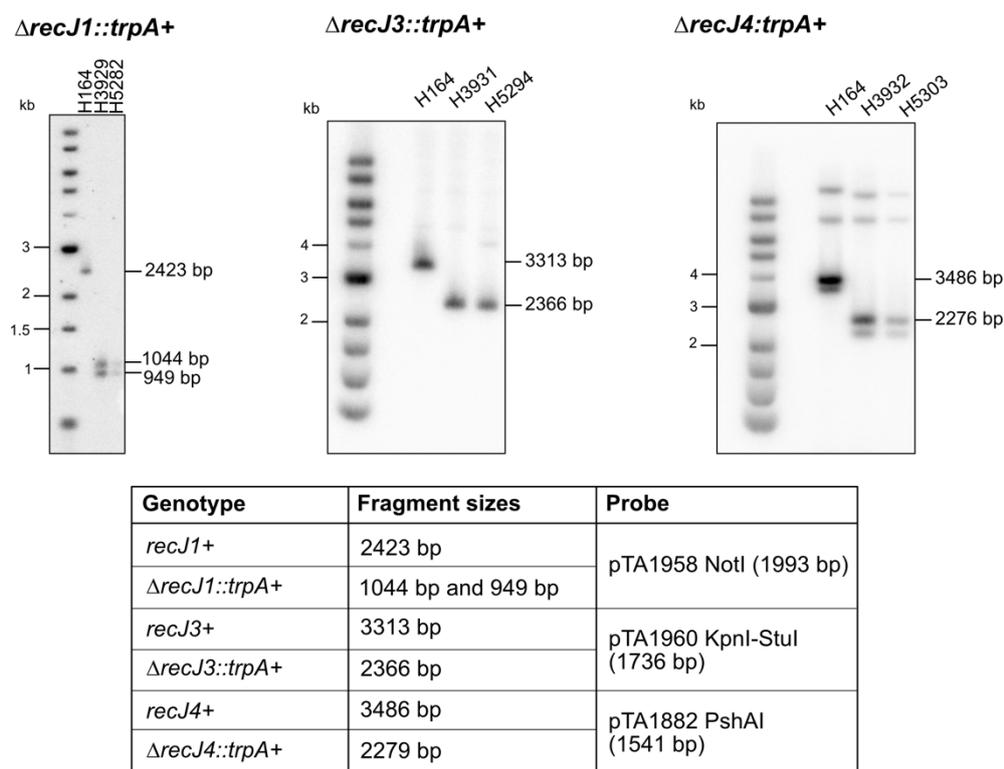


Figure 4.32: Restriction digest and Southern blotting to confirm deletion of *recJ1*, *recJ3* and *recJ4* in Δ *oriC* strain H1804. Strain H5282 is confirmed as Δ *recJ1::trpA*⁺, H5294 is confirmed as Δ *recJ3::trpA*⁺ and H5303 is confirmed as Δ *recJ4::trpA*⁺. H164 is wild type, while H3929, H3931 and H3932 were used as positive controls for Δ *recJ1::trpA*⁺, Δ *recJ3::trpA*⁺ and Δ *recJ4::trpA*⁺ respectively. Bands seen >6 kb on the *recJ4::trpA*⁺ blot likely represent non-specific binding of the probe.

Growth rate

In order to compare growth rate differences, strains were assayed for growth rate alongside the single mutants for each strain; H1804 (Δ *oriC1, C2, C3, pHV4*), H3929 (Δ *recJ1*), H3931 (Δ *recJ3*) and H3932 (Δ *recJ4*). Strains were grown for two consecutive overnights in Hv-YPC broth, ensuring on day three actively dividing cells were used for the assay. Cells were diluted in fresh Hv-YPC broth and arrayed in a 96-well plate. OD (A_{600}) was semi-continuously measured, allowing the plotting of growth curves for each strain (Figure 4.33).

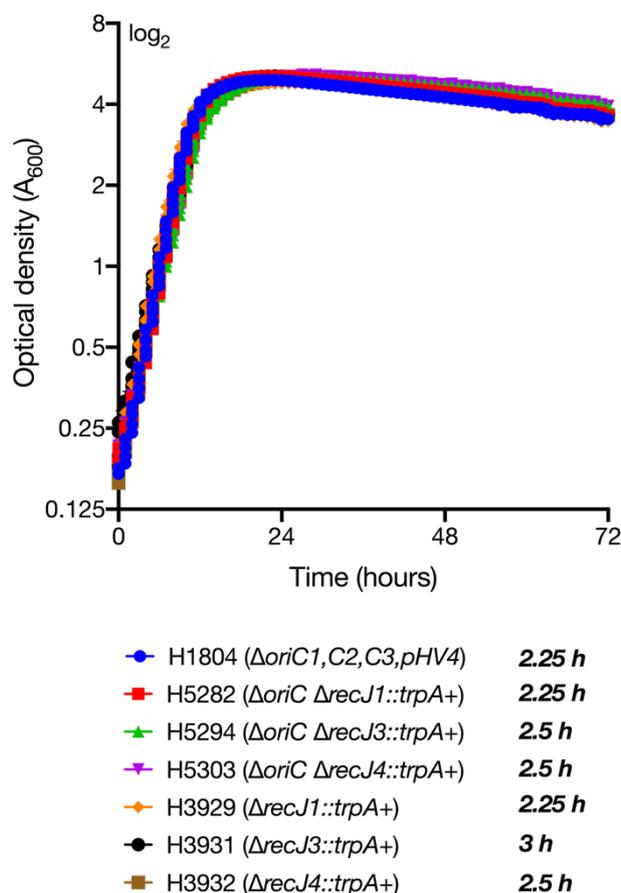


Figure 4.33: Exponential growth rate of strains deleted for *oriC1*, *oriC2*, *oriC3* and *ori-pHV4* in combination with *recJ1*, *recJ3* or *recJ4*. Generation time in hours (h) is indicated in bold beside each strain legend. Strains were grown in Hv-YPC for two consecutive overnights before being diluted and plated. All strains (n=2) were incubated on the same 96-well plate and measured simultaneously for optical density (A_{600}) using an Epoch2 Microplate Spectrophotometer (BioTek).

The growth rates of all $\Delta oriC \Delta recJ$ strains are comparable to the originless parent H1804. Their growth rates are also comparable to the origin-dependent $\Delta recJ$ mutants (H164 background), with the exception of slow-growing mutant H3931 ($\Delta recJ3$).

DNA content and cell size

In order to determine the DNA content and cell size of cells deleted for *recJ1*, *recJ3* or *recJ4* in combination with the chromosomal origins of replication ($\Delta oriC1, 2, 3, pHV4$), strains were analysed using flow cytometry (Figure 4.34).

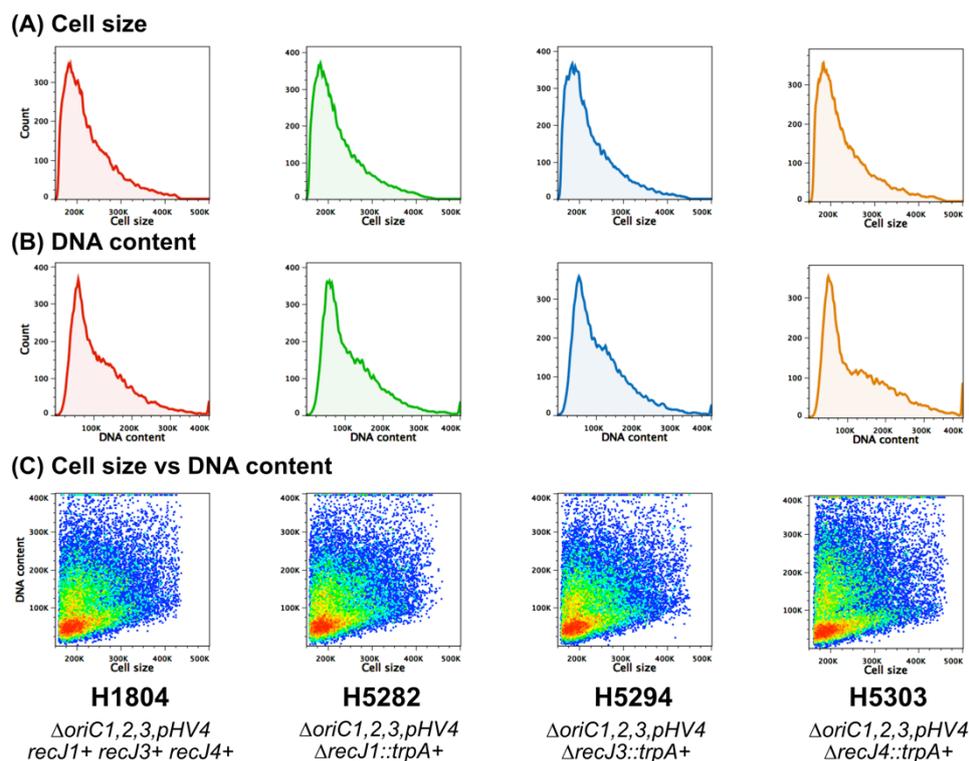


Figure 4.34: Flow cytometry analysis of strains deleted for *recJ1*, *recJ3* or *recJ4* in combination with deletion of chromosomal origins *oriC1*, *oriC2*, *oriC3* and *ori-pHV4*. (A) Determination of cell size. All strains show similar cell size profiles. (B) Determination of DNA content. Generally, all strains have comparable DNA content profiles. (C) Density dot plots displaying cell size vs DNA content. H5303 ($\Delta oriC1,2,3,pHV4 \Delta recJ4::trpA^+$) shows a wider distribution of DNA content within cells than all other strains.

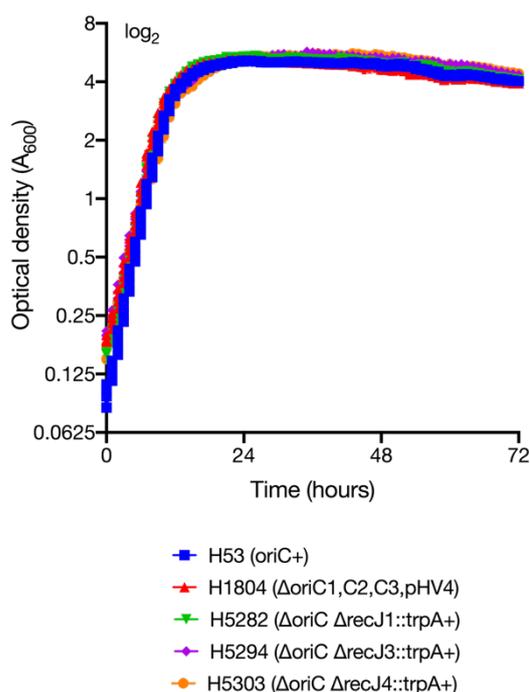
While all strains have comparable cell size and DNA content profiles, the peacock plot for H5303 ($\Delta oriC1,2,3,pHV4 \Delta recJ4::trpA^+$) suggests a wider distribution of DNA content in this strain compared to the others screened.

Aphidicolin response

Previously, origin-dependent $\Delta recJ1$ mutant H3929 was shown to exhibit an increased resistance to aphidicolin treatment (inhibitor of Family B replicative DNA polymerases). Strain H1804, lacking chromosomal origins of replication ($\Delta oriC$), has also been shown to have an increased resistance to aphidicolin. It can therefore be expected that all mutants generated in a H1804 background should show an increased resistance in comparison to their *oriC*⁺ counterparts (e.g., H53; $\Delta pyrE2 \Delta trpA$), unless the subsequent mutation affects the cellular response to aphidicolin.

In order to compare aphidicolin responses, strains H53 ($\Delta oriC^+$), H1804 ($\Delta oriC$), H5282 ($\Delta oriC \Delta recJ1::trpA^+$), H5294 ($\Delta oriC \Delta recJ3::trpA^+$) and H5303 ($\Delta oriC \Delta recJ4::trpA^+$) were grown in the presence of aphidicolin for 72 hours. H53 (*oriC*⁺) and H1804 ($\Delta oriC$) were used as controls to show the differing response shown to aphidicolin in the presence/absence of origins. Strains were grown for two consecutive overnights in Hv-YPC broth, ensuring on day three actively dividing cells were used for the assay. Cells were diluted in fresh Hv-YPC broth and arrayed in a 96-well plate. OD (A_{600}) was semi-continuously measured, allowing the plotting of growth curves for each strain (**Figure 4.35**).

(A) 0 µg/ml aphidicolin



(B) 15 µg/ml aphidicolin

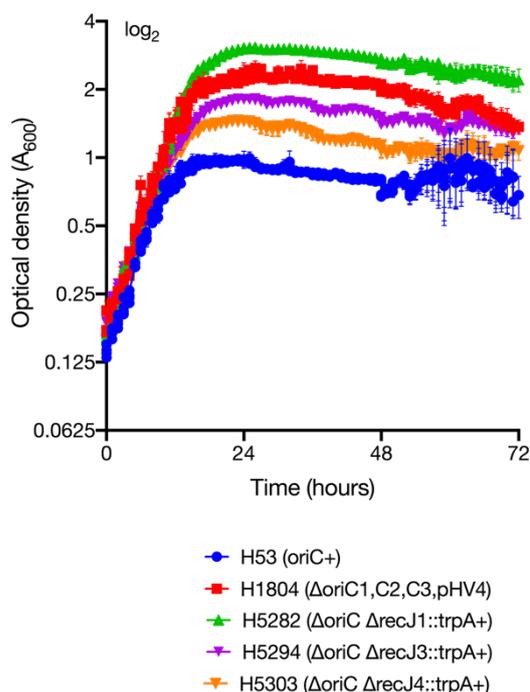


Figure 4.35: Effect of aphidicolin treatment on survival of Δ *oriC* Δ *recJ* strains. Strains H53 (*oriC*+) , H1804 (Δ *oriC*) and strains with both Δ *oriC* and Δ *recJ*1, Δ *recJ*3 or Δ *recJ*4 deletions (H5282, H5294, H5303 respectively) were monitored for growth through measurement of optical density (A_{600}). All strains were treated with either DMSO (A; control) or aphidicolin (B; final concentration 15 µg/ml) chronically for 72 hours. All strains (n=2) were incubated in a single 96-well plate and measured simultaneously using an Epoch Microplate Spectrophotometer (BioTek).

As previously observed, strains deleted for all chromosomal origins show an increased resistance to aphidicolin treatment compared to the *ori*⁺ strain (here, H53). H5294 ($\Delta oriC \Delta recJ3::trpA^+$) and H5303 ($\Delta oriC \Delta recJ4::trpA^+$) have an decreased resistance to aphidicolin treatment compared to background strain H1804. However, H5282 ($\Delta oriC \Delta recJ1::trpA^+$) shows a resistance to aphidicolin greater than that of the $\Delta oriC$ strain H1804 alone.

Both the *oriC*⁺ $\Delta recJ1$ mutant, H3929, and $\Delta oriC \Delta recJ1$ mutant tested here, H5282, show an increased resistance to aphidicolin than their background counterpart (H164 and H1804 respectively; both *oriC*⁺). This suggests the increased resistance to aphidicolin is partially due to the absence of *recJ1*.

4.3.6.3 Deletion in combination with RNase HII (*rnhB*) and Flap endonuclease 1 (*fen1*)

During lagging strand maturation, the RNA primers of Okazaki fragments must be removed to allow for ligation and the generation of a continuous DNA strand. RNase H proteins act to remove RNA primers associated with Okazaki fragments, while flap endonucleases remove any flap structures generated due to displacement of primers. RNase H proteins are also capable of degrading R-loops (DNA-RNA hybrids) in a sequence-independent manner (Hyjek *et al.*, 2019). Of the three classes of RNase H proteins, it is RNase H2 proteins (present in all domains of life) that have a specialised function for cleaving ribonucleotides within DNA (Hyjek *et al.*, 2019).

H. volcanii encodes a Fen1 flap endonuclease homologue (*fen1*; HVO_2873) and a single RNase H2 protein (*rnhB*; HVO_1978). Both *fen1* and *rnhB* encode proteins with a PIP motif, linking them to clamp protein PCNA and therefore to the replication fork (Meslet-Cladiere *et al.*, 2007). Both are non-essential in *H. volcanii*, both alone and in combination, suggesting redundancy with alternative proteins able to fulfil these roles (Lestini *et al.*, 2010, Meslet-Cladiere *et al.*, 2007).

In *T. kodakarensis*, GAN is known to possess 5'-3' exonuclease activity, which may provide an alternative route for primer removal. Genetic analysis in *T. kodakarensis* has revealed that in a strain deleted for GAN (Δgan), both RNase HII (TK0805) and Fen1 (TK1281) become essential (Burkhart, 2017), while in the presence of GAN (*gan*⁺), strains can be deleted for both Fen1 and RNase HII.

If the same logic is applied to *H. volcanii*, it could be hypothesised that the *H. volcanii* RecJ acting in replication (akin to GAN) should be essential in strains deleted for *fen1* and/or *rnhB*. However, it may be possible that *H. volcanii* encodes other redundant pathways that are not present in *T. kodakarensis*. To test this, *rnhB* and *fen1* were targeted for deletion in strains carrying *recJ* deletions in addition to *recJ1*, *recJ3* and *recJ4* being targeted for deletion in strains carrying *fen1* or *rnhB* deletions.

Deletion in combination with $\Delta rnhB$

A deletion construct targeting *rnhB* was generated prior to this study; pTA1775 (Figure 4.36; constructed by Dasha Ausiannikava, unpublished data). As a part of this study, a *trpA* marker was added to allow for direct selection of mutants. Since *rnhB* is located within an operon, the *trpA* marker does not require its own promoter and its expression can be driven via the promoter of the operon itself. The promoterless *trpA* marker was digested from pTA1166 (Stroud *et al.*, 2012) and inserted into pTA1775 at its corresponding *NdeI* site, giving *trpA*-marker *rnhB* deletion plasmid pTA2329 (Figure 4.37).

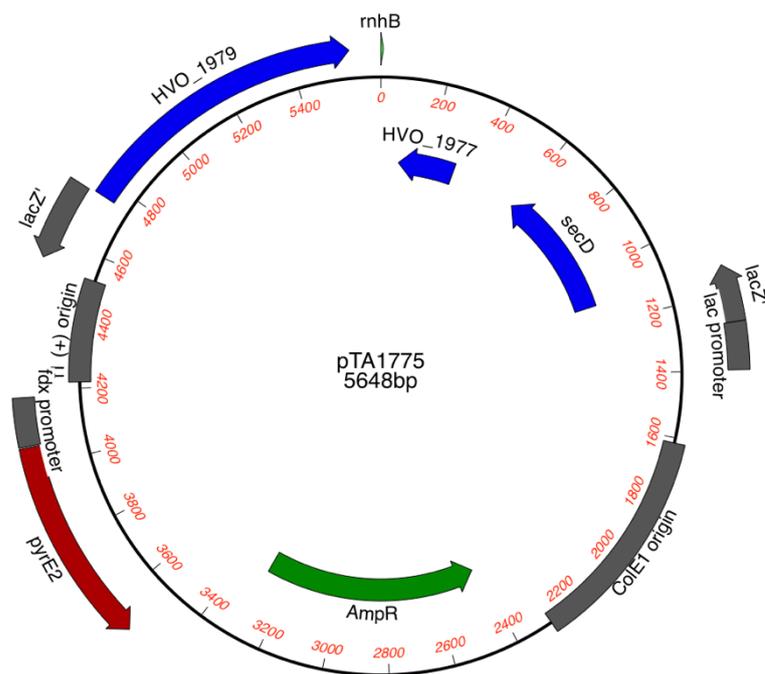


Figure 4.36: pTA1775. Deletion construct for *rnhB*. Generated by Dasha Ausiannikava.

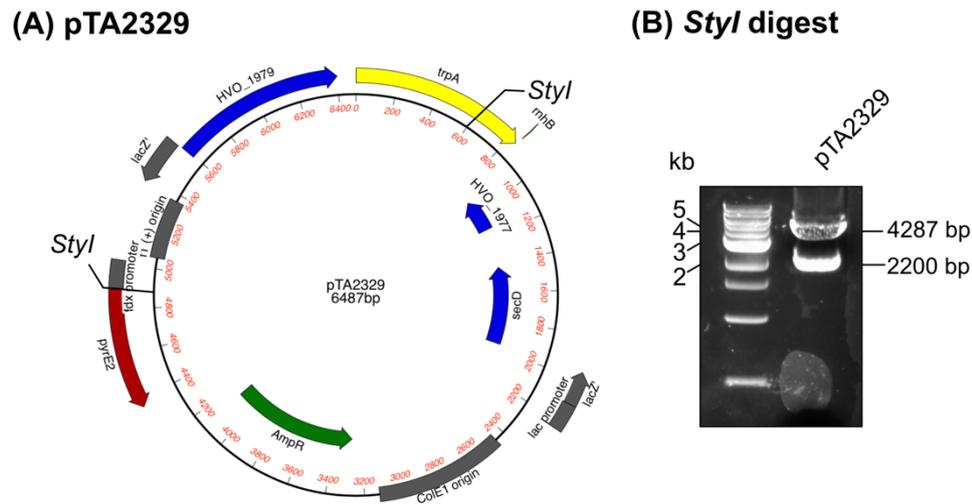
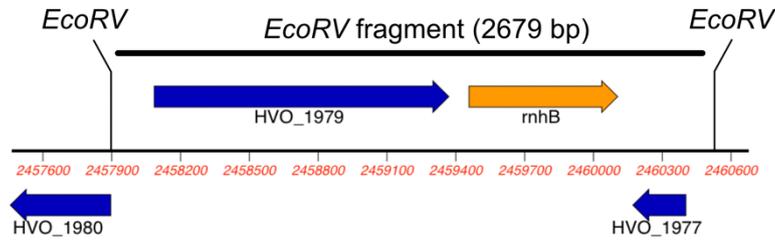


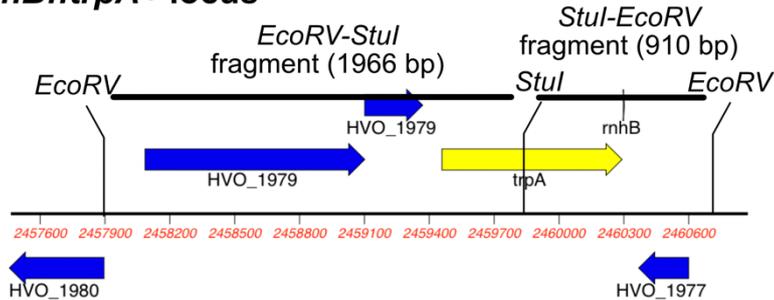
Figure 4.37: pTA2329. (A) Deletion construct for *rnhB* with a *trpA* selectable marker. (B) *StyI* digest shows bands at 4286 bp and 2200 bp, as predicted.

H53 was transformed with $\Delta rnhB::trpA+$ construct pTA2329 to give pop-in strain H4702. Pop-out gave rise to colonies on 5-FOA which were screened by both colony hybridisation and Southern blotting. Strain H4743 was confirmed as being $\Delta rnhB::trpA+$ (**Figure 4.38**).

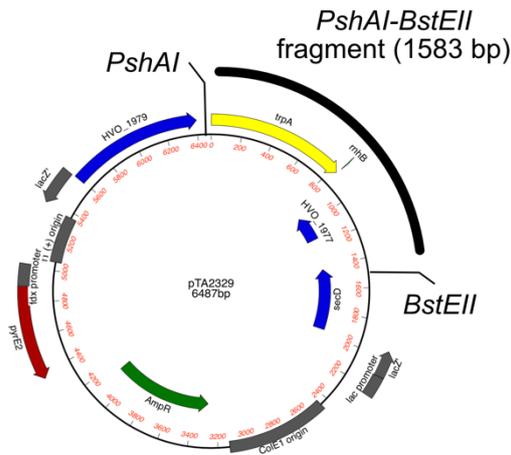
(A) Wild type *rnhB* locus



(B) $\Delta rnhB::trpA^+$ locus



(C) pTA2329 probe



(D) Southern blot

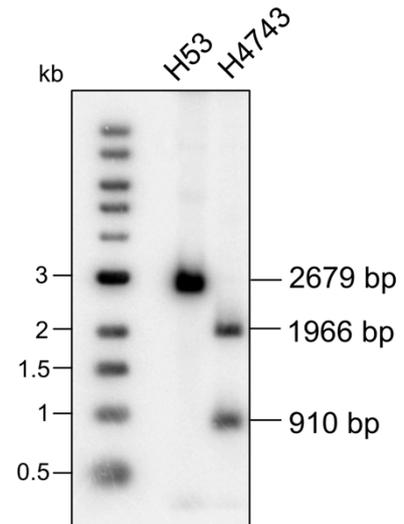


Figure 4.38: Deletion of *rnhB* (HVO_1978). (A) Expected Southern blot band sizes for wild type *rnhB* genome region digested with *EcoRV* and *StuI*. (B) Expected Southern blot band sizes for $\Delta rnhB::trpA^+$ genome region digested with *EcoRV* and *StuI*. (C) 1583 bp $\Delta rnhB::trpA^+$ Southern probe consisting of an *PshAI*-*BstEII* fragment of pTA2329. (D) Southern blot against wild type strain H53 confirming strain H4743 as $\Delta rnhB::trpA^+$.

To allow for use of the *trpA* marker in downstream deletion events, H4743 was subsequently transformed with the $\Delta rnhB$ construct lacking a *trpA* marker, pTA1775, to give pop-in strain H5359. Pop-out gave rise to colonies on 5-FOA which were then screened on plates with and without tryptophan.

Where pTA1773 had been integrated in place of pTA2329, colonies were now tryptophan auxotrophic. This gave rise to $\Delta rnhB$ strain H5382. Strain H5382 was subsequently transformed with plasmids pTA1958 ($\Delta recJ1::trpA+$), pTA1960 ($\Delta recJ3::trpA+$) and pTA1997 ($\Delta recJ4::trpA+$) to give rise to pop-in strains H5390, H5391 and H5392 respectively. Pop-outs gave rise to 5-FOA^R candidates that were patched on YPC agar and primarily screened via colony hybridisation. Strains H5404 ($\Delta rnhB \Delta recJ1::trpA+$), H5406 ($\Delta rnhB \Delta recJ3::trpA+$) and H5408 ($\Delta rnhB \Delta recJ4::trpA+$) were further verified by restriction digest and Southern blot (**Figure 4.39**).

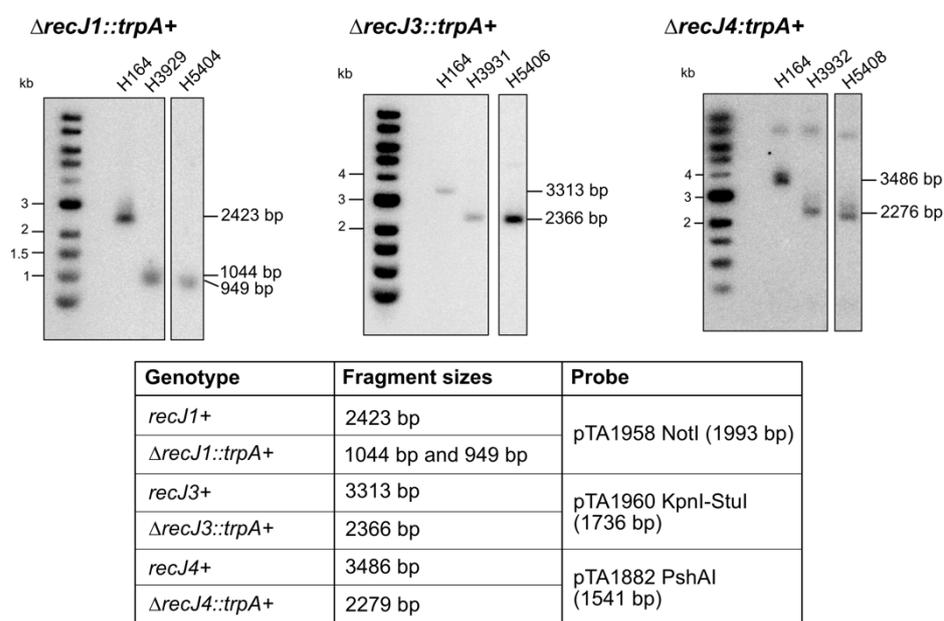


Figure 4.39: Restriction digest and Southern blotting to confirm deletion of *recJ1*, *recJ3* and *recJ4* in $\Delta rnhB$ strain H5382. Strain H5404 is confirmed as $\Delta recJ1::trpA+$, H5406 is confirmed as $\Delta recJ3::trpA+$ and H5408 is confirmed as $\Delta recJ4::trpA+$. H164 is wild type, while H3929, H3931 and H3932 were used as positive controls for $\Delta recJ1::trpA+$, $\Delta recJ3::trpA+$ and $\Delta recJ4::trpA+$ respectively. Bands seen >6 kb on the *recJ4::trpA+* blot likely represent non-specific binding of the probe.

Since it was predicted that $\Delta gan \Delta rnhB$ double mutants should be inviable, the $\Delta rnhB \Delta recJ$ strains H5404, H5406 and H5408 were further screened by Southern blot to ensure there had not been a contamination event during the replacement of the $\Delta rnhB$ construct to remove the *trpA* marker and that, in fact, these strains still carried a deletion at the *rnhB* locus (**Figure 4.40**).

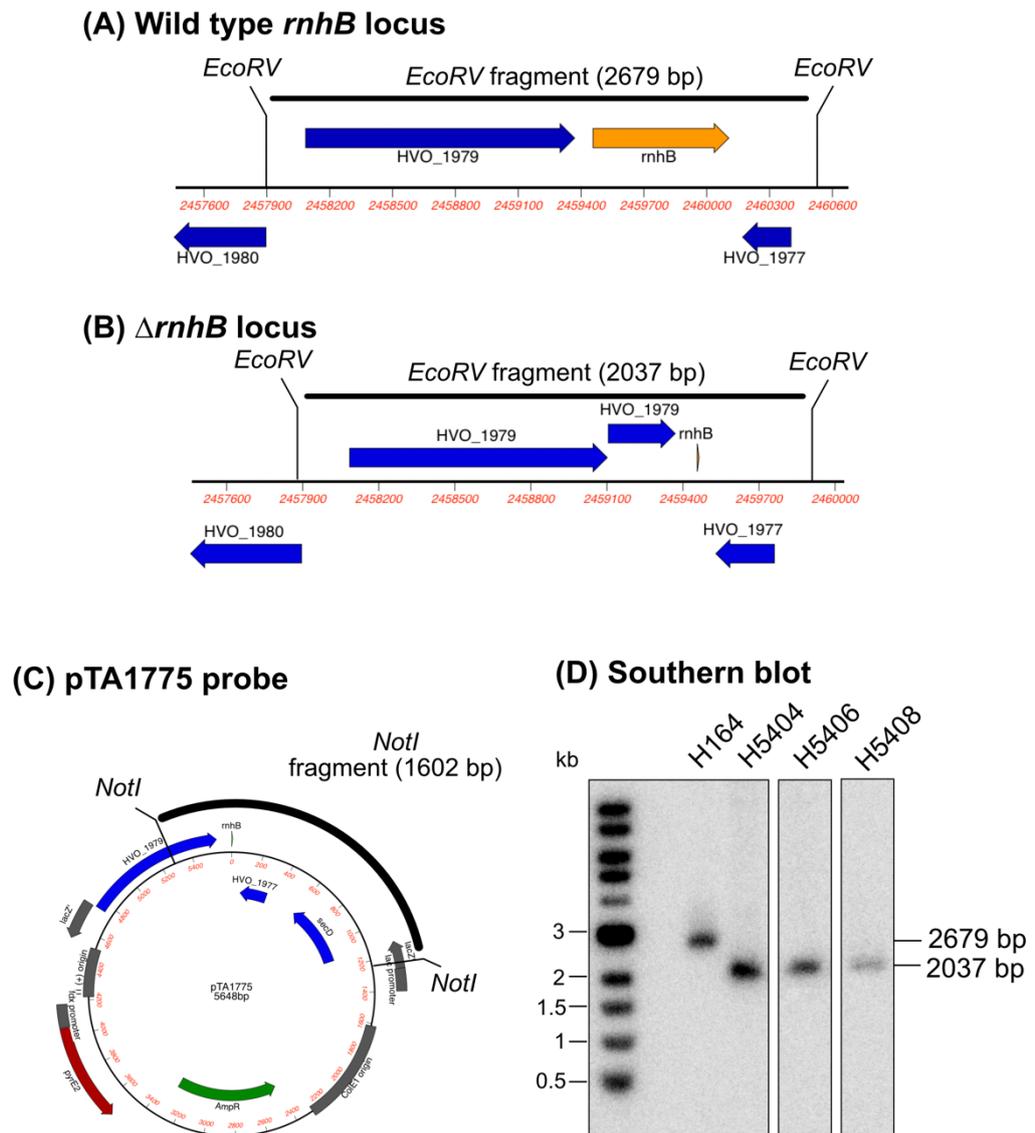


Figure 4.40: Restriction digest and Southern blotting to confirm deletion of $\Delta rnhB$ in $\Delta recJ$ strains H5404, H5406 and H5408. (A) Expected Southern blot band sizes for wild type *rnhB* genome region digested with *EcoRV*. (B) Expected Southern blot band sizes for $\Delta rnhB$ genome region digested with *EcoRV*. (C) 1602 bp $\Delta rnhB$ Southern probe consisting of an *NotI* fragment of pTA1775. (D) Southern blot confirming strains H5404, H5406 and H5408 carry $\Delta rnhB$ region.

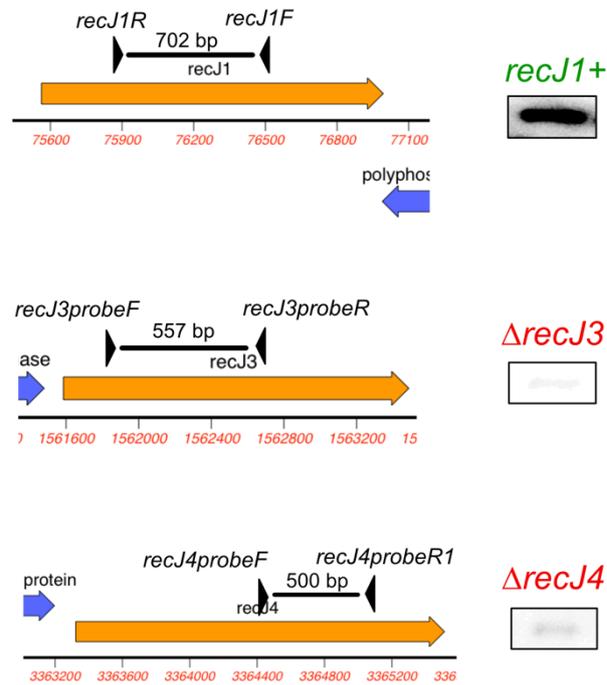
Confirmation that these strains still carried a deletion construct for *rnhB* shows deletion of the predicted GAN of *H. volcanii* is possible in combination with *rnhB* (unlike in *T. kodakarensis*).

Deletion in combination with $\Delta fen1$

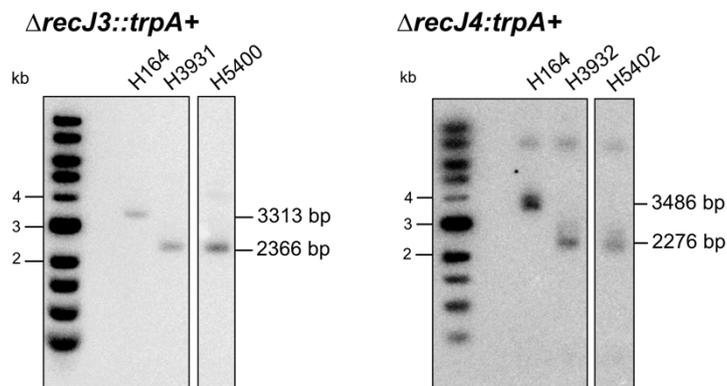
Prior to this study, deletion constructs and deletion strains targeting *fen1* were generated. Strain H588 carries a *trpA*-marked *fen1* deletion. To allow for use of the *trpA* marker in downstream deletion events, H588 was subsequently transformed with the $\Delta fen1$ construct lacking a *trpA* marker, pTA535, to give pop-in strain H5358. Pop-out gave rise to colonies on 5-FOA which were then screened on plates with and without tryptophan. Where pTA535 had been integrated in place of the *trpA*-marked construct, colonies were now tryptophan auxotrophic. This gave rise to $\Delta fen1$ strain H5381.

Strain H5381 was subsequently transformed with plasmids pTA1958 ($\Delta recJ1::trpA+$), pTA1960 ($\Delta recJ3::trpA+$) and pTA1997 ($\Delta recJ4::trpA+$) to give rise to pop-in strains H5387, H5388 and H5389 respectively. Pop-outs gave rise to 5-FOA^R candidates that were patched on YPC agar and primarily screened via colony hybridisation. Strains H5400 ($\Delta fen1 \Delta recJ3::trpA+$) and H5402 ($\Delta fen1 \Delta recJ4::trpA+$) were further verified by restriction digest and Southern blot (**Figure 4.41**). No candidates were isolated for deletion of *recJ3* or *recJ4* in the Δhjc background. Pop-outs of H5387 were repeated multiple times, with each pop-out event only giving rise to an average of 20 colonies, however all candidates screened by colony hybridisation hybridised with the probe, suggesting RecJ1 is essential in a $\Delta fen1$ strain.

A) Colony hybridisation



(B) Southern blot



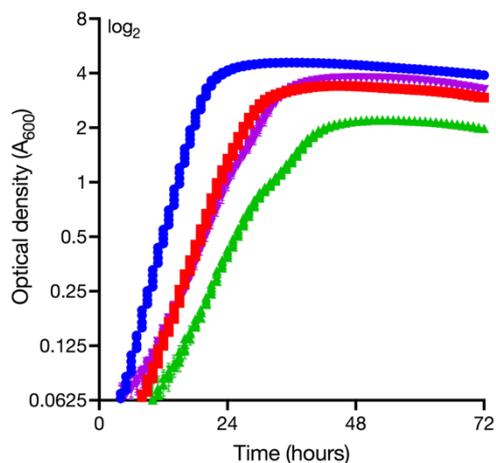
Genotype	Fragment sizes	Probe
<i>recJ3+</i>	3313 bp	pTA1960 KpnI-StuI (1736 bp)
$\Delta recJ3::trpA+$	2366 bp	
<i>recJ4+</i>	3486 bp	pTA1882 PshAI (1541 bp)
$\Delta recJ4::trpA+$	2279 bp	

Figure 4.41: Generation of $\Delta recJ$ mutants in $\Delta fen1$ strain H5381. (A) Hybridisation with probes consisting of PCR products of *recJ1* sequence (702 bp), *recJ3* sequence (557 bp) and *recJ4* sequence (500 bp) gave rise to candidates for $\Delta fen1 \Delta recJ3::trpA+$ and $\Delta fen1 \Delta recJ4::trpA+$, but all candidates probed for $\Delta fen1 \Delta recJ1::trpA+$ were wild type. (B) Southern blotting of candidates H5400 and H5402 show successful deletion of *recJ3* and *recJ4* respectively.

Growth rate

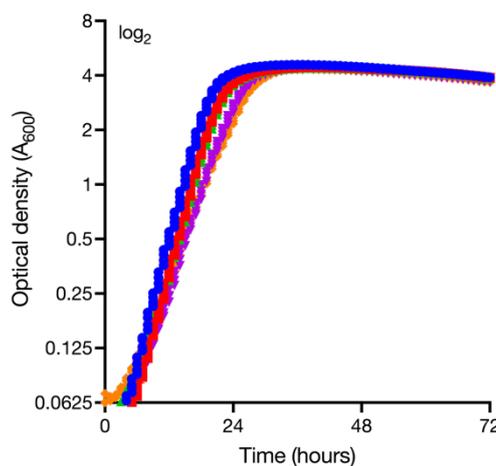
In order to compare growth rate differences, strains were assayed for growth rate alongside the single mutant for $\Delta rnhB$ or $\Delta fenI$, and a wild type control, H53 ($\Delta pyrE2 \Delta trpA$). Strains were grown for two consecutive overnights in Hv-YPC broth, ensuring on day three actively dividing cells were used for the assay. Cells were diluted in fresh Hv-YPC broth and arrayed in a 96-well plate. OD (A_{600}) was semi-continuously measured, allowing the plotting of growth curves for each strain (**Figure 4.42**).

(A) $\Delta fen1$



- H53 (*fen1*+) **2.5 h**
- H5381 ($\Delta fen1$) **3.5 h**
- ▲ H5400 ($\Delta fen1 \Delta recJ3::trpA+$) **4.5 h**
- ▼ H5402 ($\Delta fen1 \Delta recJ4::trpA+$) **4 h**

(B) $\Delta rnhB$



- H53 (*rnhB*+) **2.5 h**
- H4743 ($\Delta rnhB::trpA+$) **2.5 h**
- ▲ H5404 ($\Delta rnhB \Delta recJ1::trpA+$) **2.75 h**
- ▼ H5406 ($\Delta rnhB \Delta recJ3::trpA+$) **3 h**
- ◆ H5408 ($\Delta rnhB \Delta recJ4::trpA+$) **3.75 h**

Figure 4.42: Exponential growth rate of strains deleted for *rnhB* or *fen1* in combination with *recJ1*, *recJ3* or *recJ4*. Generation time in hours (h) is indicated in bold beside each strain legend. Strains were grown in Hv-YPC broth for two consecutive overnights before being diluted in fresh Hv-YPC broth. All strains (n=2) were incubated on the same 96-well plate and measured simultaneously for optical density (A_{600}) using an Epoch2 Microplate Spectrophotometer (BioTek).

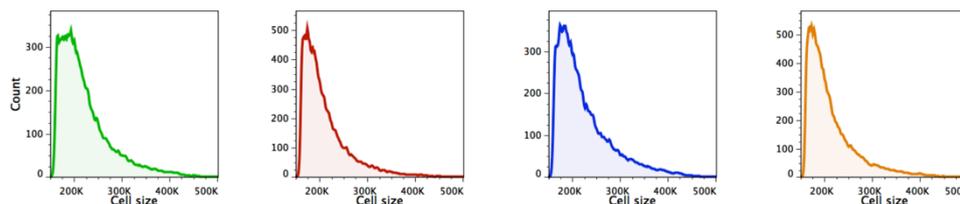
Deletion of $\Delta rnhB$ does not cause a growth defect compared to wild type. Growth rates of $\Delta rnhB \Delta recJ$ mutants are generally comparable to the single $\Delta rnhB$ deletion strain, with H5406 ($\Delta rnhB \Delta recJ3::trpA+$) and H5408 ($\Delta rnhB \Delta recJ4::trpA+$) having a slight growth defect compared to wild type and H4743 ($\Delta rnhB::trpA+$).

Deletion of *fen1* causes a minor defect, with a doubling time of 3.5 hours, compared to 2.5 hours for control strain H53. Both H5400 ($\Delta fen1 \Delta recJ3::trpA+$) and H5402 ($\Delta fen1 \Delta recJ4::trpA+$) show a defect greater than the single $\Delta fen1$ mutant alone (4.5 hours and 4 hours respectively).

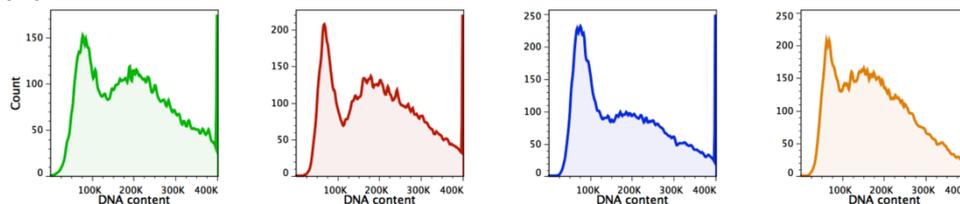
DNA content and cell size

In order to determine the DNA content and cell size of cells deleted for *recJ1*, *recJ3* or *recJ4* in combination with *rnhB*, strains were analysed using flow cytometry (**Figure 4.43**).

(A) Cell size



(B) DNA content



(C) Cell size vs DNA content

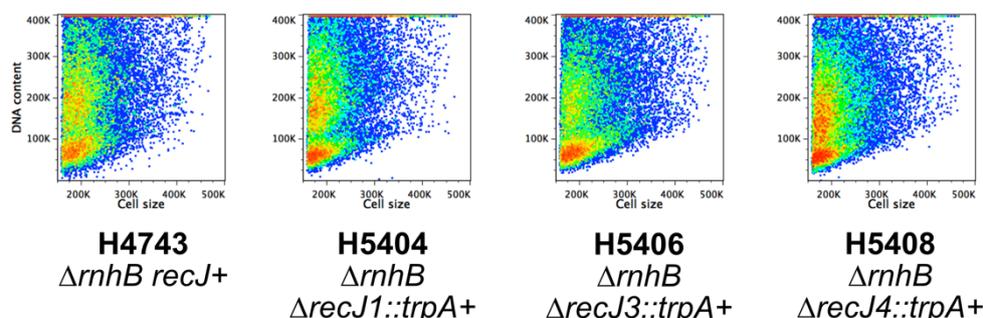


Figure 4.43: Flow cytometry analysis of strains deleted for *recJ1*, *recJ3* or *recJ4* in combination with deletion of *rn h B*. (A) Determination of cell size. Cell sizes are generally comparable for all strains tested. (B) Determination of DNA content. H5408 shows an increase in DNA content compared to the other mutants. (C) Density dot plots displaying cell size vs DNA content. H5408 shows a greater DNA content, as seen by an upward shift in the graph.

Generally, all $\Delta recJ \Delta rn h B$ mutants had cell profiles comparable to that of the $\Delta rn h B$ mutant H4743. H5408 shows a minor increase in DNA content compared to the other mutants, as seen in both the DNA content histogram and the cell size vs DNA content pseudoplot.

In order to determine the DNA content and cell size of cells deleted for *recJ3* or *recJ4* in combination with *fen1*, strains were analysed using flow cytometry (Figure 4.44).

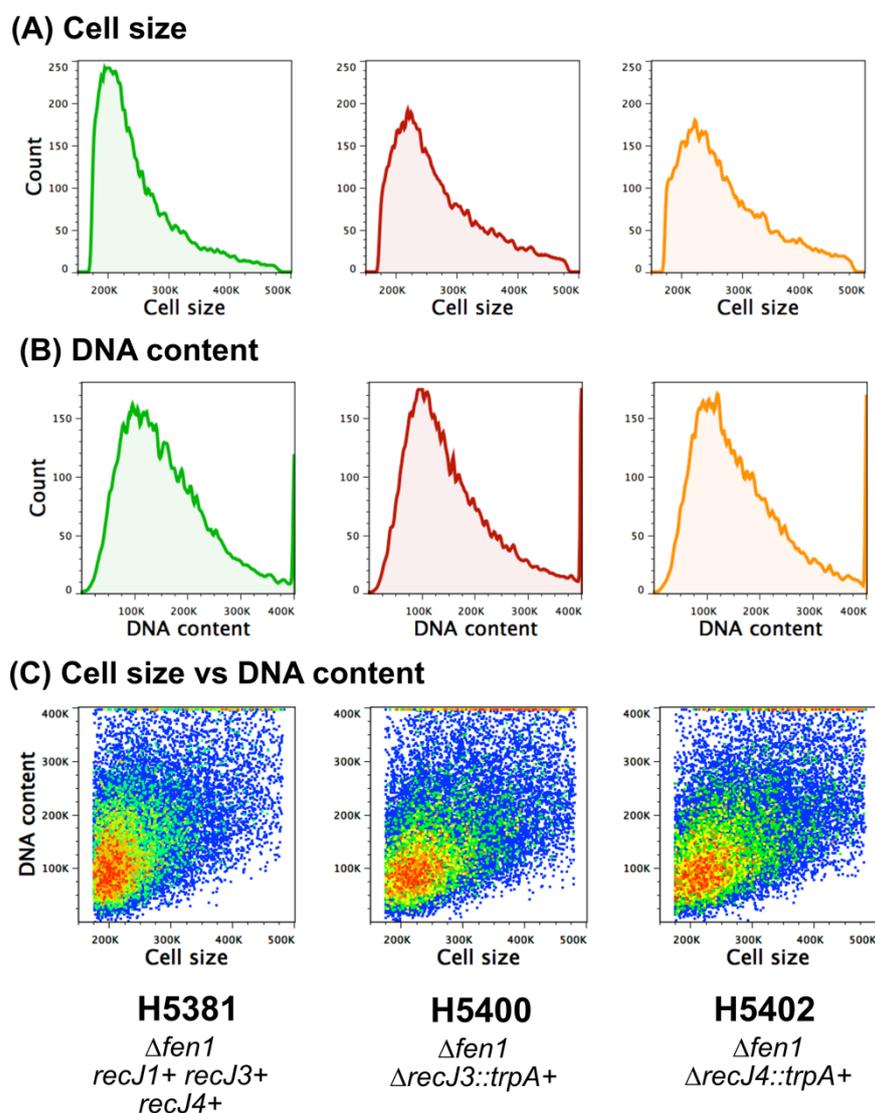


Figure 4.44: Flow cytometry analysis of strains deleted for *recJ3* or *recJ4* in combination with deletion of *fen1*. (A) Determination of cell size. H5400 ($\Delta fen1 \Delta recJ3::trpA+$) and H5402 ($\Delta fen1 \Delta recJ4::trpA+$) show a wider distribution of cell sizes compared to $\Delta fen1$ strain H5381. (B) Determination of DNA content. Generally, all strains have comparable DNA content profiles. (C) Density dot plots displaying cell size vs DNA content. H5400 and H5402 show a wider distribution of cell sizes than H5381.

Generally, the DNA content of the $\Delta fen1$ mutant strains were comparable. However, where deleted in combination with *recJ3* or *recJ4*, there is an increased distribution of cell sizes compared to H5381 ($\Delta fen1$ alone).

Aphidicolin response

Archaeal Fen1 has been predicted to have a direct role in DNA repair, as in eukaryotes, where it functions to remove displaced primers during lagging strand synthesis (Meslet-Cladiere *et al.*, 2007). RNase H proteins are predicted to remove RNA primers from fully replicated Okazaki fragments, and they are also capable of degrading R-loops (RNA-DNA hybrids) in a sequence-independent manner (Hyjek *et al.*, 2019, Chapados *et al.*, 2001). Both *fen1* and *rnhB* encode a PIP motif for interaction with PCNA, placing them at the replication fork (Meslet-Cladiere *et al.*, 2007). Thus, it was of interest to see if the mutants generated here showed an altered response to aphidicolin treatment, and whether this is changed when deleted in combination with any *recJ* gene.

In order to compare aphidicolin responses of $\Delta fen1$ mutants, strains H53 ($\Delta fen1+$), H5381 ($\Delta fen1$), H5400 ($\Delta fen1 \Delta recJ3::trpA+$) and H5402 ($\Delta fen1 \Delta recJ4::trpA+$) were grown in the presence of aphidicolin for 72 hours. Strains were grown for two consecutive overnights in Hv-YPC broth, ensuring on day three actively dividing cells were used for the assay. Cells were diluted in fresh Hv-YPC broth and arrayed in a 96-well plate. OD (A_{600}) was semi-continuously measured, allowing the plotting of growth curves for each strain (**Figure 4.45**).

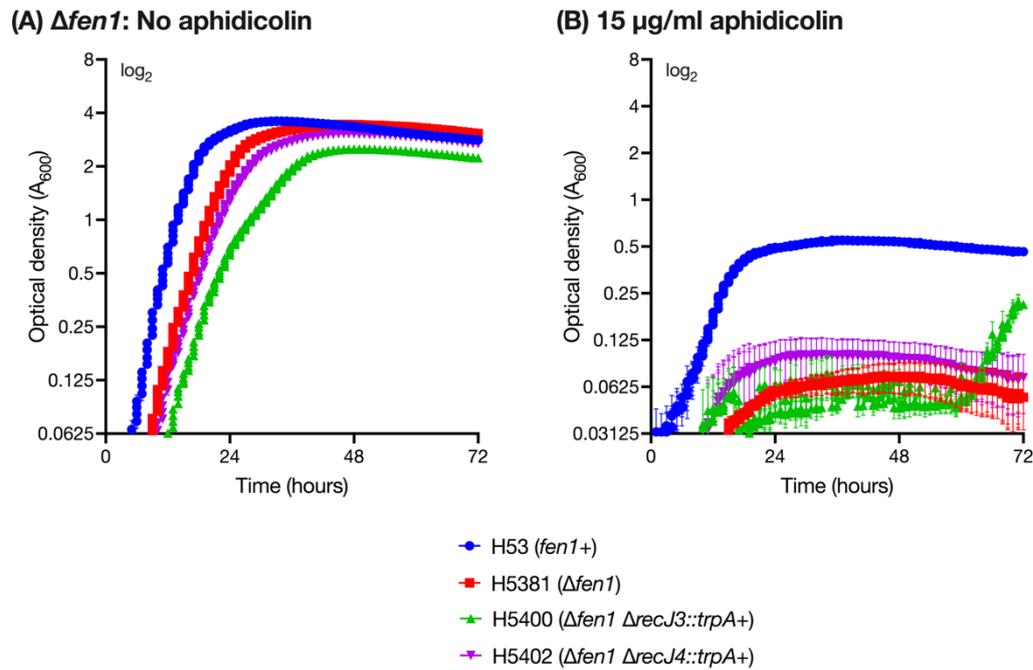


Figure 4.45: Effect of aphidicolin treatment on survival of Δ *fenI* Δ *recJ* strains. Strains H53 (*fenI*+), H5400 (Δ *fenI* Δ *recJ3::trpA*+) and H5402 (Δ *fenI* Δ *recJ4::trpA*+) were monitored for growth through measurement of optical density (A_{600}). All strains were treated with either DMSO (A; control) or aphidicolin (B; final concentration 15 μ g/ml) chronically for 72 hours. All strains (n=2) were incubated in a single 96-well plate and measured simultaneously using an Epoch Microplate Spectrophotometer (BioTek).

All strains carrying a *fenI* deletion show a significant sensitivity to aphidicolin compared to wild type H53 (*fenI*+).

In order to compare aphidicolin responses of Δ *rnhB* mutants, strains H53 (Δ *rnhB*+), H4743 (Δ *rnhB::trpA*+), H5404 (Δ *rnhB* Δ *recJ1::trpA*+), H5406 (Δ *rnhB* Δ *recJ3::trpA*+) and H5408 (Δ *rnhB* Δ *recJ4::trpA*+) were grown in the presence of aphidicolin for 72 hours. Strains were grown for two consecutive overnights in Hv-YPC broth, ensuring on day three actively dividing cells were used for the assay. Cells were diluted in fresh Hv-YPC broth and arrayed in a 96-well plate. OD (A_{600}) was semi-continuously measured, allowing the plotting of growth curves for each strain (Figure 4.46).

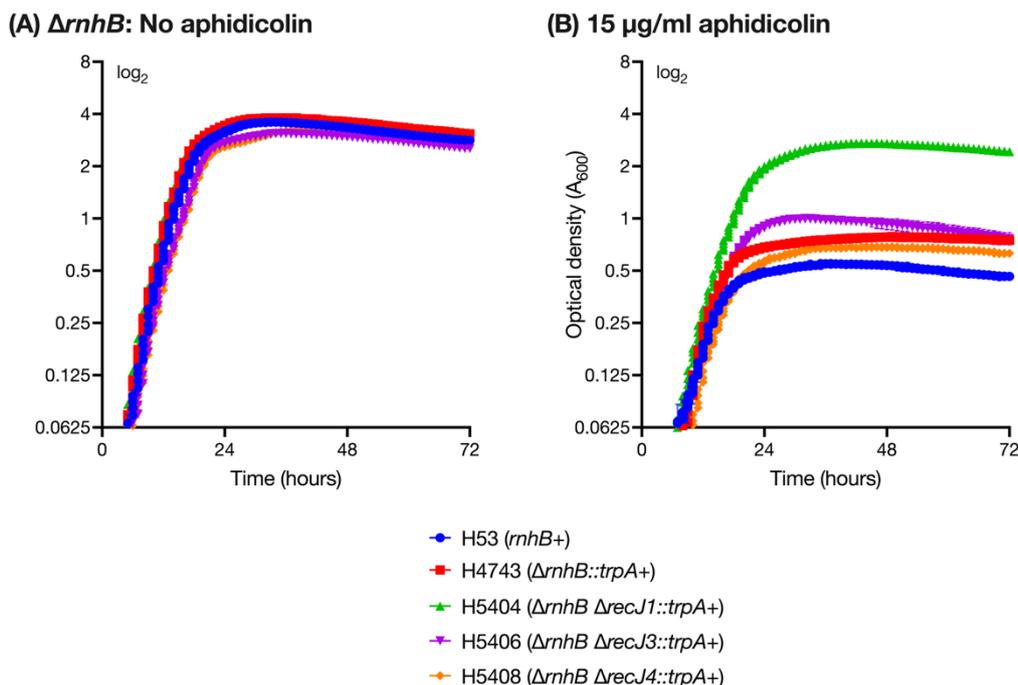


Figure 4.46: Effect of aphidicolin treatment on survival of $\Delta fen1 \Delta recJ$ strains. Strains H53 (*rnhB*+), H5404 ($\Delta rnhB \Delta recJ1::trpA+$), H5406 ($\Delta rnhB \Delta recJ3::trpA+$) and H5408 ($\Delta rnhB \Delta recJ4::trpA+$) were monitored for growth through measurement of optical density (A_{600}). All strains were treated with either DMSO (A; control) or aphidicolin (B; final concentration 15 $\mu\text{g/ml}$) chronically for 72 hours. All strains (n=2) were incubated in a single 96-well plate and measured simultaneously using an Epoch Microplate Spectrophotometer (BioTek).

All strains carrying an $\Delta rnhB$ deletion show at least a minor increase in resistance to aphidicolin compared to wild type H53 (*rnhB*+). The most prominent resistance to aphidicolin is seen in strain H5404 ($\Delta rnhB \Delta recJ1::trpA+$).

Survival following treatment with mitomycin C

Strains deleted for *fen1* are known to be sensitive to crosslinking agent MMC (Duan, 2009), while the MMC response of strains deleted for *rnhB* remains untested in *H. volcanii*. However, it is worth noting that deletion of RNase H genes in eukaryotes leads to strong sensitivity to DNA damaging agents (Arudchandran *et al.*, 2000, Lazzaro *et al.*, 2012), and thus a sensitivity would be predicted. In order to test whether there was a synthetic defect between *recJ1/recJ3/recJ4* and *fen1* or *rnhB*, $\Delta rnhB$ strains H4743 ($\Delta rnhB::trpA+$), H5404 ($\Delta rnhB \Delta recJ1::trpA+$), H5406 ($\Delta rnhB \Delta recJ3::trpA+$) and H5408 ($\Delta rnhB \Delta recJ4::trpA+$) and $\Delta fen1$ strains H5381 ($\Delta fen1$), H5400 ($\Delta fen1 \Delta recJ3::trpA+$) and H5402 ($\Delta fen1 \Delta recJ4::trpA+$)

were tested for response to MMC treatment compared to control strain H164 (*rnhB*+ *fen1*+ *recJ*+). Cultures were grown to mid-exponential phase, and replica-spotted onto Hv-YPC agar containing either water (control) or MMC. Plates were incubated for 4-7 days at 45°C, colonies were counted, and survival fractions were calculated (Figure 4.47).

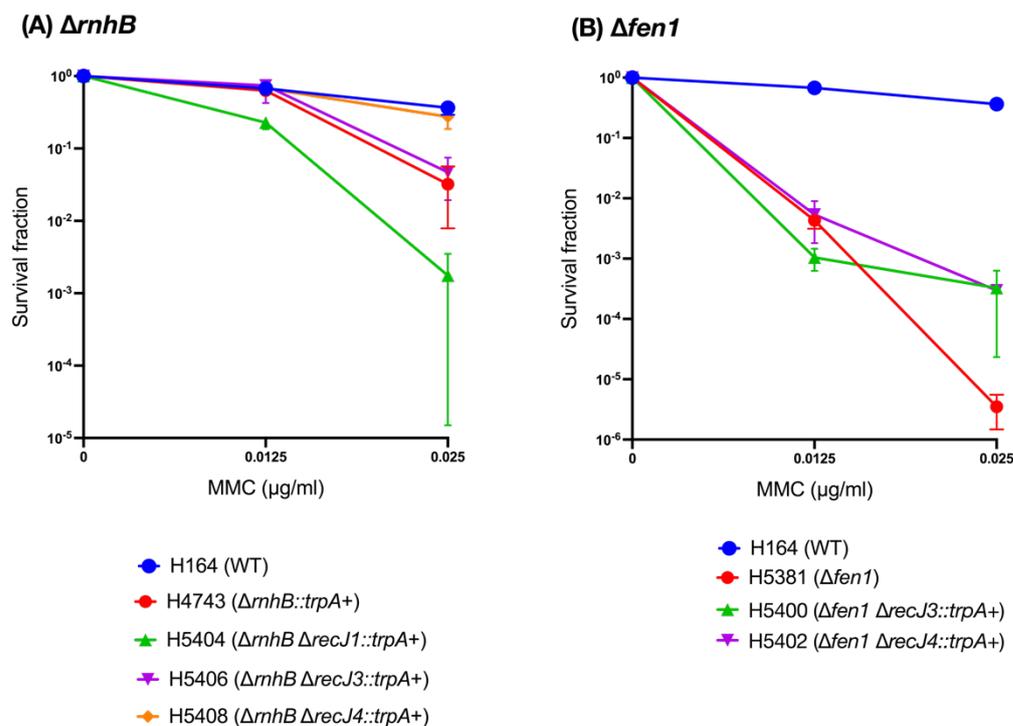


Figure 4.47: Survival frequency of (A) $\Delta rnhB$ and (B) $\Delta fen1$ strains following treatment with mitomycin C (MMC). Control strain H164, $\Delta rnhB$ strains H4743 ($\Delta rnhB::trpA+$), H5404 ($\Delta rnhB \Delta recJ1::trpA+$), H5406 ($\Delta rnhB \Delta recJ3::trpA+$) and H5408 ($\Delta rnhB \Delta recJ4::trpA+$) and $\Delta fen1$ strains H5381 ($\Delta fen1$), H5400 ($\Delta fen1 \Delta recJ3::trpA+$) and H5402 ($\Delta fen1 \Delta recJ4::trpA+$) were chronically exposed to MMC within YPC plates. (A) Strain H4743 only shows a differential response to control at the highest dose, 0.025 $\mu\text{g/ml}$. Both H5404 and H5406 show an increased sensitivity to MMC at the highest dose, with only H5404 showing an increased sensitivity at intermediate dose, 0.0125 $\mu\text{g/ml}$. However, H5408 shows no difference in MMC sensitivity compared to control strain H164. (B) Strain H5381 shows a large sensitivity to MMC compared to control strain H164. Generally, the *recJ fen1* mutants are comparable to this, with the exception of showing less sensitivity at highest dose, 0.025 $\mu\text{g/ml}$. Survival fraction is calculated relative to untreated control. Each data point is generated as an average of at least 3 independent trials. Error bars represent standard error of the mean (SEM).

Deletion of *rnhB* does not cause a major sensitivity to MMC; a minor difference in survival fraction between $\Delta rnhB$ strain H4743 and wild type strain H164 is only observed at the highest dose tested, 0.025 $\mu\text{g/ml}$. Both H5404 and H5406 show sensitivity to MMC, akin to $\Delta rnhB$ at this high dose, with H5404 showing greater sensitivity. In contrast, H5408 shows a MMC response directly comparable to that of wild type strain H164.

As previously observed, *fenI* mutants are highly sensitive to MMC. When *recJ3* or *recJ4* are deleted in combination with *fenI*, the sensitivity is still present, however there is an increased survival fraction at highest dose 0.025 $\mu\text{g/ml}$, compared to the *fenI* mutant H5381 alone.

Survival following treatment with UV

Strains deleted for *fenI* and *rnhB* have previously been reported to be sensitive to UV irradiation (Meslet-Cladiere *et al.*, 2007). In order to test whether there was a synthetic defect between *recJ1/recJ3/recJ4* and *fenI* or *rnhB*, $\Delta rnhB$ strains H4743 ($\Delta rnhB::trpA+$), H5404 ($\Delta rnhB \Delta recJ1::trpA+$), H5406 ($\Delta rnhB \Delta recJ3::trpA+$) and H5408 ($\Delta rnhB \Delta recJ4::trpA+$) and $\Delta fenI$ strains H5381 ($\Delta fenI$), H5400 ($\Delta fenI \Delta recJ3::trpA+$) and H5402 ($\Delta fenI \Delta recJ4::trpA+$) were tested for response to UV treatment compared to control strain H164 (*rnhB*+ *fenI*+ *recJ*+). Cultures were grown to mid-exponential phase, spotted onto Hv-YPC agar and treated with varying doses of UV. Plates were incubated in the dark to prevent the innate visible light-dependent action of photolyases from repairing the bulky lesions in the treated strains. After 4-7 days of incubation at 45°C, colonies were counted, and survival fractions were calculated accordingly (**Figure 4.48**).

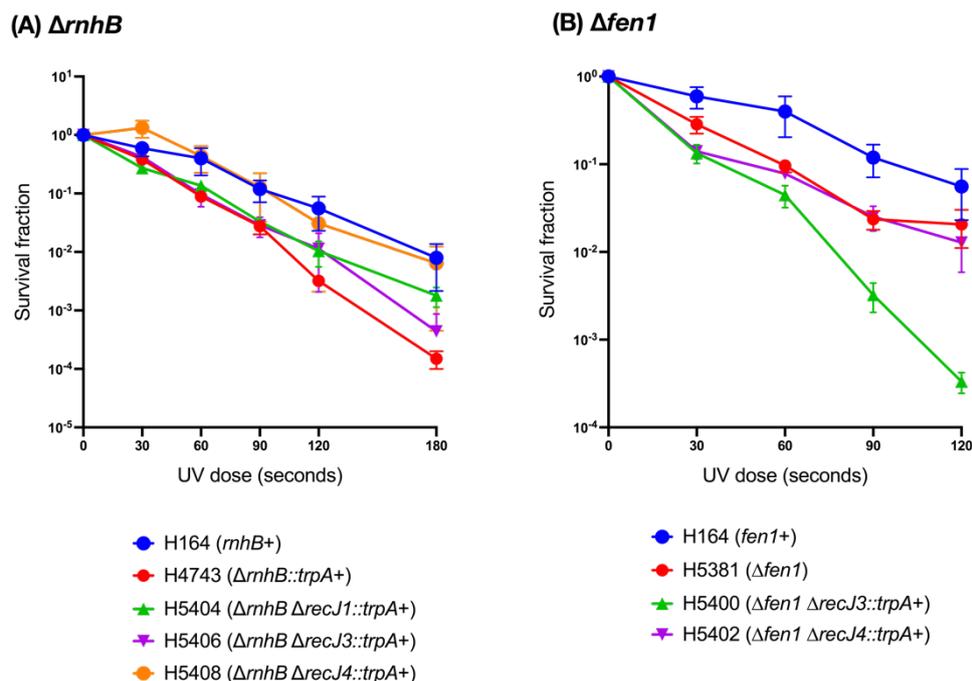


Figure 4.48: Survival frequency of (A) $\Delta rnhB$ and (B) $\Delta fen1$ strains following treatment with ultraviolet (UV) radiation. Control strain H164, $\Delta rnhB$ strains H4743 ($\Delta rnhB::trpA+$), H5404 ($\Delta rnhB \Delta recJ1::trpA+$), H5406 ($\Delta rnhB \Delta recJ3::trpA+$) and H5408 ($\Delta rnhB \Delta recJ4::trpA+$) and $\Delta fen1$ strains H5381 ($\Delta fen1$), H5400 ($\Delta fen1 \Delta recJ3::trpA+$) and H5402 ($\Delta fen1 \Delta recJ4::trpA+$) were exposed to UV irradiation and survival fractions were calculated. H4743 shows an increased sensitivity to UV compared to H164, while H5408 shows a UV sensitivity profile directly comparable to H164. Deletion of *fen1* causes mild UV sensitivity, but this is much more prominent in strain H5400. Survival fraction is calculated relative to untreated control. Each data point is generated as an average of at least 3 independent trials. Error bars represent standard error of the mean (SEM).

Strain H4743 shows an increased sensitivity to UV compared to control strain H164. Where *rnhB* is deleted in combination with *recJ4*, the UV sensitivity is no longer observed, with H5408 having a response directly comparable to H164.

Strain H5381 shows only a mild sensitivity to UV. In combination with deletion of *recJ3*, there is an increased sensitivity most prominently observed at high doses. H5402 does not show an increase in sensitivity compared to H5381.

4.3.6.4 Deletion in combination with Hel308

The Hel308 homologue from *Sulfolobus tokodaii* was shown to physically interact with Hjc and the interaction between the two inhibited the unwinding activity of the helicase (Li and Heyer, 2008).

It has been shown previously that *hel308* can be deleted from *H. volcanii*, causing a growth defect and increased sensitivity to MMC (Gamble-Milner, 2016, Lever, 2019). Additionally, affinity purification of Hel308 from *H. volcanii* identified both RecJ3 and RecJ4 as candidates for interacting partners of Hel308 and thus it is of interest to see if deletion combinations of *hel308* with these genes can shed any further information on this predicted relationship (Gamble-Milner, 2016, Lever, 2019).

To allow for use of the *trpA* marker in downstream deletion events, $\Delta hel308::trpA^+$ strain H2117 was transformed with the $\Delta hel308$ construct lacking a *trpA* marker, pTA1254, to give pop-in strain H4360. Pop-out gave rise to colonies on 5-FOA which were then screened on plates with and without tryptophan. Where pTA1254 had been integrated in place of the *trpA*-marked deletion construct, colonies were now tryptophan auxotrophic. This gave rise to $\Delta hel308$ strain H4361. Strain H4361 was subsequently transformed with plasmids pTA1958 ($\Delta recJ1::trpA^+$), pTA1960 ($\Delta recJ3::trpA^+$) and pTA1997 ($\Delta recJ4::trpA^+$) to give rise to pop-in strains H5247, H5248 and H5249 respectively. Pop-outs gave rise to 5-FOA^R candidates that were patched on YPC agar and primarily screened via colony hybridisation. Strains H5288 ($\Delta hel308 \Delta recJ1::trpA^+$), H5301 ($\Delta hel308 \Delta recJ3::trpA^+$) and H5309 ($\Delta hel308 \Delta recJ4::trpA^+$) were further verified by restriction digest and Southern blot (**Figure 4.49**).

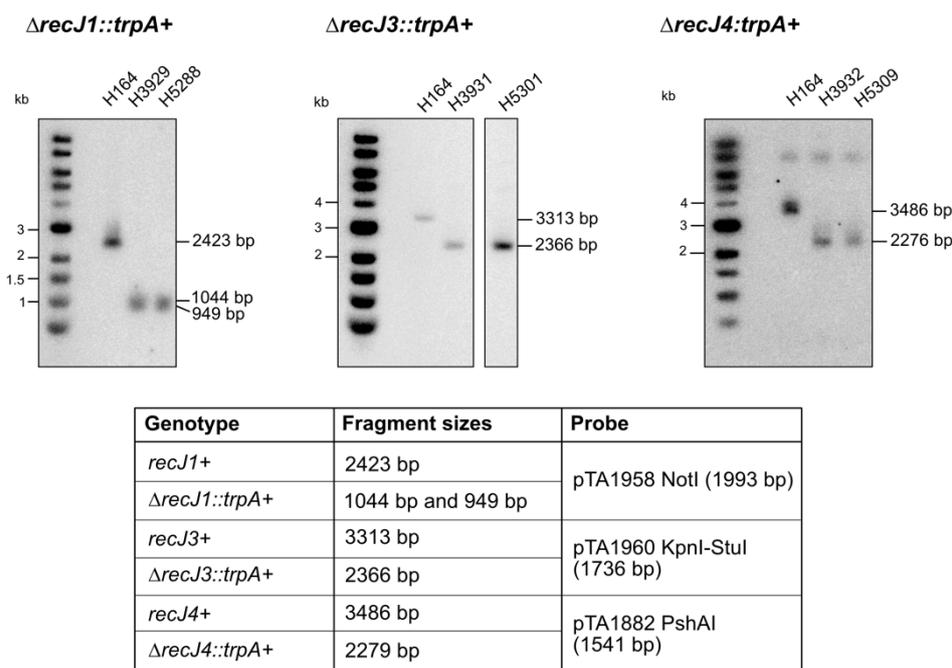


Figure 4.49: Restriction digest and Southern blotting to confirm deletion of *recJ1*, *recJ3* and *recJ4* in $\Delta hel308$ strain H4361. Strain H5288 is confirmed as $\Delta recJ1::trpA+$, H5301 is confirmed as $\Delta recJ3::trpA+$ and H5309 is confirmed as $\Delta recJ4::trpA+$. H164 is wild type, while H3929, H3931 and H3932 were used as positive controls for $\Delta recJ1::trpA+$, $\Delta recJ3::trpA+$ and $\Delta recJ4::trpA+$ respectively. Bands seen >6 kb on the *recJ4::trpA+* blot likely represent non-specific binding of the probe.

Growth rate

In order to compare growth rate differences, strains were assayed for growth rate alongside the single mutant for $\Delta hel308::trpA+$, H2117, and a wild type control, H53 ($\Delta pyrE2 \Delta trpA$). Strains were grown for two consecutive overnights in Hv-YPC broth, ensuring on day three actively dividing cells were used for the assay. Cells were diluted in fresh Hv-YPC broth and arrayed in a 96-well plate. OD (A_{600}) was semi-continuously measured, allowing the plotting of growth curves for each strain (**Figure 4.50**).

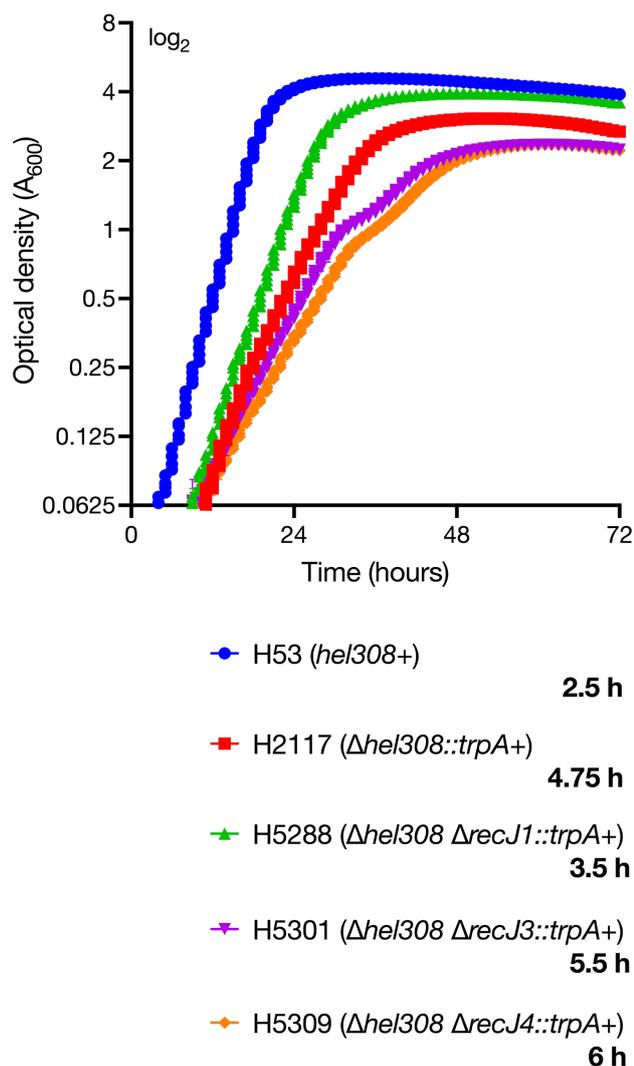


Figure 4.50: Exponential growth rate of strains deleted for *hel308* in combination with *recJ1*, *recJ3* or *recJ4*. Generation time in hours (h) is indicated in bold beside each strain legend. Strains were grown in Hv-YPC broth for two consecutive overnights before being diluted in fresh Hv-YPC broth. All strains (n=2) were incubated on the same 96-well plate and measured simultaneously for optical density (A_{600}) using an Epoch2 Microplate Spectrophotometer (BioTek).

As previously observed, deletion for *hel308* causes a growth defect compared to wild type (4.75 hours for Δ *hel308* vs. 2.5 hours for *hel308+*). In combination with deletion of *recJ3* (H5301) or *recJ4* (H5309), there is an increased attenuation of growth (5.5 hours and 6 hours respectively). In contrast, deletion of *hel308* in combination with *recJ1* (H5288) leads to a doubling time that is faster than the single Δ *hel308* deletion itself (3.5 hours).

DNA content and cell size

In order to determine the DNA content and cell size of cells deleted for *recJ1*, *recJ3* or *recJ4* in combination with *hel308*, strains were analysed using flow cytometry (Figure 4.51).

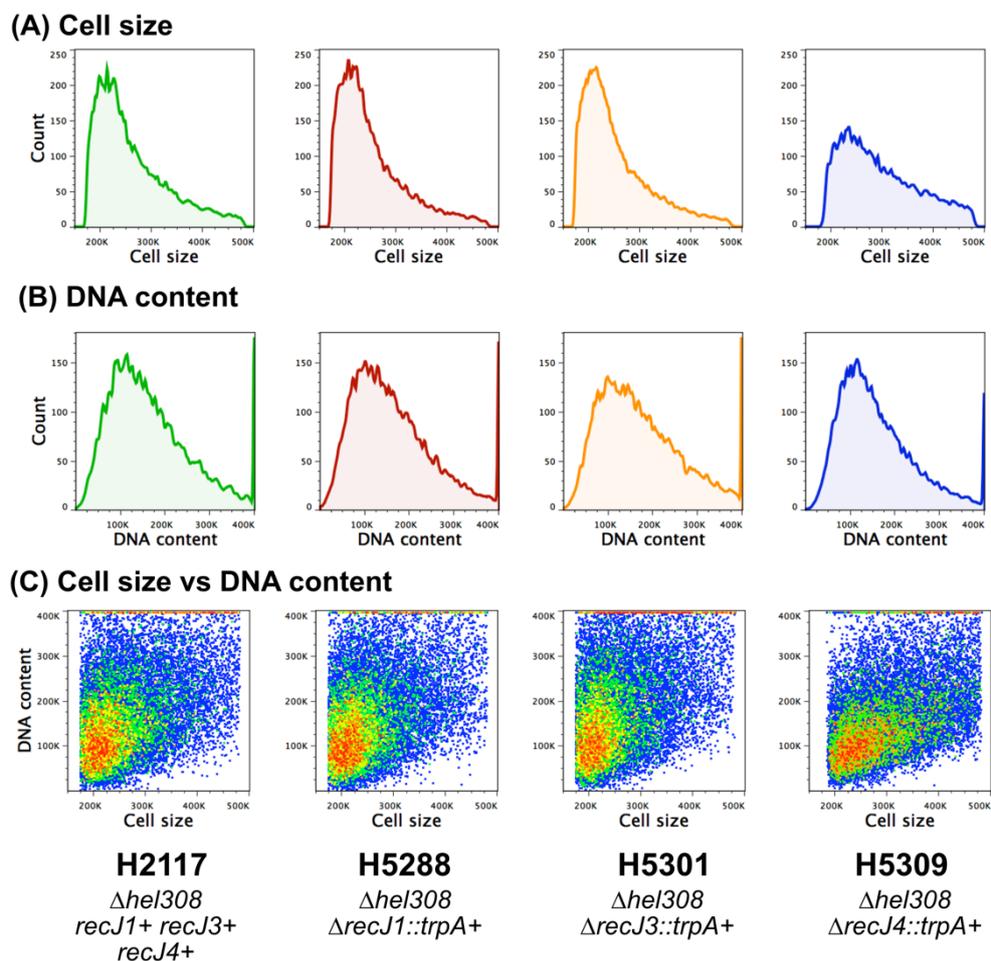


Figure 4.51: Flow cytometry analysis of strains deleted for *recJ1*, *recJ3* or *recJ4* in combination with deletion of *hel308*. (A) Determination of cell size. H5309 ($\Delta hel308 \Delta recJ4::trpA+$) show a far wider distribution of cell sizes compared to any other $\Delta hel308$ mutant analysed. (B) Determination of DNA content. Generally, all strains have comparable DNA content profiles. (C) Density dot plots displaying cell size vs DNA content. H5309 shows a wider distribution of cell sizes compared to any other strains analysed.

Cell size and DNA content for $\Delta hel308$ mutants H5288 ($\Delta hel308 \Delta recJ1::trpA+$) and H5301 ($\Delta hel308 \Delta recJ3::trpA+$) are generally comparable to the single $\Delta hel308$ mutant H2117. In contrast, H5309 ($\Delta hel308 \Delta recJ4::trpA+$) shows a much broader distribution of cell sizes, while the DNA content remains comparable to the other $\Delta hel308$ mutants.

Survival following treatment with mitomycin C

Strains deleted for *hel308* are known to be sensitive to crosslinking agent MMC (Gamble-Milner, 2016, Lever, 2019). In order to test whether there was a synthetic defect between *recJ1/recJ3/recJ4* and *hel308*, Δ *hel308* strains H2117 (Δ *hel308::trpA+*), H5288 (Δ *hel308* Δ *recJ1::trpA+*), H5301 (Δ *hel308* Δ *recJ3::trpA+*) and H5309 (Δ *hel308* Δ *recJ4::trpA+*) were tested for response to MMC treatment compared to control strain H164 (*hel308+recJ+*). Cultures were grown to mid-exponential phase, and replica-spotted onto Hv-YPC agar containing either water (control) or MMC. Plates were incubated for 4-7 days at 45°C, colonies were counted, and survival fractions were calculated (**Figure 4.52**).

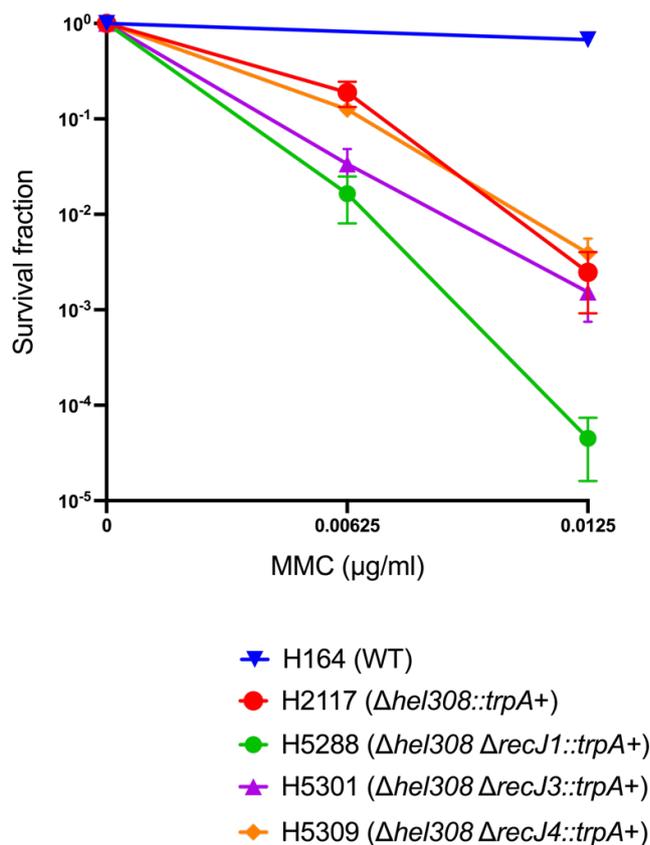


Figure 4.52: Survival frequency of $\Delta hel308$ strains in combination with deletion of *recJ1*, *recJ3* or *recJ4* following treatment with mitomycin C (MMC). Control strain H164 and $\Delta hel308$ strains H2117 ($\Delta hel308::trpA+$), H5288 ($\Delta hel308 \Delta recJ1::trpA+$), H5301 ($\Delta hel308 \Delta recJ3::trpA+$) and H5309 ($\Delta hel308 \Delta recJ4::trpA+$) were chronically exposed to MMC within YPC plates. As previously observed, even at low doses H2117 is more sensitive to MMC than H164. The survival fractions of H5301 and H5309 are generally comparable to H2117, while H5288 shows an increased sensitivity (clearest at 0.0125 $\mu\text{g/ml}$). Survival fraction is calculated relative to untreated control. Each data point is generated as an average of at least 3 independent trials. Error bars represent standard error of the mean (SEM).

As previously observed, $\Delta hel308$ strains are more sensitive to MMC than their wild type counterparts. Generally, H5301 ($\Delta hel308 \Delta recJ3::trpA+$) and H5309 ($\Delta hel308 \Delta recJ4::trpA+$) had a MMC response comparable to that of H2117 ($\Delta hel308::trpA+$). Strain H5288 ($\Delta hel308 \Delta recJ1::trpA+$) showed an increased sensitivity to MMC compared to H2117, with the clearest difference in response being seen at 0.0125 $\mu\text{g/ml}$ MMC.

4.3.4 Confirming the essentiality of *recJ2* in *H. volcanii*

Previous work has shown deletions of *recJ1*, *recJ3* and *recJ4*, either alone or in combination, are possible in *H. volcanii* with little effect on cell viability (Lever, 2019). However, repeated attempts to delete *recJ2* carried out in the Allers lab using the standard pop-in/pop-out methodology have failed.

This could be due to RecJ2 having gained a role since the multiplication event leading to four *recJ* genes occurred, classifying it as a paralogous gene. Paralogous genes are a subset of homologues related via duplication (Fitch, 1970). Generally, paralogs perform biologically distinct, even if mechanistically related, functions (Koonin, 2005).

RecJ2 is located within an operon containing numerous genes predicted to be essential, including ribosomal proteins S15 and S3Ae, a tRNA (*tRNA-AlaGGC*) and a gene predicted to encode a member of the KEOPS complex (HVO_1146, *pcc1* in *T. kodakarensis*) (**Figure 4.53**). Therefore, the standard method to determine essentiality by placing the gene under control of an inducible promoter cannot be used for *recJ2*. Integration of the tryptophan-inducible *p.tnaA* promoter upstream of *recJ2* would affect transcription of the downstream genes, of which at least *rps15* and *rps3aR* are essential (while others, e.g., HVO_1146, are predicted to be).

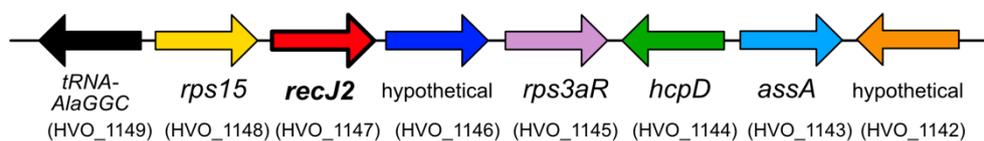


Figure 4.53: The *recJ2* operon in *Haloferax volcanii*. *recJ2* (HVO_1147) is shown in red. Genes *tRNA-AlaGGC*, *rps15*, HVO_1146 and *rps3aR* have (or are predicted to have) essential functions in transcription and/or translation.

Therefore, an alternative strategy was designed. A tryptophan-inducible *p.tnaA-recJ2* cassette is inserted at an ectopic locus and levels of (ectopic) *recJ2* expression are maintained during strain construction by the presence of tryptophan. Due to ease of screening via 5-FOA resistance, and availability of deletion constructs and lab strains, *pyrE2* was selected as the ectopic locus for integration of *recJ2*.

Ectopic expression of RecJ2 to test essentiality

Cloning of *p.tnaM3-recJ2::hdrB* for integration at *pyrE2* locus

The deletion construct for *pyrE2*, pGB68, was utilised for creation of the ectopic *p.tnaA-recJ2* gene cassette to replace *pyrE2* (Bitan-Banin *et al.*, 2003; **Figure 4.54**).

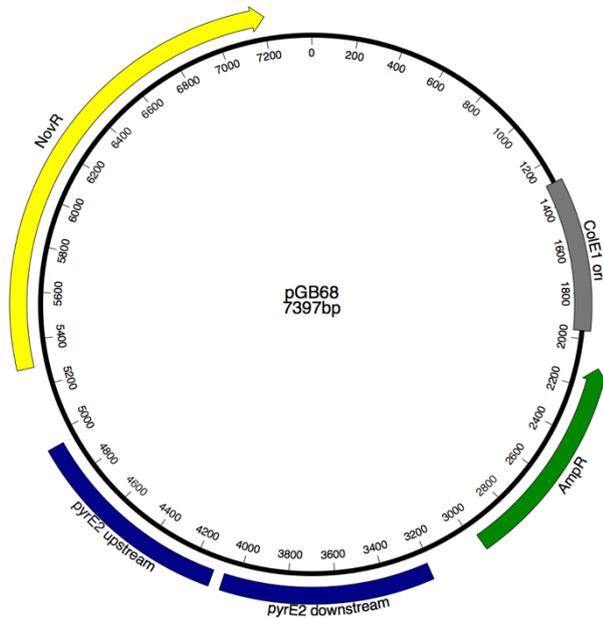


Figure 4.54: pGB68 Δ *pyrE2* construct (Bitan-Banin *et al.* 2003).

Initially, *recJ2* was placed under control of the reduced-activity variant of the tryptophan-inducible promoter, *p.tnaM3*. This was to ensure minimal-to-no expression of *recJ2* at 0 mM tryptophan, unlike the standard-activity promoter *p.tnaA* that is somewhat leaky. The use of reduced-activity *p.tnaM3* would also prevent any potentially toxic effects of RecJ2 overexpression in the pop-in strain, where both the *recJ2* cassette integrated at the *pyrE2* locus and the native *recJ2* locus are present. The coding sequence of *recJ2* was amplified from previously constructed *recJ2* genomic clone, pTA1905 (Lever, 2019), using primers *recJ2fwdNde* and *recJ2revDSBam*. These primers introduced novel 5' *Nde*I and 3' *Bam*HI sites, which were utilised to ligate *recJ2* into the *p.tnaM3* cloning plasmid pTA1451 (Braun *et al.*, 2019), generating the intermediate plasmid pTA2478 (**Figure 4.55**).

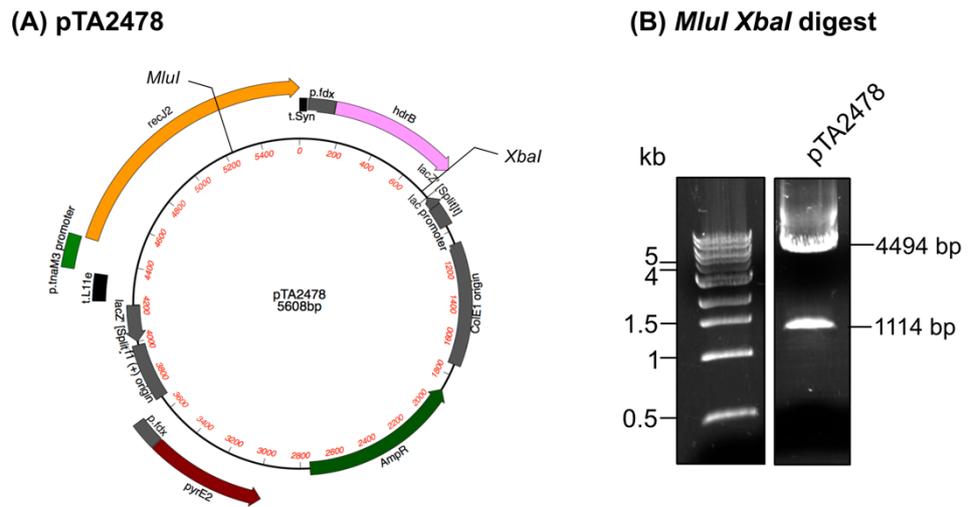


Figure 4.55: pTA2478. (A) Intermediate construct in creation of a *pyrE2* integration construct of inducible *recJ2*. (B) Restriction digest with *MluI* and *XbaI* gave bands of 4494 bp and 1114 bp, as predicted. Where the top band is running too high, this is likely due to DNA topology or overloading the gel lane with too much DNA.

The *p.tnaAM3-recJ2::hdrB* cassette was digested from pTA2478 using *BglII* and inserted into Δ *pyrE2* vector pGB68 at a compatible *BamHI* site. This generated the *pyrE2* replacement construct carrying the inducible *recJ2* expression cassette, pTA2498 (Figure 4.56).

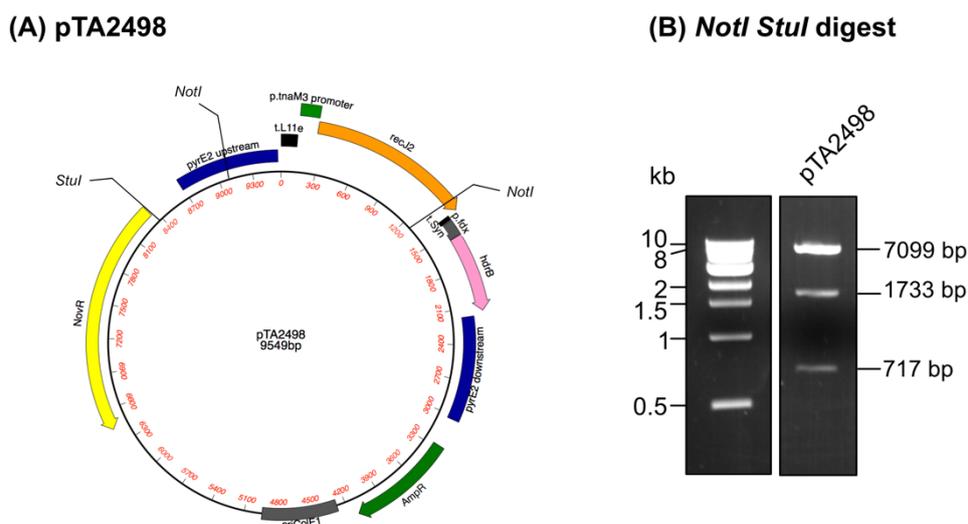
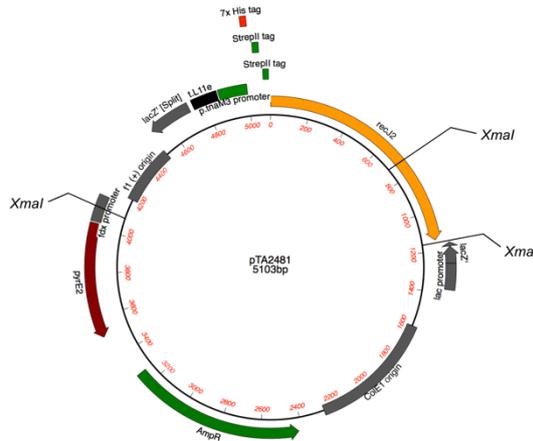


Figure 4.56: pTA2498. (A) Gene replacement construct to replace *pyrE2* with low activity tryptophan-inducible *p.tnaM3-recJ2::hdrB*. (B) Restriction digest with *NotI* and *StuI* gave bands of 7099 bp, 1733 bp and 717 bp, as predicted.

A similar cloning strategy was carried out alongside construction of pTA2498, utilising a 7xHis 2xStrepII-tagged version of the *p.tnaM3* inducible construct, pTA2096 (Lever, 2019). This would allow for the ectopically expressed *recJ2* to be purified using nickel- or streptavidin-affinity chromatography. Purification of RecJ2 would allow for identification of interacting partners via mass spectrometry and would therefore provide insight into its role within *H. volcanii*.

The genomic sequence of *recJ2* was amplified from pTA1905 using primers *RecJ2PciI_F* and *RecJ2Eco_R*, introducing novel 5' *PciI* and 3' *EcoRI* sites. The genomic sequence was digested with *PciI* and *EcoRI* and cloned in pTA2096 at corresponding *PciI/EcoRI* sites. This generated the intermediate plasmid pTA2481 (**Figure 4.57**). The tagged-inducible *recJ2* cassette (consisting of *p.tnaM3::7xHis 2xStrepII-recJ2*) was excised from pTA2481 using *XhoI* and *BamHI* and was cloned in pTA2498 in place of the *p.tnaM3-recJ2* cassette. This generated a *pyrE2* replacement construct containing 7xHis 2xStrepII-tagged inducible *recJ2*, pTA2499 (**Figure 4.58**).

(A) pTA2481



(B) *XmaI* digest

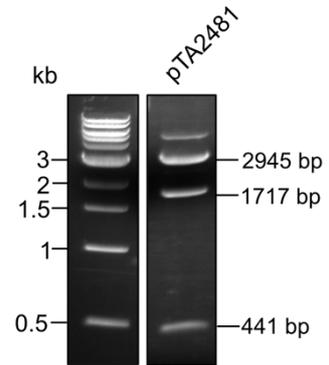
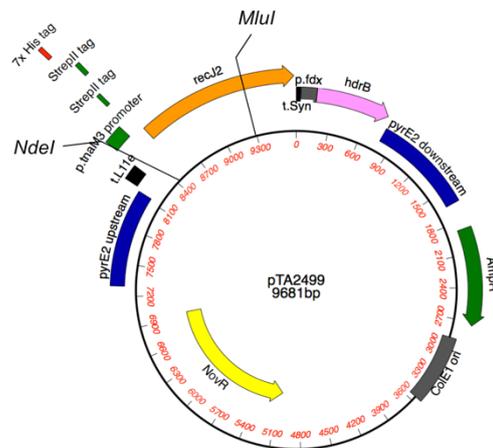


Figure 4.57: pTA2481. (A) Intermediate construct in generation of a *pyrE2* integration construct of 7xHis 2xStrepII inducible *recJ2*. (B) Restriction digest with *XmaI* gave bands of 2945 bp, 1717 bp and 441 bp, as predicted. The faint band seen at >5 kb is likely uncut plasmid; the plasmid was confirmed by sequencing.

(A) pTA2499



(B) *NdeI MluI* digest

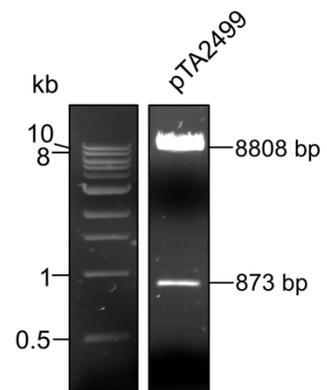


Figure 4.58: pTA2499. (A) Gene replacement construct to replace *pyrE2* with 7xHis 2xStrepII-tagged low activity tryptophan-inducible *p.tnaM3-recJ2::hdrB*. (B) Restriction digest with *NdeI* and *MluI* gave bands of 8808 bp and 873 bp, as predicted. Where the band at 8.8 kb runs high, this is likely due to DNA topology; the correct size of the smaller band suggests plasmid construction was successful. This was confirmed by sequencing.

Cloning of $\Delta recJ2::leuB$ construct

For the correct induction and application of the *p.tnaM3* promoter, the background strain needs to be capable of tryptophan synthesis; this will ensure in the absence of tryptophan, any growth defect is due to a lack of induction of the gene under *p.tnaM3*. As *recJ2* is predicted to be a hard-to-delete gene, the *recJ2* deletion construct should include a selectable marker to increase chances of a successful deletion.

Commonly used marker gene, *hdrB*, encodes for thymidine synthesis and allows for direct selection. This gene was included in the *p.tnaM3* constructs to allow selection for correct integration of the inducible cassette at the *pyrE2* locus, meaning it cannot be further utilised for the *recJ2* deletion construct. The leucine biosynthesis gene *leuB* was chosen as a selectable marker for the deletion of *recJ2* at its native locus, as it can be excluded from minimal media.

The available *recJ2* deletion construct pTA1951 (**Figure 4.59**) (Lever, 2019) does not carry a selection marker within the up- and down-stream genomic sequences, and thus *leuB* could be directly added within these sequences by restriction digest. Due to the location of *recJ2* within an operon, a promoterless version of *leuB* was integrated into the construct; the promoterless *leuB* will be transcribed using the promoter common to genes in the *recJ2* operon. Promoterless *leuB* was amplified from existing genomic clone pTA44 (Allers *et al.*, 2004) using *NdeI* and *BamHI* and inserted into pTA1951 at compatible *NdeI* and *BglIII* sites. This generated the $\Delta recJ2::leuB$ construct pTA2484 (**Figure 4.60**).

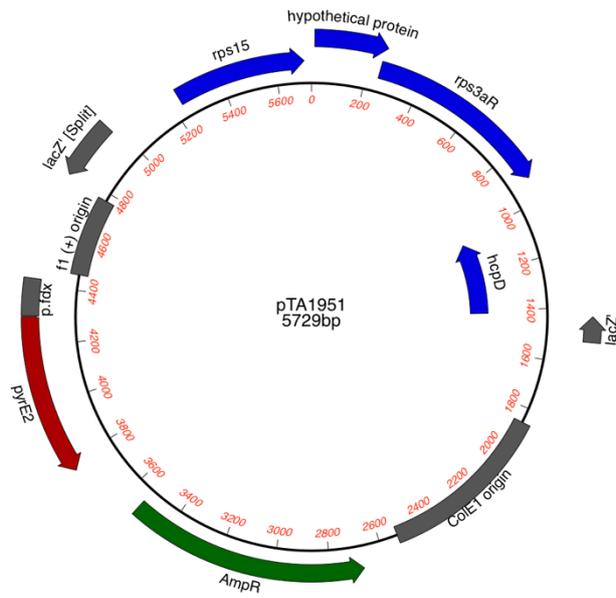
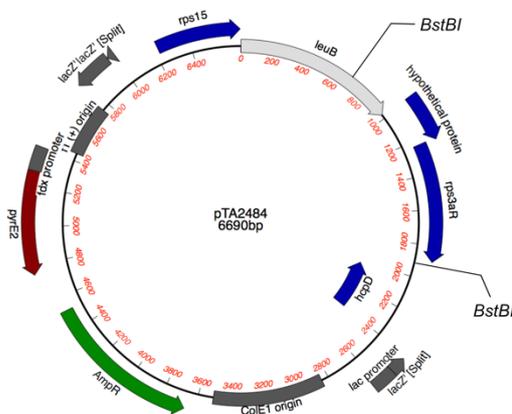


Figure 4.59: pTA1951. Deletion construct for *recJ2*. Upstream and downstream sequences meet at location 0. Generated by Rebecca Lever (Lever, 2019).

(A) pTA2484



(B) *Bst*BI digest

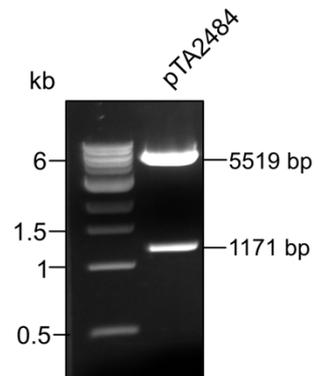


Figure 4.60: pTA2484. (A) Deletion construct for *recJ2* with *leuB* selection marker. (B) Restriction digest with *Bst*BI gave bands of 5519 bp and 1171 bp, as predicted.

Strain construction

Replacement of *pyrE2* with *p.tnaM3-recJ2*

Strain H730 ($\Delta leuB \Delta hdrB$) was transformed with pTA2498 or pTA2499 to give rise to pop-in strains H5064 and H5065 respectively. The transformation and subsequent pop-out were carried out in the absence of tryptophan to ensure expression of *recJ2* from the integrated construct was repressed, preventing any potential toxic effects of RecJ2 overexpression. Pop-out candidates were screened for 5-FOA resistance before being confirmed by Southern blot (**Figure 4.61**). H5064 gave rise to $\Delta pyrE2::p.tnaM3-recJ2::hdrB$ strain H5081. H5065 gave rise to $\Delta pyrE2::p.tnaM3-7xHis$ 2xStrepII -*recJ2::hdrB* strain H5082.

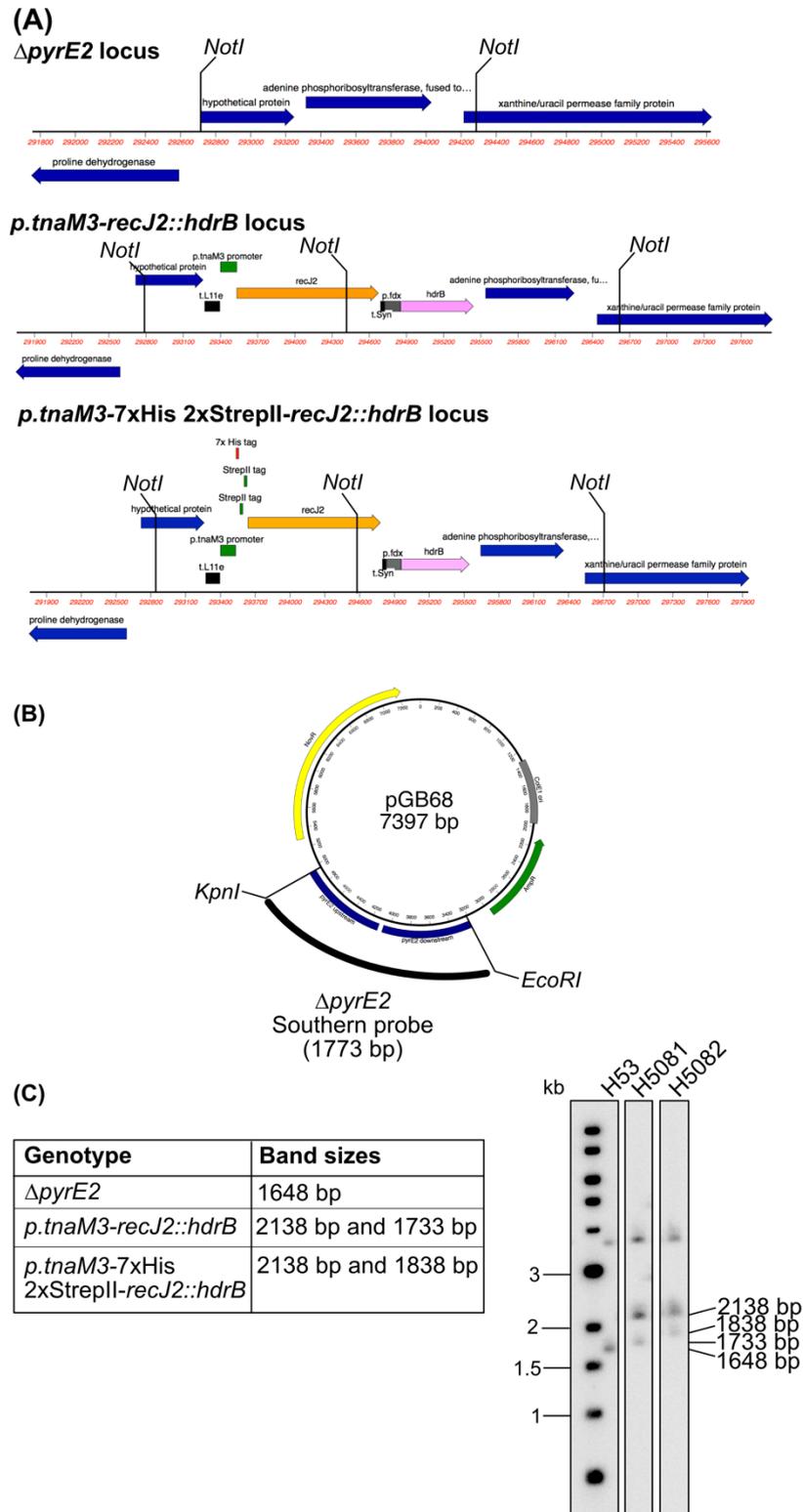


Figure 4.61: Gene replacement of *pyrE2* with *p.tnaM3-recJ2::hdrB*. (A) Genome regions showing digest sites for *NotI* digestion. (B) 1773 bp Δ pyrE2 Southern probe consisting of *KpnI-EcoRI* fragment of pGB68. (C) Southern blot confirming strain H5081 has integrated *p.tnaM3-recJ2::hdrB* and strain H5082 has integrated *p.tnaM3-7xHis 2xStrepII-recJ2::hdrB*. The band seen for all at ~3.5 kb is likely non-specific binding of the probe.

Deletion of wild type RecJ2

Inducible *recJ2* strains H5081 and H5082 were transformed with pTA2484 to generate pop-in strains H5131 and H5132 respectively. Pop-outs were performed in minimal media lacking leucine to select for the *leuB* marker within the deletion construct and plated on 5-FOA agar lacking leucine. Candidates were primarily screened by PCR using two reactions: the first PCR (PCR 1) confirmed the presence of the *recJ2* gene at the wild type locus, while the second (PCR 2) checked for the integration of the *leuB* marker within the deletion construct (**Figure 4.62 A**). This allowed for screening of merodiploid *leu+* candidates. Of 120 candidates screened, 12 gave no band for PCR 1 and the correct band for PCR 2. These candidates were taken forward for further screening by restriction digest and Southern blotting, however all were identified to be merodiploid (**Figure 4.62 B**).

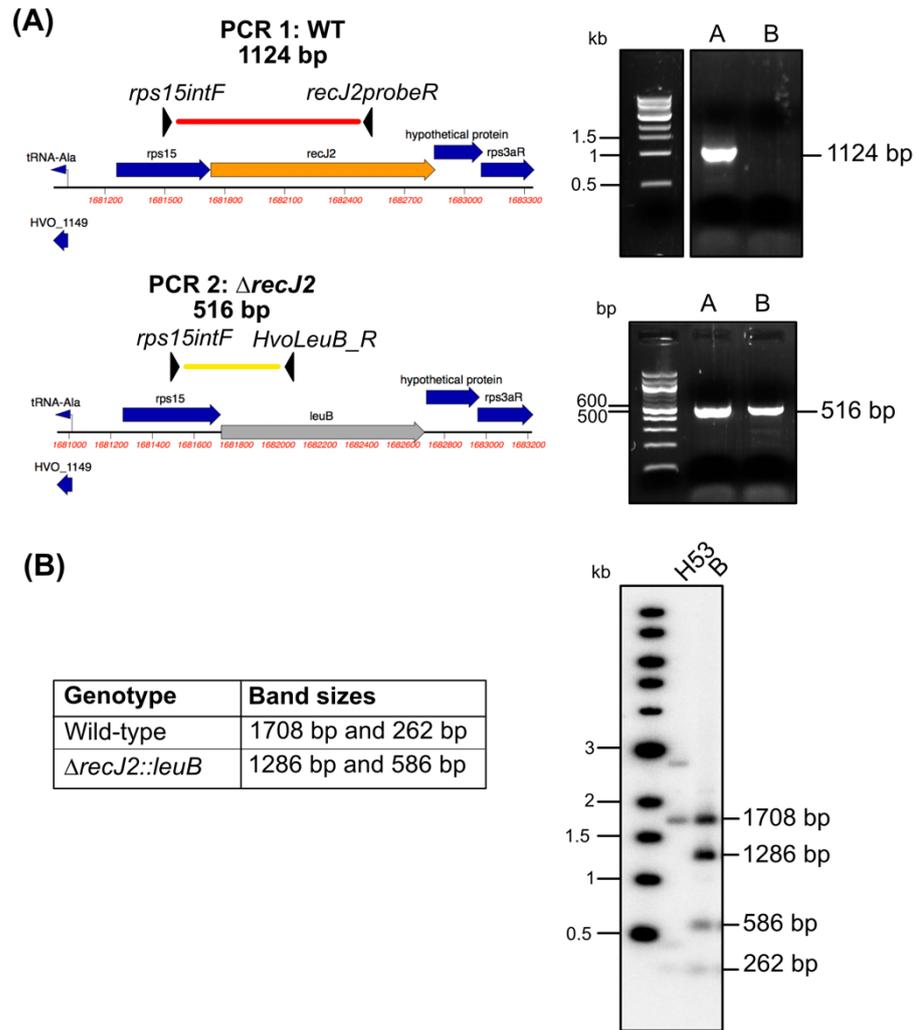


Figure 4.62: Screening of ectopic *p.tnaM3-recJ2* Δ *recJ2::leuB* candidates. (A) Genome regions showing diagnostic PCRs used for identifying deletion candidates. PCR 1, using primers *rps15intF* and *recJ2probeR* gives a product of 1124 bp if the wild type gene is present at the *recJ2* locus. PCR 2, using primers *rps15intF* and *HvoLeuB_R* gives a band of 516 bp if the deletion construct is present at the *recJ2* locus. Results show candidate A is merodiploid (gives bands for both WT and Δ *recJ2::leuB*), while candidate B only gives a band for Δ *recJ2::leuB* and thus was further screened by Southern blot. (B) Southern blot of candidate B shows predicted bands for both wild-type and Δ *recJ2::leuB*, confirming its status as merodiploid. Genomic DNA was digested with *Eco*NI and *Bsp*HI and probed with pTA2484 *Bst*EII-*Sty*I fragment (data not shown). The band for H53 at ~3 kb is due to the strain H53 retaining *leuB* (part of the probe).

The transcript expression level of native *recJ2*, along with other genes in its co-transcribed operon, is known to be relatively high; RNA-seq data from strain H53 (unpublished data) shows expression of *recJ2* at a level comparable to highly transcribed recombinase *radA*. Therefore, one hypothesis

to explain why only merodiploid candidates were seen is that the ectopic expression of *recJ2* from the *pyrE2* locus is not strong enough to fully compensate for deletion of *recJ2* from the wild type locus, due to the lower activity of the mutant *p.tnaM3* promoter. The pop-outs with H5131 and H5132 were therefore repeated in the presence of increased levels of tryptophan to increase the induction level of *recJ2*, however resulted again in only merodiploid products. Such a phenomenon has been previously observed for highly expressed gene, *radA*, where it cannot be induced to a sufficient level for viability under the weak *p.tnaM3* promoter and requires induction from the full-strength *p.tnaA* promoter (Thorsten Allers, personal communication). Therefore, *recJ2* was placed under the control of the stronger promoter *p.tnaA* at the *pyrE2* locus, to determine if increased ectopic expression allows for deletion of *recJ2* from its native locus.

Cloning of $\Delta pyrE2$ inducible gene cassettes for future use

Due to the ease of creating the *pyrE2* replacement inducible *p.tnaM3-recJ2::hdrB* strain (100% of candidates screened had correctly integrated the inducible *recJ2* cassette at the *pyrE2* locus), the additional *hdrB* selection for the integration at *pyrE2* was abandoned in all future attempts. Absence of *hdrB* selection would allow a greater number of background strains to be utilised for testing in the future (strains need only be $\Delta leuB$, as opposed to $\Delta leuB \Delta hdrB$).

Primarily, empty cassettes containing *pyrE2* US and DS sequences with the inserted inducible promoters (either *p.tnaA* or *p.tnaM3*) and associated terminator sequences were cloned, providing a resource for future inducible alleles for integration at *pyrE2*. To generate the empty constructs, PCR was used to amplify the *t.11e-p.tnaA/M3-t.syn* cassettes from existing plasmids pTA1369 and pTA1451, respectively (Braun et al., 2019), using primers *RBDX1* and *tsynBglR* (**Figure 4.63 A**). The PCR product was digested with *BglIII* and inserted into $\Delta pyrE2$ construct pGB68 at its compatible *BamHI* site. This gave generated $\Delta pyrE2$ gene replacement constructs containing the *p.tnaA* cassette, pTA2553 (**Figure 4.63 B**), or the *p.tnaAM3* cassette, pTA2546 (**Figure 4.63 C**).

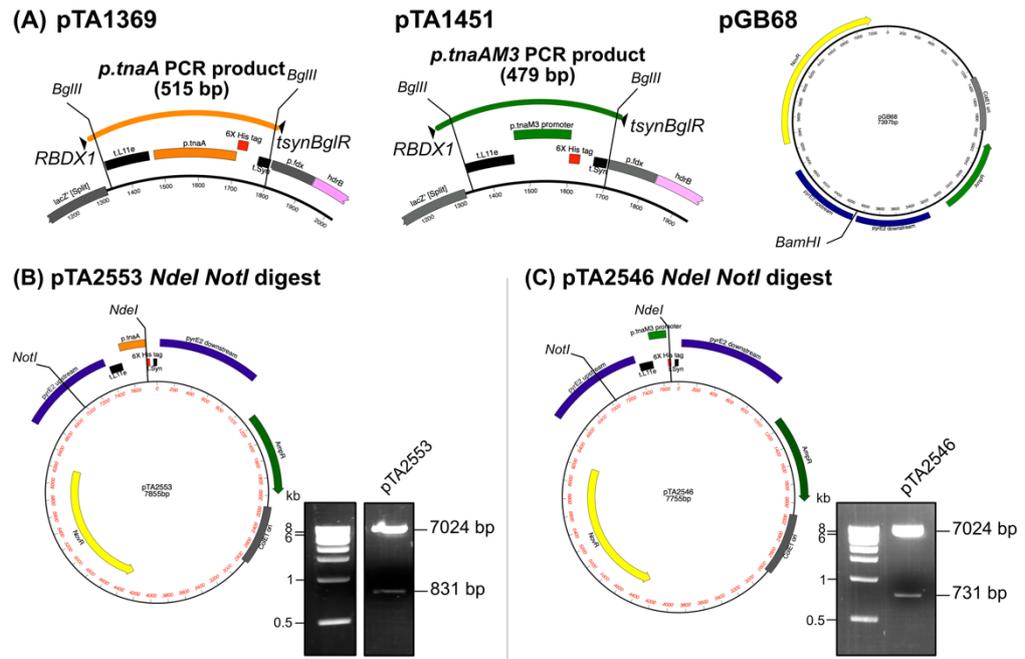


Figure 4.63: Cloning of Δ *pyrE2* replacement constructs introducing tryptophan-inducible cassettes. (A) PCR using primers *RBDX1* and *tsynBglR* on plasmids pTA1369 and pTA1451 isolate tryptophan-inducible promoters (and associated terminator sequences) *p.tnaA* and *p.tnaM3* respectively. Digestion of the PCR products with *BglIII* allows insertion of the cassette into Δ *pyrE2* vector pGB68. (B) Δ *pyrE2* *p.tnaA* replacement construct pTA2553. Digestion with *NdeI* and *NotI* gave bands of 7024 bp and 831 bp as expected. (C) Δ *pyrE2* *p.tnaM3* replacement construct pTA2546. Digestion with *NdeI* and *NotI* gave bands of 7024 bp and 731 bp as expected.

Cloning of p.tnaA-recJ2 cassette for integration at the pyrE2 locus

To generate a Δ *pyrE2* replacement construct containing *recJ2* under the full-strength promoter *p.tnaA*, the coding sequence of *recJ2* was amplified by PCR. Primers *recJ2fwdNde* and *recJ2revDSBam* amplified the 1147 bp product from genomic clone pTA1905, introducing novel 5' *NdeI* and 3' *BamHI* sites. Digestion of the *recJ2* PCR with *NdeI* and *BamHI* allowed its insertion into Δ *pyrE2* inducible construct pTA2553 at *NdeI* and *BamHI* sites. This generated the *pyrE2* replacement construct containing *p.tnaA-recJ2*, pTA2561 (Figure 4.64).

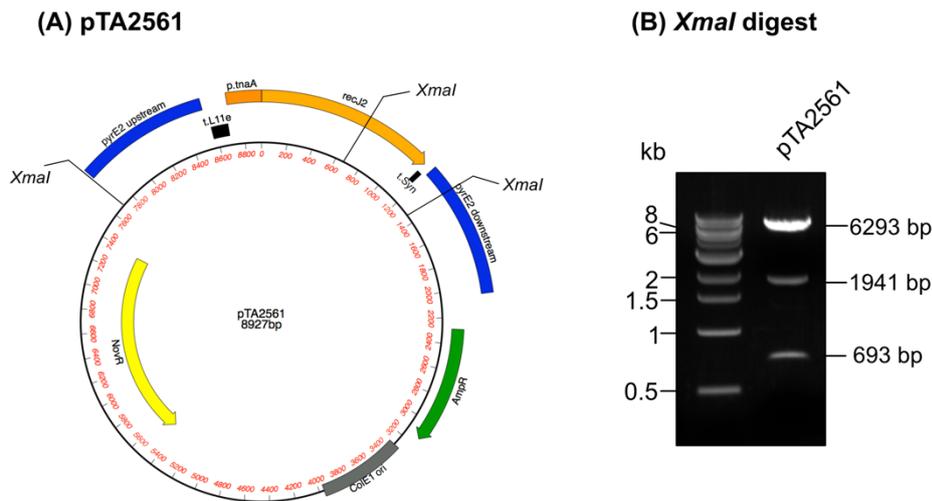


Figure 4.64: pTA2561. (A) Gene replacement construct to replace *pyrE2* with full activity tryptophan-inducible *p.tnaA-recJ2*. (B) Restriction digest with *XmaI* gave bands of 6293 bp, 1941 bp and 693 bp, as predicted.

To generate a tandem 7xHis 2xStrepII-tagged version of the full strength *p.tnaA* construct, the tags and 5' end of *recJ2* was digested from pTA2499 and replaced the corresponding *p.tnaA*-5' *recJ2* fragment of pTA2561, to generate *pyrE2* replacement construct *p.tnaA*-7xHis 2xStrepII pTA2563 (Figure 4.65).

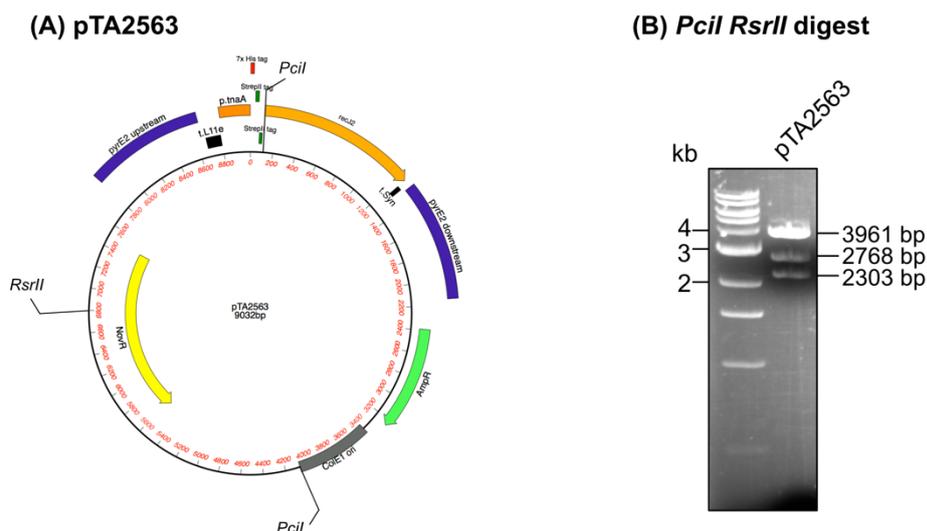


Figure 4.65: pTA2563. (A) Gene replacement construct to replace *pyrE2* with full activity tryptophan-inducible *p.tnaA*-7xHis 2xStrepII-*recJ2*. (B) Restriction digest with *PciI* and *RsrII* gave bands of 3961 bp, 2768 bp and 2303 bp, as predicted.

Strain construction

Replacement of *pyrE2* with *p.tnaA-recJ2*

Strain H37 ($\Delta pyrE2 \Delta leuB$) was transformed with pTA2561 and pTA2563 to generate pop-in strains H5217 and H5218, respectively. The transformation was carried out in the absence of tryptophan to ensure expression of *recJ2* from the integrated construct was minimal, preventing any potential toxic effects of RecJ2 overexpression. Pop-out candidates were screened for 5-FOA resistance before being confirmed by restriction digest and Southern blot (**Figure 4.66**). H5217 gave rise to $\Delta pyrE2::p.tnaA-recJ2$ strain H5235 while H5218 gave rise to $\Delta pyrE2::p.tnaA-7xHis\ 2xStrepII-recJ2$ strain H5237.

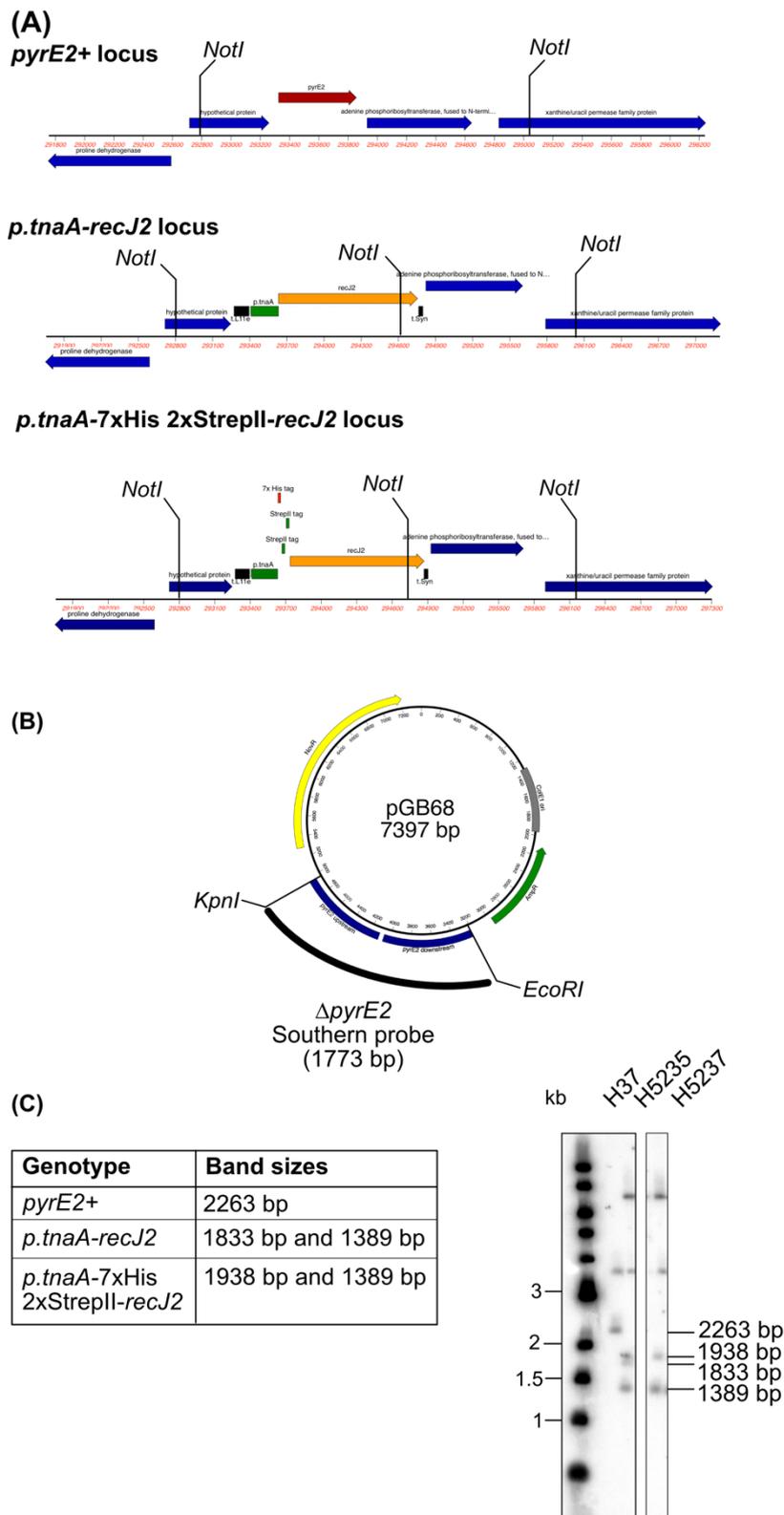


Figure 4.66: Gene replacement of *pyrE2* with *p.tnaA-recJ2*. (A) Genome regions showing digest sites for *NotI* digestion. (B) 1773 bp Δ *pyrE2* Southern probe consisting of *KpnI-EcoRI* fragment of pGB68. (C) Southern blot confirming strain H5235 has integrated *p.tnaA-recJ2* and strain H5237 has integrated *p.tnaA-7xHis 2xStrepII-recJ2*. The bands seen at ~3.5 kb and 7 kb are likely due to non-specific binding of the probe.

Deletion of wild type RecJ2

Inducible *recJ2* strains H5235 and H5237 were transformed with pTA2484 to generate pop-in strains H5250 and H5252 respectively. Pop-outs were performed in minimal media lacking leucine to select for the *leuB* marker within the deletion construct and plated on 5-FOA agar lacking leucine. As previously, candidates were primarily screened by PCR before deletion candidates were further screened by Southern blot (**Figure 4.67**).

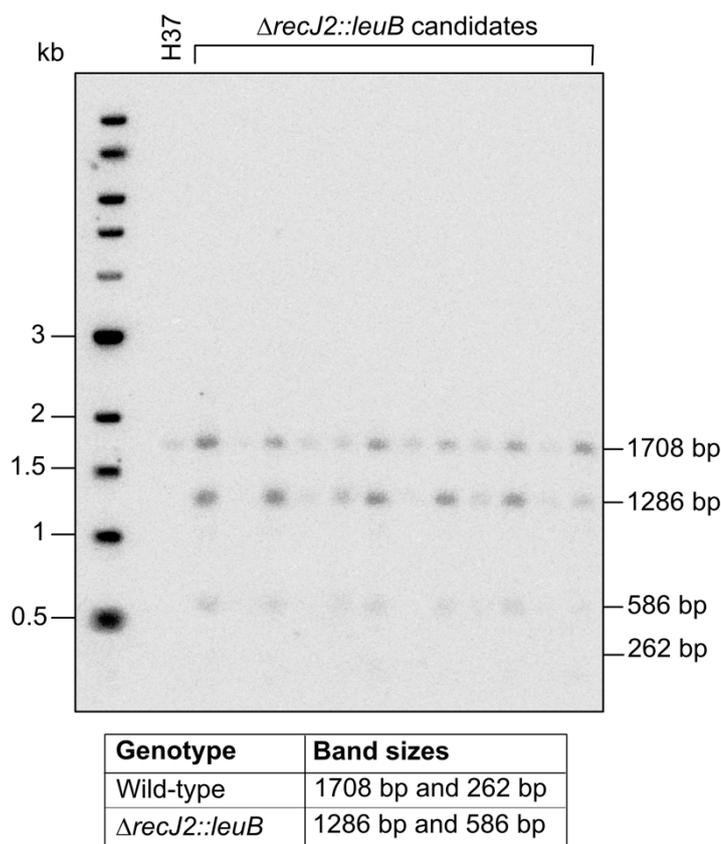


Figure 4.67: Screening of ectopic *p.tnaA-recJ2* $\Delta recJ2::leuB$ candidates. Southern blot of $\Delta recJ2::leuB$ candidates shows predicted bands for both wild-type and $\Delta recJ2::leuB$, confirming its status as merodiploid. Genomic DNA was digested with *Eco*NI and *Bsp*HI and probed with pTA2484 *Bst*EII-*Sty*I fragment (data not shown).

All strains screened by Southern blot were merodiploid, even with the stronger promoter *p.tnaA* being induced during pop-out. Pop-out events were repeated at varying levels of tryptophan (from 1x (0.25 mM) to 8x (2 mM)), to ensure ectopic *recJ2* expression was high enough to allow for deletion, but still all candidates were merodiploid.

To confirm *recJ2* was being overexpressed, as predicted, the low-activity promoter strain H5081 ($\Delta pyrE2::7xHis$ 2xStrepII *p.tnaM3-recJ2*) and

the full-activity promoter strain H5237 (Δ *pyrE2*::7xHis 2xStrepII *p.tnaA-recJ2*) were grown for two overnights in Hv-Cas +Ura media with varying levels of tryptophan (ranging from 0 mM to 2 mM). Cells were spun down, resuspended in water and DNase treated before being resuspended in protein loading buffer. Samples were then run on an SDS-PAGE gel to assess whether induction using tryptophan was in-fact increasing the amount of RecJ2 protein within cells (**Figure 4.68**).

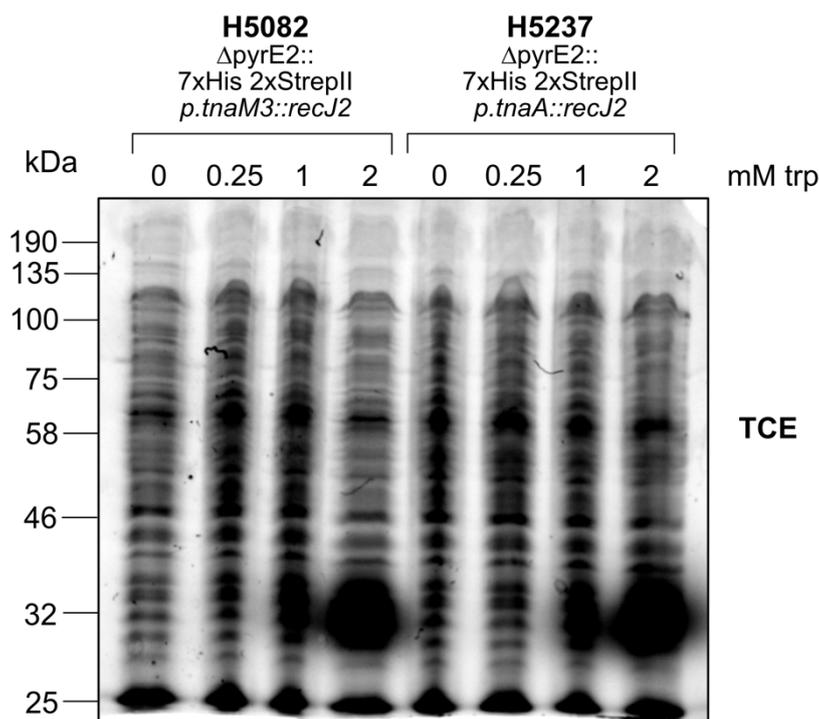


Figure 4.68: SDS-PAGE gel to assess extent of overexpression of RecJ2 in inducible strains H5082 (*p.tnaM3* promoter) and H5237 (*p.tnaA* promoter). Gel contains 0.5% TCE and was imaged following one-minute UV exposure. A band is seen to increase in intensity with tryptophan concentration at ~35 kDa. Size markers based on previously run comparable 10% gels.

Staining of total loaded protein by TCE shows a clear overexpression of a protein at a low molecular weight; this likely represents RecJ2, which has a predicted molecular weight of 37.4 kDa. Western blotting was attempted against these lysates with an anti-6xHis antibody; however, no signal was seen.

4.3.4.1 Purification of RecJ2

While it proved unsuccessful to delete wild type *recJ2*, it was possible to induce overexpression of RecJ2 while still expressing the protein from the wild type locus. This should allow utilisation of the 7xHis 2xStrepII tags within strain H5237 to purify overexpressed RecJ2 and its interacting partners via affinity purification. By identifying its interactors, this could shed light on the function of RecJ2 in *H. volcanii*. Initially, H5237 was assayed for growth rate compared to ensure overexpression of RecJ2 is not having a detrimental effect on the cell and its ability to divide. Strain H5237 was grown for two consecutive overnights in Hv-Cas +Ura broth containing varying amounts of tryptophan (up to 2 mM), ensuring on day three that actively dividing cells were used for the assay. Cells were diluted in fresh Hv-Cas +Ura (+varying [Trp] where required) broth and arrayed in a 96-well plate. Optical density (OD; A₆₀₀) was semi-continuously measured, allowing the plotting of growth curves for each concentration tested (**Figure 4.69**).

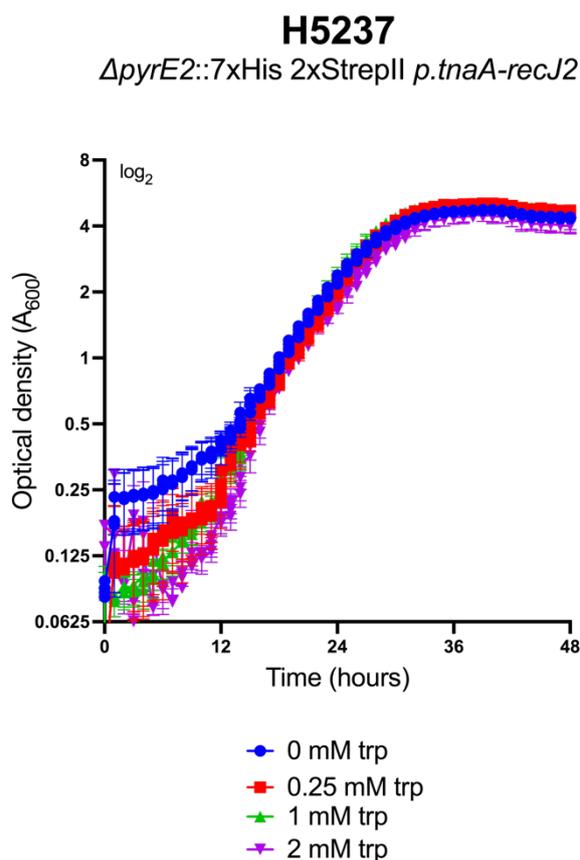


Figure 4.69: Exponential growth rate of $\Delta pyrE2::7xHis$ 2xStrepII *p.tnaA-recJ2* strain H5237 at varying levels of tryptophan induction. Strain H5237 was grown in Hv-Cas +Ura broth containing various concentrations of tryptophan (trp) for two consecutive overnights before being diluted and plated. All dilutions ($n=2$) were incubated on the same 96-well plate and measured simultaneously for optical density (A_{600}) using an Epoch2 Microplate Spectrophotometer (BioTek). Error bars represent standard error of the mean (SEM).

Generally, the induction of RecJ2 expression does not affect the growth rate, with the exponential rate of growth for H5237 being comparable at all concentrations tested. This suggested the strain remains viable and a pulldown utilising induced RecJ2 should be viable.

Isolation of proteins by Strep-tactin affinity chromatography from strain H5237 was carried out, using the culture growth in Hv-YPC with tryptophan added to a final concentration of 2 mM (to ensure a high proportion of tagged RecJ2 is present in the cell). Pulldown products were run on an SDS-PAGE gel (Figure 4.70).

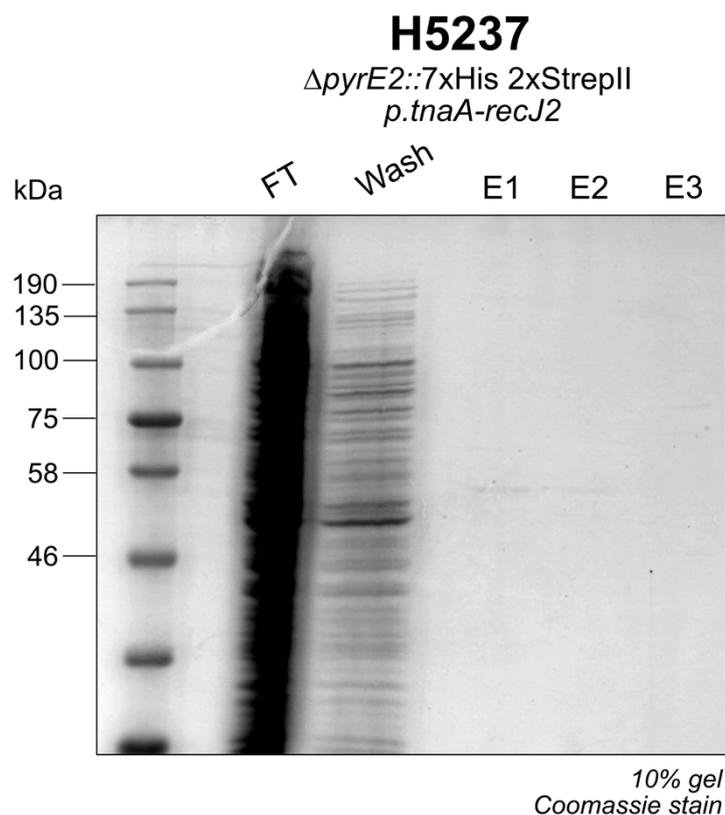


Figure 4.70: Purification of N-terminally 7xHis 2xStrepII-tagged RecJ2 via Streptavidin affinity from strain H5237. Total protein-stained gel of StrepTactin purification from strain H5237 (*ΔpyrE2::7xHis 2xStrepII p.tnaA-RecJ2*). FT refers to flowthrough from the column, wash refers to the washes following lysate loading and E refers to stepwise elutions with 5 mM D-desthiobiotin. Faint bands can be seen in E1 and E2 at ~50 kDa.

A faint band is visible at ~50 kDa, however the staining is very weak and thus samples were precipitated with TCA to aid visualisation (**Figure 4.71 A**). Western blotting was also performed on the pulldown samples using an anti-6xHis antibody to confirm the presence of 7xHis 2xStrepII-RecJ2 (**Figure 4.71 B**).

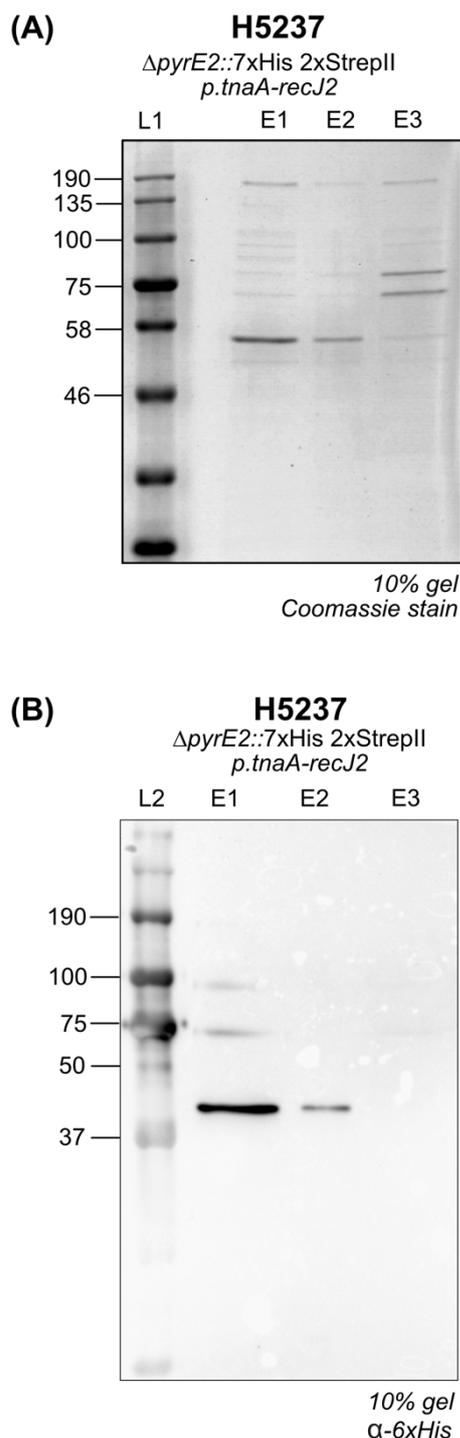


Figure 4.71: Purification of N-terminally 7xHis 2xStrepII-tagged RecJ2 via Streptavidin affinity from strain H5237 (A) Total protein-stained gel of StrepTactin purification from strain H5237 (*7xHis 2xStrepII-RecJ2*) precipitated using TCA. E refers to stepwise elutions with 5 mM D-desthiobiotin. Discrete bands are seen in lane E1 at ~55 kDa and ~190 kDa. Discrete bands are seen in lane E3 at ~70 kDa and ~80 kDa. (B) An anti-6xHis antibody detects a strong band at ~40 kDa, with weak bands seen at ~75 kDa and ~100 kDa. L1: Blue Prestained Protein Standard, Broad Range (11-190 kDa). L2: BioRad Precision Plus Protein Kaleidoscope.

Discrete bands are seen in lane E1 at ~55 kDa and ~190 kDa. Discrete bands are seen in lane E3 at ~70 kDa and ~80 kDa. There are also a number of less discrete bands present between ~75 kDa and ~120 kDa, most prominently observed in lane E1. Western blotting shows a strong discrete band in lanes E1 and E2 at ~40 kDa (RecJ2 predicted size: 37.4 kDa), as well as weaker bands at ~75 kDa and ~100 kDa.

To identify the pulldown products associating with RecJ2, bands were sent for analysis by mass spectrometry. **Figure 4.72** is annotated to show which bands were sent for analysis, while **Table 4.3** lists the key peptides identified for each band by mass spectrometry.

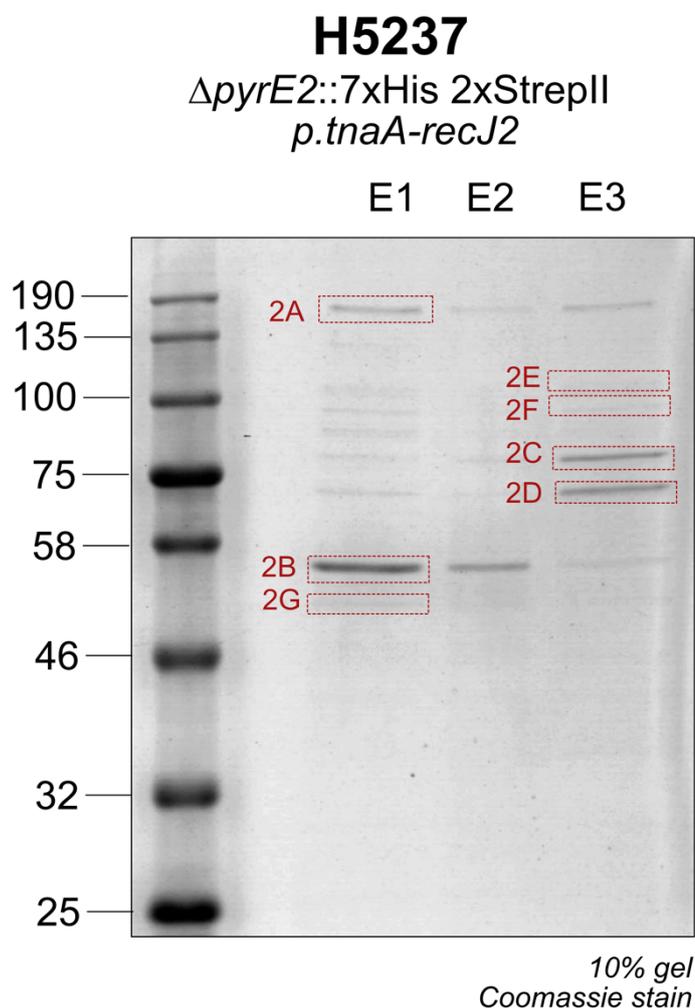


Figure 4.72: Affinity purification products from H5237 submitted for mass spectrometry. Bands highlighted in red boxes were submitted for analysis by mass spectrometry. These represent the clearly discrete bands on the gel, while banding at ~75-100 kDa was not clear to cut and likely would have high background.

Table 4.3: Summary of key peptides identified by mass spectrometry of strain H5237. All proteins listed below indicated identity or extensive homology ($p < 0.05$); they are listed in order of score (highest hits at the top). Peptides commonly identified in pulldowns or in control strain H164 were excluded. Proteins coloured red represent those with a strong link to translation (tRNA machinery and amino acid biosynthesis, part of the translation machinery). MW, molecular weight.

Band	2A	2B	2C	2D	2E	2F	2G
Approximate MW (kDa)	190	55	75	70	120	100	50
Proteins identified	RecJ2 LeuS IleS RpoH	RecJ2 Rnj Rnr ArgG HisS AroB MetB1 Tef2 MetE1 GatA HemL RtcB Tef1a1 AspC1	GyrB RpoB1 GltS	AspS SerS GatB SerA3 PyrG ProS RecJ2 PheS LysS RpoA1 Ref2	RecJ4 Tef2 GatE RecJ2 Top6B GlyS RpoA1 Hel308	Cdc48a Cdc48b Mcm RecJ2 RecJ4 Tef2 Rpap1 NrdJ	Tef1a1 RecJ2 GINS CetZ1 RpoA2 ThrC3 Top6A ArgD FtsZ1 SufS Tef1a2 Arf1 Rnj Srp54

See **Appendix 1** for complete table listing all proteins identified and associated MASCOT scores.

The results of the mass spectrometry confirmed successful purification of RecJ2, with peptides corresponding to its sequence being found in six of the seven bands isolated.

The top hit for band 2E (~120 kDa) was fellow RecJ protein, RecJ4. Interaction between RecJ2 and RecJ4 has not previously been shown and was somewhat unexpected, since RecJ4 is predicted to have a HAN-like role, while RecJ2 falls into a GAN-like grouping. In contrast, RecJ4 has previously been shown to interact with Hel308 (Gamble-Milner, 2016, Lever, 2019), and therefore it was unsurprising to identify peptides of Hel308 in the band identified as RecJ4. Purification of RecJ2 and RecJ4 was also seen within band 2F.

Band 2E also contained peptides of type II topoisomerase, Topo6B, known to relieve torsional stress during canonical replication. Fellow type II topoisomerases, GyrB and Top6A, were also purified within bands 2C and 2G respectively. Co-purification of topoisomerases and RecJ2 suggests localisation of RecJ2 at the replication fork. This is further supported by isolation of proteins known to be located at the replication fork; MCM helicase

and Cdc48a (band 2F). The identification of GINS in band 2G also provides evidence for localisation of RecJ2 at the replication fork.

While this may support a hypothesis of RecJ2 acting as GAN, it is important to note that, alongside co-purifying with replication proteins, RecJ2 was isolated in combination with a large number of proteins acting in translation (highlighted red in the table). The co-purification data alone cannot provide evidence for the role of RecJ2, however the large number of translation proteins identified could suggest an alternate role in translation (perhaps defined by its localisation within a co-transcribed operon of genes acting in said process). The operon containing RecJ2 encodes a member of the KEOPS complex (Kinase, Endopeptidase and Other Proteins of small Size; HVO_1146), along with other critical translation proteins (ribosomal proteins, tRNAs). The KEOPS complex is involved in tRNA modification and is conserved in eukaryotes with some bacterial and archaeal species carrying homologues (Srinivasan *et al.*, 2011, Wan *et al.*, 2016, Naor *et al.*, 2012); should RecJ2 associate with the KEOPS complex, this could partially explain the large number of translation-related proteins purified with RecJ2.

4.3.5 Identification of RecJ protein-protein interactions using affinity purification

Cloning of tagged RecJ constructs*5' 6xHis 3' StrepII tagged constructs*

Constructs for integration of 5' 6xHis and 3' StrepII tags were constructed prior to this project by Rebecca Lever (Lever, 2019), listed below in **Table 4.4**. The map of plasmid pTA2090 (6xHis-*recJ1*-StrepII) is shown as an example (**Figure 4.73**).

Table 4.4: Existing gene replacement constructs for N-terminal Histidine tag and C-terminal Streptavidin tag integration. All constructs listed constructed by Rebecca Lever (Lever, 2019). * represents a construct not yet appropriate for integration onto the chromosome.

Plasmid number	Description
pTA2090	6xHis- <i>recJ1</i> -StrepII gene replacement construct
pTA2091	6xHis- <i>recJ2</i> -StrepII gene replacement construct
pTA2084*	6xHis- <i>recJ3</i> -StrepII gene replacement construct
pTA2095	6xHis- <i>recJ4</i> -StrepII gene replacement construct

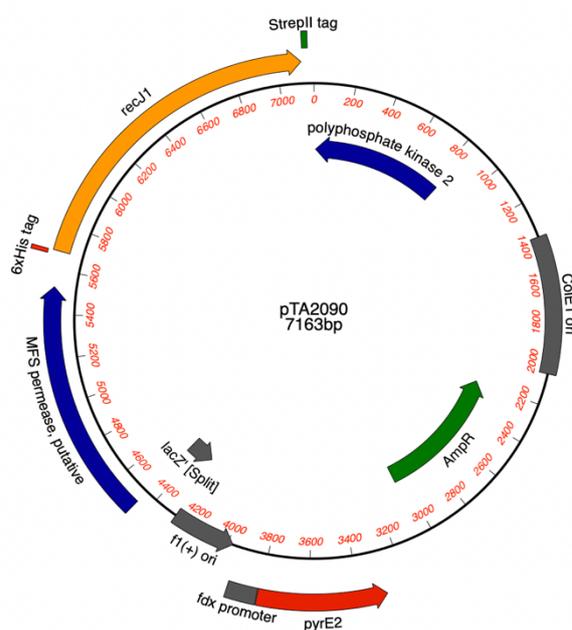


Figure 4.73: pTA2090. 6xHis-*recJ1*-StrepII gene replacement construct (Lever, 2019).

RecJ3 gene replacement construct pTA2084 contains 6xHis-*recJ3*-StrepII but does not carry the surrounding up- and down-stream sequences required for integration onto the genome. To address this, the downstream sequence of *recJ3* was digested from pTA1927 ($\Delta recJ3::trpA+$) using *Bgl*III-*Not*I and inserted into pTA2084 at compatible *Bam*HI-*Not*I sites. This gave rise to intermediate plasmid pTA2130, which still lacked the downstream genome sequence. PCR of pTA1913, the *recJ3* genomic clone, with primers *RBDX1* and *delrecJ3NsiR* isolated the upstream sequence of *recJ3*, while integrating a 3' *Nsi*I site into the product. Digestion with the existing *Pst*I site, found within the 5' end of the product and *Nsi*I allowed the insertion of *recJ3* upstream sequence into pTA2130 at compatible *Nsi*I site. This gave rise to the completed 6xHis-*recJ3*-StrepII gene replacement construct, pTA2564 (Figure 4.74).

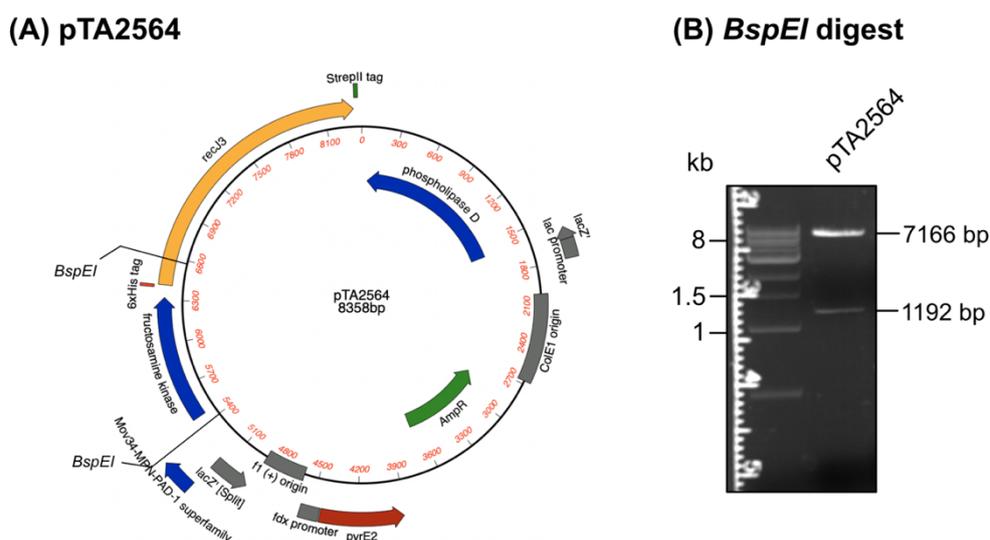


Figure 4.74: Construction of pTA2564. (A) Gene replacement construct pTA2564, introducing N-terminal 6xHis and C-terminal StrepII tags to RecJ3. (B) *Bsp*EI digest shows bands at 7166 bp and 1192 bp, as predicted.

Numerous proteins have a hidden C-terminus when folded or are unable to tolerate a tag at the C-terminus. Therefore, should it be possible to integrate tags into the genome both N- and C-terminally, the C-terminal streptavidin tag may be hidden. Not only does the streptavidin tag give a pulldown product with fewer contaminants than the histidine tag, the availability of only one type of tag limits what experiments can be done; for example, should a protein be isolated using streptavidin-affinity chromatography, the presence of both types of tag would allow verification for which band represents the tagged protein using a Western blot with an anti-His

tag antibody. For these reasons, tandem N-terminal tags, containing a single 7xHis tag coupled with two StrepII tags were engineered for tagging of all RecJ proteins.

Tandem 7xHis 2xStrepII-tagged RecJ proteins

A similar cloning strategy was utilised for tagging genes *recJ1-recJ4*, involving PCR and digest into pTA1771, 7xHis 2xStrepII cloning vector (Wardell *et al.*, 2017). This will be explained in detail for *recJ1*, with the same general cloning strategy applying for the remaining *recJ* genes, unless stated otherwise.

Tagged RecJ1 (HVO_0073)

The *recJ1* genomic clone pTA1912 (**Figure 4.75**) was used as a template to amplify the 1476 bp coding sequence of *recJ1* by PCR, using the primers *recJ1NcoI_F* and *recJ1Eco_R* (**Figure 4.76 A**). These primers introduced a *NcoI* and an *EcoRI* site respectively, allowing digestion and ligation of the *recJ1* coding sequence into the compatible *PciI* and *EcoRI* sites of 7xHis 2xStrepII tag cloning vector pTA1771 (constructed by Thorsten Allers, unpublished data; **Figure 4.76 B**).

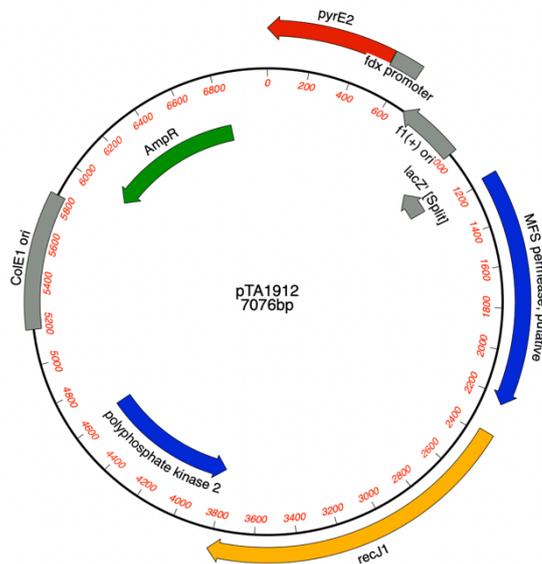
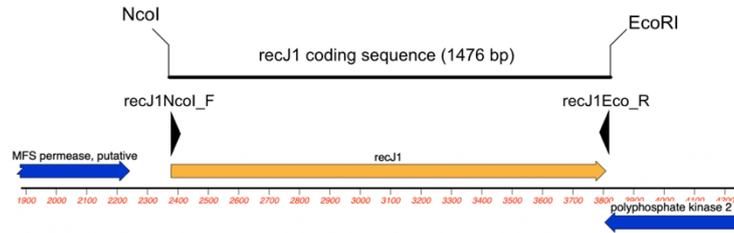
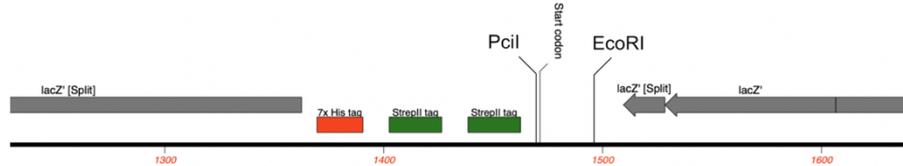


Figure 4.75: pTA1912. Genomic clone of *recJ1* (Lever, 2019).

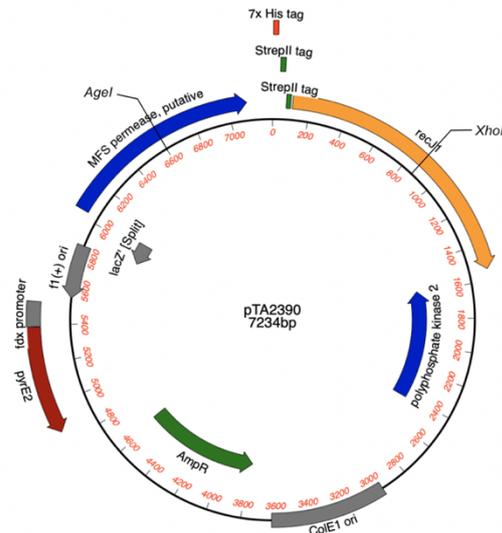
(A) PCR amplified *recJ1*



(B) pTA1771 restriction sites



(C) pTA2390



(D) *AgeI XhoI* digest

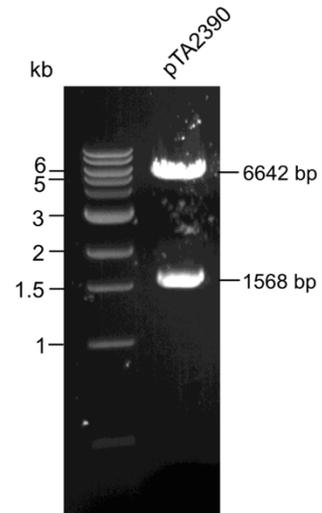


Figure 4.76: Construction of pTA2390 (A) The *recJ1* coding sequence was isolated from its genomic clone pTA1912 using primers *recJ1NcoI_F* and *recJ1EcoI_R*, which introduced novel restriction sites. (B) 7xHis 2xStrepII (tandem tag) cloning vector pTA1771 showing restriction sites utilised for insertion of *ginS* genomic sequence. (C) Gene replacement construct pTA2390, introducing an N-terminal 7xHis 2xStrepII tag at the 5' end of RecJ1. (D) *AgeI XhoI* digest shows bands at 6642 bp and 1568 bp, as predicted.

This gave rise to intermediate plasmid pTA2387, containing tandem tagged *recJ1*. To allow for integration of the tag onto the chromosome, upstream and downstream sequences were added by restriction digest. Tandem tagged *recJ1* was cut from pTA2387 using *NdeI* and *BamHI* and was inserted into *RecJ1* deletion vector, pTA1924 (Lever, 2019) at its *BamHI* and *NdeI* sites. This gave rise to plasmid pTA2390, His₇ 2xStrepII-*recJ1* gene replacement construct (**Figure 4.76 C**).

Tagged *RecJ2* (HVO_1147)

The *recJ2* genomic clone pTA1905 was used as a template to amplify the 1166 bp coding sequence of *recJ2* by PCR, using the primers *recJ2PciI_F* and *recJ2Eco_R*. These primers introduced a *PciI* and an *EcoRI* site respectively, allowing digestion and ligation of the *recJ2* coding sequence into the compatible *PciI* and *EcoRI* sites of 7xHis 2xStrepII tag cloning vector pTA1771. This gave rise to intermediate plasmid pTA2385, containing tandem tagged *recJ2*. To allow for integration of the tag onto the chromosome, upstream and downstream sequences were added by restriction digest. Tandem tagged *recJ2* was cut from pTA2385 using *NdeI* and *BamHI* and was inserted into *RecJ2* deletion vector, pTA1951 (Lever, 2019) at its compatible *BglII* and *NdeI* sites. This gave rise to 7xHis 2xStrepII-*recJ2* gene replacement construct, pTA2392 (**Figure 4.77**).

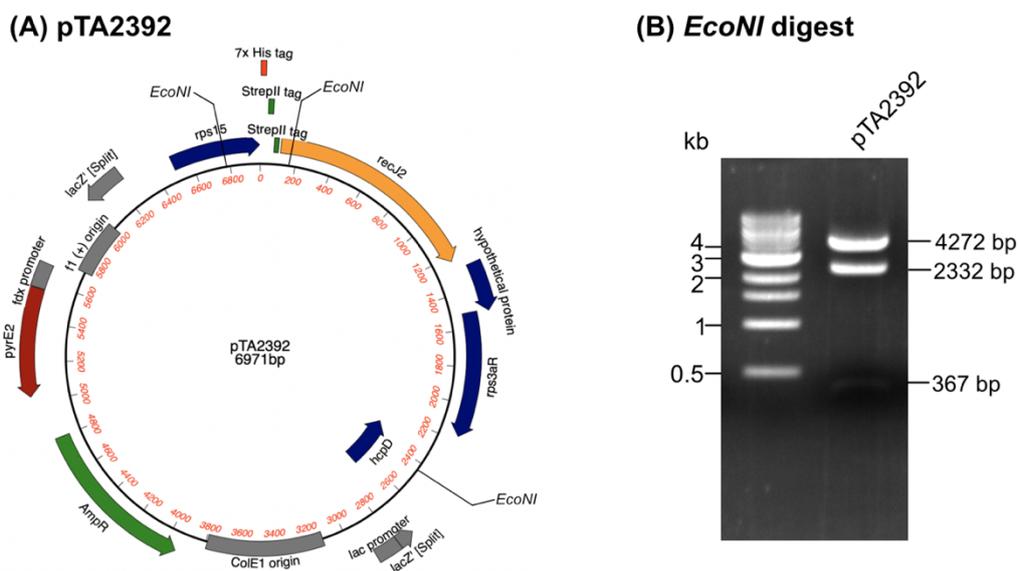


Figure 4.77: Construction of pTA2392 (A) Gene replacement construct pTA2392, introducing an N-terminal 7xHis 2xStrepII tag at the 5' end of *RecJ2*. (B) *EcoNI* digest shows bands at 4272 bp, 2332 bp and 367 bp, as predicted.

Tagged RecJ3 (HVO_1018)

The *recJ3* genomic clone pTA1913 (**Figure 4.78**) was used as a template to amplify the 1940 bp coding sequence of *recJ3* by PCR, using the primers *recJ3BspHI_F* and *recJ3Bam_R*. These primers introduced a *BspHI* and a *BamHI* site respectively, allowing digestion and ligation of the *recJ3* coding sequence into the compatible *PciI* and *BamHI* sites of 7xHis 2xStrepII tag cloning vector pTA1771. This gave rise to intermediate plasmid pTA2388, containing tandem tagged *recJ3*.

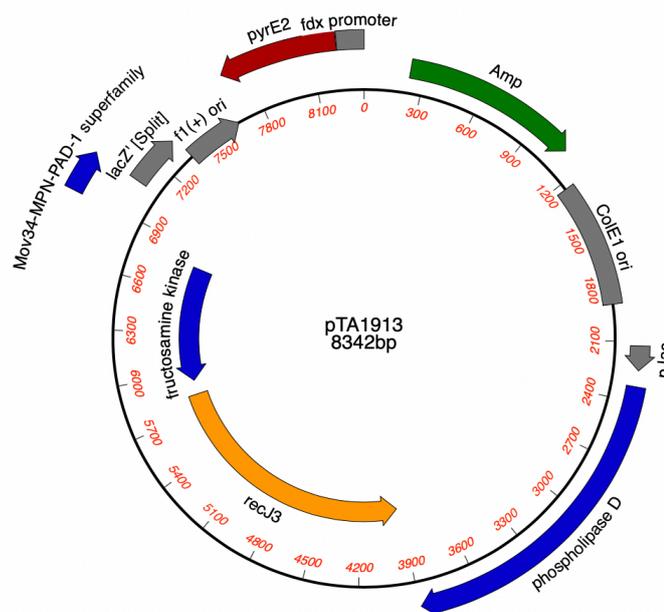


Figure 4.78: pTA1913. Genomic clone of *recJ3*. Constructed by Rebecca Lever (Lever, 2019).

Due to a lack of restriction sites available within the deletion construct for *recJ3*, the upstream and downstream genomic sequences were added using restriction digest from the *recJ3* genomic clone, pTA1913. The 1236 bp of upstream sequence of *recJ3* was cut from pTA1913 using *HindIII* and was inserted into intermediate plasmid pTA2388 at its *HindIII* site. This gave rise to plasmid pTA2391, 7xHis 2xStrepII-*recJ3* with upstream genomic sequence only. The 1.6 kb of downstream sequence was digested from pTA1913 using *SphI* and *NotI* and was inserted into pTA2391 at corresponding *SphI* and *NotI* sites. This gave rise to plasmid pTA2551, 7xHis 2xStrepII-*recJ3* gene replacement construct (**Figure 4.79**).

To allow for integration of the tag onto the chromosome, upstream and downstream sequences were added by restriction digest. Initial attempts to insert the tagged cassette into RecJ4 deletion vector, pTA1894 (Lever, 2019) at its *Bam*HI and *Nde*I sites failed; upon further screening, the expected *Nde*I site of the construct was not present.

Therefore, a two-step cloning process was carried out to add the surrounding genomic sequences. The DS sequence of RecJ4 was isolated from Δ *recJ4* construct pTA1894 by digest with *Bam*HI and *Not*I and was inserted into pTA2386 at its corresponding *Bam*HI/*Not*I sites to give plasmid pTA2633. The US sequence of RecJ4 was isolated from genomic clone pTA1882 by digest with *Kpn*I and *Nde*I and was inserted into pTA2633 at its corresponding *Kpn*I/*Nde*I sites to give 7xHis 2xStrepII-*recJ4* plasmid pTA2648 (**Figure 4.81**).

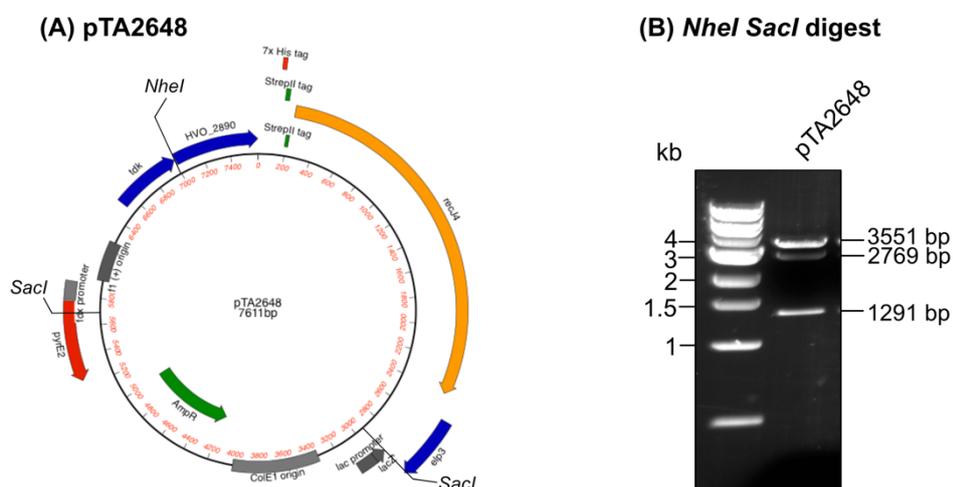


Figure 4.81: pTA2648. (A) Gene replacement construct pTA2648, introducing an N-terminal His₇ 2xStrepII tag at the 5' end of RecJ4. (B) *Nhe*I and *Sac*I digest shows bands at 3551 bp, 2769 bp and 1291 bp, as predicted.

Strain construction

The availability of *trpA*-marked RecJ deletion strains for RecJ1, RecJ3 and RecJ4 allows for the direct screening of pop-out colonies for integration of the tagged RecJ via screening of their ability to synthesise tryptophan. Where the pop-out leads to the replacement of the *trpA* marker at the site of the gene deletion with the gene (and corresponding protein tags), this strain will now be

auxotrophic for tryptophan. Whereas should the strain revert back to carrying the *trpA*-marked deletion, it will be capable of tryptophan synthesis. Therefore, screening of pop-outs in the presence and absence of tryptophan should elucidate which clones should be screened further as candidates for successful integration of the tagged gene construct. Since a *trpA*-marked deletion strain for RecJ2 was not available, efforts were focussed on RecJ1, RecJ3 and RecJ4. Primarily, strains were constructed with tandem 7xHis 2xStrepII tags at the N-terminus.

Tandem tagged RecJ strain generation

Deletion strains H3929 ($\Delta recJ1::trpA+$), H3931 ($\Delta recJ3::trpA+$) and H3932 ($\Delta recJ4::trpA+$) were transformed with tandem 7xHis 2xStrepII constructs pTA2390 (7xHis 2xStrepII-*recJ1*), pTA2551 (7xHis 2xStrepII-*recJ3*) and pTA2648 (7xHis 2xStrepII-*recJ4*) to give rise to pop-in strains H5042, H5043 and H5273 respectively. Pop-outs (5-FOA^R) were screened in the presence and absence of tryptophan (**Figure 4.82 C**), before *trp*- candidates were further screened by Southern blot to confirm genotype. H5042 pop-outs gave rise to 7xHis 2xStrepII::RecJ1 strain H5199 (**Figure 4.82**), H5043 pop-outs gave rise to 7xHis 2xStrepII::RecJ3 strain H5200 (**Figure 4.83**) and H5273 pop-outs gave rise to 7xHis 2xStrepII::RecJ4 strain H5313 (**Figure 4.84**).

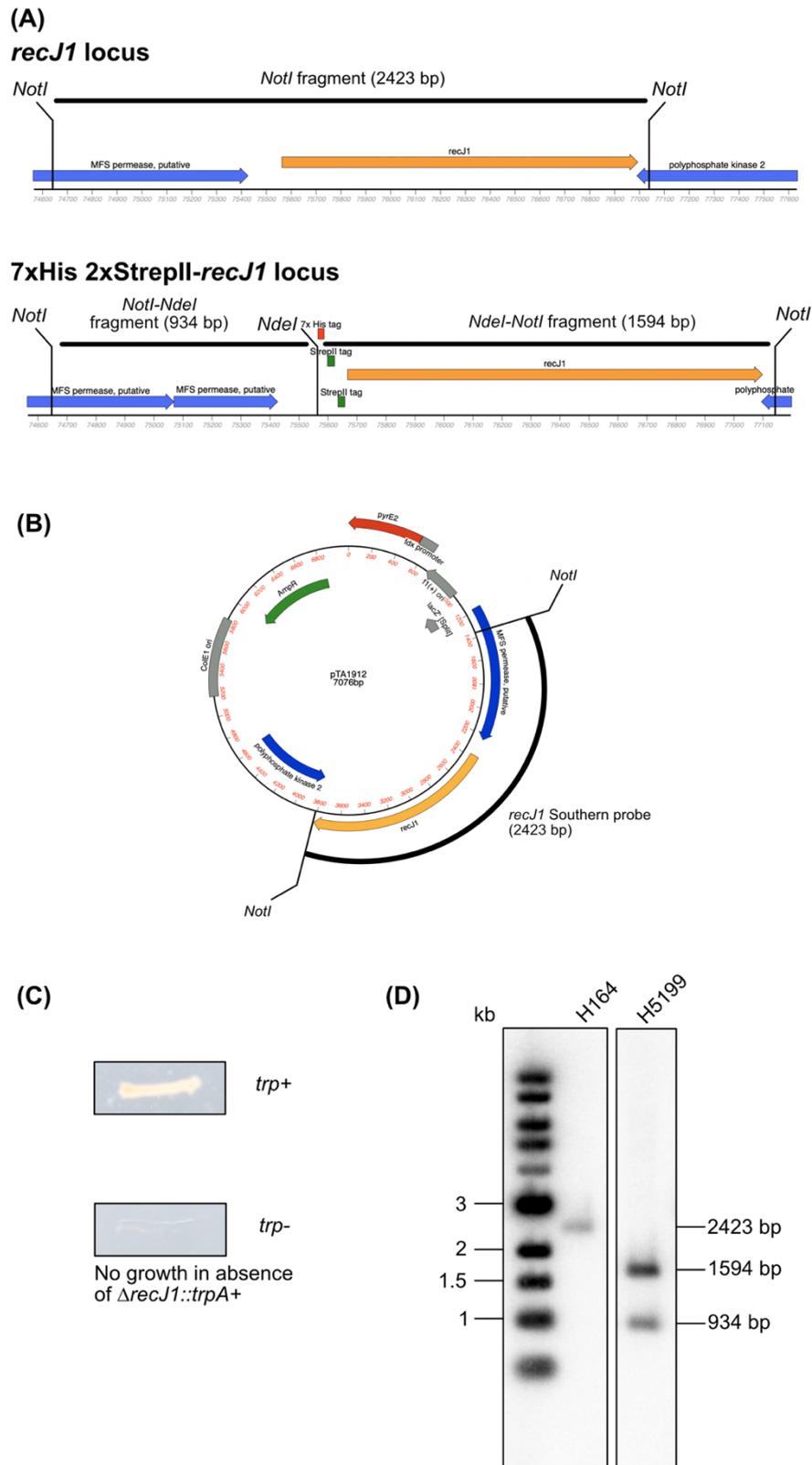


Figure 4.82: Gene replacement of $\Delta recJ1::trpA$ with 7xHis 2xStrepII-*recJ1*. (A) Expected Southern blot band sizes for *NotI*-*NdeI* digested genomic DNA. (B) 2423 bp *RecJ1* Southern probe consisting of *NotI* fragment of pTA1912. (C) Example of *trp*⁺/₋ screening, where strains now carrying 7xHis 2xStrepII-*recJ1* are tryptophan auxotrophic. (D) Southern blot confirming strain H5199 has integrated 7xHis 2xStrepII-*recJ1*.

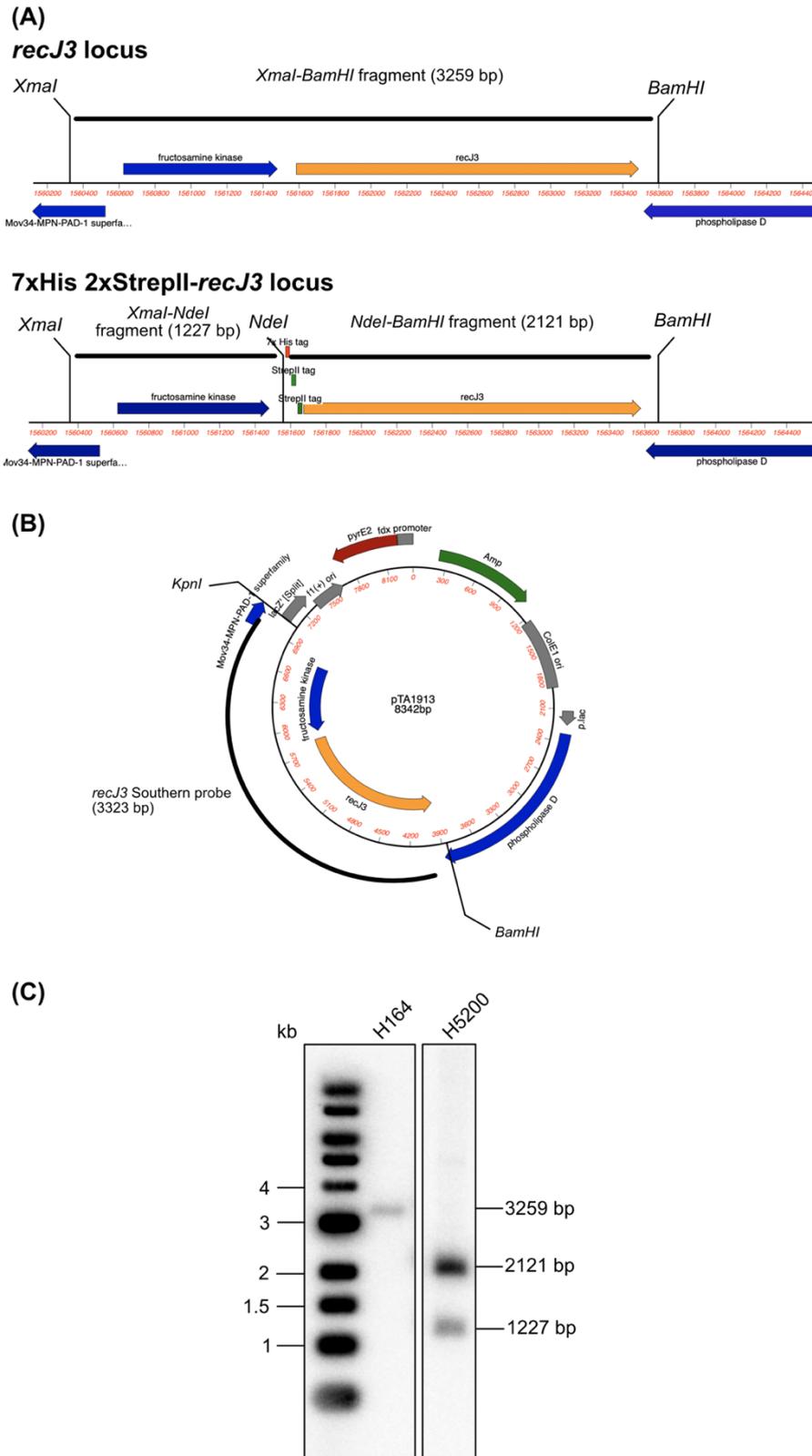


Figure 4.83: Gene replacement of $\Delta recJ3::trpA$ with 7xHis 2xStrepII-*recJ3*. (A) Expected Southern blot band sizes for *Bam*HI-*Nde*I-*Xma*I digested genomic DNA. (B) 3323 bp *RecJ3* Southern probe consisting of *Bam*HI-*Kpn*I fragment of pTA1913. (C) Southern blot confirming strain H5200 has integrated 7xHis 2xStrepII-*recJ3*.

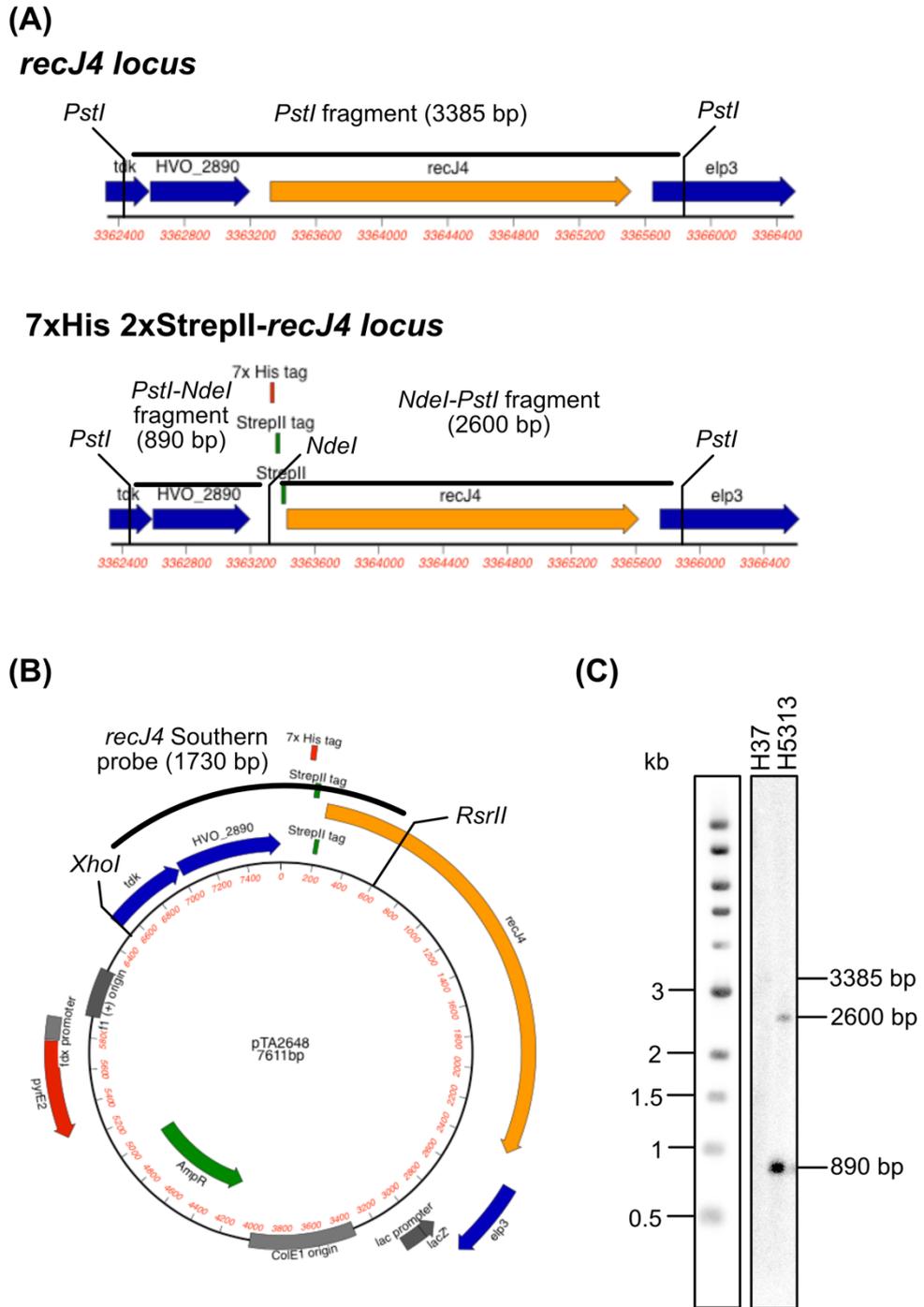


Figure 4.84: Gene replacement of $\Delta recJ4::trpA$ with 7xHis 2xStrepII-*recJ4*. (A) Expected Southern blot band sizes for *Pst*I-*Nde*I digested genomic DNA. (B) 1730 bp *RecJ4* Southern probe consisting of *Xho*I-*Rsr*II fragment of pTA2648. (C) Southern blot confirming strain H5313 has integrated 7xHis 2xStrepII-*recJ4*.

Purification of RecJ proteins *in vivo*

Initially, tagged-RecJ strains were compared to the parent of the RecJ deletion strains, H164, for growth differences using an Epoch spectrophotometer (BioTek) to ensure the integration of N-terminal tags are not affecting strain viabilities. Strains were grown for two consecutive overnights in Hv-YPC, ensuring on day three actively dividing cells were used for the assay. Cells were diluted and placed in a 96-well plate. Optical density (OD; A_{600}) was continuously measured, allowing the plotting of growth curves for each strain (**Figure 4.85**).

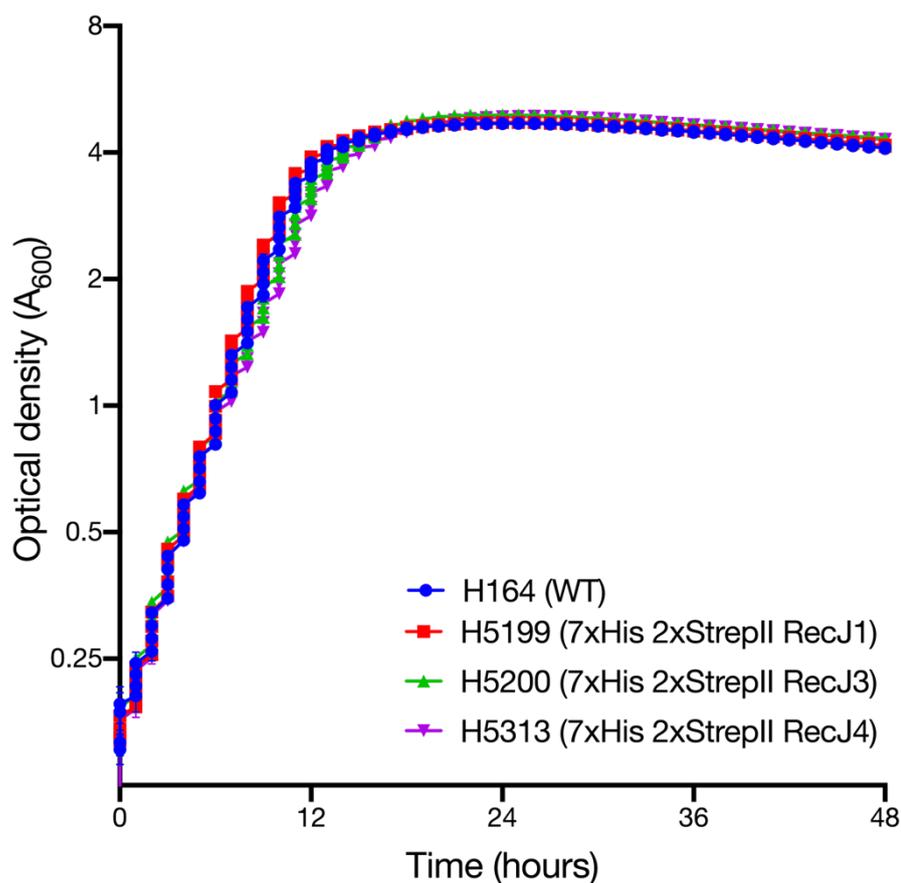


Figure 4.85: Exponential growth rate of strains with tagged RecJ1, RecJ3 or RecJ4. H164 (*recJ*⁺), H5199 (7xHis 2xStrepII-RecJ1), H5200 (7xHis 2xStrepII-RecJ3) and H5313 (7xHis 2xStrepII-RecJ4) were grown in Hv-YPC for two consecutive overnights before being diluted and plated. All strains (n=2) were incubated on the same 96-well plate and measured simultaneously for optical density (A_{600}) using an Epoch2 Microplate Spectrophotometer (BioTek).

Figure 4.85 shows the introduction of 7xHis 2xStrepII tags to RecJ1, RecJ3 and RecJ4 has not significantly impacted growth rate compared to H164.

When purifying proteins from strains with integrated tandem tags, H164 was subjected to the same pulldown conditions to act as an empty vector control. Proteins were purified from cell lysate using gravity columns packed with StrepTactin Sepharose. Proteins were eluted from the column using D-desthiobiotin and samples were subjected to precipitation using acetone.

4.3.5.1 Purification of RecJ1

Isolation of proteins by Strep-tactin affinity chromatography from strain H5199 was carried out and gave rise to 3 discrete bands when visualised on an SDS-PAGE gel (**Figure 4.86 A**). The three discrete bands seen are specific to H5199 and were not seen in the H164 control sample. Western blotting was performed on the pulldown samples using an anti-6xHis antibody to confirm the presence of 7xHis 2xStrepII-RecJ1, which gave rise to a band of ~50 kDa (RecJ1 predicted size: 50.6 kDa) (**Figure 4.86 B**).

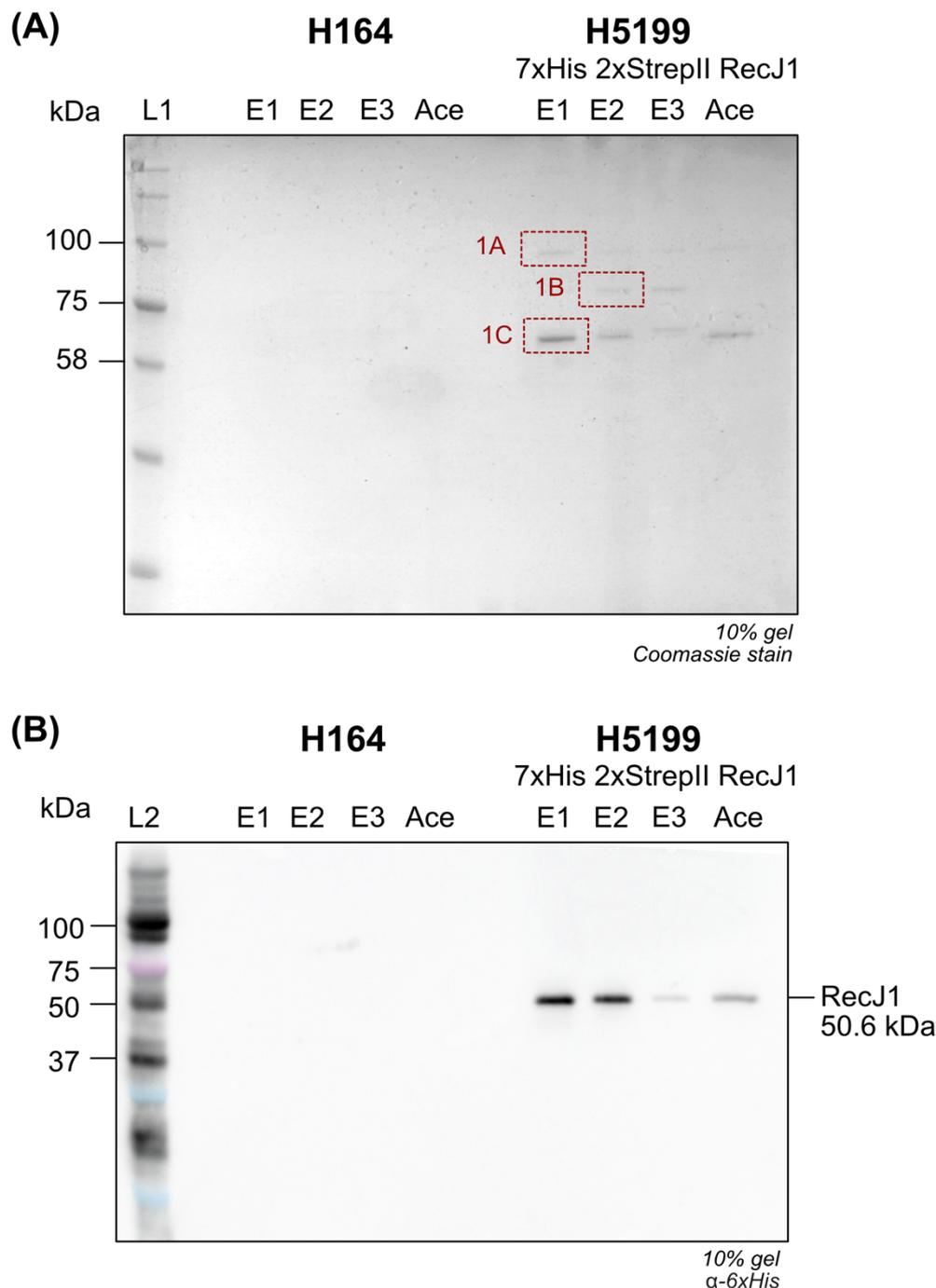


Figure 4.86: Purification of N-terminally 7xHis 2xStrepII-tagged RecJ1 via Streptavidin affinity from strain H5199 (A) Total protein-stained gel of StrepTactin purification from wild type (H164) and 7xHis 2xStrepII-RecJ1 (H5199) strains. E refers to stepwise elutions with 5 mM D-desthiobiotin. Ace refers to samples precipitated using acetate. Bands 1A, 1B and 1C (red) were submitted for analysis by mass spectrometry., alongside empty vector controls (gel cut at same location for H164 and submitted for analysis). (B) An anti-6xHis antibody detects a band in H5199 samples at ~50 kDa, presumed to be 7xHis 2xStrepII-RecJ1, corresponding to the lowest band in (A). L1: Blue Prestained Protein Standard, Broad Range (11-190 kDa). L2: BioRad Precision Plus Protein Dual Colour Standards.

Bands 1A, 1B and 1C and the corresponding bands from empty vector control strain H164 were submitted for analysis by mass spectrometry (**Table 4.5**).

Table 4.5: Summary of key peptides identified by mass spectrometry of strain H5199. All proteins listed below indicated identity or extensive homology ($p < 0.05$); they are listed in order of score (highest hits at the top). Peptides commonly identified in pulldowns or in control strain H164 were excluded. MW, molecular weight.

Band	1A	1B	1C
Approximate MW (kDa)	100	80	60
Proteins identified	Cdc48a	RecJ1	RecJ1
	RecJ4	PheT	ProS
	RecJ1	Cdc48a	RpoB2
	Cdc48b	RecJ3	PheS
			Cdc48a

See **Appendix 2** for complete table listing all proteins identified and associated MASCOT scores.

Peptides corresponding to the sequence of RecJ1 were found in all three bands sent for analysis, suggesting successful purification of RecJ1. While interactions with GINS and MCM were not observed (as would be predicted if RecJ1 was acting as GAN within the archaeal CMG complex), interactions were seen with Cdc48a in all bands; Cdc48a is known to act at the replication fork where it ensures protein turnover occurs in a timely manner when a replication fork stalls and requires recruitment of repair proteins.

RecJ1 was shown here to co-purify with both RecJ3 and RecJ4; such an interaction has not been identified previously. This suggests interplay between the RecJ proteins in *H. volcanii* and could explain the difficulty in defining the roles of these proteins where only species with two RecJ proteins have been characterised in detail.

A number of translation proteins were also co-purified; however, it remains to be confirmed experimentally if these bear any relation to the function of RecJ1 in *H. volcanii*.

4.3.5.2 Purification of RecJ3

Attempts were made to purify RecJ3 by Strep-tactin affinity chromatography from strain H5200, however no clear products were isolated. Additionally, precipitation of samples using TCA failed and therefore successful purification of RecJ3 was not possible during the timescale of this project.

4.3.5.3 Purification of RecJ4

Isolation of proteins by Strep-tactin affinity chromatography from strain H5313 was carried out and samples were visualised on an SDS-PAGE gel (Figure 4.87).

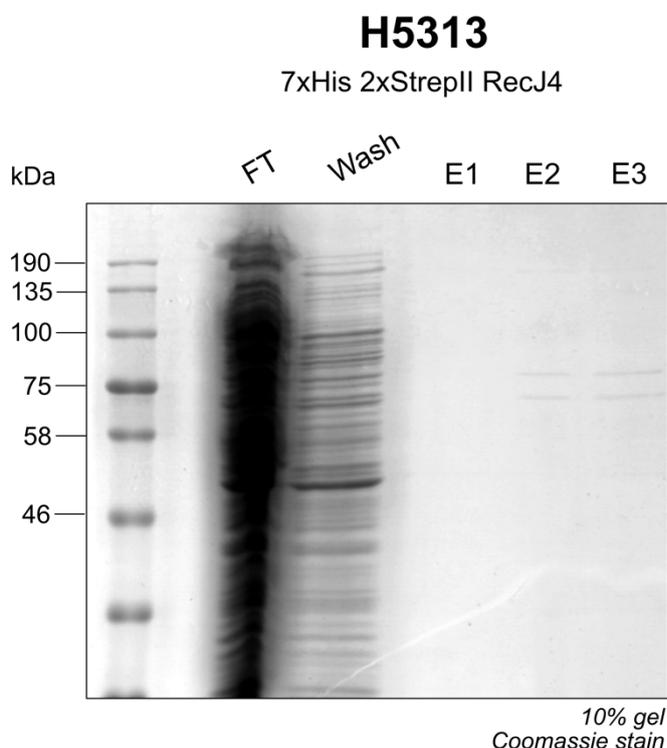


Figure 4.87: Purification of N-terminally 7xHis 2xStrepII-tagged RecJ4 via Streptavidin affinity from strain H5313. Total protein-stained gel of StrepTactin purification from strain H5313 (7xHis 2xStrepII-RecJ4). FT refers to flowthrough from the column, wash refers to the washes following lysate loading and E refers to stepwise elutions with 5 mM D-desthiobiotin. Faint bands can be seen in E2 and E3 at ~70 kDa, ~80 kDa and ~190 kDa.

Faint bands were observed at ~70 kDa, ~80 kDa and ~190 kDa, most prominently seen in lanes E2 and E3. However, these bands were faint, especially that at ~190 kDa, and as such samples were precipitated using TCA and rerun on an SDS-PAGE gel (Figure 4.88 A). Western blotting was performed on the precipitated pulldown samples using an anti-6xHis antibody to confirm the presence of 7xHis 2xStrepII-RecJ4 (Figure 4.88 B).

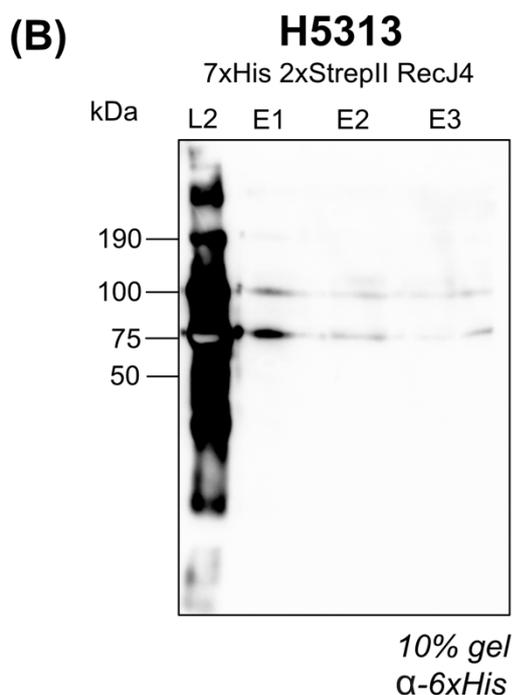
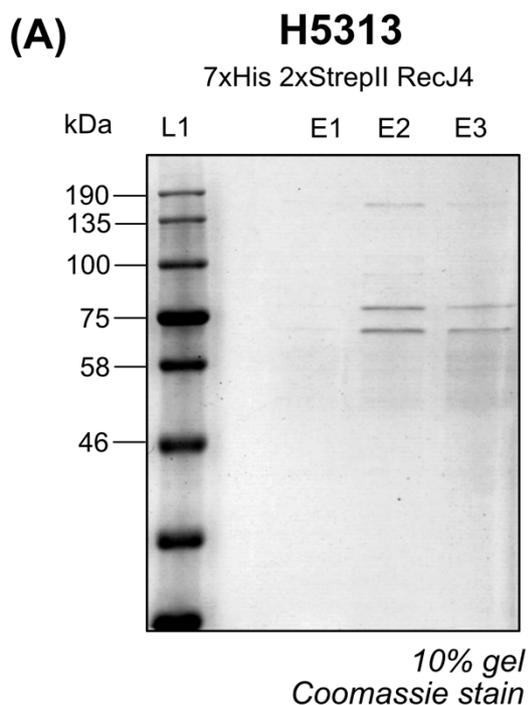


Figure 4.88: Purification of N-terminally 7xHis 2xStrepII-tagged RecJ4 via Streptavidin affinity from strain H5313. (A) Total protein-stained gel of StrepTactin purification from strain H5313 (7xHis 2xStrepII-RecJ4) precipitated using TCA. E refers to stepwise elutions with 5 mM D-desthiobiotin. Three clear bands are observed in lanes E2 and E3, at ~70 kDa, ~80 kDa and ~190 kDa. Faint, less discrete bands are present between ~90-110 kDa. (B) An anti-6xHis antibody detects bands at ~75 kDa and ~100 kDa. L1: Blue Prestained Protein Standard, Broad Range (11-190 kDa). L2: BioRad Precision Plus Protein Kaleidoscope.

Discrete bands are seen in lane E2 at ~70 kDa, ~80 kDa and ~190 kDa. Discrete bands are seen in lane E3 at ~70 kDa and ~80 kDa. Western blotting gave rise to two bands at ~70 and ~80 kDa (RecJ4 predicted size: 79.1 kDa)

To identify the pulldown products associating with RecJ4, bands were sent for analysis by mass spectrometry. **Figure 4.89** is annotated to show which bands were sent for analysis, while **Table 4.6** list the peptides identified for each band by mass spectrometry.

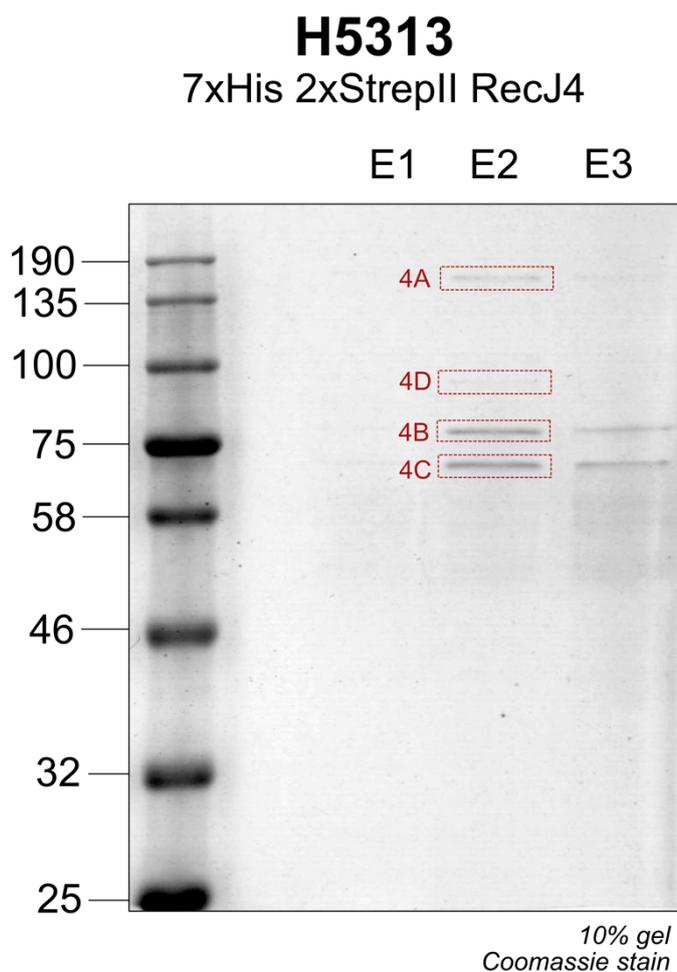


Figure 4.89: Affinity purification products from H5313 submitted for mass spectrometry. Bands highlighted in red boxes were submitted for analysis by mass spectrometry.

Table 4.6: Summary of key peptides identified by mass spectrometry of strain H5313. All proteins listed below indicated identity or extensive homology ($p < 0.05$); they are listed in order of score (highest hits at the top). Peptides commonly identified in pulldowns or in control strain H164 were excluded. MW, molecular weight.

Band	4A	4B	4C	4D
Approximate MW (kDa)	150	75	65	90
Proteins identified	LeuS	RecJ3	Cdc48a	RecJ4
	IleS	RpoB1	RecJ3	Top6B
	RecJ2	GyrB	RpoA1	TopA
		Cdc48a		

See **Appendix 3** for complete table listing all proteins identified and associated MASCOT scores.

RecJ4 was isolated and its presence was confirmed within band 4D. It was shown to co-purify with both RecJ3 (same arCOG grouping, HAN-like) and RecJ2 (different arCOG grouping, GAN-like). This, along with the previous data on the pulldown of RecJ1, suggests interplay between all RecJ proteins and adds a layer of complexity regarding defining their individual roles. Cdc48a was, again, isolated alongside the RecJ proteins, suggesting their localisation at the replication fork.

The purification of three different topoisomerases (TopA, Top6B and GyrB) provides additional evidence for a link between the RecJ proteins and DNA replication. Further experimental work will be required to define in what manner such an interaction occurs.

A number of translation proteins were also purified, including subunits of the RNA polymerase itself (RpoA1 and RpoB1); however, it remains to be confirmed experimentally if these bear any relation to the function of RecJ1 in *H. volcanii*.

4.3.6 Confirmation of RecJ protein-protein interactions using Split-GFP

Using Split-GFP to demonstrate *in vivo* protein-protein interactions

Split-GFP relies on two fragments of GFP, ^NGFP and ^CGFP, that alone do not assemble a fluorescent GFP protein when produced *in trans*, but assemble a fluorescent GFP protein when fused to interacting proteins. This methodology has been adapted for use in *H. volcanii*, whereby salt-stable smRS-GFP is utilised and was split between amino acid residues 157 and 158, leading to ^NGFP (17.7 kDa) and ^CGFP (9.0 kDa) (Reuter and Maupin-Furlow, 2004, Winter *et al.*, 2018).

Split-GFP allows for confirmation of protein interactions *in vivo*. Therefore, interactions previously observed between RecJ1, RecJ3 and RecJ4 from mass spectrometry can be further confirmed using this methodology. Alongside constructs for RecJ1, RecJ3 and RecJ4, constructs were also generated for RecJ2 to test for any interaction(s) with the other RecJ proteins.

Split-GFP cloning constructs

Episomal plasmids harbouring ^NGFP or ^CGFP fragments, with novobiocin resistance (Nov^R) and mevinolin resistance (Mev^R) genes respectively, are the start-point for generation of ^N-/^C-GFP fusion proteins. The differential resistance genes allow for selection of both episomes simultaneously within *H. volcanii*. Constructs are available for integration of the GFP fragments both N- and C-terminally of the selected gene target. A short spacer sequence (14-16 amino acids) is located between the GFP fragment and the gene target. The gene fusion will be expressed using promoter *p.fdx* and the episome is maintained in *H. volcanii* using the pHV2 origin (absent from laboratory strain H26). The plasmids are listed below in **Table 4.7**. Plasmids pJAS-NGFP-Nterm and pWL-CGFP-Nterm are shown in **Figure 4.90** as an example.

Table 4.7: Episomal plasmids harbouring NGFP or CGFP fragments for generation of GFP fusion proteins. Constructs provided by (Winter *et al.*, 2018).

Plasmid number	Description
pJAS-NGFP-Nterm	For construction of N-terminal ^N GFP fusions, Nov ^R
pJAS-NGFP-Cterm	For construction of C-terminal ^N GFP fusions, Nov ^R
pWL-CGFP-Nterm	For construction of N-terminal ^C GFP fusions, Mev ^R
pWL-CGFP-Cterm	For construction of C-terminal ^C GFP fusions, Mev ^R

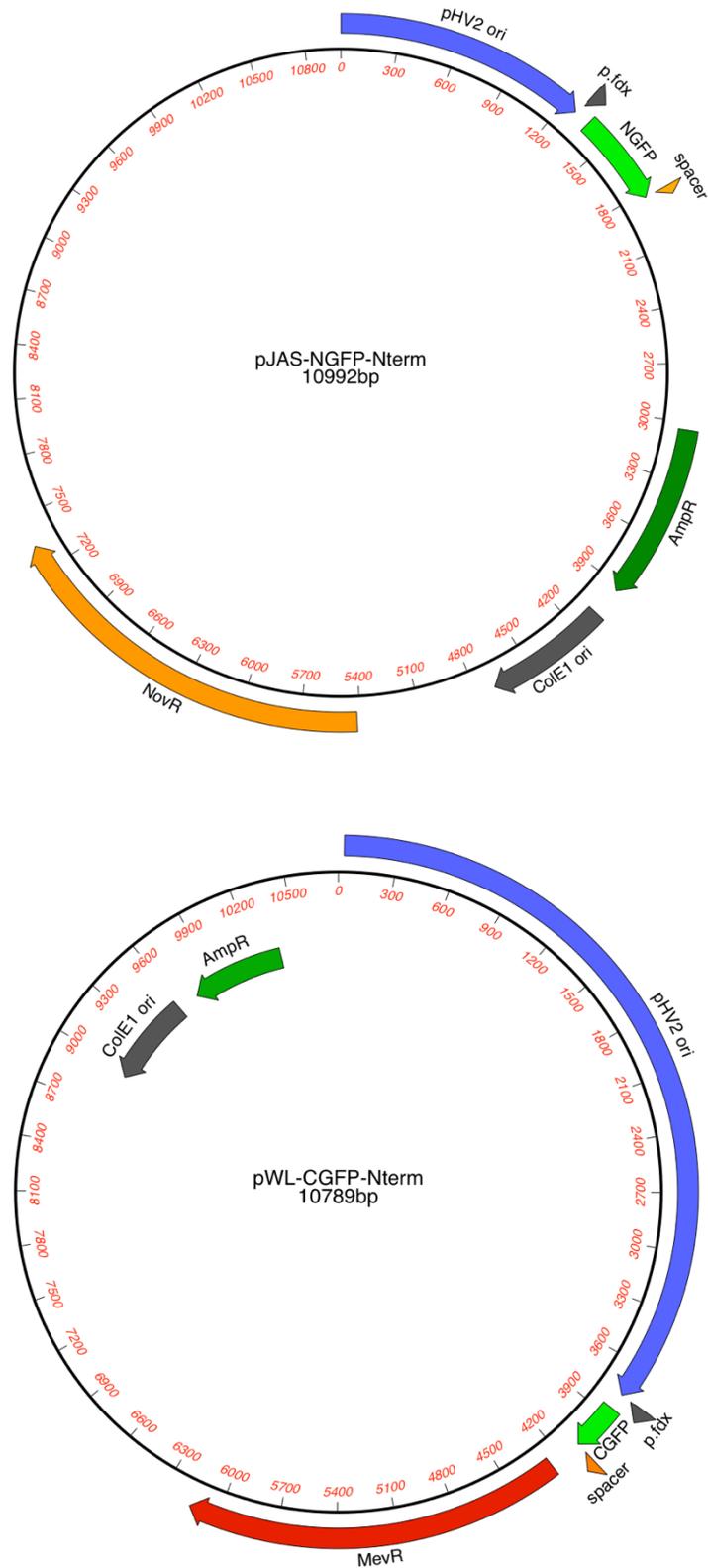


Figure 4.90: Plasmids pJAS-NGFP-Nterm and pWL-CGFP-Nterm. Constructs provided by Winter *et al.*, 2018. Episomes maintained by pHV2 origin. Gene fusions driven by *p.fdx*. ^NGFP constructs encode the novobiocin resistance gene (Nov^R), while ^CGFP constructs encode the mevinolin resistance gene (Mev^R).

Cloning of RecJ1-RecJ4 split-GFP episomal constructs

For each *RecJ* gene targeted, a similar cloning strategy was utilised to generate all four iterations of ^NGFP and ^CGFP fusion; here, the generation of *RecJ1* fusion constructs will be explained in detail, with a similar strategy being applied to the remaining gene targets. All constructs were confirmed to be in-frame with the GFP fragment using the MacVector translation tool.

For in-frame integration of *recJ1* into N-terminal ^NGFP plasmid pJAS-NGFP-Nterm, primers were designed integrating a *Bsp*HI site overlapping the start codon of the gene and a *Kpn*I site downstream of the stop codon. The coding sequence of *recJ1* was isolated from genomic clone pTA1912 using primers *RecJ1_F_Bsp*HI and *RecJ1_R_Kpn*. This product was then digested with *Bsp*HI and *Kpn*I and inserted into vector pJAS-NGFP-Nterm at its compatible *Nco*I and *Kpn*I sites, to give N-terminal ^NGFP-*RecJ1* episome pTA2586 (**Figure 4.91**).

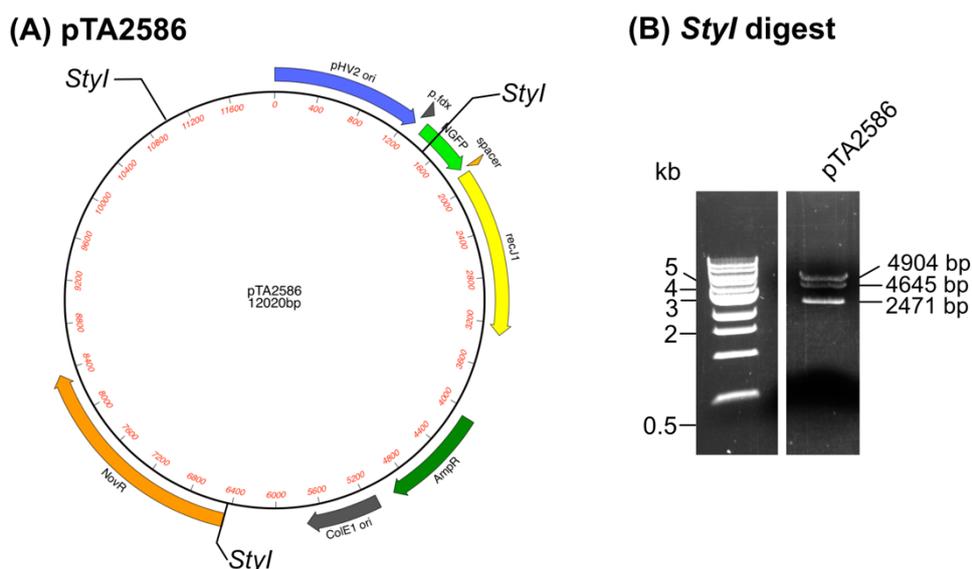


Figure 4.91: pTA2586. (A) Episomal N-terminal ^NGFP *RecJ1* construct pTA2586. (B) *Sty*I digest shows bands at 4904 bp, 4645 bp and 2471 bp, as predicted.

For in-frame integration of *recJ1* into C-terminal ^NGFP plasmid pJAS-NGFP-Cterm, primers were designed integrating a *Bsp*HI site overlapping the start codon of the gene and a *Blp*I site mutating the stop codon of *recJ1* to allow continuous transcription past the end of *recJ1*. The coding sequence of *recJ1* was isolated from genomic clone pTA1912 using primers *RecJ1_F_Bsp*HI and *RecJ1_R_Blp*. This product was then digested with *Bsp*HI and *Blp*I and inserted into vector pJAS-NGFP-Cterm at its compatible *Nco*I and *Blp*I sites, to give C-terminal ^NGFP-*RecJ1* episome pTA2587 (**Figure 4.92**).

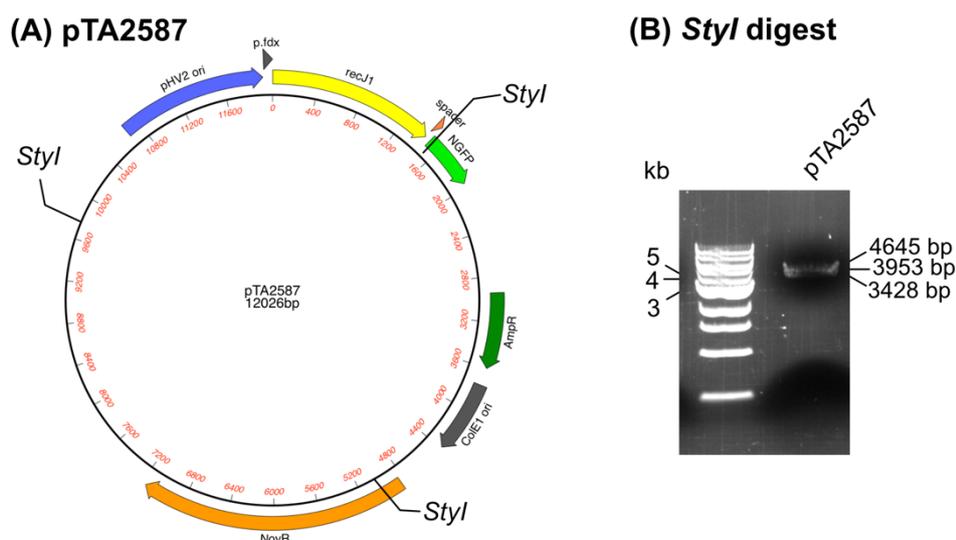


Figure 4.92: pTA2587. (A) Episomal C-terminal^NGFP RecJ1 construct pTA2587. (B) *StyI* digest shows bands at 4645 bp, 3953 bp and 3428 bp, as predicted. Bands are likely running high due to supercoiling; plasmid was further confirmed as correct by sequencing.

For in-frame integration of *recJ1* into N-terminal^CGFP plasmid pWL-CGFP-Nterm, primers were designed integrating a *Bam*HI site overlapping the start codon of the gene and a *Kpn*I site downstream of the stop codon of *recJ1*. The coding sequence of *recJ1* was isolated from genomic clone pTA1912 using primers *RecJ1_F_Bam* and *RecJ1_R_Kpn*. This product was then digested with *Bam*HI and *Kpn*I and inserted into vector pWL-CGFP-Nterm at its compatible *Bam*HI and *Kpn*I sites, to give N-terminal^CGFP-RecJ1 episome pTA2588 (Figure 4.93).

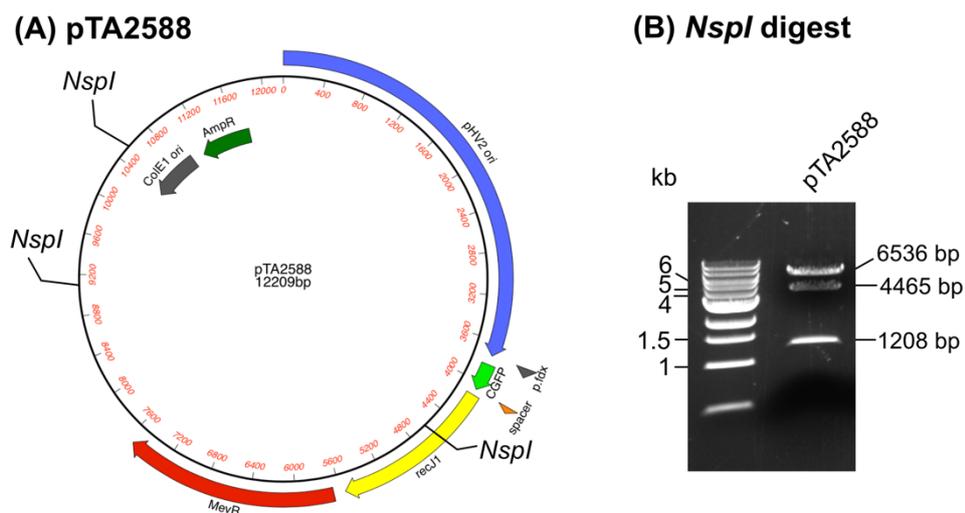
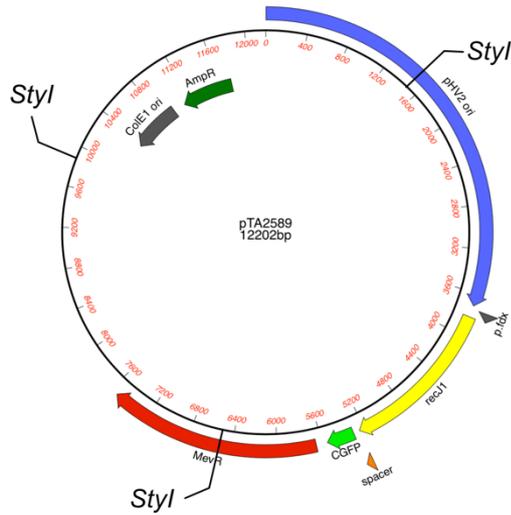


Figure 4.93: pTA2588. (A) Episomal N-terminal ^CGFP RecJ1 construct pTA2588. (B) *NspI* digest shows bands at 6536 bp, 4465 bp and 1208 bp, as predicted.

For in-frame integration of *recJ1* into C-terminal ^CGFP plasmid pWL-CGFP-Cterm, primers were designed integrating a *Bsp*HI site overlapping the start codon of the gene and a *Bam*HI site mutating the stop codon of *recJ1* to allow continuous transcription past the end of *recJ1*. The coding sequence of *recJ1* was isolated from genomic clone pTA1912 using primers *RecJ1_F_Bsp*HI and *RecJ1_R_Bam*. This product was then digested with *Bsp*HI and *Bam*HI and inserted into vector pWL-CGFP-Cterm at its compatible *Nco*I and *Bam*HI sites, to give C-terminal ^CGFP-RecJ1 episome pTA2589 (Figure 4.94).

(A) pTA2589



(B) *StyI* digest

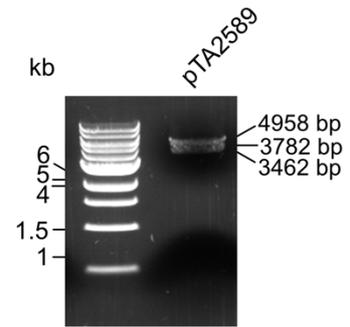


Figure 4.94: pTA2589. (A) Episomal C-terminal^CGFP RecJ1 construct pTA2589. (B) *StyI* digest shows bands at 4958 bp, 3782 bp and 3462 bp, as predicted.

Primers used for generation of split-GFP constructs of RecJ1-RecJ4 are listed in **Table 4.8**. Plasmid numbers for the resulting split-GFP episomes are listed in **Table 4.9**.

Table 4.8: Oligonucleotides used to generate split-GFP plasmids for RecJ1, RecJ2, RecJ3 and RecJ4. Lower case letters represent mismatches against the template. Restriction sites within the primer are highlighted in red.

Primer	Sequence
<i>RecJ1_F_BspHI</i>	CACACtcATGaACGGACCCG
<i>RecJ1_R_Kpn</i>	CGGTaccCGATTAGTCCGCG
<i>RecJ1_R_Blp</i>	CGGTgctCagcGAGTCCGCGTTTTTCAGCC
<i>RecJ1_F_Bam</i>	TACCggAtCcATGGACGGACCCGTCC
<i>RecJ1_R_Bam</i>	GGTCGggatccAGTCCGCGTTTTTCAGC
<i>RecJ2_F_BspHI</i>	ATAACTtcATGaCCGTGAGCC
<i>RecJ2_R_Kpn</i>	CCGCAGGtaCcCACGCCGGCTCATCG
<i>RecJ2_R_Blp</i>	ACGCtcagcGATCGGGCGCACCTCCC
<i>RecJ2_F_Bam</i>	AGAATggaTccATGTCCGTGAGCC
<i>RecJ2_R_Bam</i>	GCACGCgGatcCATCGGCCGACC
<i>RecJ3_F_BspHI</i>	GGGAtcATGAGCGACGAGCACGCC
<i>RecJ3_R_Kpn</i>	AGAGTgGtACcCCGGCTTACGCC
<i>RecJ3_R_Blp</i>	AACGCtcagcCACGCCGTCGTCGACAGC
<i>RecJ3_F_Bam</i>	AGCGGatCcATGAGCGACGAGCACG
<i>RecJ3_R_Bam</i>	AACGCgGatccACGCCGTCGTCGACAGC
<i>RecJ4_F_BspHI</i>	CAACGtcATGaATTGGATTACGCACG
<i>RecJ4_R_Kpn</i>	TCTGGtacCGGCTTAGAACTGC
<i>RecJ4_R_Blp</i>	TGGAgctcagcGAGAACTGCTCGGCGG
<i>RecJ4_F_Bam</i>	TGCTCggatCcATGGATTGGATTACGC
<i>RecJ4_R_Bam</i>	GATTgGatccAGAAGCTGCTCGGCGGC

Table 4.9: Episomes containing split-GFP fragments for screening for interactions between RecJ proteins.

Gene targeted	Plasmid number	GFP fragment	N- or C-terminal
RecJ1 (HVO_0073)	pTA2586	^N GFP	N
	pTA2587		C
	pTA2588	^C GFP	N
	pTA2589		C
RecJ2 (HVO_1147)	pTA2592	^N GFP	N
	pTA2593		C
	pTA2617	^C GFP	N
	pTA2594		C
RecJ3 (HVO_1018)	pTA2595	^N GFP	N
	pTA2618		C
	pTA2619	^C GFP	N
	pTA2620		C
RecJ4 (HVO_2889)	pTA2601	^N GFP	N
	pTA2602		C
	pTA2603	^C GFP	N
	pTA2604		C

Strain generation

To test for interaction between two proteins, wild type strain H26 (Δ *pyrE2*) was transformed with one ^NGFP and one ^CGFP construct simultaneously, before being plated on Hv-Cas +Ura +Mev +Nov. Transformants were restreaked on the same media and then screened for the presence of both ^NGFP and ^CGFP episomes by colony PCR (**Figure 4.95**). Screening for interactions between all N-terminal and C-terminal ^NGFP and ^CGFP episome combinations gave rise to a total of 48 strains harbouring varying pairs of plasmids (**Table 4.10**).

(A) Predicted PCR sizes

	RecJ1	RecJ2	RecJ3	RecJ4
Primers	<i>fdx-prom-F</i> <i>recJ1intR</i>	<i>fdx-prom-F</i> <i>recJ2intR</i>	<i>fdx-prom-F</i> <i>recJ3probeR</i>	<i>fdx-prom-F</i> <i>recJ4intR</i>
Product size (bp)				
^N GFP-x	744	1223	1441	970
^C GFP-x	532	1011	1229	758
x- ^{N/C} GFP	228	707	925	454

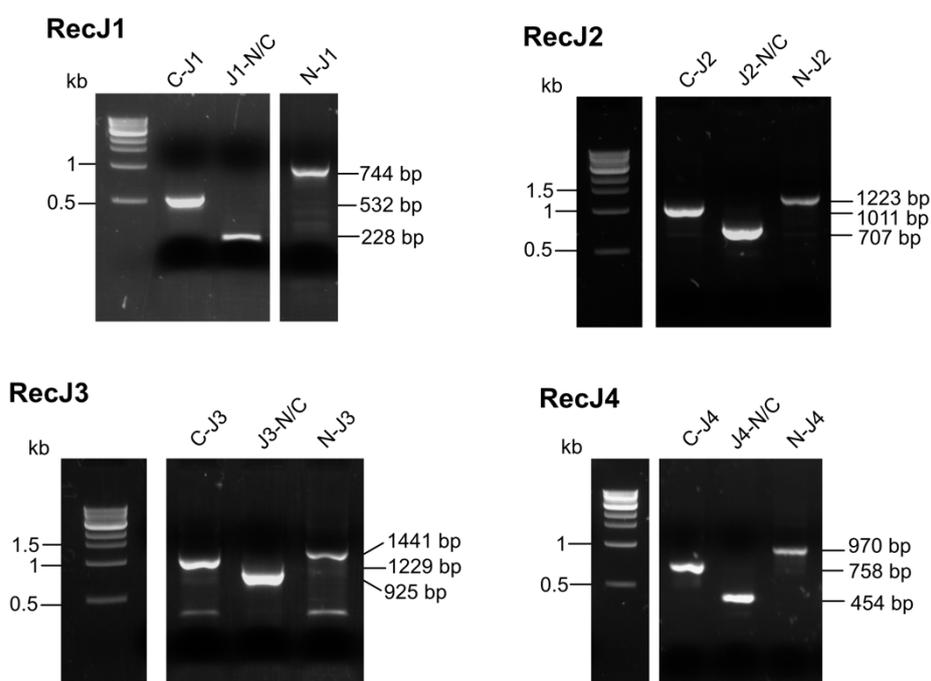
(B) Colony PCR results

Figure 4.95: Colony PCR to confirm presence of split-GFP episomes within transformed H26 candidates. (A) PCR using primers *fdx-prom-F* and *recJ1intR* (RecJ1 episomes), *recJ2intR* (RecJ2 episomes), *recJ3probeR* (RecJ3 episomes) and *recJ4intR* (RecJ4 episomes) gives products of the listed sizes depending on location of the split-GFP fragment. (B) Gels confirming the presence of the differing episomes within candidate strains. For simplicity only one example for each has been shown here.

Table 4.10: Combinations of split-GFP constructs transformed into H26. RecJ1-4 have been simplified to J1-4 for ease. ^N refers to fragment NGFP, while ^C refers to CGFP. Placement of the ^{N/C} represents the terminus to which the GFP fragment has been fused. x represents a pair of plasmids targeting the same protein (and therefore these transformations were not performed).

	^N GFP RecJ1		^N GFP RecJ2		^N GFP RecJ3		^N GFP RecJ4	
^C GFP RecJ1	x	x	^N -J2 ^C -J1	J2- ^N ^C -J1	^N -J3 ^C -J1	J3- ^N ^C -J1	^N -J4 ^C -J1	J4- ^N ^C -J1
	x	x	^N -J2 J1- ^C	J2- ^N J1- ^C	^N -J3 J1- ^C	J3- ^N J1- ^C	^N -J4 J1- ^C	J4- ^N J1- ^C
^C GFP RecJ2	^N -J1 ^C -J2	J1- ^N ^C -J2	x	x	^N -J3 ^C -J2	J3- ^N ^C -J2	^N -J4 ^C -J2	J4- ^N ^C -J2
	^N -J1 J2- ^C	J1- ^N J2- ^C	x	x	^N -J3 J2- ^C	J3- ^N J2- ^C	^N -J4 J2- ^C	J4- ^N J2- ^C
^C GFP RecJ3	^N -J1 ^C -J3	J1- ^N ^C -J3	^N -J2 ^C -J3	J2- ^N ^C -J3	x	x	^N -J4 ^C -J3	J4- ^N ^C -J3
	^N -J1 J3- ^C	J1- ^N J3- ^C	^N -J2 J3- ^C	J2- ^N J3- ^C	x	x	^N -J4 J3- ^C	J4- ^N J3- ^C
^C GFP RecJ4	^N -J1 ^C -J4	J1- ^N ^C -J4	^N -J2 ^C -J4	J2- ^N ^C -J4	^N -J3 ^C -J4	J3- ^N ^C -J4	x	x
	^N -J1 J4- ^C	J1- ^N J4- ^C	^N -J2 J4- ^C	J2- ^N J4- ^C	^N -J3 J4- ^C	J3- ^N J4- ^C	x	x

Once confirmed to carry both GFP fragments, strains were assayed for GFP signal. Single colonies were inoculated into Hv-Cas +Ura +Mev +Nov media. Lower temperatures have been shown to enhance successful folding and subsequent fluorescence of GFP (Winter *et al.*, 2018), thus instead of growth at 45°C as standard, cultures were grown to A₆₅₀ ~1 at 37°C for 1 day to obtain sufficient cell mass, followed by incubation of the culture at 30°C overnight.

2 ml of culture was then washed with 18% salt water (SW) prior to being resuspended in 500 µl 18% SW. Samples were then loaded in duplicate into a 96-well plate and measured for both optical density (BioTek Epoch 2 spectrophotometer) and GFP signal (GE Healthcare Typhoon; excitation wavelength at 488 nm). Negative control strains carrying combinations of empty vectors (pairs of ^NGFP and ^CGFP plasmids) and positive control strains (confirmed interaction via mass spec and split-GFP) were tested for GFP signal, alongside untransformed H26 and 18% SW alone. The relative fluorescence (*rf*) was calculated using the formula given below (**Equation 4.1**).

Equation 4.1: Calculation of relative fluorescence (*rf*) for split-GFP strains.

$$rf = \frac{\text{transformant} - \text{untransformed H26}}{\text{untransformed H26}}$$

Positive and negative control strains are listed in **Table 4.11**. **Figure 4.96** shows interactions with strains harbouring RecJ1 episomes encoding ^NGFP in combination with ^CGFP plasmids encoding RecJ2, RecJ3 and RecJ4, alongside control strains.

Table 4.11: Control strains for split-GFP assay. Strains H5109-H5111 are GFP-negative (negative controls) while H5334 and H5335 are GFP-positive (positive controls).

Strain	Genotype
H5109	$\Delta\text{pyrE2} \{p.fdx::^{\text{N}}\text{GFP-spacer Nov}^{\text{R}}\} \{p.fdx::^{\text{C}}\text{GFP-spacer Mev}^{\text{R}}\}$
H5110	$\Delta\text{pyrE2} \{p.fdx::^{\text{N}}\text{GFP-spacer Nov}^{\text{R}}\} \{p.fdx::\text{spacer-}^{\text{C}}\text{GFP Mev}^{\text{R}}\}$
H5111	$\Delta\text{pyrE2} \{p.fdx::\text{spacer-}^{\text{N}}\text{GFP Nov}^{\text{R}}\} \{p.fdx::^{\text{C}}\text{GFP-spacer Mev}^{\text{R}}\}$
H5112	$\Delta\text{pyrE2} \{p.fdx::\text{spacer-}^{\text{N}}\text{GFP Nov}^{\text{R}}\} \{p.fdx::\text{spacer-}^{\text{C}}\text{GFP Mev}^{\text{R}}\}$
H5334	$\Delta\text{pyrE2} \{p.fdx::^{\text{N}}\text{GFP-hel308 Nov}^{\text{R}}\} \{p.fdx::\text{recJ3-}^{\text{C}}\text{GFP Mev}^{\text{R}}\}$
H5335	$\Delta\text{pyrE2} \{p.fdx::^{\text{N}}\text{GFP-hel308 Nov}^{\text{R}}\} \{p.fdx::\text{recJ4-}^{\text{C}}\text{GFP Mev}^{\text{R}}\}$

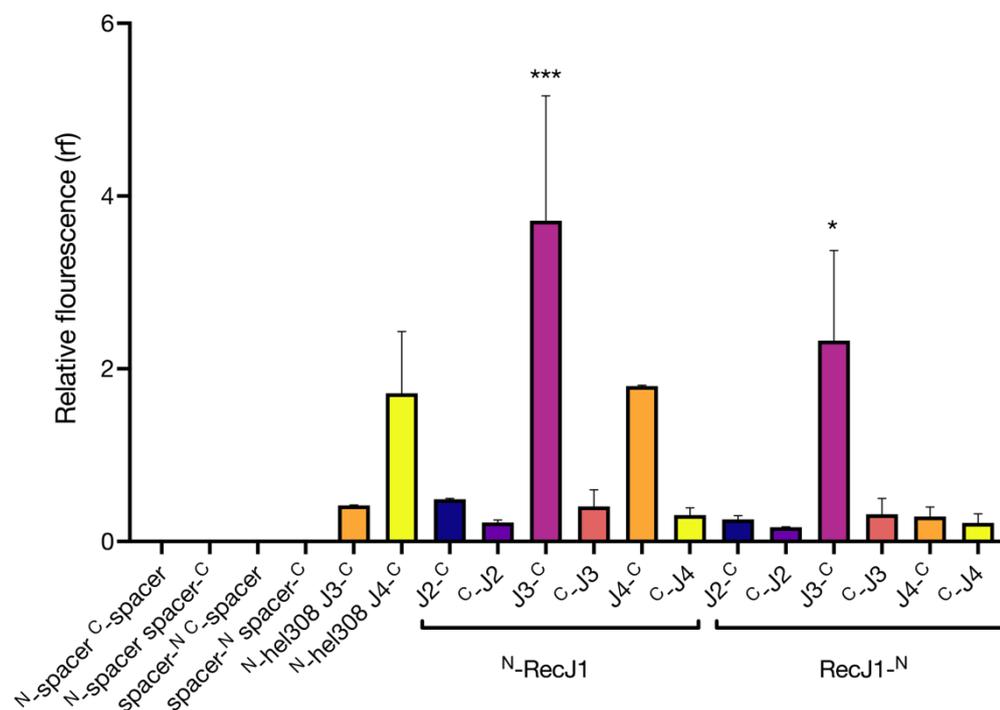


Figure 4.96: Relative fluorescence (rf) values for interactions between ^NGFP-tagged RecJ1 and ^CGFP-tagged RecJ2, RecJ3 and RecJ4. Negative control strains H5109-H5111 are annotated as ^N-spacer ^C-spacer, ^N-spacer spacer-^C, spacer-^N ^C-spacer and spacer-^N spacer-^C respectively. Positive controls H5334 and H5335 are annotated as ^N-hel308 J3-^C and ^N-hel308 J4-^C respectively. RecJ1-4 have been simplified to J1-4 for ease. Relative fluorescence (rf) was calculated against untransformed H26. Fluorescence is observed for interactions between ^NGFP-tagged RecJ1 and ^CGFP-tagged RecJ2, RecJ3 and RecJ4, with RecJ1::RecJ3 giving the highest signal. A parametric one-way ANOVA test was performed followed by Dunn's multiple comparisons test against untransformed H26. Error bars represent standard error of the mean (SEM). Results are from two independent trials. (*) $p < 0.05$, (***) $p < 0.001$.

As expected, negative control strains did not fluoresce, while positive control strains were GFP-positive (with H5334 showing a lower level of fluorescence than H5335; previously observed upon strain generation (Ambika Dattani, personal communication)). Fluorescence was observed between ^NGFP-tagged RecJ1 and ^CGFP-tagged RecJ2, RecJ3 and RecJ4. Interaction between RecJ1 and RecJ3 was the strongest tested, with both ^N-RecJ1 and RecJ1-^N constructs showing a significant difference to H26 when paired with RecJ3-^C.

Figure 4.97 shows interactions with strains harbouring RecJ2 episomes encoding ^NGFP in combination with ^CGFP plasmids encoding RecJ1, RecJ3 and RecJ4.

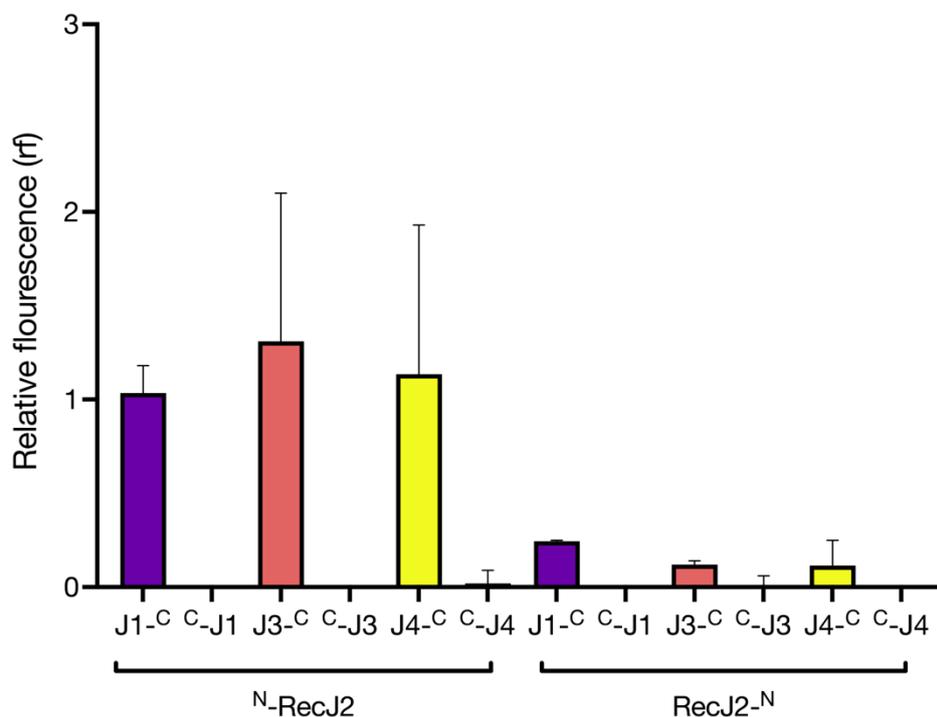


Figure 4.97: Relative fluorescence (rf) values for interactions between ^NGFP-tagged RecJ2 and ^CGFP-tagged RecJ1, RecJ3 and RecJ4. RecJ1-4 have been simplified to J1-4 for ease. Relative fluorescence (rf) was calculated against untransformed H26. Fluorescence is observed between ^NGFP-tagged RecJ2 and ^CGFP-tagged RecJ1, RecJ3 and RecJ4. A parametric one-way ANOVA test was performed followed by Dunn's multiple comparisons test against untransformed H26. Error bars represent standard error of the mean (SEM). Results are from two independent trials.

Fluorescence was observed between ^NGFP-tagged RecJ2 and ^CGFP-tagged RecJ1, RecJ3 and RecJ4. For all interactions, fluorescence was only ever observed when the interacting ^CGFP-tagged protein was carrying the GFP fragment at its C-terminus. A higher fluorescence was associated with the N-terminally tagged ^N-RecJ2 than its C-terminal RecJ2-^N counterpart.

Figure 4.98 shows interactions with strains harbouring RecJ3 episomes encoding ^NGFP in combination with ^CGFP plasmids encoding RecJ1, RecJ2 and RecJ4.

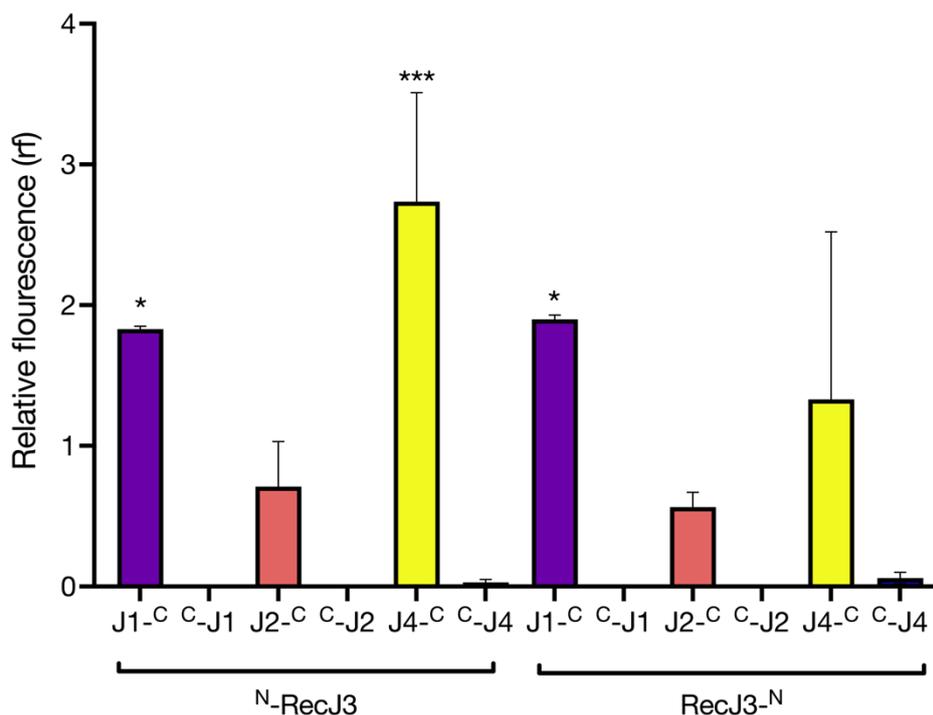


Figure 4.98: Relative fluorescence (rf) values for interactions between ^NGFP-tagged RecJ3 and ^CGFP-tagged RecJ1, RecJ2 and RecJ4. RecJ1-4 have been simplified to J1-4 for ease. Relative fluorescence (rf) was calculated against untransformed H26. Fluorescence is observed between ^NGFP-tagged RecJ3 and ^CGFP-tagged RecJ1, RecJ2 and RecJ4. A parametric one-way ANOVA test was performed followed by Dunn's multiple comparisons test against untransformed H26. Error bars represent standard error of the mean (SEM). Results are from two independent trials. (*) $p < 0.05$, (***) $p < 0.001$.

Fluorescence was observed between ^NGFP-tagged RecJ3 and ^CGFP-tagged RecJ1, RecJ2 and RecJ4. For all interactions, fluorescence was more commonly observed when the interacting ^CGFP-tagged protein was carrying the GFP fragment at its C-terminus. Signal for interactions of RecJ3 with RecJ1 and RecJ4 are stronger than those observed for RecJ2.

Figure 4.99 shows interactions with strains harbouring RecJ4 episomes encoding ^NGFP in combination with ^CGFP plasmids encoding RecJ1, RecJ2 and RecJ3.

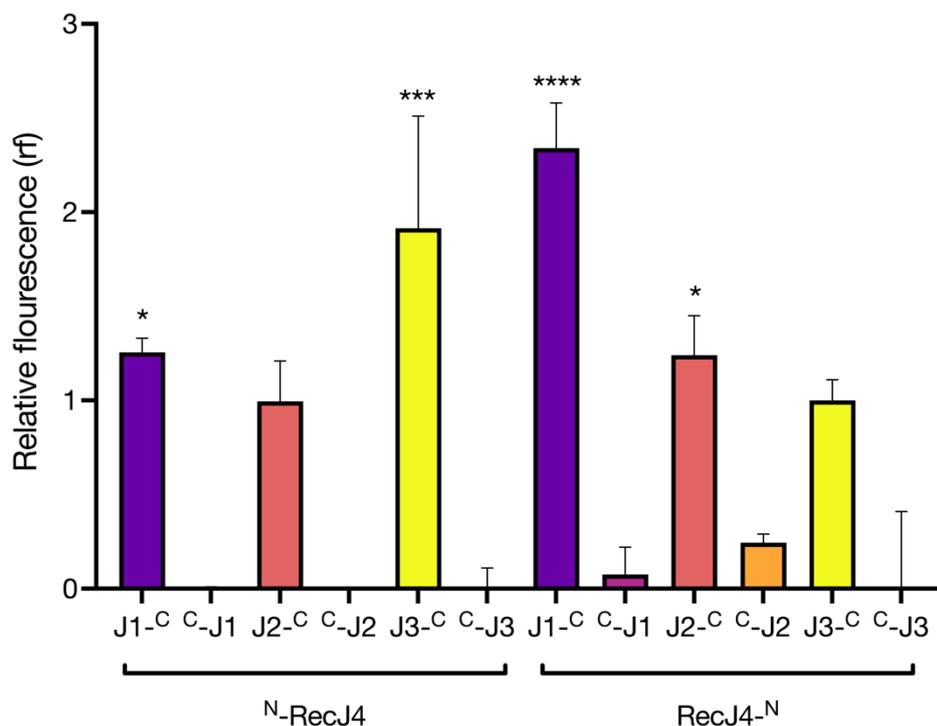


Figure 4.99: Relative fluorescence (rf) values for interactions between ^NGFP-tagged RecJ4 and ^CGFP-tagged RecJ1, RecJ2 and RecJ3. RecJ1-4 have been simplified to J1-4 for ease. Relative fluorescence (rf) was calculated against untransformed H26. Fluorescence is observed between ^NGFP-tagged RecJ4 and ^CGFP-tagged RecJ1, RecJ2 and RecJ3. A parametric one-way ANOVA test was performed followed by Dunn’s multiple comparisons test against untransformed H26. Error bars represent standard error of the mean (SEM). Results are from two independent trials. (*) $p < 0.05$, (***) $p < 0.001$, (****) $p < 0.0001$.

Fluorescence was observed between ^NGFP-tagged RecJ4 and ^CGFP-tagged RecJ1, RecJ2 and RecJ3. At least one combination of each interaction (RecJ1::RecJ4, RecJ2::RecJ4, RecJ3::RecJ4) were statistically different to untransformed H26.

A summary of RecJ interactions observed via split-GFP are found in **Table 4.12**.

Table 4.12: Summary of fluorescence measured as result of split-GFP interaction assay between RecJ1, RecJ2, RecJ3 and RecJ4. Levels of interaction represented as low, medium and high. Low represents relative fluorescence (rf) level <0.75 , medium represents $rf >0.75$ and <1.5 , high represents $rf >1.5$.

Interaction	RecJ2	RecJ3	RecJ4
RecJ1	Low	High	High
RecJ2		Medium	Medium
RecJ3			High

Where differing levels of GFP signal were observed, this could be due to a number of reasons:

- overexpression of the protein target is not stable, and the target protein is therefore present at lower amounts than expected, meaning only a certain amount of interaction (and therefore GFP signal) can be observed
- the interaction is less common than those with a stronger signal
- weak interaction is due to interactions only occurring at certain loci, while strong interactions occur more globally within the cell

Due to the unavailability of antibodies against RecJ proteins, or protein tags within the construct, we cannot test whether the difference in signal is due to amount of protein. However, assessing the signal via microscopy will allow definition of whether ‘low signal’ is due to localisation at foci vs cell wide.

One strain for each interacting pair were selected for microscopy, with *rf* values ranging between 0.49 and 3.72. These strains were grown for two subsequent overnights in Hv-YPC +Mev +Nov at 37°C, followed by one overnight incubation at 30°C, reaching a final OD of ~ 1 . Cells were then spun down, resuspended in 18% salt water, stained with DAPI (final concentration 2.5 $\mu\text{g/ml}$) and incubated in the dark for 10 minutes. Cells were then washed of excess dye, resuspended in fresh 18% salt water and placed onto prepared agarose pads containing 18% salt water. Cells were imaged using a Nikon Ti-E inverted fluorescence microscope (**Figure 4.100**).

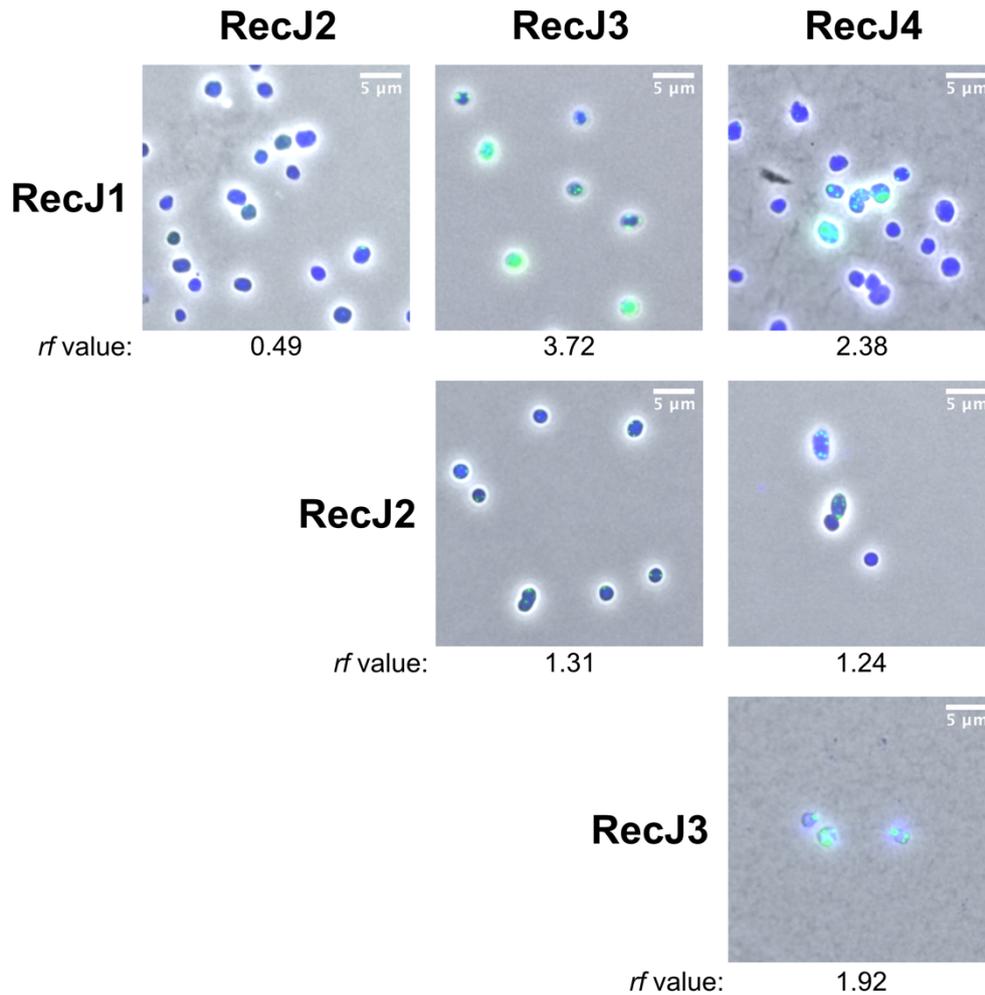


Figure 4.100: Microscopy of split-GFP interacting strains and their associated *rf* values. GFP interactions are observed for all strains tested, as expected. Where a weak *rf* value is observed, GFP signal is generally low and only associated with a subset of cells, while high *rf* values show brighter foci. DNA stained using DAPI. Fluorescence was acquired with a one second exposure. Composite images are shown. Scale bars are 5 μm .

All strains analysed showed GFP signal, as expected. For the lowest *rf* tested (RecJ1:RecJ2; $rf=0.49$), a weak signal was seen in only a subset of cells. For the highest *rf* tested (RecJ1:RecJ3; $rf=3.72$), a strong signal was seen; some cells show cell-wide expression while others form foci, however all cells showed some level of GFP expression. Where foci are seen, there appears to be a level of compartmentalisation within the cell, where the boundary of GFP signal is highly defined. It is unknown whether such compartments are membrane-bound or are an artefact of protein localisation.

For *rf* values between 1-3, generally GFP expression is not observed in all cells. This could be due to certain requirements within each cell for the interaction to occur or could be due to protein instability when being

Chapter 4: Genetic analysis of *recJ* genes in *Haloferax volcanii*

overexpressed; it would be pertinent to test strains for protein stability over time.

4.4 Discussion

Summary of phenotyping data for $\Delta recJ$ mutants

Strains harbouring deletions of *recJ1* (HVO_0073), *recJ3* (HVO_1018) and *recJ4* (HVO_2889) had previously been generated and compared for growth rate prior to commencement of this study, but not been phenotyped any further. Here, these strains were assessed for response to DNA damaging agents and replication stressors. The results of these trials are summarised in **Table 4.13**.

Table 4.13: Summary of phenotyping data for $\Delta recJ$ strains in H164 background. All strains are compared against the wild type response (H164). An empty box means the response shown was comparable to wild type. Where an increase in resistance is seen (i.e., cells respond better than wild type), this is marked as a positive (+) response, and where an increase in sensitivity is seen (i.e., cells respond worse than wild type), this is marked as a negative (–) response.

	<i>$\Delta recJ1$</i>	<i>$\Delta recJ3$</i>	<i>$\Delta recJ4$</i>
H₂O₂ response		+	+
Phleomycin response		-	-
MMC response	-	-	
UV response	-		
Aphidicolin response	+		
Recombination rate	-	+	+
Replication profile	Shallow peaks		

When screened for response to Family B DNA polymerase inhibitor, aphidicolin, only the *$\Delta recJ1$* mutant H3929 showed a response that differed from wild type. However, this altered response was not caused by a difference in origin use (as shown via replication profiling). This strain also showed a sensitivity to UV not seen with either *recJ3* or *recJ4* mutants. Generally, the responses of the *recJ3* mutant strain, H3931, and *recJ4* mutant strain, H3932, were comparable, with the exception of MMC response, where only the *recJ3* mutant showed sensitivity. Additionally, only the *recJ3* mutant had an altered replication profile, where the peaks (representing active origins) were shallower than those of the wild type.

Attempts to delete the *recJ* genes were then made in combination with predicted (or previously confirmed) interactors to check for any synthetic lethality or strong phenotypes; these included *hef*, *hjc*, *hel308*, *rnhB* and *fen1*. The results of these genetic combination mutants and their associated phenotypes are summarised in **Table 4.14**. Any strains that were unable to be generated are coloured red.

Table 4.14: Summary of mutants (and associated phenotypes) generated in combination with deletion of *recJ1*, *recJ3* or *recJ4*. A red X denotes the inability to generate this combinatorial mutant. All double mutant strains are being compared against the single mutant (as listed in the left-hand column). Aph, aphidicolin; ^R, resistant; ^S, sensitive.

	<i>ΔrecJ1</i>	<i>ΔrecJ3</i>	<i>ΔrecJ4</i>
<i>Δhjc</i>	MMC ^S	X	X
<i>Δhef</i>		MMC ^S	
<i>ΔoriC</i>	Aph ^R		
<i>ΔrnhB</i>	Aph ^R MMC ^S		
<i>Δfen1</i>	X	UV ^S	
<i>Δhel308</i>	MMC ^S		

It was not possible to delete *recJ3* or *recJ4* in combination with *hjc*, while deletion of *recJ1* and *hjc* was possible. All *recJ* genes were able to be deleted in combination with *hef*, however it is worth noting that preliminary attempts to delete *hef* in combination with both *recJ3* and *recJ4* proved impossible. It was not possible to delete *fen1* in combination with *recJ1*. Often, combinatorial deletions with *recJ1* showed an increased sensitivity to MMC (*hjc*, *rnhB* and *hel308*). Deletion of *recJ1* consistently showed an increased resistance to aphidicolin compared to any single mutant screened.

Figure 4.101 shows a summary of protein interactions observed as part of this chapter, including results of direct pulldowns of RecJ proteins and identification of interacting proteins via mass spectrometry, as well as interactions observed using split GFP methodology.

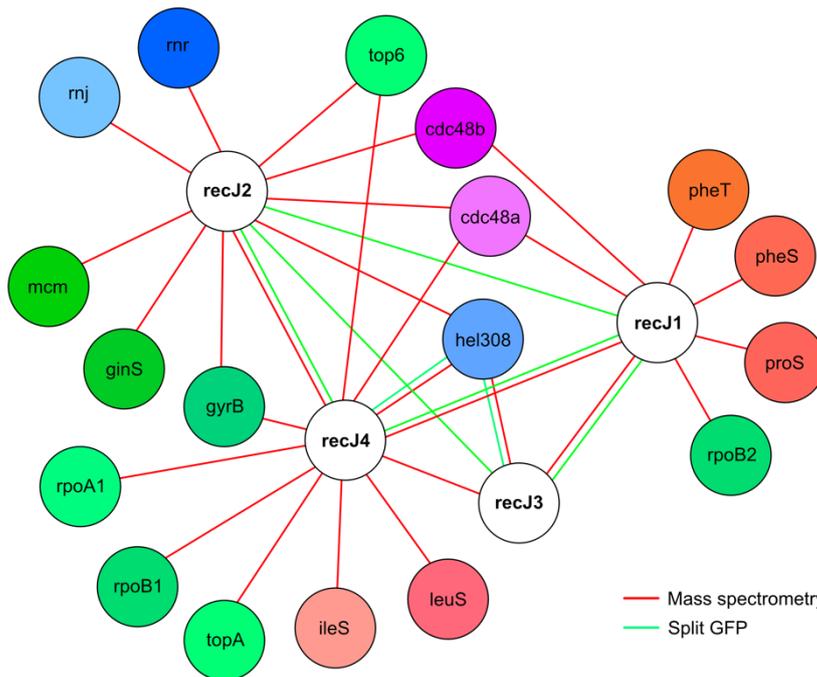


Figure 4.101: Summary of protein:protein interactions observed as a result of pulldown experiments and identification of proteins via mass spectrometry (red connecting lines) or via split GFP methodology (green connecting lines). RecJ proteins (the targets for this chapter) are coloured in white. All other proteins have been coloured according to their known or predicted functions. Green: DNA replication, red: translation, purple: protein turnover, blue: DNA damage repair.

In terms of scientific relevance of these results, this will be discussed in detail below with discussions being grouped in terms of predicted functions based on both bioinformatics and these results; where RecJ1 is acting during DNA replication as GAN, RecJ3 and RecJ4 are acting with Hef at stalled interstrand crosslinks as HAN proteins, and RecJ2 plays an unknown function.

RecJ1 as the GINS-associated nuclease (GAN) candidate

Structurally, the RecJ1 protein of *H. volcanii* (HVO_0073) is predicted to be very similar to solved structures of GAN proteins in both *T. kodakarensis* and *P. furiosus* (Oyama *et al.*, 2016, Li *et al.*, 2017). RecJ1 also falls in same arCOG grouping (arCOG00427) as GAN proteins in *Thermococcus* and *Pyrococcus* species, among others (Huerta-Cepas *et al.*, 2019).

However, work by Makarova *et al.* (2012) predicted the GAN protein should be found in a specific gene neighbourhood, containing essential genes involved in transcription and translation (including ribosomal proteins and tRNAs); in *H. volcanii* RecJ1 does not match the operonic layout suggested to be linked with GAN, and instead RecJ2 does (Makarova *et al.*, 2012). However, when comparing the nucleotide sequence of *recJ2* to well-characterised *TkoGAN*, there is no similarity. This suggests that categorisation of GAN based solely on gene neighbourhood is somewhat misleading, especially in species where the *recJ* genes have undergone a recent duplication event (as in *H. volcanii*).

While Δ *gan* strains in *T. kodakarensis* have been shown to suffer obvious growth retardation (Nagata *et al.*, 2017a), no delay in growth rate was seen for *H. volcanii* strains deleted for *recJ1*. However, the DNA damage response to UV seen in *T. kodakarensis* Δ *gan* and *H. volcanii* Δ *recJ1* is comparable (Nagata *et al.*, 2017a). All Δ *recJ1* mutant strains (with the exception of the Δ *hef* Δ *recJ1* double mutant) showed an increased sensitivity to MMC treatment, compared to the parent strain (even when these were associated with a level of MMC sensitivity already). This additive sensitivity may suggest RecJ1 plays a role away from these proteins, many of which are linked to DNA damage repair. Whether this is due to a role in DNA replication is not proven, but hypothetically this could be the cause of the increased sensitivity of RecJ1 mutants.

Of the strains deleted for a single RecJ gene in a background in which no genes known or suspected to function in DNA replication or repair are mutated, only the *H. volcanii* Δ *recJ1* mutant (H3929) showed an altered response to aphidicolin (PolB replicative polymerase inhibitor). While the biology behind this altered response is as yet unknown, this differential response of H3929 links RecJ1 to a role in DNA replication. It was shown elsewhere in this study (see **Chapter 3**) that the response to aphidicolin treatment is also altered in the absence of replication origins.

However, this study showed that the *recJ1* mutant is not dependent on RadA (as is the case in the absence of origins) and its replication profile did not differ from wild type (i.e., it actively utilises its origins). It therefore remains unknown what specific role RecJ1 is playing and why its deletion

causes aphidicolin resistance: could it be that PolB has a role outside of canonical replication (e.g., in DNA repair) where it works alongside RecJ1? In this study H3929 showed sensitivities to UV and MMC. The same sensitivities were observed in *T. kodakarensis*, with a strain deleted for *polB* (Kushida *et al.*, 2019). If PolB1 is acting with RecJ1 in DNA repair, loss of RecJ1 would then alter the function of PolB1 in repair and this could be observed as an increased resistance to aphidicolin. Further work is needed to strengthen this hypothesis. The $\Delta polB$ *T. kodakarensis* strain was also sensitive to γ - irradiation and MMS; it would be interesting to see if these phenotypes were also matched by the $\Delta recJ1$ strain.

When *recJ1* was deleted in a strain already deleted for all origins, an additive resistance to aphidicolin treatment was observed. This suggests RecJ1 is acting to alter aphidicolin response (and therefore likely PolB activity) in a manner not determined by the presence of origins, supporting the previous theory that it may be acting with PolB away from origins. It remains to be determined whether an originless $\Delta recJ1$ mutant strain would therefore have a decreased requirement for PolB1; the OD of $\Delta oriC \Delta recJ1$ strain H5282 does not reach the final OD of the untreated strain, suggesting there may still be a requirement for PolB1, but this could warrant further investigation.

In *T. kodakarensis*, GAN has been implicated in primer removal; in a strain deleted for GAN (Δgan), both RNase HIII and Fen1 become essential (Burkhart, 2017), while in the presence of GAN (*gan+*), strains can be deleted for both Fen1 and RNase HIII. Deletion of *recJ1* was attempted in combination with *rnhB* and *fen1*. While it was not possible to generate a $\Delta recJ1 \Delta fen1$ mutant (as in *T. kodakarensis*), it proved possible to generate an $\Delta recJ1 \Delta rnhB$ mutant. Does this mean RecJ1 is not actually GAN? While, at first, this may seem the obvious conclusion, this observation can potentially be explained by analysing the differences between *H. volcanii* and *T. kodakarensis*; while *T. kodakarensis* only encodes a single RNase H (Type II; *rnhB*, TK0805), *H. volcanii* encodes five RNase H homologues; three of type 1 (*rnhE*, HVO_0732; *rnhA*, HVO_2438; *rnhC*, HVO_A0463) and a single type 2 protein (*rnhB*, HVO_1978). Gene *rnhD* (HVO_A0277) does not fit clearly into either group and its function remains unknown. The role of type 1 RNase H genes remains unknown; however, it is possible that in the absence of *rnhB*, either a type 1 RNase H, or *rnhD*, are capable of compensating for its function and this is why deletion of *rnhB* and *recJ1* is possible.

Isolation of RecJ1 and its interacting partners through Strep-tactin affinity purification of strain H5199 (7xHis 2xStrepII RecJ1) was successful; RecJ1 was isolated and confirmed as present by Western blotting. RecJ1 was shown to interact with fellow RecJ proteins, RecJ3 and RecJ4. This was

somewhat unexpected as interaction between GAN and HAN has not, to date, been reported. Interaction was also observed with Cdc48 proteins, which are known to act at the replication fork to recycle proteins should DNA damage repair proteins need access (Dantuluri, 2018). Interaction between RecJ1 and GINS and/or MCM would be expected, should RecJ1 be the true GAN; however, this was not observed in this pulldown experiment. It would be worthwhile repeating this experiment, along with the negative control, to ensure all interacting partners are identified. Since DNA replication proteins are associated directly with DNA, it could also be worthwhile to alter the pulldown protocol as to reduce the number of DNA-bound proteins being lost by centrifugation following sonication. This could be via use of salt-active nucleases or loading of the sonicated lysate directly onto the column.

RecJ3 and RecJ4 as Hef-associated nuclease (HAN) candidates

The grouping of RecJ3 (HVO_1018) and RecJ4 (HVO_2889) within arCOG00429 implicates them as HAN-like proteins in *H. volcanii* (Huerta-Cepas *et al.*, 2019). Their protein domain organisation and structure are comparable to that of other archaeal HAN proteins (Nagata *et al.*, 2017a, Feng *et al.*, 2018). Regarding catalytic activity, RecJ3 is predicted to possess activity while RecJ4 has lost critical motifs, however this is yet to be proven experimentally. In both *T. kodakarensis* and *P. furiosus*, HAN has been shown to have single-strand specific 3' -5' exonuclease activity, which is stimulated when interacting with Hef (Nagata *et al.*, 2017b, Feng *et al.*, 2018).

Feng *et al.* (2018) have previously used the logic of HAN being catalytically active to designate RecJ3 as the sole HAN protein in *H. volcanii* (Feng *et al.*, 2018). However, the structural similarities of RecJ3 and RecJ4 means it cannot be ruled out that these two proteins may work in parallel in their function to assist Hef; this function may be unique to the Haloarchaea, where numerous duplication events of RecJ genes have occurred and multiple arCOG00429 members are found (unlike the well-studied HAN proteins of *T. kodakarensis* and *P. furiosus*, of which each species only encodes a single gene of arCOG00429) (Makarova *et al.*, 2012, Feng *et al.*, 2018, Nagata *et al.*, 2017b).

Generally, the phenotyping of single mutants for $\Delta recJ3$ (H3931) and $\Delta recJ4$ (H3932) were comparable to one another. The one major difference observed was in response to crosslinking agent MMC, where H3931 showed an increased sensitivity compared to wild type that was not observed for H3932. This difference in response could be due to the predicted catalytic activity of RecJ3 (where RecJ4 is predicted to be inactive based on residue analysis). This question could potentially be answered via generation and MMC screening of a catalytically dead mutant of RecJ3.

Interestingly, both RecJ3 and RecJ4 cannot be deleted in a Δhjc background. It is well documented that Hef cannot be deleted from Δhjc strain (Lestini *et al.*, 2010), and so it is of interest to see this potential link between RecJ3, RecJ4 and Hef. RecJ3 and RecJ4 were both dispensable in combination with Hef. The $\Delta hef \Delta recJ3$ strain showed an increased sensitivity to MMC greater than the Δhef mutant alone, potentially suggesting an alternate role for RecJ3 aside from ICL repair. While both $\Delta recJ3 \Delta hef$ and $\Delta recJ4 \Delta hef$ double mutants were easily generated, it is also interesting that it proved impossible to generate a $\Delta recJ3 \Delta recJ4 \Delta hef$ mutant. This suggests that either RecJ3 or RecJ4 are needed in a Δhef strain.

Hef is well studied and known to be involved in interstrand crosslink (ICL) repair (Fujikane *et al.*, 2010, Lestini *et al.*, 2010). Both $\Delta recJ3$ and $\Delta recJ4$ strains do not show a major sensitivity to MMC (akin to that associated with a Δhef mutant). However, this does not subtract from the prediction of these proteins as HAN-like; a similar response is seen in *T. kodakarensis* Δhan strains, which show only a slight sensitivity to MMC compared to Δhef cells (Nagata *et al.*, 2017b). This suggests Hef has a function specific to ICL repair that is independent of its role with HAN.

Isolation of RecJ4 and its interacting partners through Strep-tactin affinity purification of strain H5313 (7xHis 2xStrepII RecJ4) was successful; RecJ4 was isolated and confirmed as present by Western blotting. Pulldown of RecJ4 identified peptides of both RecJ2 and RecJ3. This suggests interplay between the RecJ proteins, and this was further backed up using the split GFP assay. A number of proteins localised at the replication fork were also identified, including topoisomerases and Cdc48a. These interaction data suggest all are found at the replication fork, at some point or some level, however it remains to be confirmed what the role of each is here.

Unknown role of RecJ2

In *H. volcanii*, RecJ2 (HVO_1147) is found within a conserved operon containing key genes encoding ribosomal proteins, tRNA and a predicted KEOPS complex subunit. Makarova *et al.* (2012) showed that this gene neighbourhood is seen in numerous archaeal species surrounding the GAN gene (including well-studied model *T. kodakarensis*) (Makarova *et al.*, 2012). Therefore, if genome neighbourhood alone dictated which protein is GAN, this would name RecJ2 as GAN in *H. volcanii*. However, this operonic layout is conserved throughout Haloarchaea, specifically where genes of arCOG00428 are found. Its positioning within a strongly conserved operon containing multiple essential genes may mean it has been maintained purely due to its

operon being selected for as a unit (Gabaldon and Huynen, 2004, Korbel *et al.*, 2004, Wolf *et al.*, 2001).

In haloarchaeal species lacking members of arCOG00427, arCOG00428 proteins have been shown to facilitate the role of GAN (Makarova *et al.*, 2012). However, *H. volcanii* encodes members of both arCOG00427 (*recJ1*) and arCOG00428 (*recJ2*). It therefore is not clear which arCOG gene is here playing the role of GAN; does arCOG00428 only gain the role of GAN where arCOG00427 is absent or can both carry out the role simultaneously?

With respect to protein sequence, when compared to the other *recJ* genes in *H. volcanii*, *recJ2* has no identifiable domains (as identified by Pfam) and has lost all key motifs required for catalytic activity. It is possible that, should arCOG00427 (*recJ1*) be playing the role of GAN, arCOG00428 members have gained novel paralogous roles. It still remains unknown what these roles are and if they are essential.

RecJ2 was previously shown to be essential and could not be deleted using the standard pop-in/pop-out methodology (Lever, 2019). Since *recJ2* is located within a co-transcribed operon of essential genes, it was not possible to place *recJ2* under an inducible promoter to assess its requirement without altering the expression of downstream genes. Instead, integration of an inducible cassette at an ectopic locus (here, *pyrE2*) allowed for induction of ectopic *recJ2* expression to complement the deletion at the wild type locus. Initial attempts to perform the deletion with $\Delta pyrE2::p.tnaM3-recJ2$ failed. It was predicted that this was due to the low activity of the mutant promoter; the high expression of RecJ2 required utilisation of the full-strength promoter *p.tnaA*.

While this strain was stable during high-level induction of RecJ2 (as seen by total protein staining and growth curves), it was still not possible to delete the wild type *recJ2* gene. Numerous merodiploid candidates were isolated which, when screened by PCR, appeared to carry only the deletion cassette. PCR involves amplification, which gives a non-linear readout and fails to reflect the presence of minority alleles in a merodiploid polyploid genome. The PCR method is non-quantitative and often fails to identify merodiploid cells, which can be mistaken for true gene deletions. However, Southern blotting revealed a subset of genome copies still encoded *recJ2*. It is likely the level of expression is high enough to complement the wild type expression level, and therefore the failure to delete *recJ2* is likely due to the deletion construct itself. It could be that integration of the deletion construct is somehow affecting downstream transcription of essential genes or deleting processing signals for surrounding genes within the coding sequence of *recJ2*.

Going forward, it would be pertinent to redesign the deletion construct, potentially making varying sizes of truncations.

Isolation of RecJ2 and its interacting partners through Strep-tactin affinity purification of strain H5237 ($\Delta pyrE2::p.tnaA-7xHis$ 2xStrepII RecJ2) was successful; RecJ2 was isolated and confirmed as present by Western blotting and mass spectrometry. RecJ2 co-purified with both GINS and MCM; this may be strong evidence for RecJ2 having a GAN-like role in *H. volcanii*. However, pulldowns of RecJ1 and RecJ4, plus split GFP data, showed a high level of interaction between all RecJ proteins so this cannot be confirmed using pulldown data alone. Alongside purification of these critical replication factors, pulldown of RecJ2 also isolated a large number of proteins related to translation (including tRNA ligases, proteins involved in amino acid metabolism, and subunits of the core translation machinery itself). The *recJ2* operon encodes a predicted member of the KEOPS complex, which acts in tRNA processing, and as such a link can be directly made between products of the RecJ2 pulldown and other members of the operon; it remains to be confirmed whether there is a direct link between RecJ2 and the KEOPS complex, but this warrants further work. Currently, it cannot be ruled out that, while some RecJ2 is highly likely to be localised at the replication with MCM and GINS, RecJ2 may be playing a role in translation. Whether this is the case, and a consequence of its positioning within a co-transcribed operon containing numerous essential translation proteins, remains to be defined.

Split-GFP interactions between RecJ proteins

Split-GFP utilises two fragments of full-length GFP, each tagged onto a protein of interest, which alone would not fluoresce to show *in vivo* interactions. The two fragments, if brought into proximity, can fold into full-length GFP and fluoresce, and this acts as a direct readout for protein interaction within a cell. This method was recently optimised for *H. volcanii* (Winter *et al.*, 2018) and was utilised here to investigate interactions between the four RecJ proteins.

Using a combination of protein pulldowns and split-GFP, we were able to identify interactions *in vivo* between RecJ1, RecJ2, RecJ3 and RecJ4. Interactions between RecJ3 and RecJ4 have been previously observed (Julie Maupin-Furlow, personal communication), but here we have shown a strong link between these proteins, RecJ1 (via pulldown and split-GFP), and RecJ2 (via split-GFP only). Interactions between all RecJ proteins in *H. volcanii* was somewhat unexpected; while RecJ3 and RecJ4 have been identified as interacting previously, this was somewhat expected due to their shared structure (and potentially function). Both HAN and GAN are predicted to act

at the replication fork, however interaction between HAN and GAN has not previously been reported. The most surprising result, however, was the interaction of RecJ2 with RecJ1, RecJ3, and RecJ4.

Visualisation of these interactions via microscopy revealed compartmentalised GFP signals, forming multiple distinct foci within each cell. It will be pertinent to explore whether such compartments are membrane-bound; this could easily be carried out using a membrane stain, such as FM464. Should these compartments exist, it will be important to assess whether DNA replication is occurring within these compartments and assess the impact this may have on our understanding of replication in *H. volcanii*.

While split-GFP revealed interactions, the protocol itself was not without its flaws. The lack of protein tags within the construct or antibodies specific to *H. volcanii* RecJ proteins meant analysis of protein levels in strains carrying ^NGFP and ^CGFP episomes was not possible. Since we have no measure of protein level, it could be that overexpression of the protein is not tolerated by the cell. In this case, a lack of GFP signal may not be due to lack of interaction, but due to lack of stable protein expression. Quantitative analysis of GFP signal, in the absence of protein stability data, is therefore subject to question.

In addition, split-GFP episomes were transformed into wild type strain H26 (Δ *pyrE2 recJ*⁺). The gene was placed under strong constitutive promoter *p.fdx* and therefore the levels of each protein within the cell would be elevated above the physiological levels. Even if the protein is non-toxic when overexpressed, strong interaction (as defined by the relative fluorescence) may be a consequence of overexpression; the proteins are able to interact when overexpressed, but this interaction may not commonly occur normally, or to a level as reported here. However, aside from such issues, split GFP is a relatively easy experiment to carry out, is relatively quick and requires few resources (importantly does not require radiation for strain confirmation). Therefore, it likely would act as a suitable way to confirm results previously observed by mass spectrometry, as opposed to being standalone proof of interaction.

Correlating flow cytometry data with microscopy

When performing flow cytometry, the suitable control strain was used (i.e., the mutant lacking the additional deletion of the targeted *recJ*). However, differences in culturing between experiments led to questions as to the validity of results and ability to directly compare these. Going forward it would be pertinent to include a wild type control (e.g., H26 [Δ *pyrE2*]) for all flow

cytometry experiments to ensure any differences observed are true and not due to altered preparations on the day. Due to time constraints, it was not possible to confirm any changes in phenotypes observed by flow cytometry by microscopy. This would provide further spatial information on the DNA conformation and thus would be worthwhile to carry out.

4.5 Conclusion

H. volcanii encodes four RecJ proteins (named RecJ1-RecJ4). RecJ1, RecJ3 and RecJ4 are dispensable, while attempts to delete RecJ2 with ectopic expression failed. Affinity purification and split-GFP methodologies suggest a complex relationship between the RecJ proteins of *H. volcanii*, where all interact with one another.

Genetic and bioinformatic analyses suggest RecJ1 is acting during DNA replication as GAN, RecJ3 and RecJ4 are acting with Hef at stalled interstrand crosslinks as HAN proteins, and RecJ2 plays an unknown function.

Chapter 5: Deciphering the role of the CMG replicative helicase complex in *Haloferax volcanii*

5.1 Background

In eukaryotes, the CMG protein complex, consisting of Cdc45, MCM helicase and GINS, is required for active progression of the replicative helicase along DNA during replication (Pacek *et al.*, 2006, Gambus *et al.*, 2006). Interaction of MCM with accessory factors Cdc45 and tetrameric protein complex GINS (consisting of subunits Sld5 and Psf1-3) is required to increase the efficiency with which MCM helicase unwinds the duplex; experiments performed both *in vitro* and *in vivo* have shown such interactions increase the biochemical ability of MCM helicase to unwind DNA (Ilves *et al.*, 2010, Labib and Gambus, 2007, Moyer *et al.*, 2006).

Homologues for all members of the CMG complex have been identified in Archaea (**Table 5.1**). Due to the similarity of replication proteins and core mechanisms utilised by eukaryotes and archaea, it has been proposed that archaea utilise a CMG complex during replication. This is in contrast to bacteria, where DnaB is sufficient to act as replicative helicase.

Table 5.1: Comparison of CMG replicative helicase complex members across all three domains of life. The number of proteins forming the active unit is marked in parentheses. Information taken from (Kelman *et al.*, 2020).

CMG complex member	Bacteria	Eukaryotes	Archaea
Helicase	DnaB (1)	MCM (6)	MCM (1)
Cdc45	-	Cdc45 (1)	GAN/RecJ (1)
GINS	-	GINS (4)	GINS (1-2)

MCM helicase

Eukaryotes utilise six paralogous MCM proteins to form the active MCM(2-7) hexamer, while archaea generally encode a single *mcm* gene and form homohexamers. MCM helicase homologues have been identified in all archaeal species, which is thought to act as core of the replicative helicase (Kelman and Kelman, 2003, Sakakibara *et al.*, 2009b). Where archaeal species encode more than one MCM protein, only one will be essential: the essential MCM protein shares structural and sequence similarity with other MCM proteins known to be essential for viability (e.g., those found in species encoding a single *mcm* gene), while non-essential accessory MCM proteins

Chapter 5: Deciphering the role of the CMG replicative helicase complex in *Haloferax volcanii*

may carry N-terminal extensions (Ishino *et al.*, 2011, Pan *et al.*, 2011b). For example, *T. kodakarensis* encodes three MCM genes (TK0096/*mcm1*, TK1361/*mcm2*, and TK1620/*mcm3*). MCM1 and MCM2 proteins have long N-terminal extensions and are non-essential while MCM3 does not carry an extension, is essential, and is predicted to be the main replicative helicase in this species (Pan *et al.*, 2011b).

Activity

Archaeal and eukaryotic MCM helicases bind the leading strand and travel 3'-5' along DNA, unlike bacterial helicase DnaB which migrates 5'-3' on the lagging strand. Archaeal MCM can bind both ssDNA and dsDNA in its main channel, ssDNA in its side channels, and is able to displace proteins from DNA to continue translocation (Shin *et al.*, 2003, Shin *et al.*, 2007). It has also been shown archaeal MCM can displace RNA from DNA:RNA hybrids (Shin and Kelman, 2006). These are properties expected of a *bona fide* replicative helicase, providing further evidence for MCM acting as the replicative helicase in archaea.

It is well documented that the eukaryotic MCM(2-7) complex does not exhibit helicase activity without the association of Cdc45 and GINS proteins. In archaea several *in vitro* studies have shown archaeal MCM to possess robust helicase activity without the need for additional protein:protein interactions (Kelman *et al.*, 1999a, Carpentieri *et al.*, 2002, Marinsek *et al.*, 2006). However, this is not the case for all species: some MCM proteins, for example that of *P. furiosus*, have very weak helicase activity alone (Yoshimochi *et al.*, 2008). Association of Cdc45 and GINS with MCM in eukaryotes leads to a substantial increase in enzyme activity and processivity, thus the same would be predicted for the archaeal MCM (Ilves *et al.*, 2010). In some species, this applies: *P. furiosus* MCM activity is stimulated upon interaction with GINS (Yoshimochi *et al.*, 2008) and *T. kodakarensis* MCM activity increases upon interaction with its predicted Cdc45 orthologue GAN (GINS-associated nuclease) and GINS (Nagata *et al.*, 2017a). However, *in vitro* studies of *S. solfataricus* have shown no stimulation of helicase activity in the presence of GINS or Cdc45 (Marinsek *et al.*, 2006).

Structure

Compared to eukaryotic MCM, archaeal MCM proteins are shorter (~650 residues in archaea vs. ~900 residues for eukaryotes). Archaeal MCM comprises three sections: a ~250 amino acid N-terminus (consisting of subdomains A-C), a ~300 amino acid catalytic region and a C-terminal helix-turn-helix motif (HTH) domain of ~100 amino acids (**Figure 5.1**) (Sakakibara *et al.*, 2009b, Brewster and Chen, 2010). The N-terminal and catalytic regions

can be mapped to the eukaryotic counterpart, however the HTH motif appears to be specific to archaea (Sakakibara *et al.*, 2009b).

Similarity in structure between archaeal and eukaryotic MCM proteins

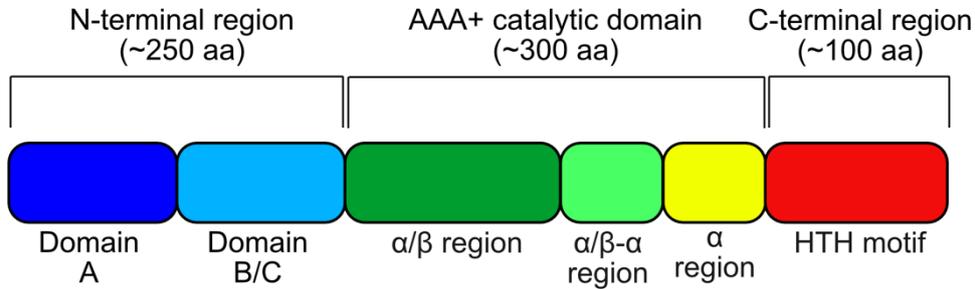


Figure 5.1: Domain organisation of archaeal MCM proteins. The N-terminal region is involved in DNA binding and multimerization, the AAA+ region encompasses the catalytic domain and the C-terminus is unique to archaeal MCM and encodes a predicted helix-turn-helix (HTH) motif. Broad regions are noted at the top, while structural motifs are annotated by coloured blocks. aa, amino acid. Adapted from (Kelman *et al.*, 2020).

allows prediction of the role of the archaeal MCM: the N-terminus predicted to act in DNA binding and multimerization, and the catalytic domain predicted to facilitate the unwinding ability of MCM. The N-terminal domain and AAA+ helicase domain have been shown to interact with one another via a conserved loop structure, known as the allosteric control loop (ACL) (Brewster *et al.*, 2008, Sakakibara *et al.*, 2009b). The archaeal-specific C-terminal HTH structure has been proposed to have a regulatory role and has been implicated in recruitment of MCM to origins of replication (Jenkinson and Chong, 2006, Barry *et al.*, 2007, Samson *et al.*, 2016).

Archaeal MCM proteins are able to form a range of structures in solution, however the hexameric MCM is usually associated with unwinding; in species *Methanothermobacter thermautotrophicus*, only hexameric MCM possesses helicase activity (Shin *et al.*, 2009). It is known that MCM hexamers are loaded at origins by the origin recognition complex (ORC) as a double hexamer, specifically in a head-to-head orientation (Evrin *et al.*, 2009, Noguchi *et al.*, 2017). Electron microscopy has revealed interactions between ORC and the C-terminus of MCM facilitate loading of two hexamers of MCM to an origin (Miller *et al.*, 2019). Here, the recruitment of one MCM molecule instigates recruitment of the second hexamer in the correct orientation through formation of a temporary interaction site in DNA (Miller *et al.*, 2019).

Chapter 5: Deciphering the role of the CMG replicative helicase complex in *Haloferax volcanii*

The near-full length structure of *S. solfataricus* MCM provided further evidence for archaeal MCM sharing structural similarity with the eukaryotic counterpart (Brewster *et al.*, 2008). The AAA+ catalytic domain was intact, with all conserved motifs being present. Each monomer was also shown to encode four β -hairpins per monomer, three positioned within the main channel and one externally. These hairpins have not been observed in eukaryotic MCM(2-7) and are predicted to be a feature specific to archaeal MCM (Brewster *et al.*, 2008). Mutational analysis has since confirmed the hairpins play a key role in DNA binding and helicase activity (Brewster and Chen, 2010). Whether this difference in structure explains the reduced requirement for priming by Cdc45/GINS remains, as yet, unknown.

The crystal structure of full-length MCM of *S. solfataricus* has now been solved (**Figure 5.2**) (Meagher *et al.*, 2019) and confirmed that ATP binding and hydrolysis occurs at an interface formed between two monomers of MCM; one provides the tri-phosphate binding loop (P-loop; including the Walker A/B motifs) to bind ATP and the other contributes residues *in trans* to interact with ATP (Brewster *et al.*, 2008, Iyer *et al.*, 2004). Akin to bacterial helicase DnaB, it has been suggested the helicase moves with a step of two nucleotides per MCM subunit (and thus 16 nucleotides per double hexamer).

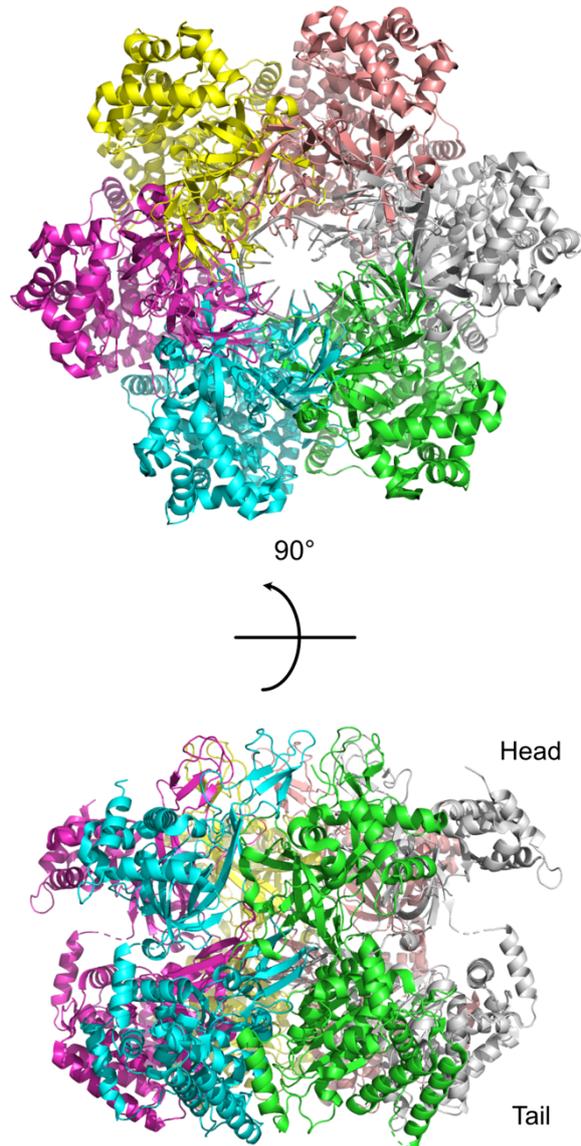


Figure 5.2: Crystal structure of an MCM hexamer from *Sulfolobus solfataricus* in the presence of ssDNA. Ribbon diagram of *S. solfataricus* MCM (PDB ID: 6MII). The ssDNA is found within the central channel and is coloured in grey. Each monomeric chain is coloured differently. The head region is where the double hexamer interactions occur, with the double hexamer having a head-to-head orientation.

GINS

GINS complex (named after Japanese numbers 5-1-2-3 Go-Ichi-Ni-San, representing subunits Sld5, Psf1, Psf2 and Psf3 respectively) is known to play a key role in eukaryotic replication. It has been seen to directly associate with key eukaryotic replication components, including Pol α , Pol ϵ , Pol δ and

Chapter 5: Deciphering the role of the CMG replicative helicase complex in *Haloferax volcanii*

MCM helicase (Aparicio *et al.*, 2009, Bermudez *et al.*, 2011, De Falco *et al.*, 2007).

The four subunits of eukaryotic GINS are encoded by four distinct genes, are distantly related to one another at the protein sequence level, and are predicted to be paralogous (Makarova and Koonin, 2005). The four subunits can be clustered into groups depending on arrangement of specific protein sequences; these regions map to distinct domains in the crystal structure of the proteins (MacNeill, 2010). Within a single subunit there are two domains: an A-domain and a B-domain. The A-domain has been structurally defined as groups of α -helices, while B-domains are smaller and rich in β -strands. The domain order changes depending on subunit type; α subunits usually have the order AB, while β subunits are BA-type. The A- and B-domains are separated by an interdomain loop, the length of which varies depending on species (MacNeill, 2010).

Archaeal GINS, as with MCM, is a simplified form of its eukaryotic counterpart. Archaeal GINS complex forms a tetrameric structure similar to that of eukaryotes, however it is encoded by fewer genes (only one or two). All archaeal species encode at least one GINS protein, Gins51, with numerous species encoding a second GINS protein, Gins23, which differ in domain organisations (MacNeill, 2011, Marinsek *et al.*, 2006, Makarova and Koonin, 2005). In archaea, two types of GINS tetramers exist (**Table 5.2**):

- i) Heterotetramers, where two GINS genes are present: Gins51 (homologous to eukaryotic Sld5 and Psf1) and Gins23 (homologous to Psf2 and Psf3), forms a dimer of dimers ($\alpha_2\beta_2$) GINS complex
- ii) Homotetramers, where only a single GINS gene is present. The encoded protein shares similarity to the Gins51 subunit and forms a homotetramer (α_4) GINS complex

Table 5.2: Domain organisation of GINS in human (*Homo sapiens*), *Thermococcus kodakarensis* and *Haloferax volcanii*. Long rectangles represent A-domains, composed of α -helices. Arrows represent B-domains, rich in β -strands. Domains are separated by an interdomain loop, represented by a black line.

	Tetramer form	Subunits	
		α (AB-type)	β (BA-type)
Human	$\alpha\alpha'\beta\beta'$	Sld5 	Psf2 
		Psf1 	Psf3 
<i>T. kodakarensis</i>	$\alpha_2\beta_2$	Gins51 	Gins23 
<i>H. volcanii</i>	α_4	GINS 	

The crystal structure of archaeal GINS was solved for species *T. kodakarensis* and is directly comparable to that of human GINS complex in terms of protein structure and planes of symmetry (**Figure 5.3**) (Oyama *et al.*, 2011, Kamada *et al.*, 2007).

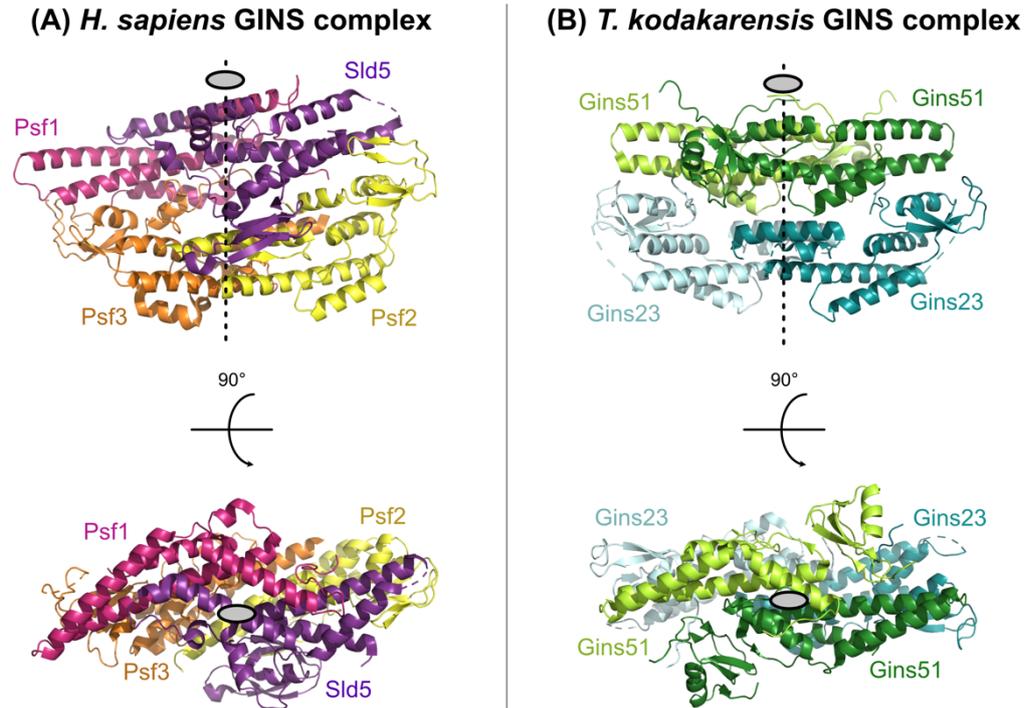


Figure 5.3: Crystal structure of *Homo sapiens* and *Thermococcus kodakarensis* GINS complexes. Ribbon representation of the tetrameric structure of GINS in *Homo sapiens* (PDB ID: 2E9X) and *Thermococcus kodakarensis* (PDB ID: 3ANW). Subunits are coloured as labelled in **Table 5.2** (Psf1, pink; Sld5, purple; Psf3, orange; Psf2, yellow; Gins51, green; Gins23, blue). Dashed lines represent missing parts of the structure. The two-fold axis of symmetry is marked by a grey oval and dotted line. The eukaryotic and archaeal GINS complexes are directly comparable in terms of protein structure, subunit assembly and symmetry planes.

In *T. kodakarensis*, GINS forms a tetrameric structure consisting of Gins51 and Gins23 dimers. The crystal structures of both human and archaeal GINS suggest the C-terminal domains of Psf1/Gins51 are mobile (Oyama *et al.*, 2011, Kamada *et al.*, 2007). This mobility, along with extended interdomain connecting loops, are likely to aid GINS complex in its functions during replication (Sengupta *et al.*, 2013, Oyama *et al.*, 2016). However, it is worth noting that while eukaryotic and archaeal GINS complexes are highly similar, some of the specific contacts between subunits and positions differ (Bell, 2011, Oyama *et al.*, 2011).

Interaction between MCM and GINS in archaeal species encoding both Gins51 and Gins23 has been shown to be mediated by BA-type subunit Gins23, where interaction between MCM and GINS boosts the ATPase and helicase activities of MCM (Marinsek *et al.*, 2006, Yoshimochi *et al.*, 2008). Conversely, species encoding only Gins51-type subunits, for example

Thermoplasma acidophilum, carry out this interaction in a Gins51-dependent manner (Ogino *et al.*, 2011, Ogino *et al.*, 2014). Such a Gins51-dependent interaction is yet to be observed in species encoding the two-subunit GINS complex.

Cdc45/RecJ

Cell division cycle protein 45 (Cdc45) is an essential replication factor in eukaryotes that together with MCM(2-7) and GINS, forms the active CMG replicative helicase complex. Archaeal homologues for MCM and GINS are identified relatively easily, but no obvious Cdc45 homologue is found in archaea.

Bioinformatic studies have elucidated that Cdc45 is the eukaryotic orthologue of bacterial and archaeal RecJ family phosphodiesterase nucleases (Makarova *et al.*, 2012, Pellegrini, 2017). Specifically, the N-terminus of Cdc45 shows amino acid sequence similarity to the DHH phosphodiesterase domain of bacterial and archaeal RecJ proteins, which are 5'-3' exonucleases that function in DNA repair (Sanchez-Pulido and Ponting, 2011, Krastanova *et al.*, 2012). However, the DHH domain of Cdc45 differs from that of canonical RecJ proteins in that it has lost key residues required for catalytic activity (Krastanova *et al.*, 2012, Makarova *et al.*, 2012). Like RecJ exonucleases, Cdc45 can bind ssDNA but has lost the ability to bind dsDNA (Krastanova *et al.*, 2012, Szambowska *et al.*, 2014). The crystal structure of Cdc45 has now been solved (Simon *et al.*, 2016). All archaeal species to date encode at least one RecJ gene, with some species containing multiple distinct RecJ genes.

T. kodakarensis encodes two RecJ-like proteins, known as GINS-associated nuclease (GAN) and Hef-associated nuclease (HAN). In any given species, it is likely that only one RecJ protein will act in the role of Cdc45. GAN was primarily identified as an interaction partner of GINS (Li *et al.*, 2011, Li *et al.*, 2010), while HAN was shown not to interact with GINS, instead interacting with repair protein Hef (Nagata *et al.*, 2017b, Fujikane *et al.*, 2010). GAN, MCM and GINS have since been shown to form a complex *in vitro* (Nagata *et al.*, 2017a). This links GAN to the replication fork, with MCM and GINS as part of the archaeal CMG complex.

GAN has since had its catalytic activity characterised and its crystal structure solved (Nagata *et al.*, 2017a, Oyama *et al.*, 2016). Unlike its eukaryotic counterpart Cdc45, GAN is a catalytically active RecJ protein, with processive 5'-3' ssDNA exonuclease activity (Li *et al.*, 2011). Structurally, it shares similarity with both bacterial RecJ and eukaryotic Cdc45 (**Figure 5.4**), supporting the relatedness of RecJ/DHH proteins across all three domains of

Chapter 5: Deciphering the role of the CMG replicative helicase complex in *Haloferax volcanii*

life. The structural relationship between Cdc45 and GAN, in particular, is very strong, while bacterial RecJ carries an extra bacterial RecJ-specific OB-fold domain not seen in the archaeal and eukaryotic counterparts.

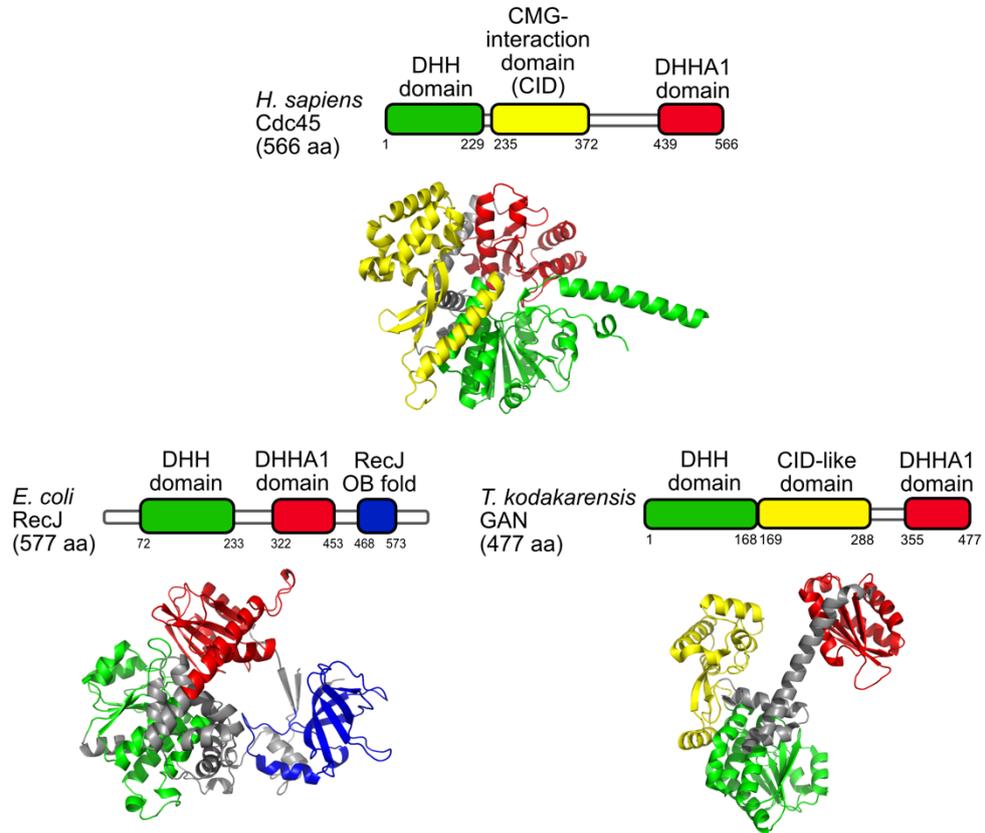


Figure 5.4: Structural comparison of *Homo sapiens* Cdc45, *Escherichia coli* RecJ and *Thermococcus kodakarensis* GAN proteins. Structure and domain information for *H. sapiens* Cdc45 (PDB ID: 5DGO) and *T. kodakarensis* GAN (PDB ID: 5GHT) were annotated according to papers defining their crystal structures (Simon *et al.*, 2016, Oyama *et al.*, 2016). Protein structure for *E. coli* RecJ was mapped using Phyre2 protein modelling software (Kelley *et al.*, 2015). Domain analysis of *E. coli* RecJ protein was carried out using Pfam, and domains were coloured (as above). Domain boxes are not to scale. Cdc45, RecJ and GAN all show structural similarity. However, the lack of OB-fold in both Cdc45 and GAN mean they are more similar in structure to each other compared to RecJ.

Co-crystallisation of GAN with Gins51 showed a strong interaction between the B-domain of Gins51 and the DHH domain of GAN (Oyama *et al.*, 2016). Similar results have been seen in *P. furiosus*, where its GAN-like RecJ protein interacts with Gins51 (at its B-domain) via the DHH domain (Li *et al.*, 2017). This differs from Cdc45, which has been shown to interact with GINS

Chapter 5: Deciphering the role of the CMG replicative helicase complex in *Haloferax volcanii*

complex via its CMG-interaction domain (CID) (Simon *et al.*, 2016). Based on structure alone, GAN does carry a CID, but this domain seems to play a different function in archaeal Cdc45-like proteins as no interactions were mapped to this zone (Oyama *et al.*, 2016).

While there are many similarities between Cdc45 and GAN, there are also key differences. In sharp contrast to eukaryotic Cdc45, GAN has been shown to be non-essential in *T. kodakarensis* (Nagata *et al.*, 2017a, Burkhart, 2017). The nuclease activity of GAN, compared to inactive Cdc45, also raises questions about the function this nuclease may play at an active replication fork.

In crenarchaeon *S. solfataricus*, the GINS complex, specifically Gins51, co-purifies with protein RecJdbh (RecJ DNA-binding homologue), homologous to the DNA-binding domain of bacterial RecJ (Marinsek *et al.*, 2006). In contrast to *T. kodakarensis* GAN and *P. furiosus* RecJ, RecJdbh contains a degenerate DHH domain and lacks catalytic activity. Various *Sulfolobus* species have now been shown to form CMG complexes *in vitro* using these ‘inactive’ RecJ-like proteins, which act to stimulate MCM (Xu *et al.*, 2016, Marinsek *et al.*, 2006), akin to eukaryotic Cdc45. *Sulfolobales* species have been shown to have defined cell cycles and there is an enrichment for RecJdbh at replication origins in early S phase, with origin enrichment decreasing as the replication fork progresses (Xu *et al.*, 2016); the same expression pattern has been observed for Cdc45, GINS and MCM in eukaryotes. RecJdbh shares little similarity to *T. kodakarensis* GAN or *P. furiosus* RecJ, suggesting there may be more than one type of RecJ-like protein acting in replication across archaeal species.

The CMG complex in *Haloferax volcanii*

Haloferax volcanii is known to encode all components required for formation of a CMG replicative helicase complex; MCM, GINS, and RecJ in place of Cdc45/GAN.

Gene *mcm* (HVO_0220) encodes a homo-hexameric MCM helicase that has previously been shown to be essential (Marriott, 2017). By placing the *mcm* gene under low-activity tryptophan-inducible promoter *p.tnaM3*, the level of *mcm* expression required by cells was directly compared between strains replicating with and without origins. Interestingly, it was shown the requirement for MCM helicase altered depending on the mode of replication used: in the absence of origins, where replication is predicted to initiate

randomly throughout the genome, there is an increased requirement for MCM (Figure 5.5) (Marriott, 2017).

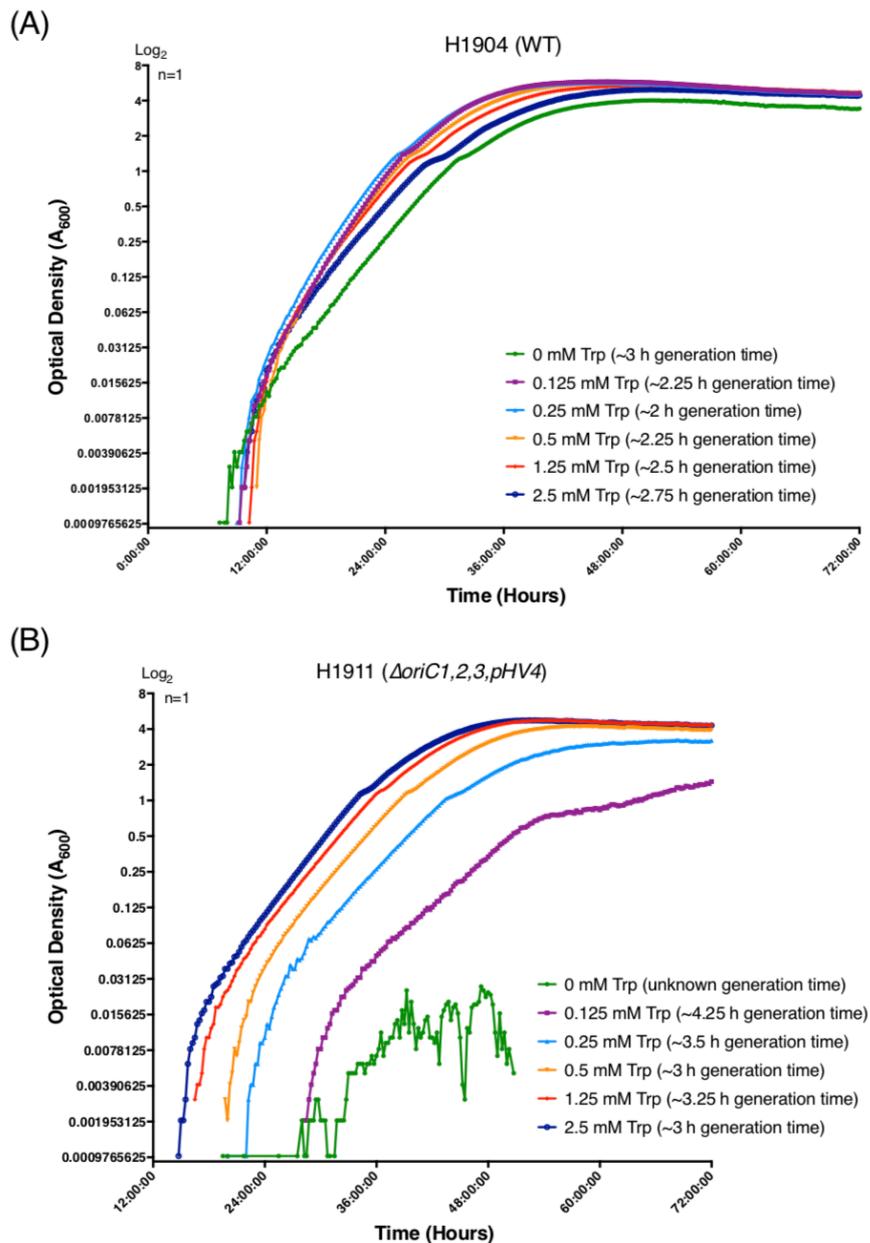


Figure 5.5: Increased requirement for MCM helicase in the absence of origins in *Haloflex volcanii*. Gene *mcm* was placed under the tryptophan-inducible promoter *p.tnaM3* and integrated into strains with (H1904, *oriC*⁺) and without replication origins on its main chromosome (H1911, $\Delta oriC1,2,3,pHV4$). Generation times are shown in hours. (A) The WT strain H1904 shows little change in growth in the absence of origins, suggesting low requirement for MCM during origin-dependent replication. (B) The originless strain H1911 is not viable in the absence of tryptophan (and therefore the absence of *mcm* induction) suggesting originless replication requires an increased level of *mcm* expression. Taken from Marriott, 2017.

This may be explained thus: if the four origins on the main chromosome fire once per round of replication in a wild-type strain, each origin would require two MCM double hexamers (total of 8 hexamers) to unwind DNA bidirectionally away from the origin. This is a relatively low requirement for MCM and thus may explain why growth is still seen with no induction; the *p.tnaM3* promoter is slightly leaky and thus is likely to fulfil this relatively low requirement for MCM molecules in the absence of tryptophan. This is in contrast to originless replication, where replication is primed randomly around the chromosome. Here each unidirectional priming event would only require one double hexamer MCM molecule, however the number of priming events would be predicted to be higher than that of origin-dependent cells, as originless strains have been shown to grow 7.5% than their origin-dependent counterparts. Therefore the requirement for MCM is elevated and, here, greater tryptophan induction of *p.tnaM3::mcm* is required.

H. volcanii encodes a homotetrameric GINS complex (gene *ginS*, HVO_2698), predicted to be structurally similar to the Gins51 subunit of *T. kodakarensis* (**Figure 5.6**). Generally, halophilic GINS proteins are found to be larger than other archaeal and eukaryotic counterparts due to a sequence insertion between the A and B domains, the function of which remains unknown (MacNeill, 2009). In eukaryotes, GINS is well established as being an essential component of replication (Takayama *et al.*, 2003, Kanemaki *et al.*, 2003) and thus essentiality is predicted in *H. volcanii*.

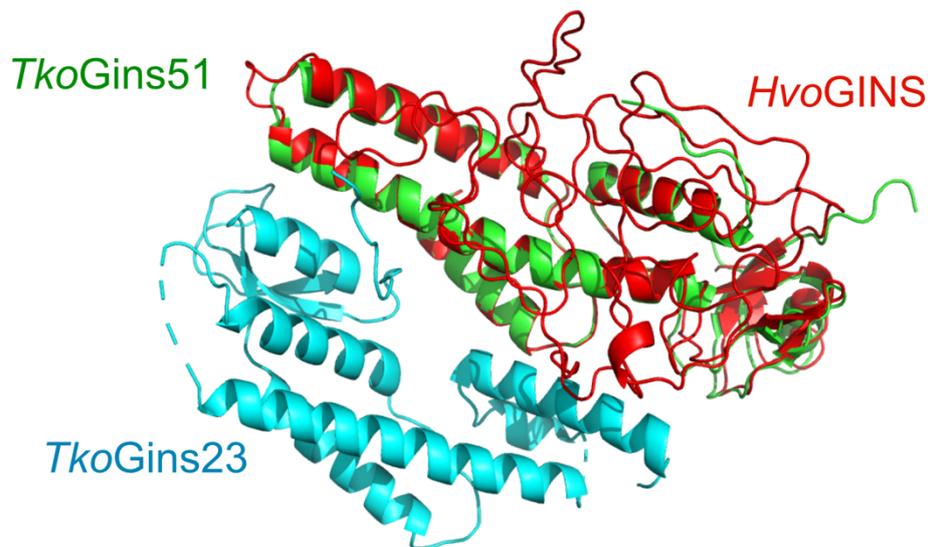


Figure 5.6: Comparison of protein structures of GINS monomers in *Thermococcus kodakarensis* (*Tko*) and *Haloferax volcanii* (*Hvo*). Crystal structures of *TkoGins51* (green) and *TkoGins23* (teal) (PDB ID: 3ANW) compared to the Phyre2 predicted structure of *HvoGINS* show *HvoGINS* is more similar to Gins51 (AB-type subunit) than Gins23 (BA-type).

The situation regarding the putative Cdc45-like protein in *Haloferax* is more complex. *H. volcanii* encodes four RecJ-like genes (*recJ1-recJ4*, with locus tags HVO_0073, HVO_1147, HVO_1018 and HVO_2889, respectively). As previously discussed in *Chapter 4*, these RecJ genes differ in catalytic ability and domain structure and work is ongoing to elucidate which RecJ(s) play the role of GINS/GAN in *H. volcanii*.

5.2 Aims and Objectives

While it is assumed that *Haloferax volcanii* utilises the CMG replicative helicase complex to unwind the DNA duplex, this has not been proven experimentally. The differing requirement for MCM helicase in the absence of origins hints at alternative methods of DNA replication being utilised in the absence of origins. The objectives of this chapter are:

- Confirm the presumed essentiality of GINS
- If shown to be essential, test whether the requirement for *ginS* increases in the absence of origins, using a tryptophan-inducible system (akin to MCM)
- Generate tagged inducible systems to correlate induction level with protein level by Western blot
- Use protein affinity purification to isolate GINS-interacting partners, with the aim of defining which RecJ protein acts at the replication fork

5.3 Results

5.3.1 Deletion of *ginS*

GINS is a key component of the CMG complex and is thus predicted to be essential; this is the case in eukaryotes where interaction between MCM and GINS is critical for helicase activity. The assumption would be that *ginS* is also essential in *H. volcanii*, however this has yet to be tested.

To test the presumed essentiality of *ginS*, deletion of *ginS* (HVO_2698) was attempted using tryptophan-marked deletion construct pTA2335 (Figure 5.7). This construct was previously constructed by Rebecca Lever in 2019 (Lever, 2019).

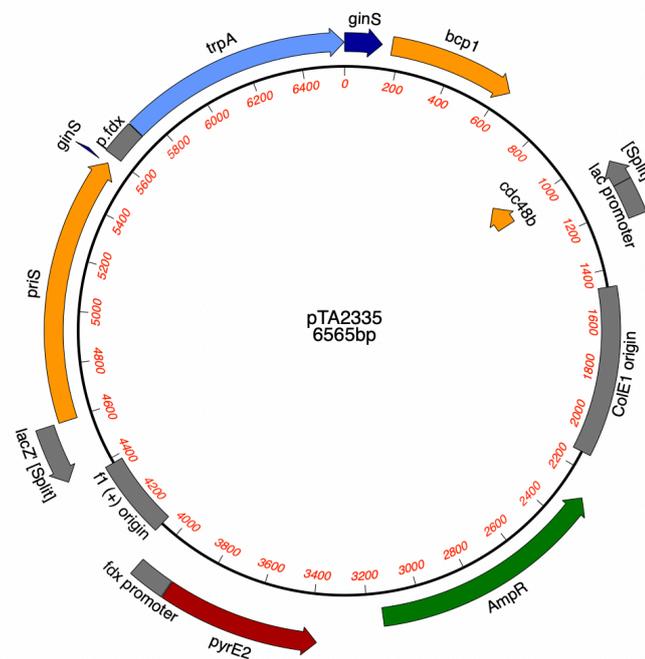


Figure 5.7: pTA2335. Tryptophan-marked deletion construct for *ginS* (Lever, 2019).

GINS is co-transcribed with genes *priS* (upstream) and *bcp1* (downstream) and, as such, the deletion construct had to be designed to ensure integration of the deletion construct would not affect the surrounding genes. The coding sequences of *priS* (HVO_2697) and *ginS* (HVO_2698) overlap by four base pairs. Gene *priS* encodes a subunit of primase, a known essential gene, and thus it is critical its reading frame is maintained. The deletion construct for *ginS* ensured that integration of the deletion construct maintains the complete *priS* coding sequence, this was accomplished by duplicating the overlapping *priS* stop codon/*ginS* start codon within the upstream sequence portion of the construct (Figure 5.8). Additionally, the terminal 136 bp of *ginS*

was included in the construct as it likely contains the promoter for downstream *bcp1*. This gene encodes peroxiredoxin and is predicted to be essential, thus it was important to include its promoter in the attempt to delete *ginS*.

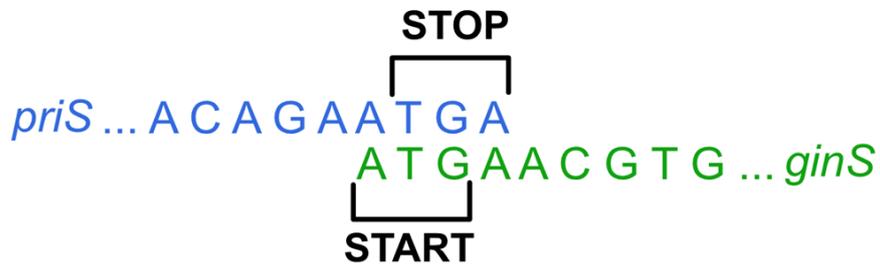
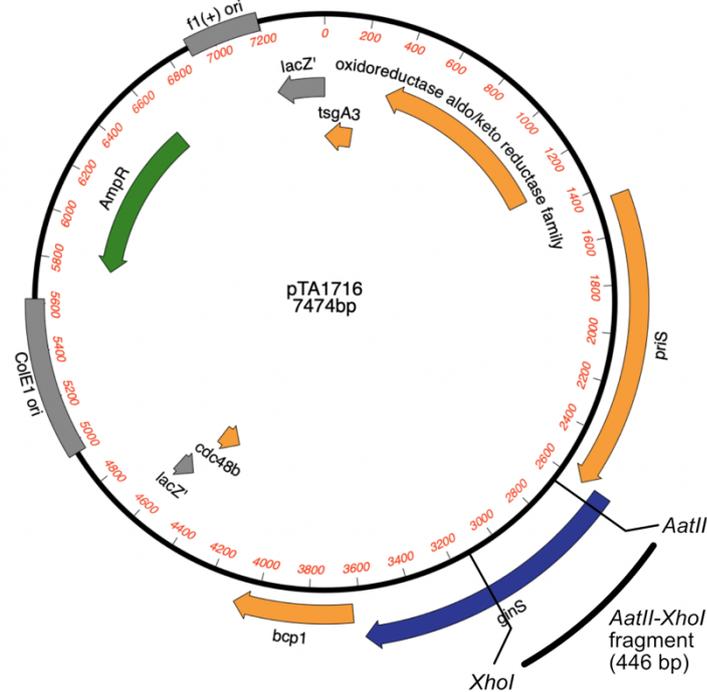


Figure 5.8: *priS*-*ginS* sequence overlap. The stop codon of *priS* overlaps with the start codon of *ginS*, meaning deletion of the entirety of *ginS* coding sequence will result in the mistranslation of upstream gene *priS*.

Strains H4045 (*oriC*⁺) and H4598 (Δ *oriC*) were transformed with pTA2335 to generate pop-in strains H4730 and H4732 respectively. Pop-out colonies were screened by colony hybridisation with a *ginS*-specific probe. Each strain was subjected to pop-out and 160 colonies per strain were screened to ensure 90% confidence of isolating a rare deletion mutant, assuming a rate of 1/70 for hard-to-delete genes (as previously explained in **Chapter 3 [Equation 3.1]**). The *ginS*-specific probe was generated by restriction digest of genomic clone pTA1716 using enzymes *Aat*II and *Xho*I (**Figure 5.9**). No deletion candidates for GINS were isolated, suggesting GINS is essential in *H. volcanii*.

(A) GINS probe



(B) Colony hybridisation



Figure 5.9: Colony hybridisation of $\Delta ginS::trpA+$ candidates. (A) Patches were probed with the *AatII-XhoI* fragment of pTA1716 (*ginS* genomic clone; Marriott, 2017). (B) All pop-outs hybridised with the probe, suggesting all clones carry at least one copy of the wild type *ginS* sequence.

5.3.2 Assessing the requirement for *ginS* expression using inducible promoters

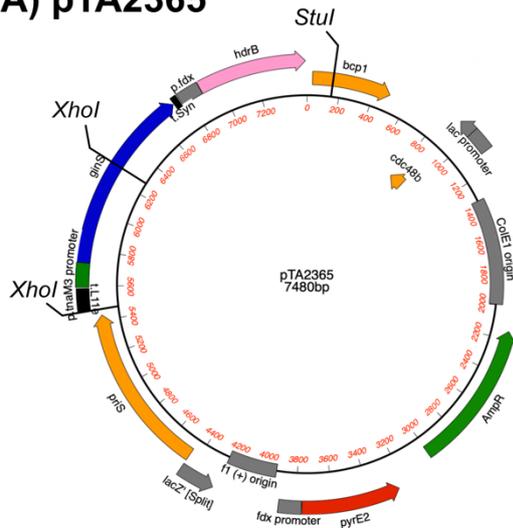
Due to the inability to delete *ginS*, it was assumed its expression was essential. As a component of the CMG complex in *H. volcanii* this was unsurprising and fits with the essentiality of the core complex protein *mcm* (Marriott, 2017). Previous work has shown the requirement for MCM, the main replicative helicase in *H. volcanii*, is increased in the absence of origins (Marriott, 2017). As MCM and GINS are both members of the CMG complex it could be predicted the requirement for expression of *ginS* will match that of *mcm*, with the requirement for the CMG complex as a whole increasing in the absence of origins. Alternatively, *mcm* alone could have gained a novel, *ginS*-

independent role in originless strains and therefore no change in the requirement for *ginS* would be expected. To test this, a plan was designed to place *ginS* under the control of the low-activity tryptophan-inducible promoter, *p.tnaM3*, both with and without protein tags (as these would allow additional correlation of protein expression with induction).

Generation of an inducible *ginS* construct

The coding sequence of *ginS* was amplified by PCR from genomic clone pTA1716 (Marriott, 2017). Primers *ginSfwdNde* and *ginSrevDSBam* amplified the 1016 bp of *ginS* sequence, integrating a 5' *NdeI* site and 3' *BamHI* site for downstream cloning. The PCR product was digested with *NdeI* and *BamHI* and inserted into *p.tnaM3* cloning vector pTA1451 (Braun *et al.*, 2019) at *NdeI/BamHI* sites, giving rise to intermediate product pTA2361 (**Figure 5.10 A**). The 2020 bp *p.tnaM3-ginS-hdrB* cassette was digested from pTA2361 using *BglIII* and was inserted into Δ *ginS* vector pTA2184 at its compatible *BamHI* site to add US/DS genomic sequences to facilitate integration onto the chromosome (**Figure 5.10 B**). This gave rise to the final *p.tnaM3-ginS::hdrB* construct, pTA2365 (**Figure 5.11**).

(A) pTA2365



(B) *StuI* *XhoI* digest

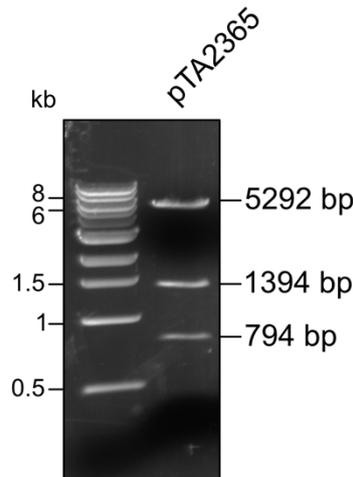
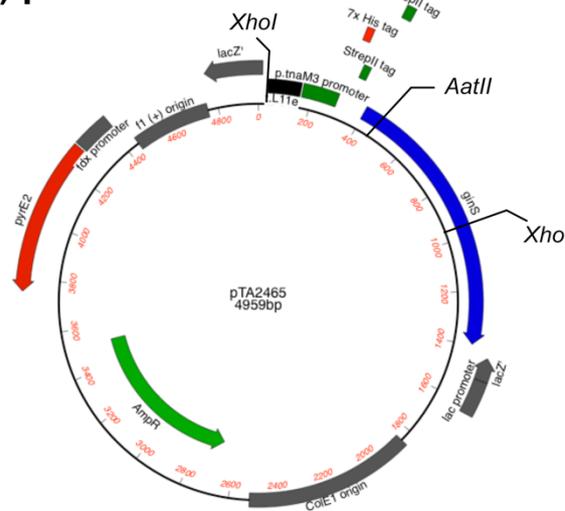


Figure 5.11: pTA2365. (A) pTA2365 was constructed by placing the *p.tnaM3::ginS::hdrB* cassette from pTA2361 into Δ *ginS* vector pTA2184. (B) Digestion of pTA2365 with *StuI* and *XhoI* gave rise to bands at 5292 bp, 1394 bp and 794 bp, as expected.

Generation of a 7xHis 2xStrepII-tagged inducible *ginS* construct

To correlate the induction of *p.tnaM3-ginS* with GINS protein levels, cloning was performed to place a tandem 7xHis 2xStrepII protein tag at the N-terminus of *ginS*, downstream of the inducible promoter, to allow direct visualisation of GINS expression levels via Western blotting. The coding sequence of *ginS* was amplified by PCR from genomic clone pTA1716 ((Marriott, 2017)). Primers *gintagGBspF* and *gintagGEcoR* amplified the 1016 bp of *ginS* sequence, integrating novel 5' *Bsp*HI and 3' *Eco*RI sites. The PCR product was digested with *Bsp*HI and *Eco*RI and inserted into *p.tnaM3-7xHis 2xStrepII* cloning vector pTA2096 (Lever, 2019) at compatible *PciI/EcoRI* sites, giving rise to intermediate product pTA2465 (**Figure 5.12**). The 1385 bp *p.tnaM3-7xHis 2xStrepII-ginS* cassette was digested from pTA2465 using *Bgl*II and *Bam*HI and was inserted into Δ *ginS* vector pTA2184 at its compatible *Bam*HI site, to add US/DS genomic sequences that enable integration onto the chromosome. This gave rise to *p.tnaM3-7xHis 2xStrepII-ginS* construct, pTA2479 (**Figure 5.13**).

(A) pTA2465



(B) *AatII XhoI* digest

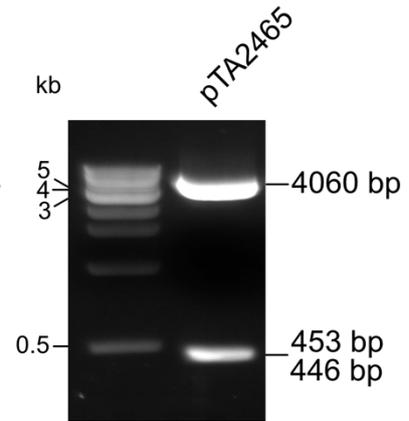
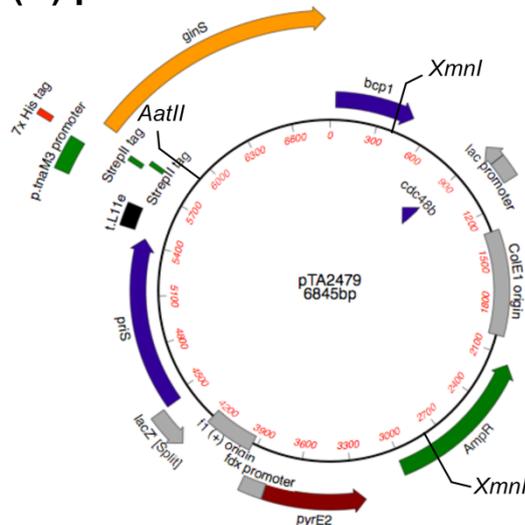


Figure 5.12: Construction of pTA2465 (A) Intermediate construct pTA2465, placing *ginS* under control of *p.tnaM3*, with 7xHis 2xStrepII N-terminal tags. (B) *AatII XhoI* digest shows bands at 4060 bp, 453 bp and 446 bp, as predicted.

(A) pTA2479



(B) *AatII XmnI* digest

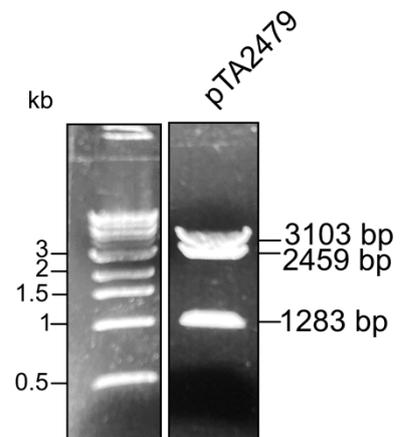


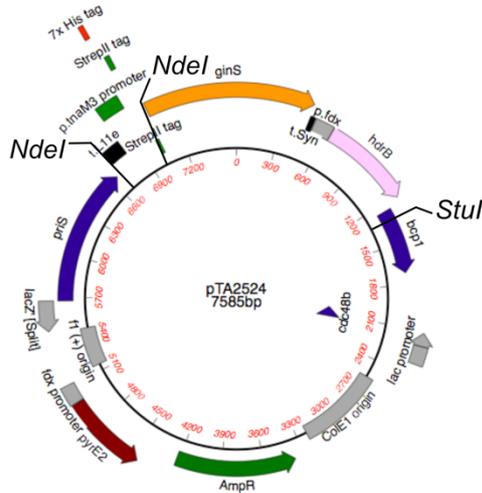
Figure 5.13: Construction of pTA2479 (A) Intermediate construct pTA2479, containing *p.tnaM3*-7xHis 2xStrepII-*ginS* with US and DS genomic sequence. (B) *AatII XmnI* digest shows bands at 3103 bp, 2459 bp and 1283 bp, as predicted.

Due to the absence of the selectable *hdrB* marker in *p.tnaM3*-7xHis 2xStrepII cloning vector pTA2096, this was added subsequently to construct pTA2479. Plasmid pTA2479 was digested with *XhoI* to isolate the 899 bp *p.tnaM3*-7xHis 2xStrepII cassette and was swapped for the 794 bp *XhoI* fragment of the untagged *hdrB*-marked inducible *ginS* construct pTA2365.

Chapter 5: Deciphering the role of the CMG replicative helicase complex in *Haloferax volcanii*

This final step gave rise to the final construct for *p.tnaM3-7xHis 2xStrepII-ginS*, pTA2524 (**Figure 5.14**).

(A) pTA2524



(B) *NdeI StulI* digest

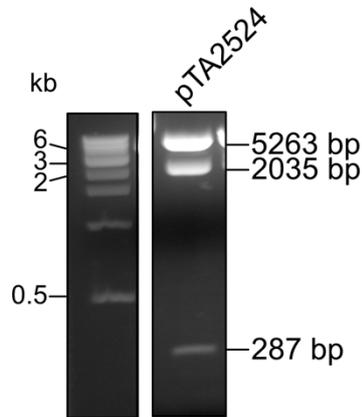


Figure 5.14: pTA2524. (A) Gene replacement construct for placing tandem tagged His₇ 2xStrepII-*ginS* under the control of inducible promoter *p.tnaM3*. (B) Diagnostic digest with *NdeI* and *StulI* gives bands of 5263 bp, 2035 bp and 287 bp, as predicted.

Inducible *ginS* strain construction

To ensure correct integration of the inducible *p.tnaM3-7xHis 2xStrepII-ginS::p.fdx-hdrB* cassette and loss of the wild type *ginS* promoter, pop-in orientation must be screened prior to pop-out. Where an upstream (US) pop-in has occurred, only a DS pop-out event will be capable of thymidine synthesis (*hdrB*⁺), and result in correct integration of the construct. Due to the orientation of the US pop-in, a downstream pop-out has an increased likelihood. Candidates were screened for orientation by colony PCR using primers *priSintF* and *ginSintR* to isolate US pop-ins. An US pop-in of *p.tnaM3-ginS* gives a product of 991 bp, *p.tnaM3-7xHis 2xStrepII-hdrB* gives a product of 1096 bp, while a DS or gene pop-in for both constructs would give a product of 694 bp (untagged example shown in **Figure 5.15**).

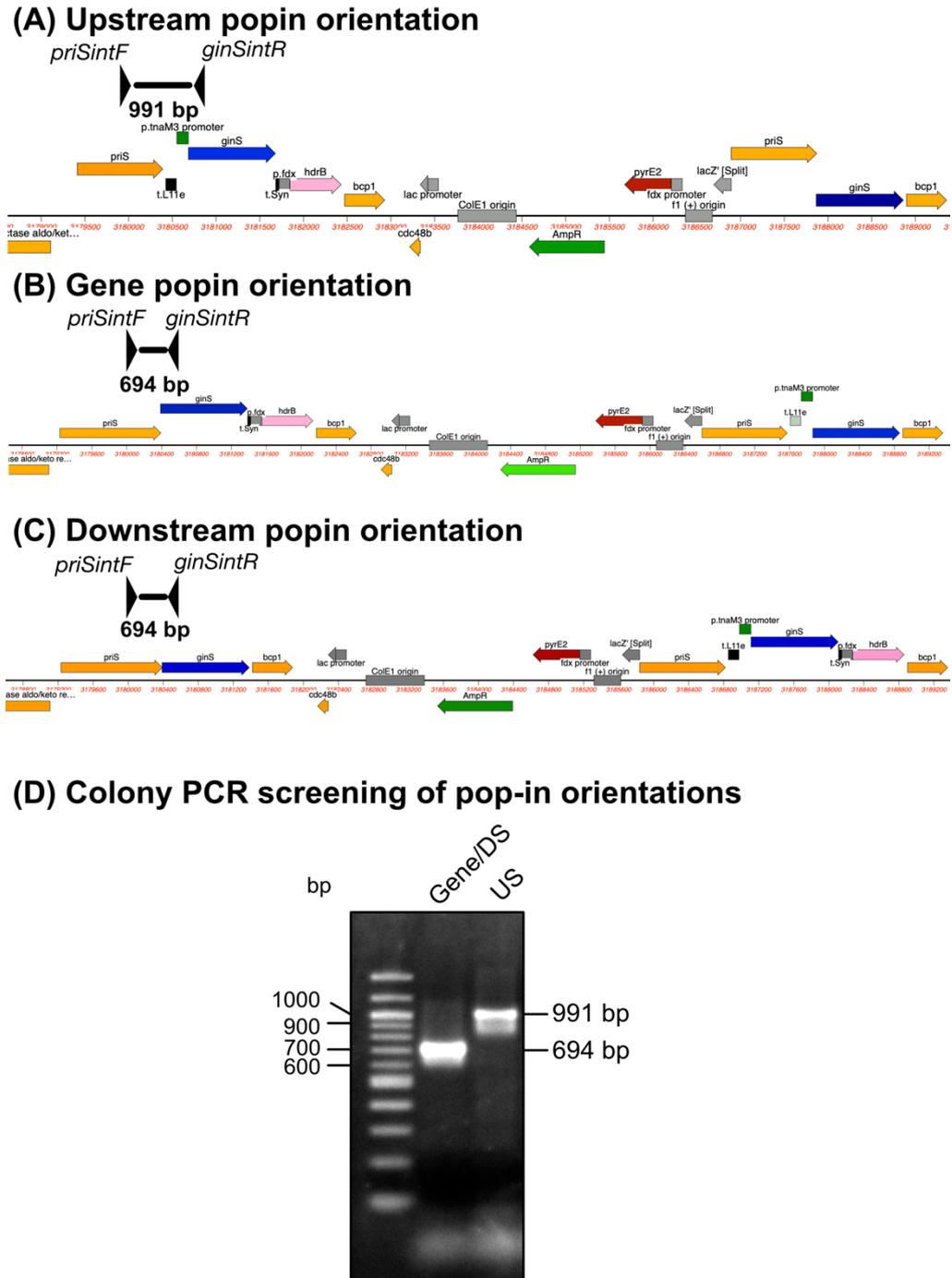


Figure 5.15: Genotyping of pop-in candidates for *p.tnaM3-ginS::hdrB*. Colony PCR performed against pop-in candidates with primers *priSintF* and *ginSintR*. (A) A pop-in event between the chromosome and the US sequence within pTA2365 gives rise to a PCR product of 991 bp. (B) A pop-in event between the chromosome and the *ginS* sequence within pTA2365 gives rise to a PCR product of 694 bp. (C) A pop-in event between the chromosome and the DS sequence within pTA2365 gives rise to a PCR product of 694 bp. (D) An example of the PCR screen showing a DS or gene pop-in product of 694 bp and an US pop-in product of 991 bp.

Inducible *ginS* constructs pTA2365 and pTA2524 were transformed into wild-type (*ori+*; H4829) and originless (*ori-*; H4832) strains. Screening of pop-in candidates for pTA2365 generated pop-in strains H4920 and H5271 respectively. Screening of pop-in candidates for pTA2524 generated pop-in strains H5129 and H5272 respectively.

Pop-outs gave rise to colonies on 5-FOA, which were primarily screened using tryptophan selection, due to the expected essentiality of *ginS*. Where the *p.tnaM3* promoter has correctly integrated onto the chromosome, growth would be expected on the *trp+* plates where *ginS* induction occurs, while in the absence of induction *ginS* expression does not occur and thus the strain should not grow on *trp-* plates.

Repeated pop-out events yielded only colonies that were capable of growth on both *trp+* and *trp-* plates. Due to the repeatability of the result, candidates were further screened by colony PCR as previously, using primers *priSintF* and *ginSintR*. Of 80 candidates screened for pop-out of *p.tnaM3-ginS::hdrB*, only 12 were wild type while the remaining 68 carried the promoter (85% success). This suggested, unexpectedly, that cells remained viable in the absence of *ginS* induction. Integration was further confirmed by Southern blot (**Figure 5.16**). Pop-in strain H4920 gave rise to strain H5017 (*oriC+*), while pop-in strain H5271 gave rise to strain H5383 ($\Delta oriC1,2,3,pHV4$). For tagged inducible strains, pop-in H5129 gave rise to strain H5312 (*oriC+*), while pop-in strain H5272 gave rise to strain H5385 ($\Delta oriC1,2,3,pHV4$).

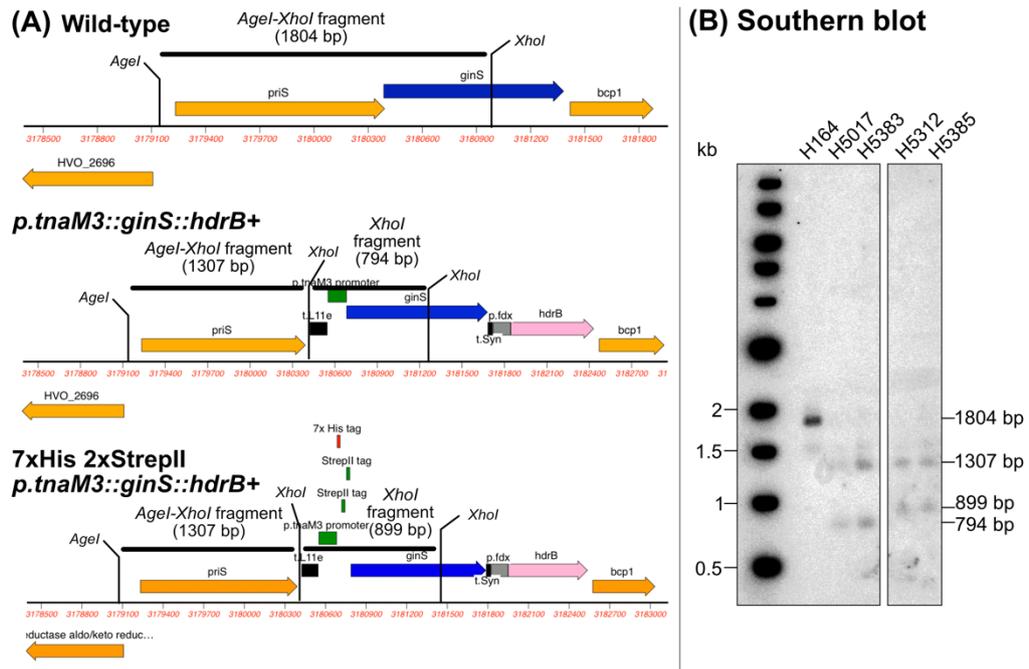


Figure 5.16: Confirmation of integration of *p.tnaM3-ginS::hdrB*. (A) Expected Southern blot band sizes for *AgeI-XhoI* digested genomic DNA. The Southern probe used was a 1932 bp was a *KpnI-SfiI* fragment of pTA2365 (not shown). (B) Southern blot confirming strains H5017 and H5383 have integrated *p.tnaM3-ginS::hdrB+* and H5312 and H5385 have integrated *7xHis 2xStrepII-p.tnaM3-ginS::hdrB+*.

Assessing requirement for GINS using inducible promoters

Inducible GINS strains H5017, H5383, H5312 and H5385 were plated onto tryptophan gradient plates, which forms a concentration gradient for the level of tryptophan (trp) across the plate from 0 mM to any selected concentration at the other; here 0.25 mM (**Figure 5.17**). Should *ginS* expression prove essential, the inducible GINS strains should not be able to grow on the side of the plate lacking tryptophan. If non-essential, they should be able to grow across the plate in the absence of tryptophan (and therefore *ginS* expression). It should be noted that parental strains H4829 and H4832 are capable of tryptophan synthesis (*trpA+*) and thus should have the ability to grow at 0 mM trp where altered gene expression remains viable.

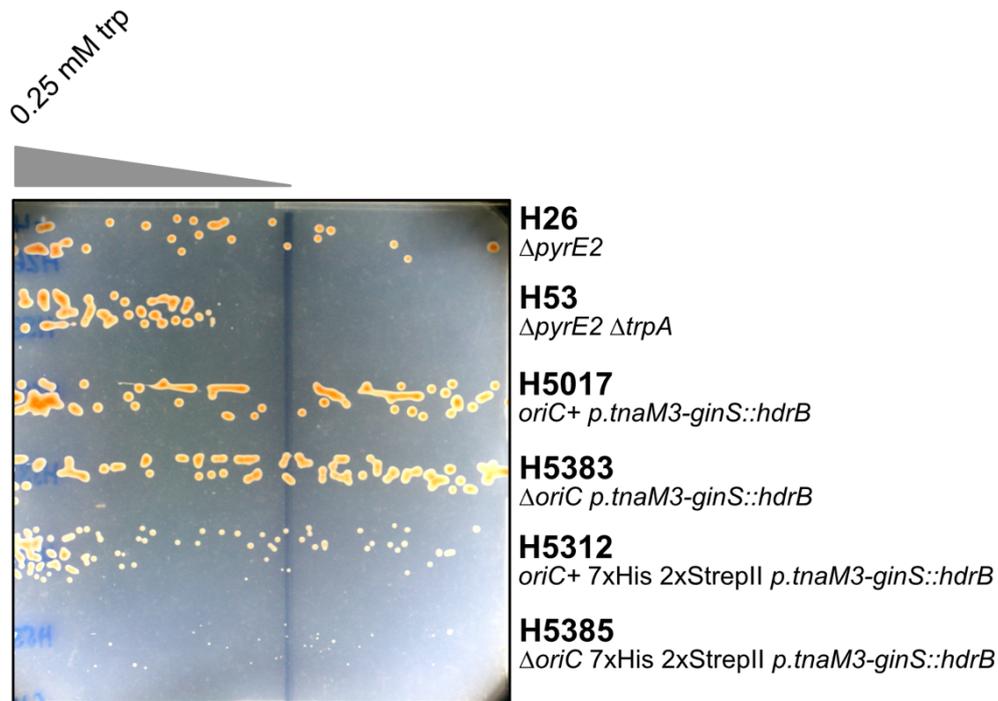


Figure 5.17: Tryptophan gradient plates with *p.tnaM3::ginS* strains. Strains H5017 ($oriC+ p.tnaM3-ginS::hdrB+$), H5312 ($\Delta oriC p.tnaM3-ginS::hdrB+$), H5312 ($oriC+ 7xHis 2xStrepII p.tnaM3-ginS::hdrB+$) and H5385 ($\Delta oriC 7xHis 2xStrepII p.tnaM3-ginS::hdrB+$) were painted onto tryptophan gradient plates with a maximal tryptophan concentration of 0.25 mM. The untagged *p.tnaM3-ginS* clones, H5017 and H5383, remain viable across the gradient. However, the tagged counterparts, H5312 and H5383, show attenuated growth compared to the untagged counterparts. H5385 shows a more severe growth defect than H5312. It is unclear whether growth of these strains are majorly affected by the concentration of tryptophan.

Surprisingly, both WT and Δori inducible *ginS* strains remained viable at 0 mM tryptophan, growing at a rate similar to wild type (H26; $\Delta pyrE2$). This suggests expression of *ginS* is not essential for viability, contradicting the previous failed deletion attempt. However, where 7xHis 2xStrepII tandem tags were integrated alongside the inducible promoter, cells showed a reduced growth rate compared to the untagged inducible counterparts, suggesting that the N-terminal tags on GINS impair cell growth. Growth was still seen at 0 mM trp, suggesting GINS is still non-essential in these strains, but smaller colonies were seen for both (with a more drastic phenotype seen for H5385). The difference between H5312 and H5385 hints at a differential requirement for GINS in the presence and absence of origins.

While tryptophan gradient plates give an indication of growth in the absence of promoter induction, it does not allow for direct comparison of doubling times or allow for minor differences in optical density when plating

Chapter 5: Deciphering the role of the CMG replicative helicase complex in *Haloferax volcanii*

cells. Thus, the strains were subsequently subjected to liquid growth assays. Primarily, the untagged versions were screened to ensure growth was consistent in the absence of tryptophan. Strains H5017 (*oriC*⁺) and H5383 (Δ *oriC*) were grown for two consecutive overnights in 0 mM tryptophan (to ensure no tryptophan was carried over from liquid media dilutions) to give an actively growing culture. Cell growth was measured at 0 mM and 1 mM tryptophan concentrations over a 48-hour time course using a microplate spectrophotometer. A growth curve was plotted, and relative generation times were calculated during exponential phase (**Figure 5.18**).

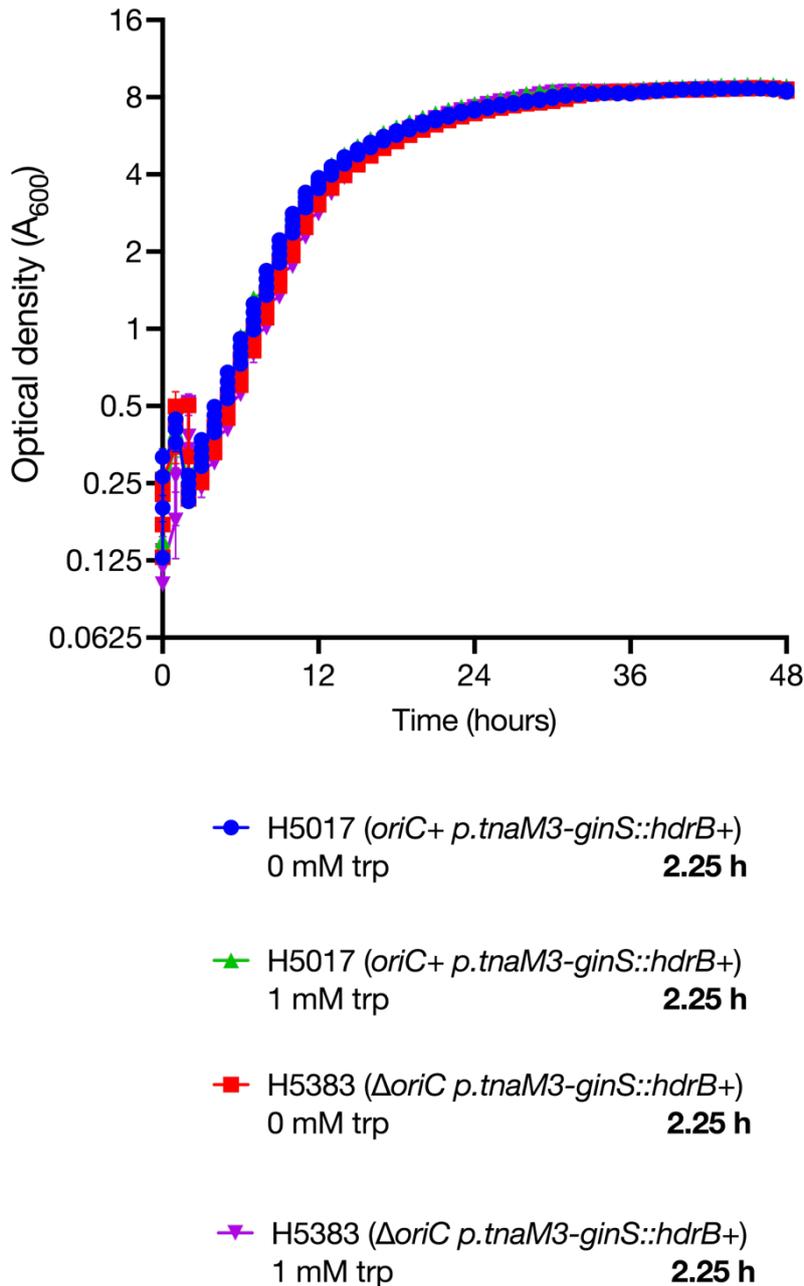


Figure 5.18: Exponential growth rate of strains with *p.tnaM3-ginS::hdrB+* integration. Generation time in hours (h) is indicated in bold in the legend. Strains were grown in Hv-Cas + ura broth for two consecutive overnights before being diluted and plated. All strains (n=2) were incubated on the same 96-well plate and measured simultaneously for optical density (A_{600}) using an Epoch2 Microplate Spectrophotometer (BioTek). Error bars represent standard error of the mean (SEM). All strains show comparable doubling times in both the presence and absence of tryptophan.

As previously observed with the gradient plates, both untagged inducible GINS strains, H5017 and H5383, remained viable in the absence of

Chapter 5: Deciphering the role of the CMG replicative helicase complex in *Haloferax volcanii*

tryptophan and no change in doubling time was observed following the induction of GINS expression. Previous work has shown that the lower activity promoter *p.tnaM3* is able to prevent gene transcription in the absence of tryptophan, suggesting a leaky promoter is not the cause for this lack of phenotype (Hawkins *et al.*, 2013a, Marriott, 2017).

Since a defect in growth was seen with the tryptophan gradient plates where strains had integrated 7xHis 2xStrepII tags alongside *p.tnaM3*, the same experiment was repeated using the tagged counterparts, H5312 (*oriC*⁺) and H5385 (Δ *oriC*) (**Figure 5.19**).

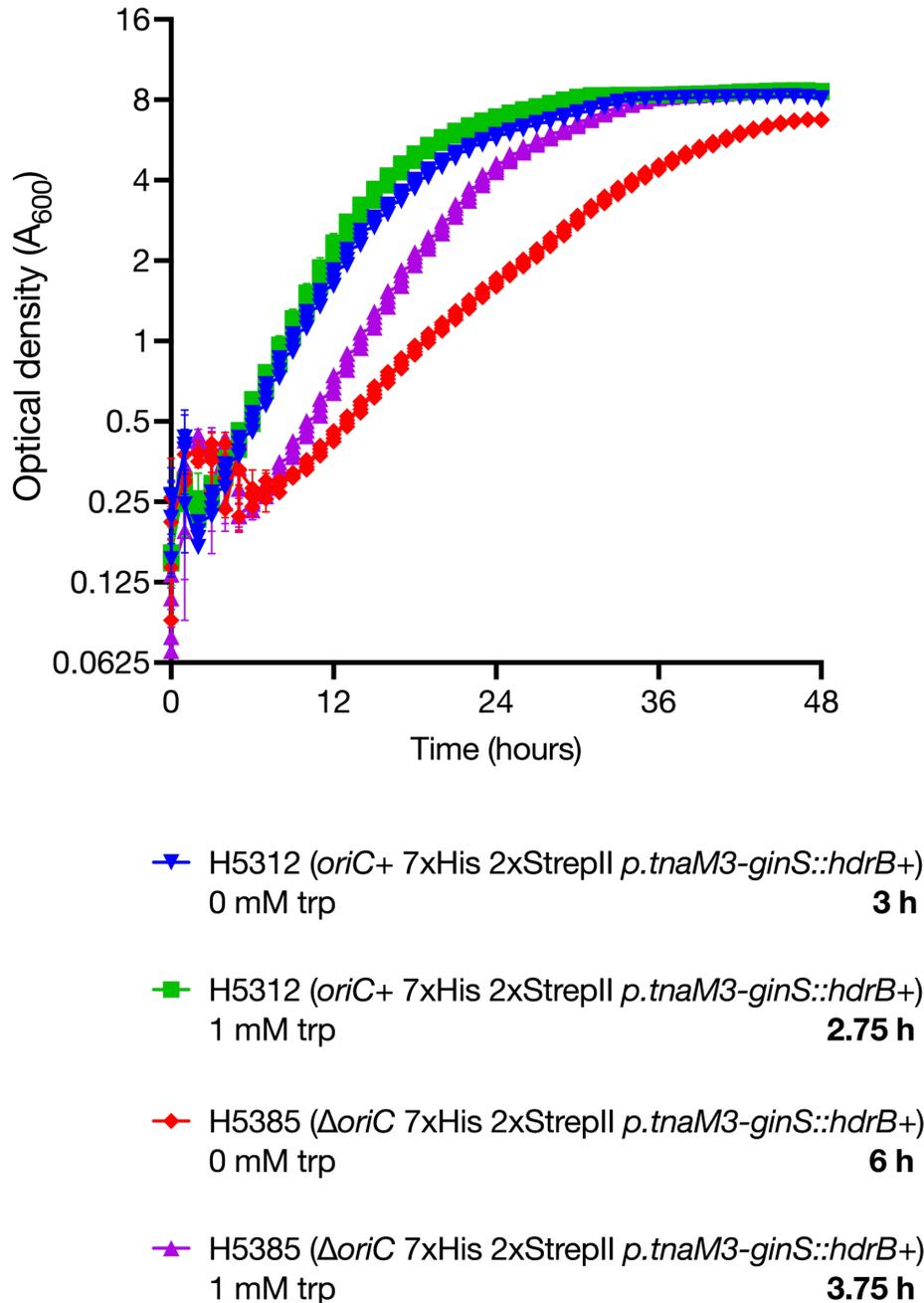


Figure 5.19: Exponential growth rate of strains with 7xHis 2xStrepII *p.tnaM3-ginS::hdrB+* integration. Generation time in hours (h) is indicated in bold in the legend. Strains were grown in Hv-YPC broth for two consecutive overnights before being diluted in Hv-Cas + ura (+ trp where needed) and plated. All strains (n=2) were incubated on the same 96-well plate and measured simultaneously for optical density (A_{600}) using an Epoch2 Microplate Spectrophotometer (BioTek). Error bars represent standard error of the mean (SEM). H5312 (*oriC+*) shows a minor defect in doubling time in the absence of tryptophan. H5385 (Δ *oriC*) shows a major defect in doubling time in the absence of tryptophan, and this defect is not fully recovered following induction with tryptophan.

To confirm that no induction of *ginS* occurs in the absence of tryptophan, strains H5312 and H5385 were assessed for GINS expression by Western blotting. H5312 and H5385 were grown for two overnights in Hv-Cas +Ura media with varying levels of tryptophan (ranging from 0 mM to 2 mM). Cells were spun down, resuspended in water and DNase treated before being resuspended in protein loading buffer. Samples were then run on an SDS-PAGE gel to assess whether induction expression of *ginS* was absent in the absence of tryptophan (**Figure 5.20**).

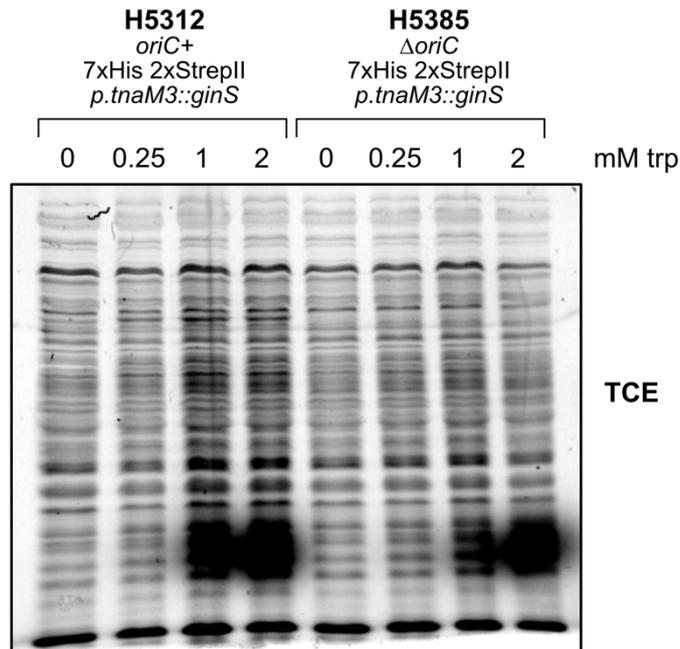


Figure 5.20: SDS-PAGE gel to assess extent of expression of GINS in 7xHis 2xStrepII-inducible strains H5312 (*oriC+*) and H5385 ($\Delta oriC$). Gel contains 0.5% TCE and was imaged following one-minute UV exposure. A band is seen to increase in intensity with tryptophan concentration at ~35 kDa.

Staining of total loaded protein by TCE (2,2,2-trichloroethanol) shows a clear overexpression of a protein at a low molecular weight; this likely represents GINS, which has a predicted molecular weight of 34.2 kDa, suggesting the constructs have integrated correctly and are able to control level of expression. This gel was probed with an anti-His6x antibody, however only non-specific binding was seen and therefore it is not possible to confirm whether 0 mM trp correlates with zero expression of GINS.

5.3.3 Deletion of *ginS* incorporating *priS* transcription termination site

Analysis of *H. volcanii* transcription and sequencing data provided by Anita Marchfelder (Babski *et al.*, 2016, Berkemer *et al.*, 2020) revealed that *ginS* was co-transcribed with essential primase gene, *priS* (Figure 5.21). Unexpectedly, there was also evidence of an additional promoter within the sequence of *priS*; the function and validity of this promoter remains unknown.

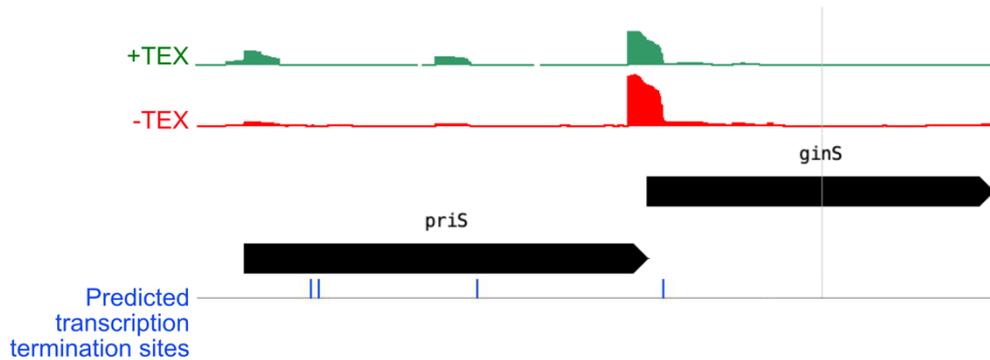


Figure 5.21: Predicted transcription start sites (TSS) and termination sites for the *priS ginS* region (HVO_2697 and H2698). Terminator Exonuclease (TEX) digestion of RNA enriches for primary transcripts (newly transcribed with no 5' processing). Comparison of +TEX (green) and -TEX (red) samples can provide a generic measure of transcription start sites, where higher +TEX measures than -TEX measures is suggestive of an active promoter. Mapping of the 3' end of TEX-treated transcripts allowed for mapping of predicted termination sites (represented here by blue lines). There is a promoter at the 5' end of *priS*, as well as within its sequence. The equally strong signal for both +/-TEX at the 5' end of *ginS* is likely a processing site. Data provided by Anita Marchfelder.

These data indicated that the transcriptional start/stop boundaries were less defined than previously assumed. The data revealed that the suspected transcription termination site (TTS) for *priS* fell 31 bp downstream of the start of *ginS*. Thus, if the *p.tnaM3-ginS* data is true, whereby cells not expressing *ginS* (or expressing GINS at a very low level due to promoter leakiness) remain viable, the failure to delete *ginS* previously could be due to the failure to terminate transcripts of essential primase subunit PriS. Therefore, the failure of pTA2335 deletion construct targeting the entirety of the 5' end of *ginS* could actually be due to the inability to delete both *ginS* and *priS*, where *priS* is known to be essential. To test this hypothesis, two new deletion constructs were designed.

Generation of new *ginS* deletion constructs

Plasmid pTA2315 was created by Rebecca Lever as an intermediate to the previously used $\Delta ginS::trpA^+$ construct pTA2335 (Figure 5.22; Lever, 2019).

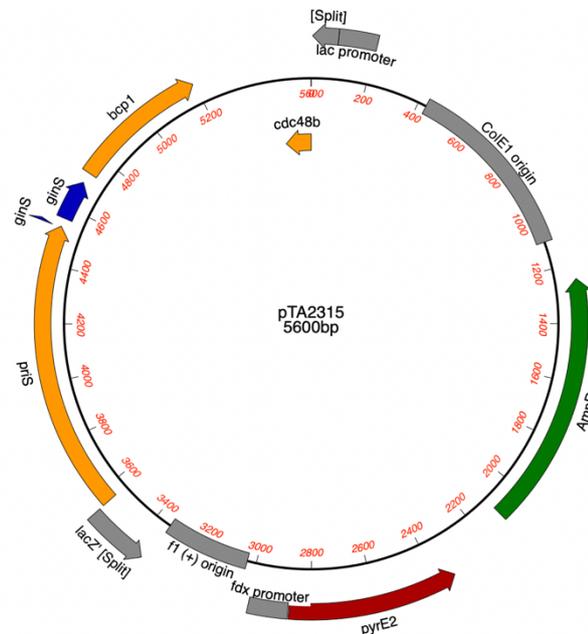


Figure 5.22: pTA2315. Intermediate plasmid in construction of $\Delta ginS::trpA^+$ deletion vector (Lever, 2019).

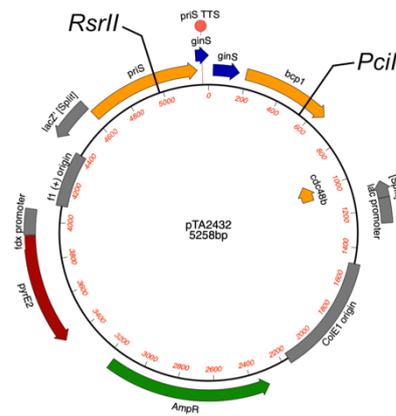
To allow for retention of the *priS* TTS in the *ginS* deletion strain, two constructs were designed incorporating the 5' end of *ginS*, which included the predicted *priS* TTS. The first construct, pTA2439, introduced a premature stop codon within the truncated *ginS*, ensuring there was no truncated *ginS* transcript formed as a fusion of the remaining 5' and 3' *ginS* sequences. The second construct, pTA2437, incorporated a larger fragment of *ginS* before insertion of a stop codon, whereby the deletion construct encodes the first alpha helix of the GINS protein. This allows for some error in the mapping of the *priS* TTS, but should still leave a non-functional segment of *ginS*, as the primary alpha helix is not known to be involved in binding interactors and is therefore unlikely to assist GINS with its key structural role in DNA replication. Primers used for creation of the new deletion constructs are shown in **Table 5.3**.

Table 5.3: Primers used for construction of new $\Delta ginS$ deletion vectors. Restriction sites are highlighted in red. Lower case letters denote mismatches in the primer compared to the template. Underlined bases are those introducing a premature stop codon.

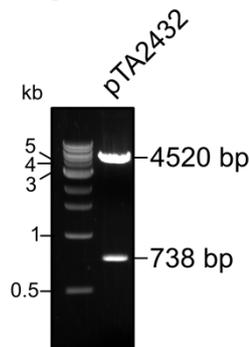
Primer	Sequence
priSKpn_F	CACGGtACcACGACCCAGCGCGTCCTCCGAACCG
ginSintNde_R	CGCGcAtaTGCTGTAGGCTGTCTTTCTaTCGCTCC
ginSint2Nde_R	TCGTcatAtGACTaGCGGAGGTGCTGTAGG

For the first construct that introduced the early premature stop codon, *priS* and the first 42 bp of *ginS* were amplified using primers *priSKpn_F* and *ginSintNde_R*, which introduced a novel 5' *KpnI* site and 3' *NdeI* site respectively. This 657 bp product was digested with *KpnI* and *NdeI* and inserted into pTA2315 in place of the existing upstream fragment at *KpnI* and *NdeI* sites. This gave rise to intermediate plasmid pTA2432 (**Figure 5.23 A**). Due to the previously assumed essentiality of *ginS*, the *trpA* marker was added for additional selection at the pop-out stage. The former deletion construct, pTA2335, carried *trpA* under the synthetic high-activity promoter *p.fdx*. Since *ginS* is co-transcribed with upstream gene *priS*, the promoter for *priS* would be sufficient to drive transcription of *trpA* in place of *ginS*. Therefore, in these constructs, *trpA* was ligated into the construct without any promoter, as *trpA* transcription should occur naturally from the promoter of the operon. Promoter-less *trpA* was digested from plasmid pTA1166 (Stroud *et al.*, 2012) using *NdeI* and inserted at the upstream-downstream boundary by digest with *NdeI*. This gave rise to the truncated $\Delta ginS::trpA+$ deletion construct pTA2439 (**Figure 5.23 B**).

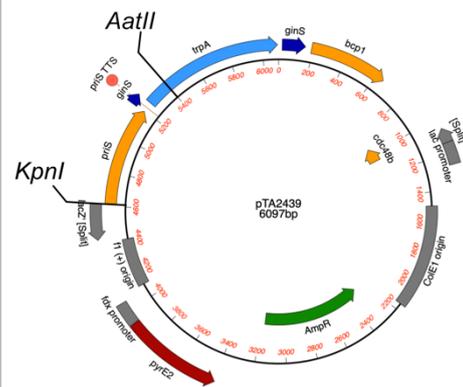
(A) pTA2432



PciI RsrII digest



(B) pTA2439



AatII KpnI digest

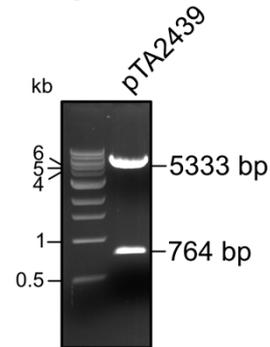


Figure 5.23: Deletion constructs targeting *ginS* introducing early premature stop codon. (A) pTA2432 Δ *ginS* deletion construct with no *trpA* marker. Digestion of pTA2432 with *PciI* and *RsrII* gives bands of 4520 bp and 738 bp as expected. (B) pTA2439 Δ *ginS* deletion construct with *trpA* marker. Digestion of pTA2439 with *AatII* and *KpnI* gives bands of 5333 bp and 764 bp as expected.

The second *ginS* construct that integrated a stop codon following the first alpha helix of *ginS* followed a similar cloning plan, only with the reverse primer differing in binding position to allow for amplification of the first 69 bp of *ginS* before incorporating the stop codon. Primers *priSKpn_F* and *ginSint2Nde_R* gave rise to a 670 bp product which was digested using *KpnI* and *NdeI* and swapped for the existing upstream sequence at *KpnI-NdeI* sites in pTA2315. This gave rise to intermediate plasmid pTA2433 (**Figure 5.24 A**). The promoter-less *trpA* gene was added in the same orientation as *ginS* by digest of pTA1166 with *NdeI* and insertion into the *NdeI* within pTA2433 at the upstream-downstream sequence boundary. This gave rise to the second *ginS* deletion construct pTA2437 (**Figure 5.24 B**).

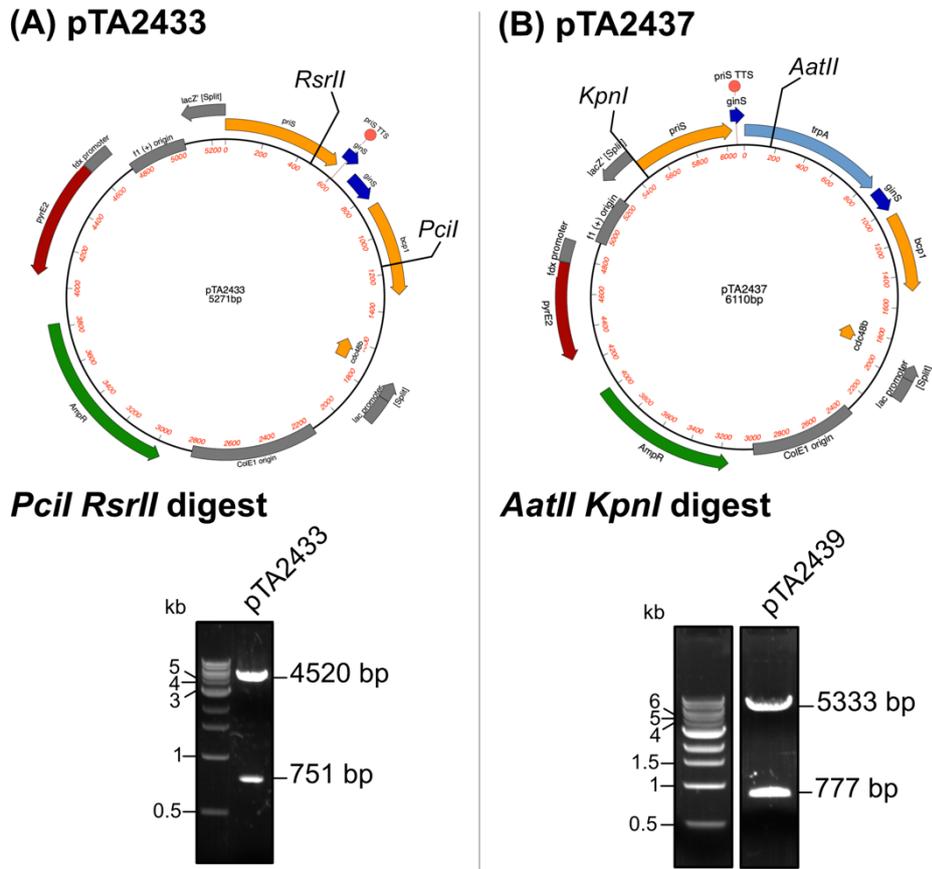
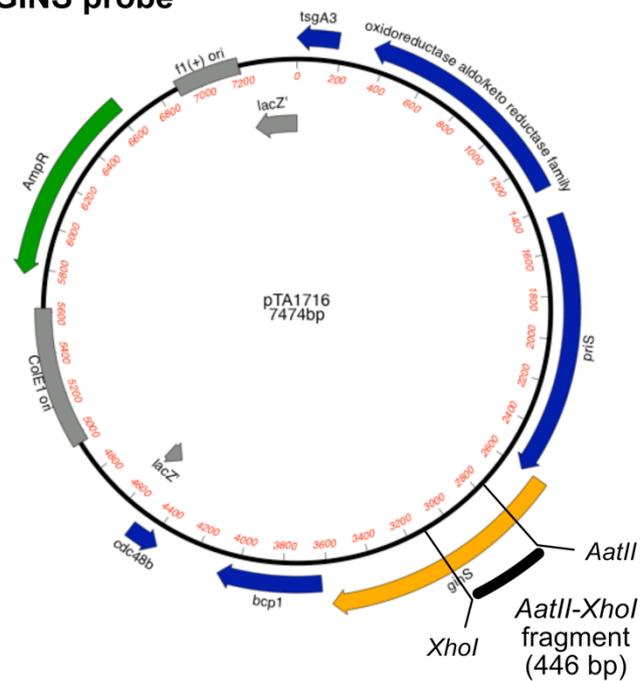


Figure 5.24: Deletion constructs targeting *ginS* integrating stop codon following first alpha helix of *ginS*. (A) pTA2433 Δ *ginS* deletion construct with no *trpA* marker. Digestion of pTA2433 with *PciI* and *RsrII* gives bands of 4520 bp and 751 bp as expected. (B) pTA2437 Δ *ginS* deletion construct with *trpA* marker. Digestion of pTA2437 with *AatII* and *KpnI* gives bands of 5333 bp and 777 bp as

Strain generation

Strains H4045 (*oriC*⁺) and H4598 (Δ *oriC*) were transformed with pTA2437 to generate pop-in strains H5038 and H5039 respectively, and with pTA2439 to generate strains H5040 and H5041 respectively. Pop-out colonies were screened by colony hybridisation with a *ginS*-specific probe. 160 colonies per strain were screened to ensure 90% confidence of isolating a rare deletion mutant, assuming a rate of 1/70 for hard-to-delete genes (as previously explained in **Chapter 3 [Equation 3.1]**). The *ginS*-specific probe was generated by restriction digest of genomic clone pTA1716 using enzymes *AatII* and *XhoI* (**Figure 5.25**). No deletion candidates for GINS were isolated suggesting GINS is essential in *H. volcanii*.

(A) GINS probe



(B) Colony hybridisation



Figure 5.25: Colony hybridisation of $\Delta ginS::trpA^+$ candidates. (A) Patches were probed with a *AatII-XhoI* fragment of pTA1716 (*ginS* genomic clone) (Marriott, 2017). (B) All pop-outs hybridised with the probe, suggesting all clones carry wild-type *ginS* sequence.

5.3.4 Identification of CMG complex members through co-purification

Co-purification of MCM and GINS

Plasmid pTA1791 was used to introduce a C-terminal StrepII-tag to *ginS* (HVO_2698), remaining on the chromosome at native expression levels (**Figure 5.26 A**). Plasmid pTA1663 was used to introduce an N-terminal 6xHis tag to *mcm* helicase (HVO_0220), again remaining on the chromosome at native expression levels (**Figure 5.26 B**). pTA1791 was transformed into H2047 (strain for protein expression), giving rise to pop-out strain H2962 ($\Delta pyrE2 \Delta trpA \Delta mrr Nph-pitA cdc48d-Ct ginS+$ -StrepII tag). H2962 was then transformed with pTA1663, and subsequent pop-out gave a double-tagged strain, H3628 ($\Delta pyrE2 \Delta trpA \Delta mrr Nph-pitA cdc48d-Ct ginS+$ -StrepII tag His₆ tag-*mcm*+). This strain allows for the investigation of direct interactions between MCM and GINS in a wild type background. Aforementioned plasmids and strains were constructed by Hannah Marriott.

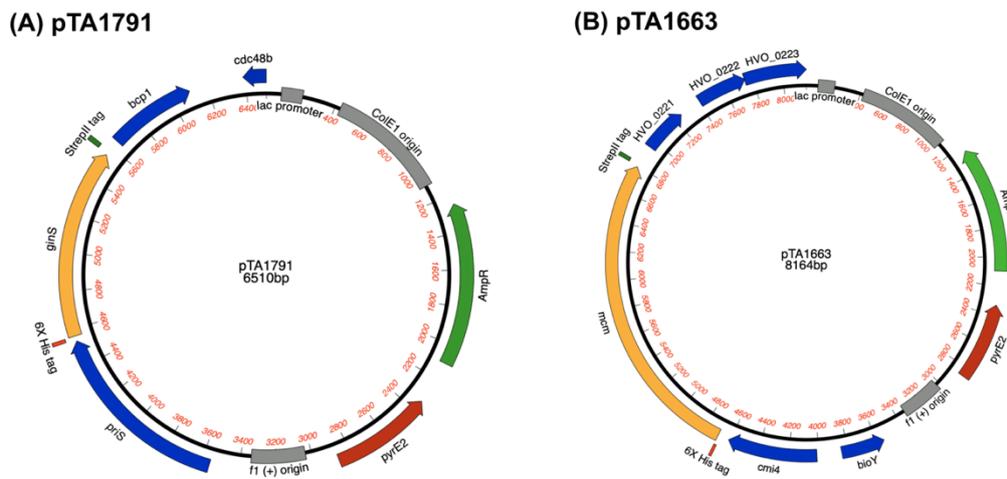


Figure 5.26: Tagged gene replacement constructs for MCM and GINS. (A) pTA1791 with 5' 6xHis and 3' StrepII tagged *ginS* with surrounding genomic region for gene replacement on the chromosome. (B) pTA1663 with 5' 6xHis and 3' StrepII tagged *mcm* with surrounding genomic region for gene replacement on the chromosome. Plasmids constructed by Hannah Marriott (Marriott, 2017).

Chapter 5: Deciphering the role of the CMG replicative helicase complex in *Haloferax volcanii*

Protein purification and interaction

Nickel affinity chromatography of strain H3628 lysate previously gave a band of ~100 kDa, suspected to be MCM helicase (predicted size: 78.8 kDa; Marriott, 2017). Since interaction between MCM and GINS has previously been reported in both eukaryotic and archaeal species (Douglas *et al.*, 2018, Xu *et al.*, 2016), the pulldown experiments were repeated as part of this study, with the aim of optimising Western blotting to reveal any interaction between MCM and GINS in *H. volcanii*. Both nickel and streptavidin-based pulldowns were performed on strain H3628 and control strain H2047.

Nickel affinity chromatography in *Haloferax* species is known to be contaminated by numerous histidine-rich proteins (Allers *et al.*, 2010). This is why direct comparison against the control strain is essential, to ensure any bands seen are products of the specific tagged protein. Ni²⁺ pulldown of H3628 revealed a band at ~100 kDa that was absent from H2047 (H2047 data not shown; **Figure 5.27 A**), and is likely to be MCM (as previously stated by Marriott, 2017). Streptavidin affinity chromatography revealed a band at ~65 kDa (**Figure 5.27 B**); the predicted size of GINS is 34 kDa, however no band of this size is seen.

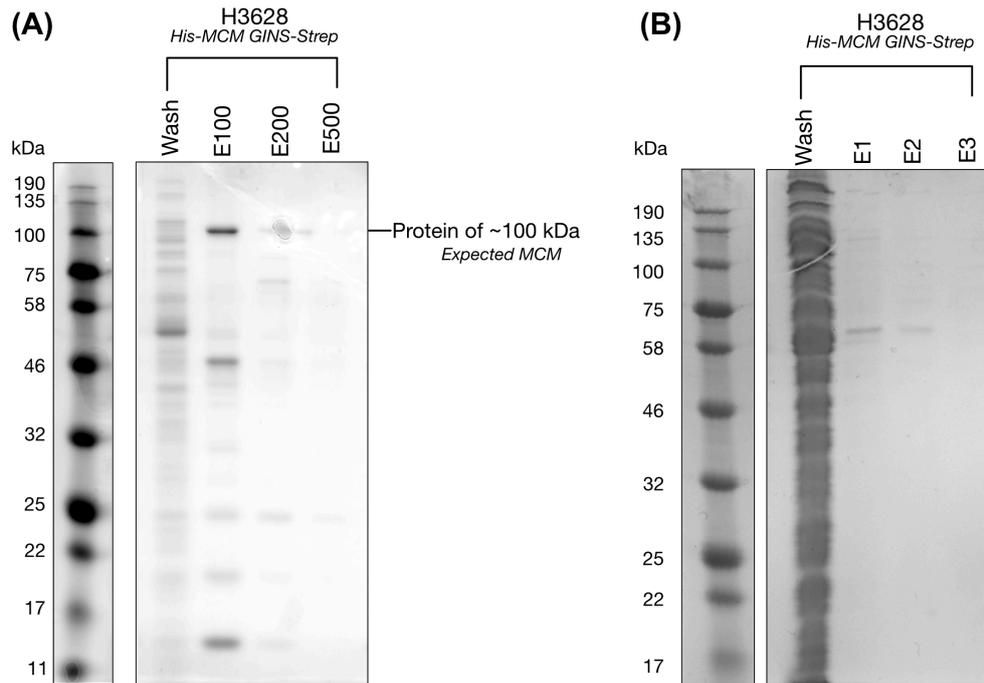


Figure 5.27: SDS-PAGE gels of nickel (A) and streptavidin (B) affinity chromatography pulldowns. (A) A protein of ~100 kDa (predicted to be helicase MCM – 76 kDa) was expressed from the chromosome at native levels from strain H3628 (6xHis-*mcm ginS*-StrepII). Elution (E) was carried out stepwise using imidazole (numbers refer to [imidazole] in mM). (B) Streptavidin-based affinity chromatography in strain H3628 was expected to pulldown GINS (34 kDa). A protein of unknown identity is visible at ~65 kDa. Elution (E) was carried out using sequential elutions with D-desthiobiotin.

Proteins were transferred to nitrocellulose membranes for probing with antibodies specific to the epitope tags on each protein (anti-6xHis antibody for MCM and anti-StrepII antibody for GINS). The eluate from Ni²⁺ affinity chromatography on strains H2047 and H3628 were probed with an anti-6xHis antibody. This showed a clear band at 100 kDa (**Figure 5.28 A**), providing further support for this band representing the successful pulldown of MCM helicase. The same nickel pulldown eluate was also probed with an anti-StrepII antibody to assess whether StrepII-tagged GINS had been co-purified during the pulldown of MCM via its 6xHis tag (**Figure 5.28 B**). A band is seen with the anti-StrepII antibody at ~50 kDa, however it is present in both control (H2047) and experimental (H3628) strains, suggesting this is a contaminant of the pulldown. Interestingly, a band is present at ~70 kDa, matching that previously observed via streptavidin pulldown with H3628 (**Figure 5.28 B**). This 70 kDa band is absent from the control strain, suggesting this protein (potentially GINS) is co-purifying with MCM in a specific manner.

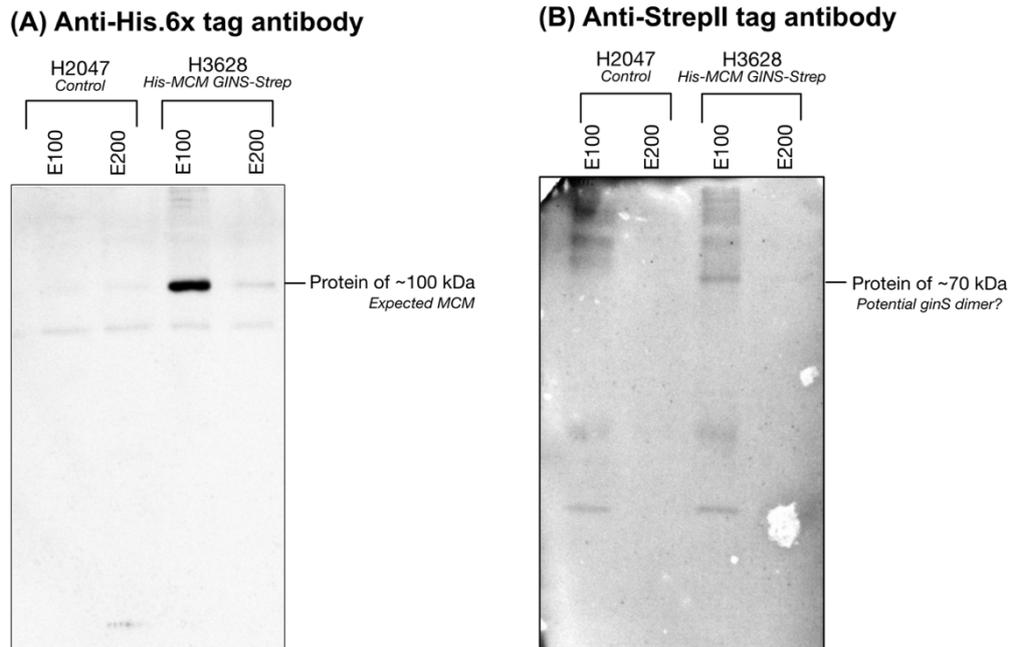


Figure 5.28: Western blots against His₆ (A) and StrepII (B) antibodies of nickel affinity chromatography in strains H2047 and H3628. (A) A 6xHis-tagged protein of ~100 kDa was detected through probing of the membrane with anti-6xHis antibody. This was seen in strain H3628 (*6xHis-mcm ginS-StrepII*), but was absent in control strain H2047, suggesting this protein could be 6xHis-tagged MCM (predicted size 76 kDa). (B) A band of ~70 kDa was seen in strain H3628 but is absent in strain H2047. The expected size of GINS is 34 kDa, therefore this band could be a potential dimer of GINS. A ~45 kDa band was seen in both H2047 and H3628, suggesting this band is a contaminant from the pulldown. There were issues with imaging ladder, however gels were run in parallel with those above and are therefore generally comparable.

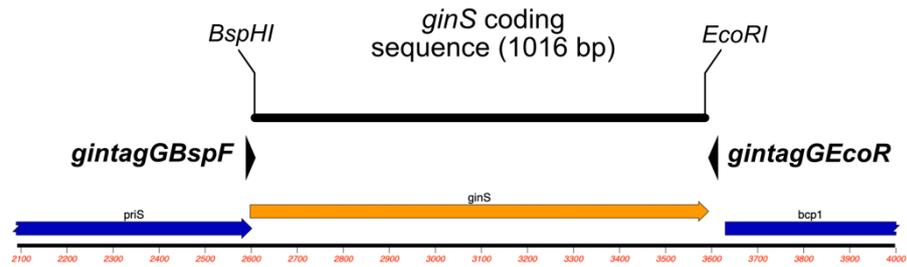
N-terminally tagged GINS

While a band was seen at ~70 kDa after purification of C-terminally tagged GINS, this is unlikely to correspond to 35 kDa GINS. C-terminal tags are known to be problematic for use with many proteins, due to their internalisation upon protein folding. Therefore, an N-terminal tag is preferential for protein pulldowns; this previously proved successful for isolation of N-terminally tagged 6xHis-MCM. A tandem tag comprising a 7xHis tag alongside 2xStrepII tags had previously been used for successful protein purification in *Haloflex volcanii* (Wardell *et al.*, 2017). A cloning strategy was designed to incorporate this tandem tag at the N-terminus of *ginS* for integration onto the chromosome.

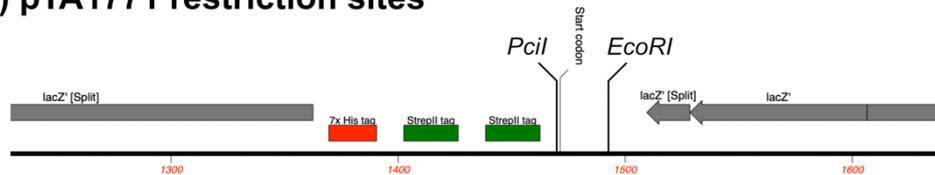
Chapter 5: Deciphering the role of the CMG replicative helicase complex in *Haloferax volcanii*

The *ginS* genomic clone pTA1716 ((Marriott, 2017)) was used as a template to amplify the 1016 bp coding sequence of *ginS* by PCR, using the primers *gintagGBspF* and *gintagGEcoR* (**Figure 5.29 A**). These primers introduced a *Bsp*HI and an *Eco*RI site respectively, allowing digestion and ligation of the *ginS* coding sequence into the compatible *Pci*I and *Eco*RI sites of 7xHis 2xStrepII tag cloning vector pTA1771 (constructed by Thorsten Allers, unpublished data; **Figure 5.29 B**). This gave rise to intermediate plasmid pTA2260, containing tandem tagged *ginS*. The US genomic sequence was isolated from the previously N- and C-terminal tagged *ginS* vector, pTA1791, by digest with *Nde*I and *Kpn*I. This was then inserted into pTA2260 at its corresponding *Nde*I and *Kpn*I sites to create the plasmid pTA2289, 7xHis 2xStrepII-*ginS* gene replacement construct (**Figure 5.29 C**, **Figure 5.30**).

(A) PCR amplified *ginS*



(B) pTA1771 restriction sites



(C) pTA2289

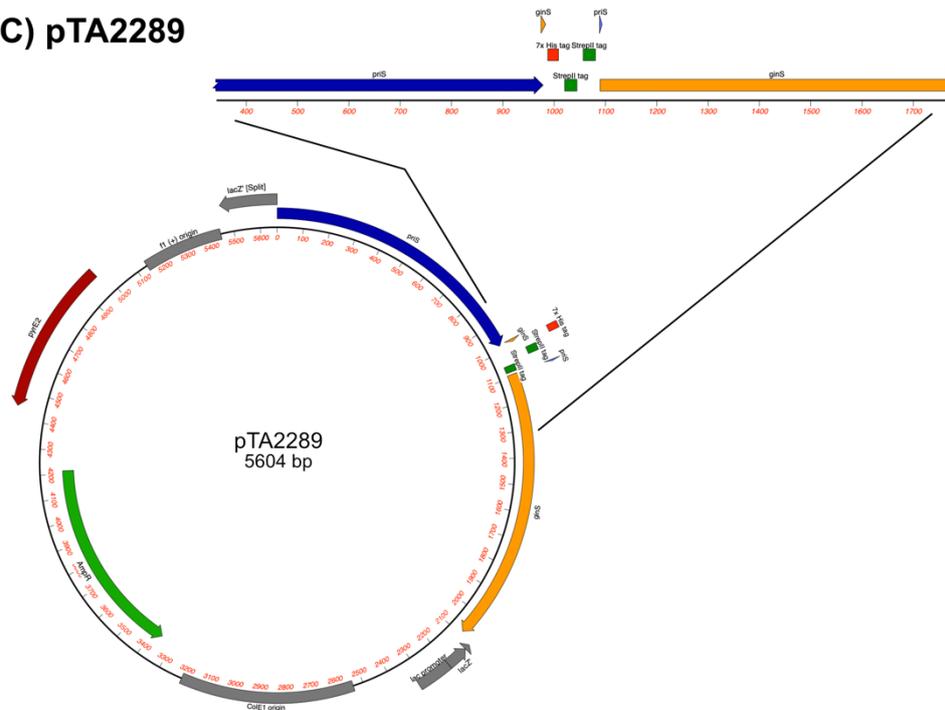


Figure 5.29: Construction of pTA2289 (A) The *ginS* coding sequence was amplified from its genomic clone pTA1791 using primers *gintagGBspF* and *gintagGEcoR*, which introduced novel restriction sites. (B) 7xHis 2xStrepII (tandem tag) cloning vector pTA1771 showing restriction sites utilised for insertion of *ginS* genomic sequence. (C) Gene replacement construct pTA2289, introducing an N-terminal 7xHis 2xStrepII tag at the 5' end of *ginS*.

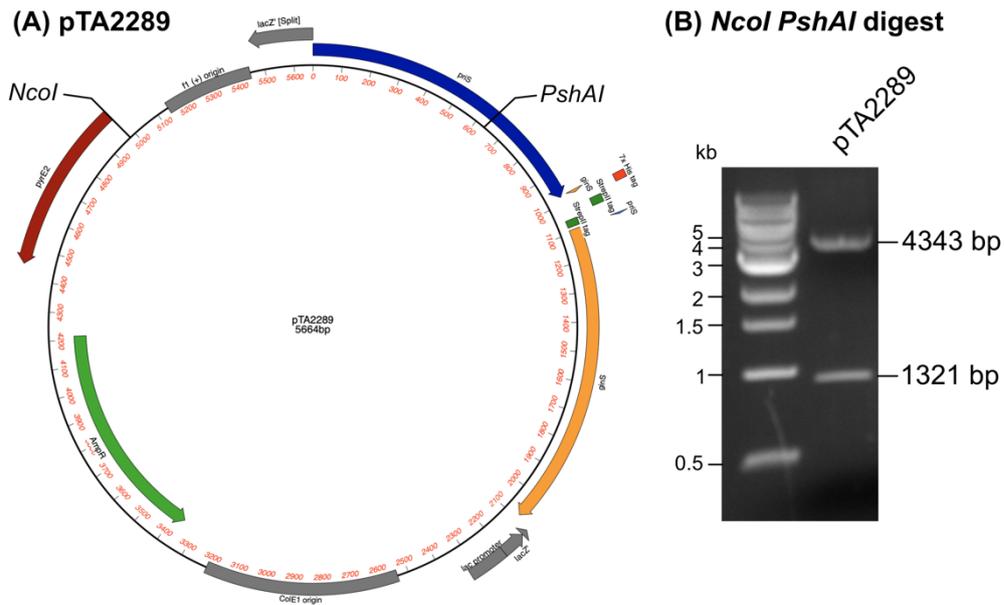


Figure 5.30: pTA2289. (A) pTA2289 represents the pop-in vector for integration of a 7xHis 2xStrepII tandem tag at the 5' end of the gene (and therefore protein). (B) *NcoI PshAI* digest shows bands at 4343 bp and 1321 bp, as predicted.

H2047 ($\Delta pyrE2 \Delta mrr \Delta trpA cdc48d-Ct Nph-pitA$) was transformed with pTA2289, introducing the 7xHis 2xStrepII-*ginS* gene replacement construct into the genome, generating pop-in strain H4531. Pop-out from this strain gave rise to 7xHis 2xStrepII-*ginS* strain H4614. Screening for integration of the N-terminal tag was primarily carried out using colony PCR (Figure 5.31 A). Primers *priSintF* and *ginSintR* amplified different sized products for wild-type and tagged *ginS* alleles (694 bp vs 809 bp respectively). Strain H4614 was confirmed to have integrated the tandem tag by Southern blot (Figure 5.31 B).

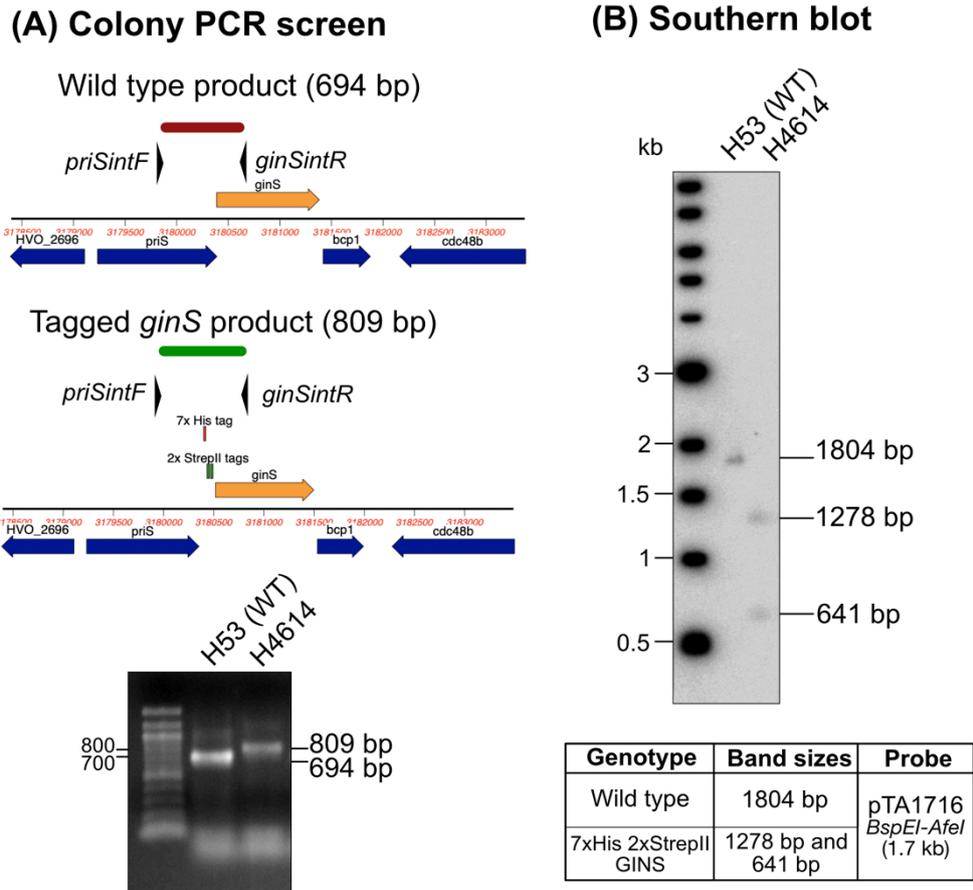


Figure 5.31: Confirmation of integration of 7xHis 2xStrepII at the *ginS* locus. (A) Colony PCR with primers *priSintF* and *ginSintR* give differential sized products depending on the presence/absence of protein tags at the *ginS* allele, where wild type (WT) gives a band of 694 bp and a tagged strain gives a band of 809 bp. (B) Genomic DNA was digested with *AgeI*, *NdeI* and *XhoI* to give a band of 1804 bp for WT and 1278 bp and 641 bp for 7xHis 2xStrepII-*ginS*. Blots were probed with genomic clone pTA1716 fragment *BspEI*-*AfeI* (1.7 kb). Colony PCR and Southern blot confirm H4614 has integrated the 7xHis 2xStrepII tag at the *ginS* locus.

Purification of N-terminal tagged GINS

Initially, the tagged *ginS* strain H4614 was compared to its parent strain, H2047, for growth differences using an Epoch spectrophotometer (BioTek); this was to ensure that integration of the N-terminal tag did not affect the strain viability. Strains were grown for two consecutive overnights in Hv-YPC broth, ensuring on day that three actively dividing cells were used for the assay. Cells were diluted and placed in a 96-well plate. Optical density (OD; A_{600}) was continuously measured, allowing the plotting of growth curves for each strain (Figure 5.32).

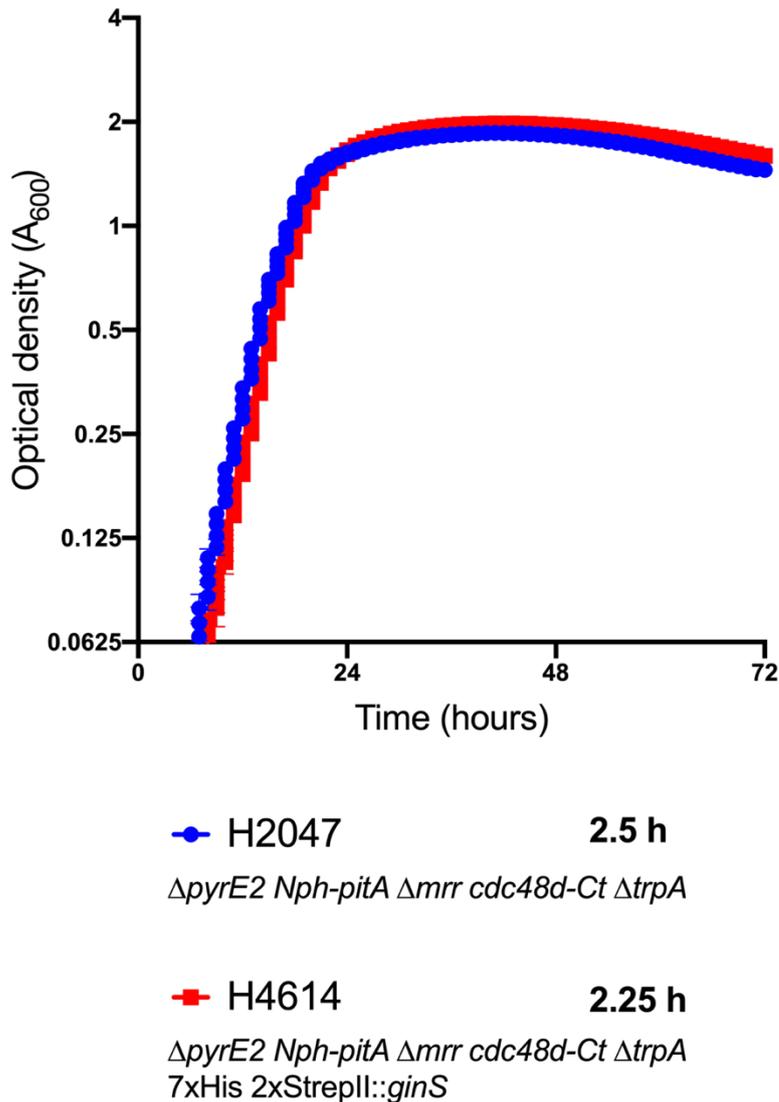


Figure 5.32: Exponential growth rate of tagged *ginS* strain H4614 and its parent H2047. Generation time in hours (h) is indicated in bold beside each strain legend. Strains were grown in Hv-YPC broth for two consecutive overnights before being diluted in fresh Hv-YPC broth. All strains (n=2) were incubated on the same 96-well plate and measured simultaneously for optical density (A_{600}) using an Epoch2 Microplate Spectrophotometer (BioTek).

The growth curve and generation time for *7xHis 2xStrepII-ginS* strain H4614 was comparable to that of parent strain H2047, suggesting integration of the N-terminal tandem tag does not affect the function of GINS.

Streptavidin affinity chromatography using strain H4614 primarily suggested the pulldown had been unsuccessful, with only one weak band being visible at ~60 kDa (**Figure 5.33**). However, it is unclear whether this itself is a

contaminant due to the presence of a faint band at ~60 kDa in the primary elution for control strain H2047.

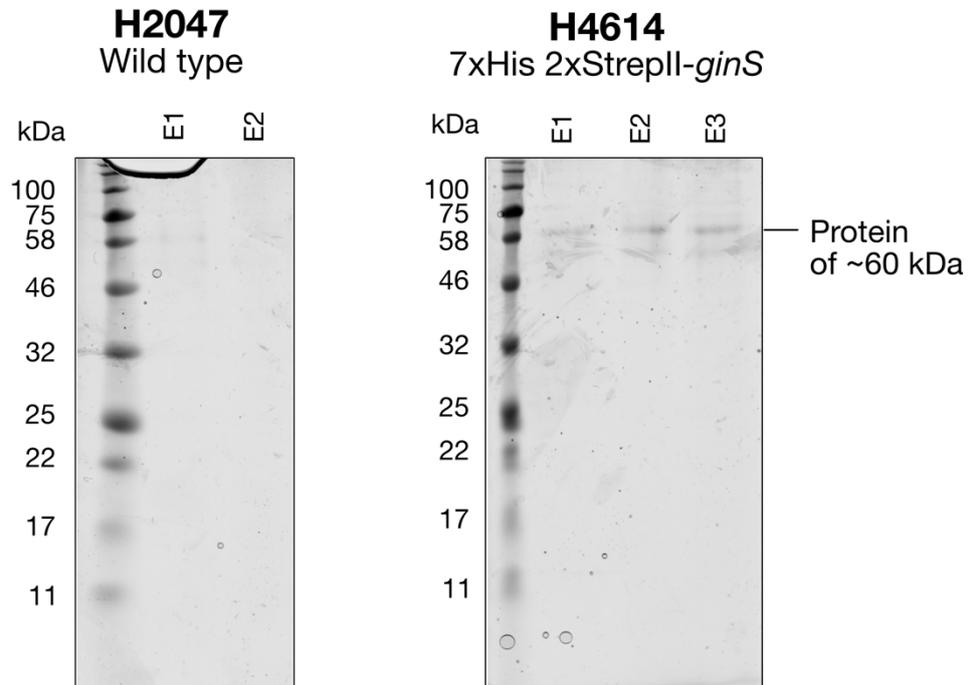


Figure 5.33: Streptavidin-based affinity chromatography of 7xHis 2xStreptII-GINS strain H4614 and empty vector control strain H2047. Bands are seen within the E1 elution at 100 kDa (potentially MCM) and ~52 kDa. The previously observed band at ~70 kDa is present but is less intense than the aforementioned two bands. Bands over 135 kDa are contaminants of the streptavidin method and have been previously observed (Rebecca Lever, personal communication). Elution (E) was carried out using sequential elutions with D-desthiobiotin.

To ensure all proteins purified were visible by SDS-PAGE, eluate samples were precipitated using trichloroacetic acid (TCA). These were run on an SDS-PAGE gel to assess if any previously unobserved proteins were present (**Figure 5.34**).

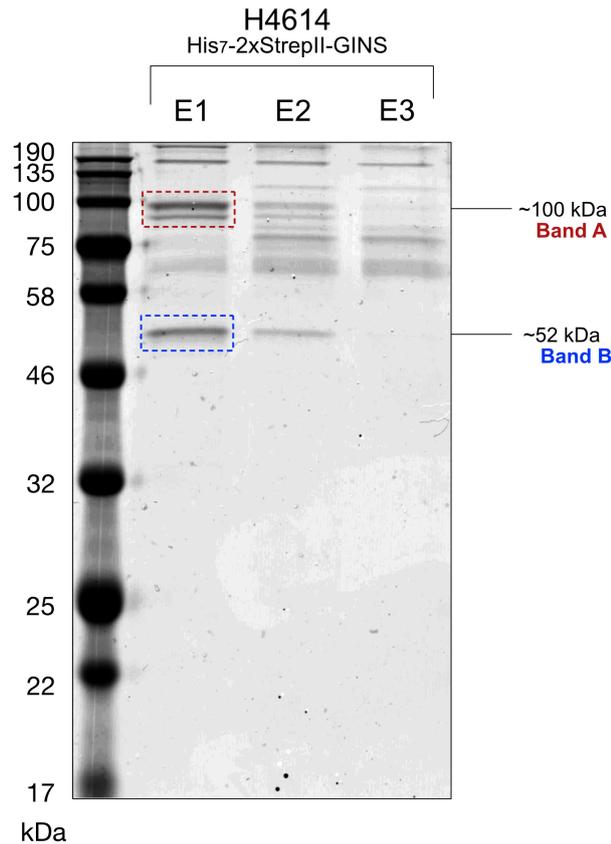


Figure 5.34: Precipitation of streptavidin-based affinity chromatography samples from 7xHis 2xStrepII strain H4614 using trichloroacetic acid (TCA). Bands are seen within the E1 elution at 100 kDa (potentially MCM) and ~52 kDa. The previously observed band at ~70 kDa is present but is less intense than the aforementioned two bands. Bands over 135 kDa are contaminants of the streptavidin method and have been previously observed (Rebecca Lever, personal communication). Elution (E) was carried out using sequential elutions with D-desthiobiotin. Bands A and B (boxed in red and blue respectively) were sent for identification by mass spectrometry.

Prominent bands were observed at ~100 kDa and ~52 kDa, as well as a less discrete band at ~70kDa along with a range of other bands spanning all three elution columns (**Figure 5.34**). Where bands are present across all three elutions, it is unlikely these are genuine pulldown products, since it would be expected that the intensity of a band would decrease with sequential elutions as the protein is released from the column. Proteins spanning all three elution lanes either represent degradation products or more likely artefacts of the precipitation process. This matches the previous prediction that the band seen at 60 kDa prior to TCA precipitation was in fact a contaminant and was not a product specific to the pulldown of GINS. Therefore, only the bands at ~100 kDa (labelled as Band A) and ~52 kDa (labelled as Band B) were assumed to

be the true pulldown products of the H4614 streptavidin affinity. These were sent for analysis by mass spectrometry to identify peptides present. **Table 5.4** shows the key proteins identified from the analysis.

Table 5.4: Summary of key proteins predicted to act at the replication fork identified from analysis of H4614 streptavidin pulldown samples.

Band	A		B	
Approximate MW (kDa)	100		52	
Proteins identified (and HVO_ locus)	Cdc48a	HVO_2380	Cdc48a	HVO_2380
	DnaK	HVO_1590	GINS	HVO_2698
	RecJ4	HVO_2889	RNJ	HVO_2724
	Cdc48b	HVO_2700	Hel308a	HVO_0014
	GINS	HVO_2698	PriS	HVO_2697
	MCM	HVO_0220		
	Hel308a	HVO_0014		

See **Appendix 4** for complete table listing all proteins identified and associated MASCOT scores.

The roles of the key proteins predicted to act at the replication fork are summarised below:

Cdc48a (HVO_2380) and *Cdc48b* (HVO_2700)

The cell division control protein 48 (CDC48) family are AAA+ class ATPases. There are four homologues in *H. volcanii*; sequentially named *cdc48a-cdc48d* (HVO_2380, HVO_2700, HVO_1327 and HVO_1907 respectively).

Cdc48a is found within a highly conserved genome neighbourhood centring on recombination mediator *RadB* (**Figure 5.35**) (Wardell, 2013).

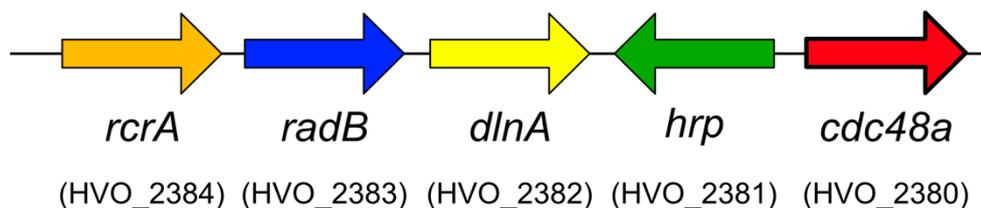


Figure 5.35: Conserved genome neighbourhood containing *cdc48a*. This genome neighbourhood is conserved and syntenic in a number of *Halobacteriales*.

In both eukaryotes and archaea, *Cdc48* targets ubiquitinated proteins for degradation by the proteasome. Ubiquitin-like proteins are found in *H.*

Chapter 5: Deciphering the role of the CMG replicative helicase complex in *Haloferax volcanii*

volcanii in the form of Small Archaeal Modifier Proteins (SAMPs), which share structural motifs with ubiquitin including a β -grasp fold and carboxy-terminal di-glycine motif (Humbard *et al.*, 2010). Cdc48a is thought to play a role in replication fork restart, whereby it acts to target sampylated components of the replisome for degradation by the proteasome allowing access of repair proteins to overcome a replication fork stall, likely by homologous recombination (Forouzan *et al.*, 2012). Both *cdc48a* and *cdc48b* can be deleted in *H. volcanii*; strains deleted for *cdc48a* are slow-growing and show increased levels of recombination (Patricia Perez, unpublished data).

Cdc48d has previously been identified as a histidine-rich contaminant often observed during nickel affinity chromatography (Allers *et al.*, 2010). Deletion of this gene is not possible; however, truncations of the histidine-rich C terminus are viable, and the background strain here contained a C-terminal truncated form (*Cdc48d-Ct*) of the protein and therefore it was unlikely to be identified as a contaminant.

DnaK (HVO_1590)

DnaK is a molecular chaperone found to interact with RcrA, a member of the aforementioned RadB/Cdc48a genome neighbourhood. It is likely involved in a protein turnover role, akin to Cdc48 proteins, and has previously been co-purified with sampylation proteins (Fu *et al.*, 2017). Additionally, the *E. coli* homologue has been suggested to be involved in replication fork restart (Goldfless *et al.*, 2006).

RecJ4 (HVO_2889)

Identification of RecJ4 as a pulldown product associated with GINS provides strong evidence of a role for RecJ4 at the replication fork. RecJ proteins have been implicated in differing roles at the replication fork, either as GINS-associated nucleases (GAN), where they are thought to be the archaeal homologue of eukaryotic Cdc45 (Nagata *et al.*, 2017a), or as Hef-associated nucleases (HAN) where it is implicated in overcoming stalled replication forks (Feng *et al.*, 2018, Nagata *et al.*, 2017b). Work in **Chapter 4** suggests RecJ4 acts as a HAN-like protein, and therefore the co-purification with Hel308a may suggest these proteins are acting together at stalled replication forks. However, no other RecJ proteins were identified here and thus the function of RecJ4 as GAN cannot be ruled out.

Chapter 5: Deciphering the role of the CMG replicative helicase complex in *Haloferax volcanii*

MCM (HVO_0220) and PriS (HVO_2697)

Identification of the replicative helicase and primase proteins increases confidence that the GINS pulldown has successfully isolated true interacting partners that cooperate at replication forks.

Hel308a (HVO_0014)

Hel308a has been previously studied in detail in *H. volcanii* (Gamble-Milner, 2016, Lever, 2019). It is believed to act at stalled replication forks, where it assists in the repair of DNA damage prior to the continuation of canonical replication. Affinity purification using Hel308 has previously identified RecJ4 and RNase J as interacting partners (Gamble-Milner, 2016, Lever, 2019), however the biological significance of these interactions remains unknown.

RNase J, rnj (HVO_2724)

Ribonuclease J (Rnj) is a RNase J-like ribonuclease. Bacterial RNase J is a member of the β -CASP family of ribonucleases involved in mRNA processing and degradation. Together with RecJ3, RecJ4 and Cdc48a (the latter two purified here), it is predicted to form an RNA degradasome complex implicated in protein turnover (Julie Maupin-Furlow, personal communication).

To ensure these protein identifications were truly the result of the tagged GINS pulldown and would not be present in an empty vector control, purification of GINS from strain H4614 was repeated. However, numerous contaminants were seen in H2047 that have not previously been observed, suggesting a technical issue with the pulldown or associated reagents (**Figure 5.36**). It was not possible to successfully repeat the pulldown with empty vector control during the timescale of this project.

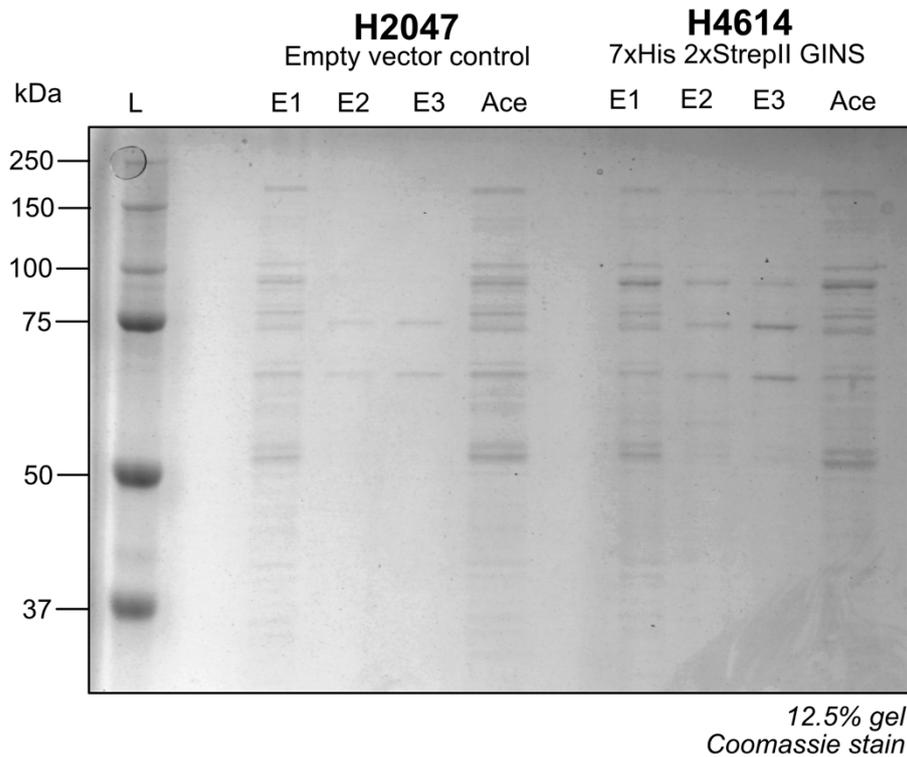


Figure 5.36: Repeat of Streptavidin-based affinity chromatography in 7xHis 2xStreptII-GINS strain H4614 and empty vector control strain H2047. Large numbers of contaminants are seen in both H2047 and H4614; previously, such contaminants have not been observed for H2047. Some bands are specific to H4614 (e.g., band at 100 kDa clearly seen for H4614 in E2 and E3), however large amounts of contamination suggests an issue with the assay itself. Elution (E) was carried out using sequential elutions with D-desthiobiotin. Ace refers to samples precipitated using Acetone.

5.3.5 Investigating the increased requirement for MCM helicase in absence of origins

Previous work has identified an increased requirement for *mcm* transcription in the absence of replication origins in *H. volcanii* (Marriott, 2017). This work utilised inducible promoter *p.tnaM3*, showing that originless cells were not viable in the absence of induction, while *oriC*⁺ cells could survive. This was predicted to be due to a small amount of leakiness occurring from the promoter of these strains. It has not yet been confirmed that this requirement for increased *mcm* transcription correlates with the MCM protein level.

Utilising existing vector pTA2096 (Figure 5.37; constructed by Rebecca Lever; Lever, 2019), containing *p.tnaM3* with downstream histidine and streptavidin protein tags, the protein level of MCM can be quantitated using

Chapter 5: Deciphering the role of the CMG replicative helicase complex in *Haloferax volcanii*

Western blotting and thereby determine if the increased requirement *mcm* transcription in originless cells is due to an increased requirement for the MCM helicase protein itself.

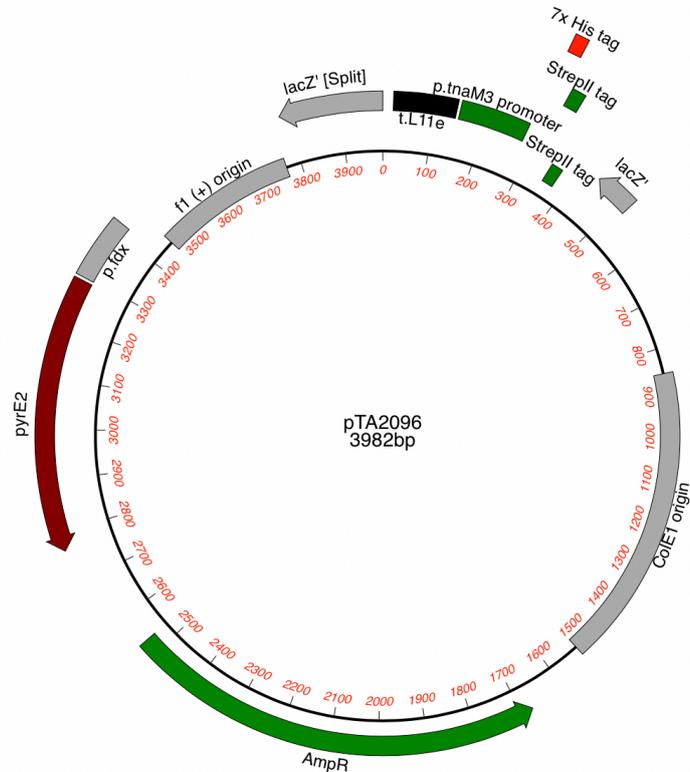


Figure 5.37: pTA2096. Vector for creating *p.tnaM3*-inducible genes with 5' protein tags, namely 7xHis and 2xStrepII tags. Created by Rebecca Lever (2019).

Cloning of *p.tnaM3*-7xHis 2xStrepII-*mcm* construct

The coding sequence of gene *mcm* (HVO_0220) was amplified from *mcm* genomic clone pTA1404 (Marriott, 2017). Primers *mcmtagGNcoF* and *mcmtagGEcoR* amplified 2136 bp of sequence, introducing a 5' *NcoI* site in-frame with the start codon and 3' *EcoRI* site following the stop codon of *mcm* into the product. The product was digested with *NcoI* and *EcoRI* and inserted into *p.tnaM3*-7xHis 2xStrepII vector pTA2096 at compatible *PciI/EcoRI* sites. This gave rise to intermediate plasmid pTA2421 (**Figure 5.38**).

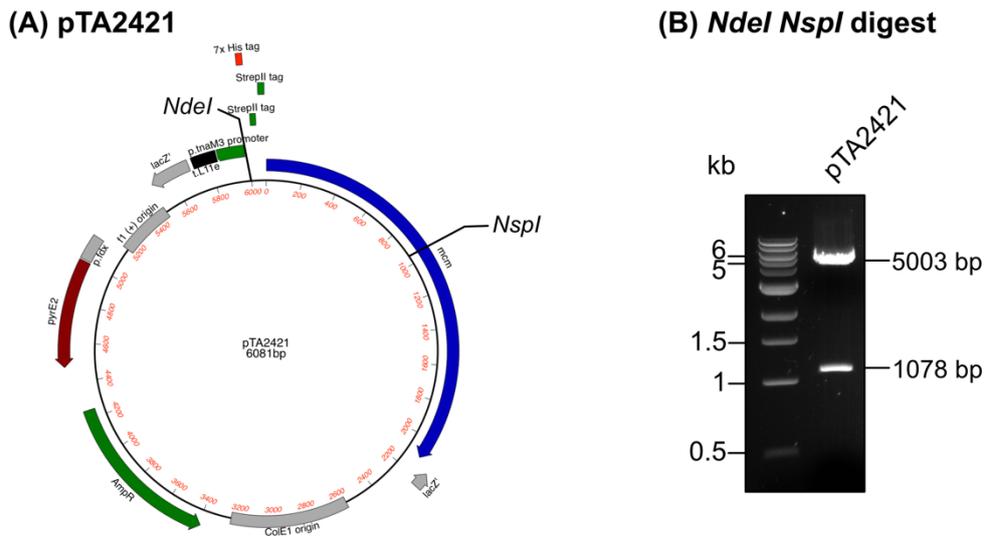
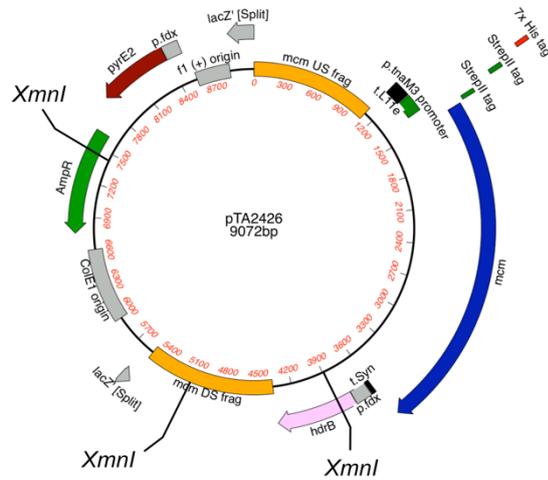


Figure 5.38: pTA2421. (A) Map of pTA2421, intermediate plasmid in construction of *p.tnaM3-7xHis 2xStrepII-mcm* for integration onto the chromosome. (B) Diagnostic digest with *NdeI* and *NspI* give bands at 5003 bp and 1078 bp, as expected.

Previously, *hdrB* selection was required for the integration of *p.tnaM3-mcm* onto the chromosome, and thus a similar strategy was used here. To add up- and downstream sequences and *hdrB* selection to pTA2421, the *p.tnaM3-7xHis 2xStrepII-mcm* cassette was digested from pTA2421 using *XhoI* and *BamHI*. This 2.5 kb cassette was ligated into previously-used inducible *mcm* construct, pTA1460, in place of its *p.tnaM3-mcm* cassette at corresponding *BamHI/XhoI* sites. This gave rise to the final plasmid pTA2426, carrying *p.tnaM3-7xHis 2xStrepII-mcm::hdrB* for integration onto the chromosome (Figure 5.39).

(A) pTA2426



(B) *XmnI* digest

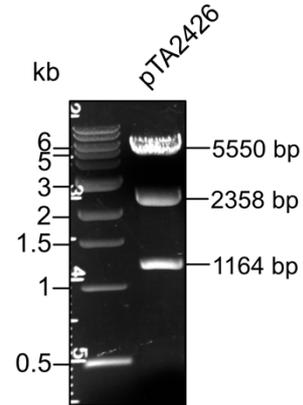


Figure 5.39: pTA2426. (A) Map of pTA2426, vector containing *p.tnaM3-7xHis 2xStrepII-mcm-hdrB* for integration onto the chromosome. (B) Diagnostic digest with *XmnI* give bands at 5550 bp, 2358 bp and 1164 bp, as expected.

Tagged inducible *mcm* strain construction

Strains H1530 ($\Delta pyrE2 \Delta hdrB$) and H1608 ($\Delta pyrE2 \Delta hdrB \Delta oriC1 \Delta oriC2 \Delta oriC3 \Delta ori-pHV4$) were transformed with pTA2426. To ensure correct integration of the inducible promoter, it is favourable to isolate an US pop-in, where a subsequent DS pop-out will leave the promoter in place. Pop-in candidates were patched on Hv-Cas +Trp and screened for pop-in orientation by PCR. Primers *mcmUSextF* and *mcmUSseqR* would give differential products depending on pop-in orientation; an US pop-in would give a product of 1726 bp while both gene and DS pop-ins would give a product of 1374 bp (**Figure 5.40**).

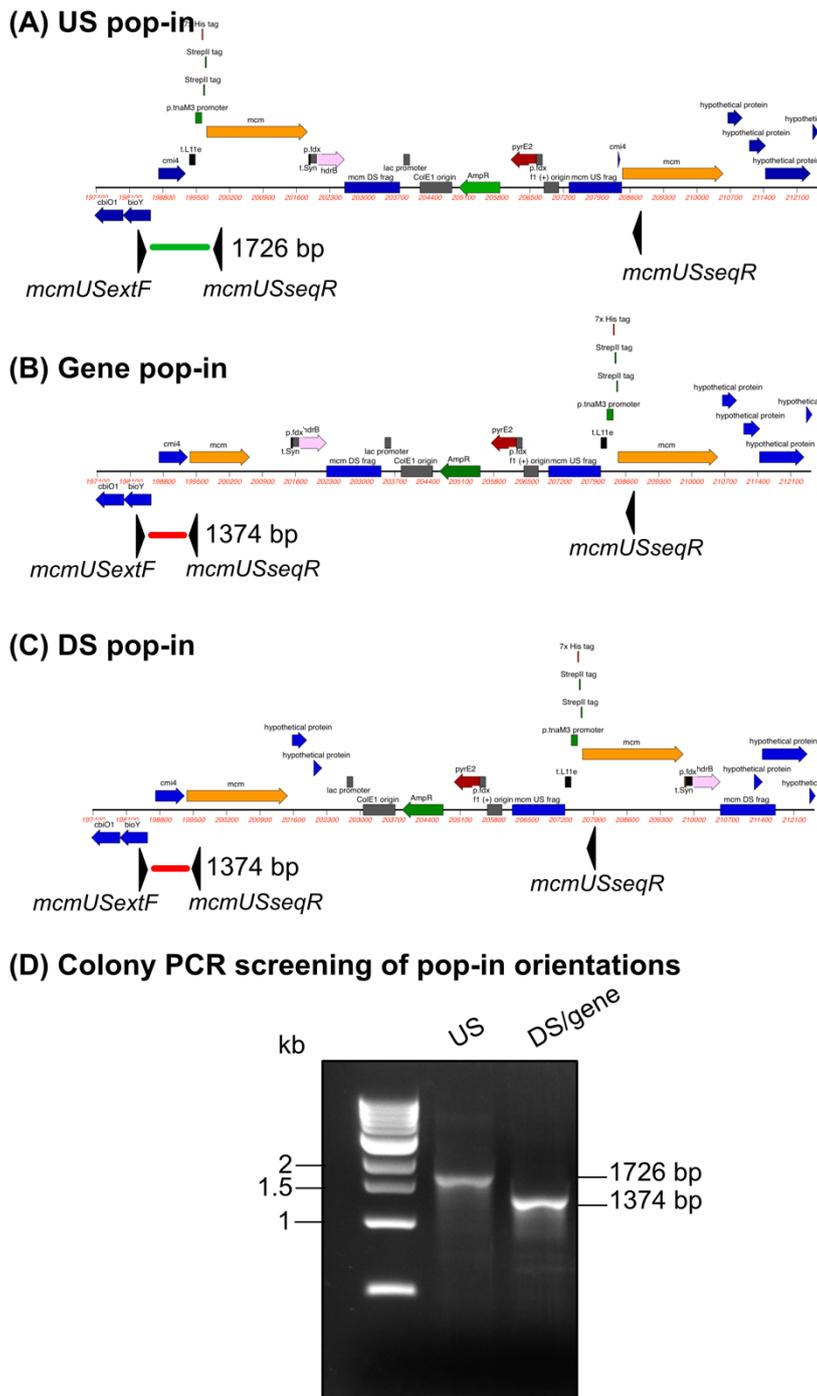


Figure 5.40: Genotyping of pop-in candidates for 7xHis 2xStrepII-*p.tnaM3-mcm::hdrB*. Colony PCR performed against pop-in candidates with primers *mcmUSextF* and *mcmUSseqR*. (A) A pop-in event between the chromosome and the US sequence within pTA2426 gives rise to a PCR product of 1726 bp. (B) A pop-in event between the chromosome and the *mcm* sequence within pTA2426 gives rise to a PCR product of 1374 bp. (C) A pop-in event between the chromosome and the DS sequence within pTA2426 gives rise to a PCR product of 1374 bp. (D) An example of the PCR screen showing an US pop-in product of 1726 bp and a DS or gene pop-in product of 1374 bp.

Chapter 5: Deciphering the role of the CMG replicative helicase complex in *Haloferax volcanii*

While it proved possible to generate US pop-ins, no *trp*- candidates were isolated following pop-out. Inducible MCM strains H1904 and H1911 are both verified as being correctly inducible for MCM and are *trp*-, therefore it may be assumed that no candidates for tagged inducible MCM were isolated. Since the previously generated strains have a strong phenotype and are inherently difficult to culture, it may be that the extra challenge of harbouring a protein tag may prevent the strain being viable, however this is not confirmed.

Assessment of inducible MCM strains by microscopy

It is expected that overexpression of a replication component as critical as the MCM helicase would impact cell morphology and division. For this reason, microscopy was used to assess the inducible MCM strains in the presence and absence of tryptophan.

Strains H1904 (*oriC*+ *p.tnaM3::mcm*) and H1911 (Δ *oriC* *p.tnaM3::mcm*) were grown for two subsequent overnights in Hv-Cas +Ura (+/- 1 mM *trp*). Cells were spun down, resuspended in 18% salt water, stained with DAPI (final concentration 2.5 μ g/ml), and incubated in the dark for 10 minutes. Cells were washed of excess dye, resuspended in fresh 18% salt water and spotted onto prepared agarose pads containing 18% salt water. Cells were imaged using a Nikon Ti-E inverted fluorescence microscope (**Figure 5.41**).

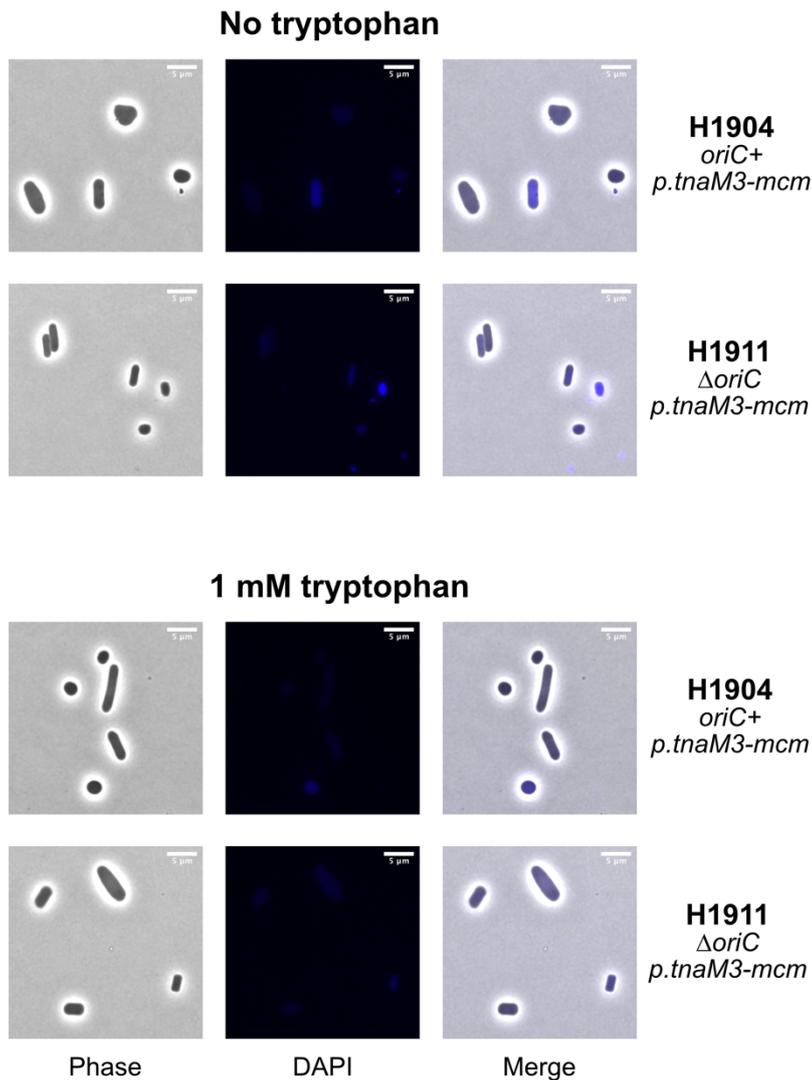


Figure 5.41: Microscopy showing cell morphology and DNA content for strains H1904 (*oriC*⁺ *p.tnaM3::mcm*) and H1911(Δ *oriC* *p.tnaM3::mcm*) in the presence and absence of tryptophan (1 mM). In both the presence and absence of tryptophan, cells of both strains show a more elongated morphology. In the presence of tryptophan, DNA within cells is diffuse. In contrast, in the absence of tryptophan, cells show a more compacted nucleoid, suggesting cell stress. Fluorescence was acquired with a one second exposure. Scale bars are 5 μ m.

In both the presence and absence of tryptophan, cells of both strains show an elongated morphology. These rod-shaped cells could indicate a lack of cell division, where box-shaped cells are dividing.

In the presence of tryptophan, DNA within cells is diffuse. This suggests overexpression of MCM is not causing major issues with canonical replication, and that staining of DNA is comparable between the *ori*⁺ and

Chapter 5: Deciphering the role of the CMG replicative helicase complex in *Haloferax volcanii*

$\Delta oriC$ strains. In contrast, in the absence of tryptophan, both H1904 and H1911 cells show a more compacted nucleoid, suggesting cell stress (Delmas *et al.*, 2013). It is expected that the lack of helicase would cause cell stress, especially when the demand for MCM is higher in $\Delta oriC$ cells than the requirement for canonical replication in *ori+* cells.

5.4 Discussion

Essentiality of GINS in H. volcanii

In eukaryotes, the GINS complex is a well-established member of the replisome and is generally thought to be essential for DNA replication. However, recent genomic analyses have revealed that parasitic metamonads carry a reduced eukaryotic replication apparatus, where some species of the genus *Carpodimonas* encode an incomplete GINS complex (1 or 2 subunits, instead of 4) (Salas-Leiva *et al.*, 2021); this may have an implication on the previously assumed essentiality of GINS. However, since these organisms cannot yet be cultured and studied in detail, the implications for this remain unknown.

While archaeal GINS differ from their eukaryotic counterpart at the sequence level, the structures are generally comparable and thus structural analyses suggest the function of GINS is conserved between archaea and eukaryotes (Oyama *et al.*, 2011). Preliminary attempts to delete GINS using construct pTA2335 failed. This result was unsurprising as GINS would be predicted to be an essential gene due to its putative role within the CMG complex. Since an increase in requirement was seen for MCM expression in the absence of origins (as shown using *p.tnaM3* tryptophan-inducible promoters; (Marriott, 2017)), it was of interest to see whether a similar pattern was observed for GINS. Should the trend also be observed for GINS, it may suggest that the canonical CMG complex is utilised even in the absence of origins.

It proved possible to generate *p.tnaM3* tryptophan-inducible GINS strains in both wild type and originless backgrounds. Phenotyping of these strains using tryptophan gradient plates, alongside growth assays, revealed that there was no difference in growth in the absence of tryptophan and this was not altered in the originless strain (unlike MCM). This suggested GINS expression, if essential as predicted, is only required at low levels and is unlikely to play a key role in originless replication. However, the question remains as to how much GINS is actually required for canonical replication?

It has previously been observed that the wild type inducible MCM strain H1904 can survive in the absence of tryptophan (and therefore the absence of MCM induction). The *mcm* gene is known to be essential (Marriott, 2017), and thus the *p.tnaM3* promoter must be leaky to some extent to ensure the strain remains viable. To assess the level of GINS expression present without induction, 7xHis 2xStrepII-tagged versions of the GINS inducible constructs were produced and strains were generated.

Here, an unexpected change in phenotype was observed. While both strains were still viable in the absence of induction, a difference in growth rate was now seen between the wild type and originless counterparts (as seen by both tryptophan gradient plates and growth assays; summarised in **Table 5.5**).

Table 5.5 Summary of phenotypes observed for inducible GINS strains. Growth defects compared to wild type strain (H26). trp, tryptophan; WT, wild type.

Strain genotype	Visible growth defect?	
	Trp gradient plates	Growth assay
<i>oriC+</i> <i>p.tnaM3-ginS</i>	No change from WT	No change from WT
Δ <i>oriC</i> <i>p.tnaM3-ginS</i>	No change from WT	No change from WT
<i>oriC+</i> 7xHis 2xStrepII- <i>ginS</i>	Minor growth defect (not specific across trp gradient)	Slower growing than WT, but no change in absence of trp
Δ <i>oriC</i> 7xHis 2xStrepII- <i>ginS</i>	Major growth defect (not specific across trp gradient)	Slow growing compared to WT, major growth defect in absence of trp

Alongside this difference between strains, all tagged inducible GINS strains had a growth defect compared to their non-tagged counterparts. The presence of an inducible promoter or protein tags alone did not affect strain viability, however the presence of both promoter and tags seems to have an additive effect whereby the strain fitness is reduced. Only once the fitness was reduced was it possible to see this minor difference in viability between wild type and originless strains, where originless strains have a higher requirement for GINS expression (akin to the result previously observed for MCM).

These results may suggest that the increased requirement for MCM in originless strains is actually an increased requirement for the CMG complex as a whole. However, unlike the MCM strains, the originless tagged inducible GINS strain, while sick, remains viable in the absence of induction. This could be explained if MCM, as the replicative helicase, is essential for replication to occur, but GINS, whose expression further boosts the activity of the helicase is not essential for replication to occur (assuming *H. volcanii* MCM alone possesses limited helicase activity, as seen in other archaeal species).

The tagged inducible strains were assessed for level of GINS expression by Western blotting with an anti-His6x antibody. The total protein stain showed induction of a protein of the correct size with increasing amounts

of tryptophan, however the Western result showed only non-specific bands and therefore no conclusion can be reached regarding amount of GINS expression at 0 mM tryptophan. It will prove important to optimise the Western blotting procedure with alternative anti-His or anti-Strep antibodies. Alternatively, performing RNAseq in the absence of GINS induction could show levels of *ginS* transcripts, as well as whether reduced expression of GINS alters the transcription of any other genes. Should the *p.tnaM3* promoter prove to be leaky at 0 mM tryptophan, this could have important implications for the supposed essentiality of other genes screened in this manner.

Since previous attempts to delete GINS had failed, deletion constructs were redesigned that allowed further 5' sequences to remain intact, ensuring correct termination of *priS* transcripts could occur (should a transcription termination site be located within the GINS coding sequence). Nevertheless, it was not possible to generate a GINS deletion strain, again suggesting the *ginS* gene is essential. While this may be the case, it would be pertinent to generate differing sizes of truncations or mutate residues predicted to be involved in key interactions, alongside generating an episome expressing GINS to allow for *in trans* complementation while the gene is deleted from the wild type locus.

Identification of proteins interacting with GINS

The first archaeal GINS homologue identified was in *S. solfataricus*, where it was shown to be a direct binding partner of MCM (Marinsek *et al.*, 2006). Since then, bioinformatic analyses have identified at least one GINS homologue encoded within every archaeal genome available. Affinity purifications of GINS in various archaeal species have identified a wide range of binding partners, including origin recognition complex protein Orc1, primase, PolD, PCNA, and GAN (Yoshimochi *et al.*, 2008, Marinsek *et al.*, 2006, Li *et al.*, 2010, Xu *et al.*, 2016, Oyama *et al.*, 2016). It was of interest to purify proteins interacting with GINS in *H. volcanii*, both to confirm previously observed interactions and potentially shed light on the identity of GAN.

Initially, previously-generated strain H3628 (6xHis-MCM GINS-StrepII) was subjected to nickel affinity chromatography. This isolated a protein of ~100 kDa, amongst many contaminants; Western blotting with an anti-His6x antibody suggested this band represented successful purification of MCM. Streptavidin affinity purification isolated a protein of ~65 kDa, which gave a signal via Western blotting with an anti-StrepII antibody, suggesting this may represent the GINS protein (however, its predicted molecular weight is only 34 kDa). While these results were encouraging, there was a high level of background for the nickel affinity chromatography due to the presence of

many histidine-rich proteins within the lysate of *H. volcanii*, therefore any interactors of MCM were unlikely to be identified via this route.

To overcome the issue of non-specific binding, histidine-based purification methods were abandoned, with attention instead being placed on the streptavidin-based purification method. An N-terminally tagged 7xHis 2xStrepII-GINS strain was generated and subjected to streptavidin affinity purification and associated bands were sent for identification by mass spectrometry. A summary of interacting proteins is shown in **Figure 5.42**.

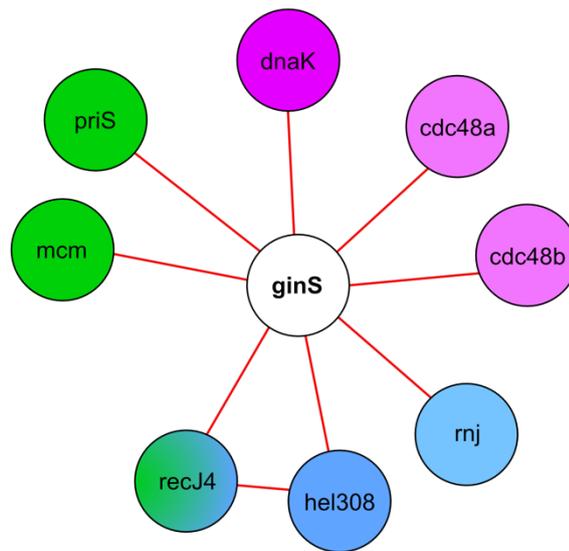


Figure 5.42: Summary of protein:protein interactions observed as a result of the streptavidin-based purification of GINS. GINS is shown in white, as the bait for the pulldown. Proteins identified as interactors are coloured according to their known or predicted functions. Green: DNA replication, purple: protein turnover, blue: DNA damage repair.

Along with isolation of GINS itself, MCM was seen as an interacting partner. This suggests the pulldown has been successful and that any other interactions are *bona fide*. Primase PriS was also observed, again an expected member of the replisome. Cdc48 proteins were also identified; while not a member of the replisome, Cdc48 proteins have been linked to the replication fork in their role as marking proteins for degradation by the proteasome (Forouzan *et al.*, 2012). Regarding the identity of GAN, it was notable to see an interaction between GINS and RecJ4. Since no other RecJ was identified, this may suggest a role for RecJ4 as GAN. However, this contradicts the data in Chapter 4. It is worth noting Hel308 was also purified, which is known to act in DNA repair at interstrand crosslinks, and Hel308 and RecJ4 have been shown to interact previously (Lever, 2019, Gamble-Milner, 2016); the presence of RecJ4 in this pulldown may be a consequence of isolating GINS

proteins at the point of acting in DNA repair. In order to clarify this, it was important to repeat this pulldown (alongside a negative control to ensure these proteins are specific to strain H4614). However, repeated attempts failed, with high background being observed, suggesting an issue with reagents. A successful repeat was not able to be completed within the timescale of this project.

When visualising the total protein yield from a pulldown of 2-3 L of culture, it is somewhat surprising how little protein is present, especially when compared to the yield of lysate from only 5 ml culture (e.g., **Figure 5.20**). Due to the high salt requirements to ensure correct protein folding, cell lysis and removal of chromosomal DNA contamination relies solely on sonication, since the high salt concentration is not compatible with DNase activity. Therefore, pelleting of sonicated lysate to remove DNA may remove a large amount of DNA-interacting proteins. To increase the yield, and potentially increase the number of interactors identified, it would be of interest to test salt-active nucleases to allow degradation of the DNA. This method could ensure more thorough degradation of DNA and release of a higher number of DNA-interacting proteins. It would also be of interest to attempt purification without pelleting of cell debris following sonication, meaning all proteins would be present. However, it is likely the amount of DNA remaining within the solution would still overwhelm the column.

Quantifying MCM expression in inducible strains

Strains encoding *mcm* under the control of promoter *p.tnaM3* were generated prior to this project. Phenotyping of these strains (in both wild type and originless backgrounds) showed a marked increase in requirement for induction of *mcm* expression in strains lacking chromosomal origins of replication. However, one major question regarding this data was why, if *mcm* is an essential gene, the wild type strain remained viable in the absence of induction? Here, constructs were generated to again place *mcm* under the *p.tnaM3* promoter, but to add 7xHis 2xStrepII protein tags to quantify protein levels via Western blot (as antibodies specific to archaeal MCM are not readily available). While pop-in strains could be generated, pop-out events never led to isolation of a tryptophan auxotrophic mutant; the existing inducible strains are unable to grow in the absence of tryptophan and thus the same would be expected of the tagged counterpart. Upon seeing the additive effect of combining *p.tnaM3* and protein tags for the inducible GINS strains, it may not be surprising such a strain could not be generated. Existing inducible MCM strains are hard to culture and require high amounts of tryptophan for subculturing; if there is an additive effect to having tags in combination with *p.tnaM3*, these strains will become even harder to culture (and may not be possible to generate at all).

Chapter 5: Deciphering the role of the CMG replicative helicase complex in *Haloferax volcanii*

It still remains of interest why wild type strains can survive in the absence of MCM induction in the wild type inducible strain, and this warrants future work. If this cannot be answered through the integration of tags, it would be worthwhile using more complex methodology, including generation of archaeal MCM-specific antibodies, RNAseq or qRT-PCR.

5.5 Conclusion

GINS is a central component of the replisome, where it interacts with MCM to aid in canonical DNA replication. Use of tryptophan-inducible promoters suggests the requirement for GINS is not as high as that for MCM helicase; it remains under question whether GINS is actually essential in *H. volcanii*, since inducible strains remained viable in the absence of tryptophan.

Chapter 6: Discussion

Non-canonical DNA replication

Bacteria

Genetic studies have shown that bacteria are capable of replication in the absence of origins, or in the absence of initiator protein DnaA, but in most cases such non-canonical replication initiation is inefficient (Ogawa *et al.*, 1984, Kogoma and Vonmeyenburg, 1983).

E. coli can utilise either D-loops or R-loops to initiate DNA replication, with both mechanisms relying on recombinase RecA (i.e., recombination-dependent replication [RDR]) (Masai and Arai, 1996, Kogoma, 1997). To utilise D-loops, RecA (along with RecBCD) facilitates strand exchange, leading to formation of a stable D-loop. This is processed by PriA to allow loading of the replication machinery and the formation of a *bona fide* replication fork (Masai *et al.*, 1994). When using R-loops to prime replication in *E. coli*, mutation of RNase HI allows for formation of persistent DNA:RNA hybrids which can prime replication, in a manner similar to RNA primers during canonical replication (Kogoma, 1997, Ogawa *et al.*, 1984). Here, RecA remains essential, while RecBCD is not required; RecA facilitates removal of the RNA primers (Asai and Kogoma, 1994). While such mechanisms exist, originless replication in *E. coli* is not well tolerated, and replication in the absence of origins had a major effect on cell growth and viability (Kogoma, 1997).

Cyanobacterial species have recently been shown to replicate asynchronously from multiple sites across the chromosome, in a manner independent of initiator protein DnaA (Ohbayashi *et al.*, 2020). This mode of replication does not utilise origins, suggesting that these species utilise RDR throughout the chromosome. Interestingly, where *dnaA* is deleted, there is an associated increase in chromosomal copy number per cell, suggesting polyploidy is inherently linked to the ability to replicate independently of origins (Ohbayashi *et al.*, 2020). This may explain the loss of viability in Δori *E. coli* mutants, which have only a few copies of the chromosome and therefore few homologous templates are available for strand exchange and RDR.

Eukaryotes

Eukaryotic replication has been intensively studied and is generally held to be a highly conserved process. Upon meeting a disruptive lesion, eukaryotes can utilise two alternative pathways to overcome the block: trans-lesion synthesis or recombination-dependent synthesis. The former utilises alternative

polymerases to bypass the lesion, while the latter uses non-homologous or homologous templates for repair.

Break induced replication (BIR) occurs in yeast distal from origins of replication and is reliant on Pif1 helicase. Resection of a DSB leads to invasion and formation of a D-loop, which is extended unidirectionally by DNA polymerases; in BIR, Pol- δ acts as the main DNA polymerase, in place of canonical leading strand polymerase Pol- ϵ (Miyabe *et al.*, 2015). R-loops have also been implicated in origin-independent replication in yeast (Michel and Bernander, 2014, Stuckey *et al.*, 2015). Novel replication origins have been observed in human cancer cells at highly transcribed oncogenes, which suggest initiation via the use of R-loops (Macheret and Halazonetis, 2018).

Eukaryotic origin-independent replication has been previously observed in origin-deficient or -depleted yeast chromosomes (Bogenschutz *et al.*, 2014). It was shown that recruitment of Orc proteins to DNA was reduced, however it remains to be shown whether deletion of all Orc genes can be tolerated. Studies in human and *Drosophila* cells have shown that Orc1- or Orc2-depleted lines are capable of replication, however the former become reliant on the ATPase activity of Cdc6 (Park and Asano, 2008, Shibata *et al.*, 2016).

Recently, comparative genomic analysis of the replication machinery in parasitic metamonads, and throughout eukaryotes, has revealed a highly reduced eukaryotic replication apparatus in various metamonad species (**Figure 6.1**) (Salas-Leiva *et al.*, 2021).

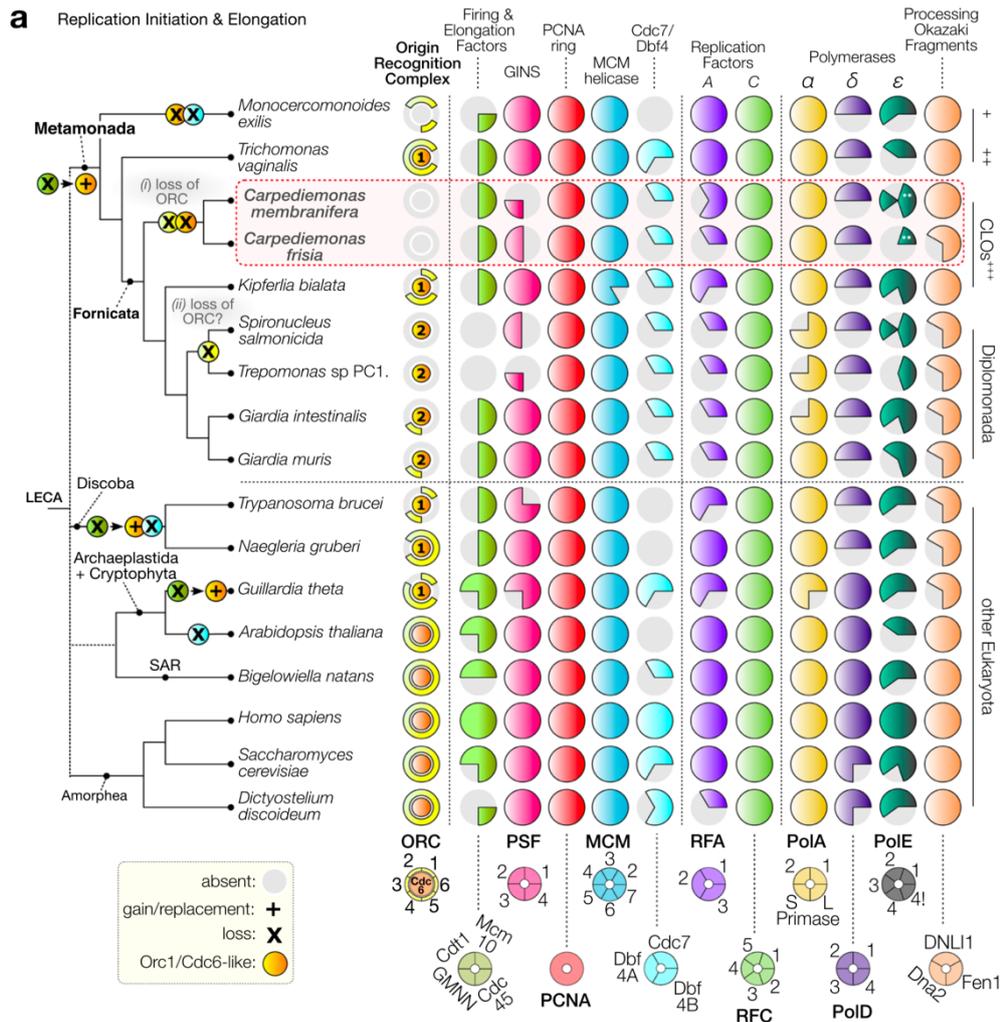


Figure 6.1: The distribution of core molecular systems in the replisome and DNA repair across eukaryotic diversity. The global eukaryotic phylogeny studied is shown on the left. The classification of major lineages is shown on the right. +, Preaxostyla; ++, Parabasalida; +++/CLOs, *Carpediemonas*-like organisms. Taken from (Salas-Leiva *et al.*, 2021).

This included the absence of all Orc1/Cdc6 initiator proteins and an incomplete GINS complex (either 1 or 2 out of 4 subunits) in both *Carpediemonas membranifera* and *C. frisia*. *Carpediemonas* species were also shown to possess a highly divergent clade of Pif1-like proteins, with multiple copies believed to have independently duplicated within each species. It is possible that Pif1 has gained a role in these species, in the absence of Orc1/Cdc6 canonical replication. Additionally, *C. membranifera* lacks a typical eukaryotic RNAse H1 but has gained two copies of prokaryotic homologues via lateral gene transfer. It is suggested that the replication in these species becomes reliant on RNA:DNA hybrids and Pif1-like helicases, however the specifics of replication mechanisms in these species remain unclear.

Archaea

A number of archaeal species have been shown to be viable in the absence of replication origins, including: *H. volcanii* (Hawkins *et al.*, 2013a), *T. kodakarensis* (Gehring *et al.*, 2017) and *T. barophilus* (Moalic, 2021). These species are all highly polyploid and therefore RDR should be able to occur with relative ease, since a high number of homologous templates are available for strand exchange.

H. volcanii encodes multiple RNase H family members. Of these, none carry mutations or have altered expression levels in the originless mutant, compared to the wild-type origin-encoding counterpart. This suggests that RNA:DNA hybrids would be no more stable in the originless strain than the wild-type, leading to the conclusion that the replication initiation mechanism is likely based on D-loops, not R-loops (Hawkins *et al.*, 2013a).

This study has shown that origin-deleted mutants of *H. volcanii* differ from their wild type counterparts in their relative use of the replicative DNA polymerases, PolB and PolD. The origin-deleted strain revealed a tolerance of the PolB-specific inhibitor, aphidicolin, which increased with each origin deletion, suggesting a reduced requirement for PolB in the absence of origins. Deletion of PolB has been successful in multiple archaeal species, including *T. kodakarensis*. It is notable that this species does not appear to utilise replication origins under laboratory conditions, as evident by its flat replication profile (Cubonova *et al.*, 2013). While PolB was found to be essential in the originless *H. volcanii*, its altered response to aphidicolin treatment suggests profound differences between origin-dependent and origin-independent replication.

The CMG complex is a key part of the eukaryotic replisome, and the same has been assumed for archaea. This work has revealed that the requirement for the CMG complex is altered in the absence of origins, and questions remain regarding the essentiality of GINS in *H. volcanii*, and archaea more widely. The discovery of highly reduced eukaryotic replication apparatus (Salas-Leiva *et al.*, 2021) lacking a complete GINS complex raises questions about the inherent essentiality of GINS; it cannot be ruled out that archaeal GINS may not be essential in all situations.

Where previous assumptions have stated that archaeal species will encode one RecJ that acts as GAN, and another that acts as HAN, this study has shown that in species encoding more than two RecJ proteins, such as *H. volcanii*, the situation can be much more complex. Protein purification of RecJ proteins and GINS, and split GFP interaction assays using RecJ proteins highlighted a network of proteins located at the replication fork (**Figure 6.2** and **Figure 6.3**).

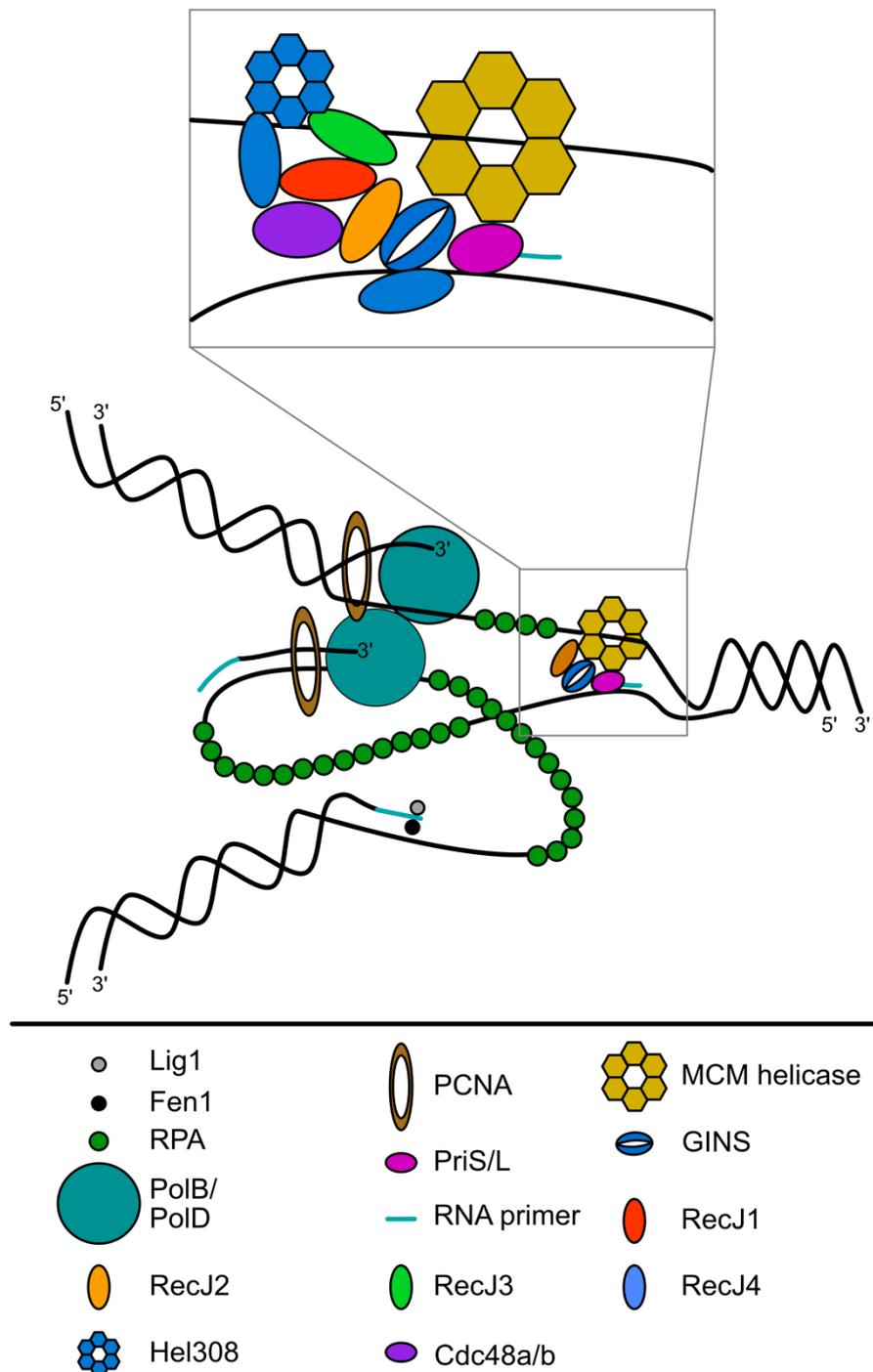


Figure 6.2: Updated summary of the replication fork in *H. volcanii*. Data retrieved during this project have allowed for an updated model of the replication fork, specifically those proteins interacting with MCM, GINS and RecJ. Where the previous version of this figure assumed a single RecJ/Cdc45 protein was acting at the CMG complex, this work has shown that there is strong interplay between all four RecJ proteins in *H. volcanii*. Interactions discovered as part of this study are shown in the highlighted section (top). Further work is required to decipher the specific roles of these proteins during canonical replication.

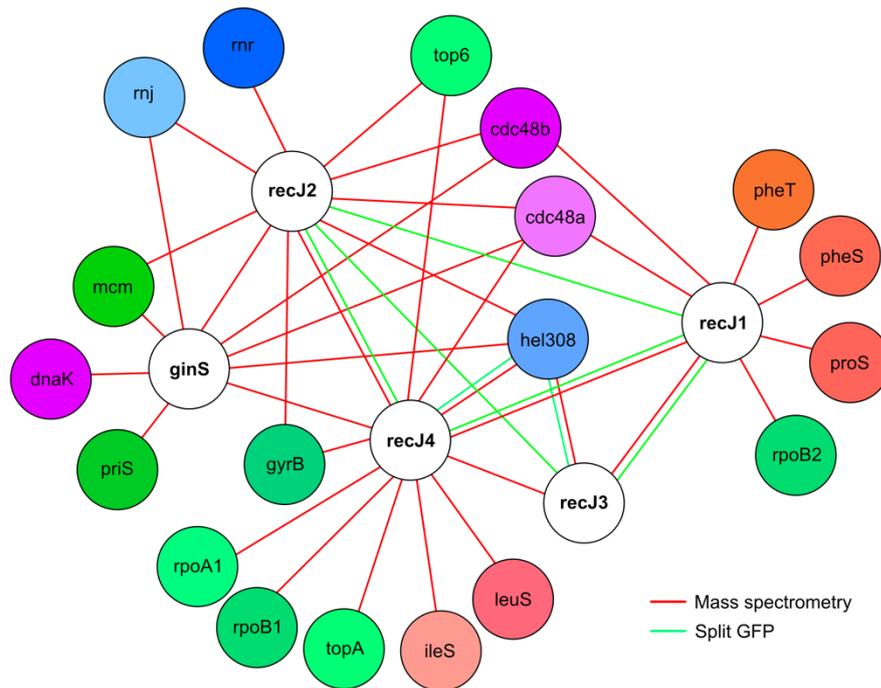


Figure 6.3: Summary of protein:protein interactions observed during this study. RecJ1, RecJ2, RecJ3, RecJ4 and GINS are shown in white, as the bait for the pulldown or split GFP interaction assays performed. Proteins identified as interactors are coloured according to their known or predicted functions. Green: DNA replication, purple: protein turnover, blue: DNA damage repair.

Phenotyping of *recJ* mutants provided further information regarding the role of RecJ proteins in *H. volcanii*. Evidence suggested a role for RecJ3 in DNA repair, which was further supported by its interaction with repair helicase Hel308. RecJ4 was shown to interact with a large number of proteins at the replication fork but showed no strong damage repair defect. Based on its region of disorder, it could be hypothesised it acts as a central scaffold protein and further work will be required to define this activity. Strains lacking *recJ1* showed an altered aphidicolin response, linking the absence of RecJ1 to an altered mode of DNA replication, providing strong evidence for RecJ1 having a GAN-like role. This was further supported by its sensitivity to certain DNA damaging agents that have previously been reported for other archaeal Δ *gan* mutants. It was previously reported that RecJ2 cannot be deleted (Lever, 2019). This study utilised ectopic expression of RecJ2 to complement the deletion, however the strain could not be generated. The seemed essentiality of RecJ2 remains under question, but its interaction with core proteins of both replication and translation could partially explain this.

This study provides the first evidence for an archaeal species where the roles of GAN and HAN cannot be readily defined using interactions or

Chapter 6: Discussion

bioinformatic analysis. To date, interactions between RecJ proteins in archaea have not been observed. Additionally, the interplay between RecJ proteins and key replication and repair proteins, such as Hel308 and Cdc48a/b, are suggestive of additional roles for RecJ proteins away from the predicted activities of GAN and HAN. It will be interesting to consider these interactions, and to further decipher the specificities of their roles in *H. volcanii* going forward.

Summary

DNA replication is a vital process, and its regulation is critical to ensure cell viability. This study has investigated numerous components of the archaeal replisome and shown that requirement for replisome components (namely PolB and GINS) differs in the presence and absence of origins. Work on the RecJ family in *H. volcanii* has shown a previously unobserved interplay of RecJ homologues, suggesting gene duplication can alter functions and relationships between these proteins. The interplay between DNA polymerases, replication origins and the CMG complex warrants further work. Differential usage of polymerases and CMG proteins in the presence or absence of origins could provide critical information on the mechanisms of both canonical and recombination-dependent DNA replication in archaea.

Chapter 7: References

- Albers, S. V., Forterre, P., Prangishvili, D. & Schleper, C. 2013. The legacy of Carl Woese and Wolfram Zillig: from phylogeny to landmark discoveries. *Nat Rev Microbiol*, 11, 713-9.
- Albers, S. V. & Meyer, B. H. 2011. The archaeal cell envelope. *Nat Rev Microbiol*, 9, 414-26.
- Allen, J. M., Simcha, D. M., Ericson, N. G., Alexander, D. L., Marquette, J. T., Van Biber, B. P., Troll, C. J., Karchin, R., *et al.* 2011. Roles of DNA polymerase I in leading and lagging-strand replication defined by a high-resolution mutation footprint of ColE1 plasmid replication. *Nucleic Acids Res*, 39, 7020-33.
- Allers, T. 2010. Overexpression and purification of halophilic proteins in *Haloferax volcanii*. *Bioeng Bugs*, 1, 288-90.
- Allers, T., Barak, S., Liddell, S., Wardell, K. & Mevarech, M. 2010. Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in *Haloferax volcanii*. *Appl Environ Microbiol*, 76, 1759-69.
- Allers, T. & Mevarech, M. 2005. Archaeal genetics - the third way. *Nat Rev Genet*, 6, 58-73.
- Allers, T. & Ngo, H. P. 2003. Genetic analysis of homologous recombination in Archaea: *Haloferax volcanii* as a model organism. *Biochem Soc Trans*, 31, 706-10.
- Allers, T., Ngo, H. P., Mevarech, M. & Lloyd, R. G. 2004. Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. *Appl Environ Microbiol*, 70, 943-53.
- Anjum, R. S., Bray, S. M., Blackwood, J. K., Kilkenny, M. L., Coelho, M. A., Foster, B. M., Li, S., Howard, J. A., *et al.* 2015. Involvement of a eukaryotic-like ubiquitin-related modifier in the proteasome pathway of the archaeon *Sulfolobus acidocaldarius*. *Nat Commun*, 6, 8163.
- Aparicio, T., Guillou, E., Coloma, J., Montoya, G. & Mendez, J. 2009. The human GINS complex associates with Cdc45 and MCM and is essential for DNA replication. *Nucleic Acids Research*, 37, 2087-2095.
- Aravind, L. & Koonin, E. V. 1998. A novel family of predicted phosphoesterases includes *Drosophila* prune protein and bacterial RecJ exonuclease. *Trends in Biochemical Sciences*, 23, 17-19.
- Arezi, B. & Kuchta, R. D. 2000. Eukaryotic DNA primase. *Trends in Biochemical Sciences*, 25, 572-576.
- Arudchandran, A., Cerritelli, S., Narimatsu, S., Itaya, M., Shin, D. Y., Shimada, Y. & Crouch, R. J. 2000. The absence of ribonuclease H1 or H2 alters the sensitivity of *Saccharomyces cerevisiae* to hydroxyurea, caffeine and ethyl methanesulphonate: implications for roles of RNases H in DNA replication and repair. *Genes Cells*, 5, 789-802.
- Asai, T. & Kogoma, T. 1994. D-loops and R-loops - Alternative mechanisms for the initiation of chromosome replication in *Escherichia coli*. *Journal of Bacteriology*, 176, 1807-1812.

- Auchtung, T. A., Takacs-Vesbach, C. D. & Cavanaugh, C. M. 2006. 16S rRNA phylogenetic investigation of the candidate division "Korarchaeota". *Appl Environ Microbiol*, 72, 5077-82.
- Ausiannikava, D. & Allers, T. 2017. Diversity of DNA Replication in the Archaea. *Genes (Basel)*, 8.
- Ausiannikava, D., Mitchell, L., Marriott, H., Smith, V., Hawkins, M., Makarova, K. S., Koonin, E. V., Nieduszynski, C. A., *et al.* 2018. Evolution of Genome Architecture in Archaea: Spontaneous Generation of a New Chromosome in *Haloferax volcanii*. *Mol Biol Evol*, 35, 1855-1868.
- Aussel, L., Barre, F. X., Aroyo, M., Stasiak, A., Stasiak, A. Z. & Sherratt, D. 2002. FtsK is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. *Cell*, 108, 195-205.
- Babski, J., Haas, K. A., Nather-Schindler, D., Pfeiffer, F., Forstner, K. U., Hammelmann, M., Hilker, R., Becker, A., *et al.* 2016. Genome-wide identification of transcriptional start sites in the haloarchaeon *Haloferax volcanii* based on differential RNA-Seq (dRNA-Seq). *Bmc Genomics*, 17.
- Bailey, R., Priego Moreno, S. & Gambus, A. 2015. Termination of DNA replication forks: "Breaking up is hard to do". *Nucleus*, 6, 187-96.
- Barrangou, R. & Marraffini, L. A. 2014. CRISPR-Cas Systems: Prokaryotes Upgrade to Adaptive Immunity. *Molecular Cell*, 54, 234-244.
- Barry, E. R. & Bell, S. D. 2006. DNA replication in the archaea. *Microbiol Mol Biol Rev*, 70, 876-87.
- Barry, E. R., Mcgeoch, A. T., Kelman, Z. & Bell, S. D. 2007. Archaeal MCM has separable processivity, substrate choice and helicase domains. *Nucleic Acids Res*, 35, 988-98.
- Bastia, D., Zzaman, S., Krings, G., Saxena, M., Peng, X. H. & Greenberg, M. M. 2008. Replication termination mechanism as revealed by Tus-mediated polar arrest of a sliding helicase. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 12831-12836.
- Bauer, R. J., Begley, M. T. & Trakselis, M. A. 2012. Kinetics and Fidelity of Polymerization by DNA Polymerase III from *Sulfolobus solfataricus*. *Biochemistry*, 51, 1996-2007.
- Baxter, J. & Diffley, J. F. X. 2008. Topoisomerase II inactivation prevents the completion of DNA replication in budding yeast. *Molecular Cell*, 30, 790-802.
- Beattie, T. R. & Reyes-Lamothe, R. 2015. A Replisome's journey through the bacterial chromosome. *Front Microbiol*, 6, 562.
- Belanger, K. G. & Kreuzer, K. N. 1998. Bacteriophage T4 initiates bidirectional DNA replication through a two-step process. *Molecular Cell*, 2, 693-701.
- Bell, S. D. 2011. DNA replication: archaeal oriGINS. *Bmc Biology*, 9.
- Bell, S. P. & Dutta, A. 2002. DNA replication in eukaryotic cells. *Annual Review of Biochemistry*, 71, 333-374.
- Bell, S. P. & Stillman, B. 1992. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature*, 357, 128-134.

Chapter 7: References

- Berezney, R., Dubey, D. D. & Huberman, J. A. 2000. Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. *Chromosoma*, 108, 471-484.
- Berghuis, B. A., Dulin, D., Xu, Z. Q., Van Laar, T., Cross, B., Janissen, R., Jergic, S., Dixon, N. E., *et al.* 2015. Strand separation establishes a sustained lock at the Tus-Ter replication fork barrier. *Nature Chemical Biology*, 11, 579-U75.
- Berkemer, S. J., Maier, L. K., Amman, F., Bernhart, S. H., Wortz, J., Markle, P., Pfeiffer, F., Stadler, P. F., *et al.* 2020. Identification of RNA 3' ends and termination sites in *Haloferax volcanii*. *Rna Biology*, 17, 663-676.
- Bermudez, V. P., Farina, A., Raghavan, V., Tappin, I. & Hurwitz, J. 2011. Studies on Human DNA Polymerase epsilon and GINS Complex and Their Role in DNA Replication. *Journal of Biological Chemistry*, 286, 28963-28977.
- Berquist, B. R., Dassarma, P. & Dassarma, S. 2007. Essential and non-essential DNA replication genes in the model halophilic Archaeon, *Halobacterium* sp. NRC-1. *BMC Genet*, 8, 31.
- Birien, T., Thiel, A., Henneke, G., Flament, D., Moalic, Y. & Jebbar, M. 2018. Development of an Effective 6-Methylpurine Counterselection Marker for Genetic Manipulation in *Thermococcus barophilus*. *Genes (Basel)*, 9.
- Bitan-Banin, G., Ortenberg, R. & Mevarech, M. 2003. Development of a gene knockout system for the halophilic archaeon *Haloferax volcanii* by use of the *pyrE* gene. *J Bacteriol*, 185, 772-8.
- Bleichert, F., Botchan, M. R. & Berger, J. M. 2015. Crystal structure of the eukaryotic origin recognition complex. *Nature*, 519, 321-6.
- Bochkarev, A., Pfuetzner, R. A., Edwards, A. M. & Frappier, L. 1997. Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature*, 385, 176-181.
- Bogenschutz, N. L., Rodriguez, J. & Tsukiyama, T. 2014. Initiation of DNA Replication from Non-Canonical Sites on an Origin-Depleted Chromosome. *Plos One*, 9.
- Braun, F., Thomalla, L., Van Der Does, C., Quax, T. E. F., Allers, T., Kaever, V. & Albers, S. V. 2019. Cyclic nucleotides in archaea: Cyclic di-AMP in the archaeon *Haloferax volcanii* and its putative role. *Microbiologyopen*, 8, e00829.
- Brendel, J., Stoll, B., Lange, S. J., Sharma, K., Lenz, C., Stachler, A. E., Maier, L. K., Richter, H., *et al.* 2014. A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (crispr)-derived rnas (crnas) in *Haloferax volcanii*. *J Biol Chem*, 289, 7164-77.
- Breuert, S., Allers, T., Spohn, G. & Soppa, J. 2006. Regulated polyploidy in halophilic archaea. *PLoS ONE*, 1, e92.
- Brewster, A. S. & Chen, X. J. S. 2010. Insights into the MCM functional mechanism: lessons learned from the archaeal MCM complex. *Critical Reviews in Biochemistry and Molecular Biology*, 45, 243-256.
- Brewster, A. S., Wang, G. G., Yu, X., Greenleaf, W. B., Carazo, J. M., Tjajadia, M., Klein, M. G. & Chen, X. J. S. 2008. Crystal structure of a near-full-length archaeal MCM: Functional insights for an AAA plus

- hexameric helicase. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 20191-20196.
- Bryson, V. & Szybalski, W. 1952. Microbial Selection. *Science*, 116, 45-51.
- Burgers, P. M. 2009. Polymerase dynamics at the eukaryotic DNA replication fork. *J Biol Chem*, 284, 4041-5.
- Burgers, P. M. J. & Kunkel, T. A. 2017. Eukaryotic DNA Replication Fork. *Annu Rev Biochem*, 86, 417-438.
- Burkhardt, B. W. C., L.; Heider, M. R.; Kelman, Z.; Reeve, J. N.; Santangelo, T. J. 2017. The GAN Exonuclease or the Flap Endonuclease Fen1 and RNase HIII are necessary for viability of *Thermococcus kodakarensis*. *J Bacteriol*, 199, e00141-17.
- Bycroft, M., Hubbard, T. J. P., Proctor, M., Freund, S. M. V. & Murzin, A. G. 1997. The solution structure of the S1 RNA binding domain: A member of an ancient nucleic acid-binding fold. *Cell*, 88, 235-242.
- Calzada, A., Hodgson, B., Kanemaki, M., Bueno, A. & Labib, K. 2005. Molecular anatomy and regulation of a stable replisome eukaryotic DNA at a paused replication fork. *Genes & Development*, 19, 1905-1919.
- Cann, I. K., Komori, K., Toh, H., Kanai, S. & Ishino, Y. 1998. A heterodimeric DNA polymerase: evidence that members of Euryarchaeota possess a distinct DNA polymerase. *Proc Natl Acad Sci USA*, 95, 14250-5.
- Cann, I. K. O. K., K.; Toh, H.; Kanai, S.; Ishino, Y. 1998. A heterodimeric DNA polymerase: Evidence that members of Euryarchaeota possess a distinct DNA polymerase. *Proc Natl Acad Sci USA*, 95, 14250-14255.
- Carles-Kinch, K. & Kreuzer, K. N. 1997. RNA-DNA hybrid formation at a bacteriophage T4 replication origin. *Journal of Molecular Biology*, 266, 915-926.
- Carpentieri, F., De Felice, M., De Falco, M., Rossi, M. & Pisani, F. M. 2002. Physical and functional interaction between the mini-chromosome maintenance-like DNA helicase and the single-stranded DNA binding protein from the crenarchaeon *Sulfolobus solfataricus*. *Journal of Biological Chemistry*, 277, 12118-12127.
- Chaban, B., Ng, S. Y. & Jarrell, K. F. 2006. Archaeal habitats--from the extreme to the ordinary. *Can J Microbiol*, 52, 73-116.
- Chaconas, G. & Kobryn, K. 2010. Structure, Function, and Evolution of Linear Replicons in *Borrelia*. *Annual Review of Microbiology*, Vol 64, 2010, 64, 185-+.
- Champoux, J. J. 2001. DNA topoisomerases: Structure, function, and mechanism. *Annual Review of Biochemistry*, 70, 369-413.
- Chapados, B. R., Chai, Q., Hosfield, D. J., Qiu, J., Shen, B. & Tainer, J. A. 2001. Structural biochemistry of a type 2 RNase H: RNA primer recognition and removal during DNA replication. *J Mol Biol*, 307, 541-56.
- Cheng, K., Zhao, Y., Chen, X., Li, T., Wang, L., Xu, H., Tian, B. & Hua, Y. 2015. A Novel C-Terminal Domain of RecJ is Critical for Interaction with HerA in *Deinococcus radiodurans*. *Front Microbiol*, 6, 1302.
- Cheng, K. Y., Xu, H., Chen, X. Y., Wang, L. Y., Tian, B., Zhao, Y. & Hua, Y. J. 2016. Structural basis for DNA 5'-end resection by RecJ. *Elife*, 5.

Chapter 7: References

- Chilkova, O., Stenlund, P., Isoz, I., Stith, C. M., Grabowski, P., Lundstrom, E. B., Burgers, P. M. & Johansson, E. 2007. The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic Acids Res*, 35, 6588-97.
- Choi, J. M., Lim, H. S., Kim, J. J., Song, O. K. & Cho, Y. J. 2007. Crystal structure of the human GINS complex. *Genes & Development*, 21, 1316-1321.
- Chow, M. H. & Courcelle, J. 2007. RecBCD and RecJ/RecQ initiate DNA degradation on distinct substrates in UV-irradiated *Escherichia coli*. *Radiation Research*, 168, 499-506.
- Christian, J. H. & Waltho, J. A. 1962. Solute concentrations within cells of halophilic and non-halophilic bacteria. *Biochim Biophys Acta*, 65, 506-8.
- Cline, S. W., Lam, W. L., Charlebois, R. L., Schalkwyk, L. C. & Doolittle, W. F. 1989a. Transformation methods for halophilic archaeobacteria. *Can J Microbiol*, 35, 148-52.
- Cline, S. W., Schalkwyk, L. C. & Doolittle, W. F. 1989b. Transformation of the archaeobacterium *Halobacterium volcanii* with genomic DNA. *J Bacteriol*, 171, 4987-91.
- Cocker, J. H., Piatti, S., Santocanale, C., Nasmyth, K. & Diffley, J. F. X. 1996. An essential role for the Cdc6 protein in forming the pre-replicative complexes of budding yeast. *Nature*, 379, 180-182.
- Cohen, S., Shilo, M. & Kessel, M. 1991. Nature of the salt dependence of the envelope of a Dead-Sea archaeobacterium, *Haloferax volcanii*. *Archives of Microbiology*, 156, 198-203.
- Connolly, B. A. 2009. Recognition of deaminated bases by archaeal family-B DNA polymerases. *Biochem Soc Trans*, 37, 65-8.
- Cooper, C. D. O. 2018. Archaeal DNA polymerases: new frontiers in DNA replication and repair. *Emerging Topics in Life Sciences*, 2, 503-516.
- Corn, J. E. & Berger, J. M. 2006. Regulation of bacterial priming and daughter strand synthesis through helicase-primase interactions. *Nucleic Acids Research*, 34, 4082-4088.
- Courcelle, J. & Hanawalt, P. C. 1999. RecQ and RecJ process blocked replication forks prior to the resumption of replication in UV-irradiated *Escherichia coli*. *Molecular and General Genetics*, 262, 543-551.
- Cubonova, L., Richardson, T., Burkhart, B. W., Kelman, Z., Connolly, B. A., Reeve, J. N. & Santangelo, T. J. 2013. Archaeal DNA polymerase D but not DNA polymerase B is required for genome replication in *Thermococcus kodakarensis*. *J Bacteriol*, 195, 2322-8.
- Cucinotta, C. E., Hildreth, A. E., Mcshane, B. M., Shirra, M. K. & Arndt, K. M. 2019. The nucleosome acidic patch directly interacts with subunits of the Paf1 and FACT complexes and controls chromatin architecture in vivo. *Nucleic Acids Research*, 47, 8410-8423.
- Da Cunha, V., Gaia, M., Gabelle, D., Nasir, A. & Forterre, P. 2017. Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS Genet*, 13, e1006810.
- Dantuluri, S. 2018. *AAA+ ATPase Cdc48a and its association with ubiquitin-like SAMP1 and DNA repair in archaea*. PhD, University of Florida.

- Davey, M. J., Fang, L. H., Mcinerney, P., Georgescu, R. E. & O'donnell, M. 2002. The DnaC helicase loader is a dual ATP/ADP switch protein. *Embo Journal*, 21, 3148-3159.
- De Falco, M., Ferrari, E., De Felice, M., Rossi, M., Hubscher, U. & Pisani, F. M. 2007. The human GINS complex binds to and specifically stimulates human DNA polymerase alpha-primase. *Embo Reports*, 8, 99-103.
- Delmas, S., Duggin, I. G. & Allers, T. 2013. DNA damage induces nucleoid compaction via the Mre11-Rad50 complex in the archaeon *Haloferax volcanii*. *Mol Microbiol*, 87, 168-79.
- Delmas, S., Shunburne, L., Ngo, H. P. & Allers, T. 2009. Mre11-Rad50 promotes rapid repair of DNA damage in the polyploid archaeon *Haloferax volcanii* by restraining homologous recombination. *PLoS Genet*, 5, e1000552.
- Depamphilis, M. L. & Bell, S. D. 2011. *Genome duplication*, New York ; Abingdon, New York ; Abingdon : Garland Science.
- Devbhandari, S., Jiang, J. Q., Kumar, C., Whitehouse, I. & Remus, D. 2017. Chromatin Constrains the Initiation and Elongation of DNA Replication. *Molecular Cell*, 65, 131-141.
- Dewar, J. M., Budzowska, M. & Walter, J. C. 2015. The mechanism of DNA replication termination in vertebrates. *Nature*, 525, 345-50.
- Dewar, J. M., Low, E., Mann, M., Raschle, M. & Walter, J. C. 2017. CRL2(Lrr1) promotes unloading of the vertebrate replisome from chromatin during replication termination. *Genes & Development*, 31, 275-290.
- Dewar, J. M. & Walter, J. C. 2017. Mechanisms of DNA replication termination. *Nat Rev Mol Cell Biol*, 18, 507-516.
- Dianov, G., Sedgwick, B., Daly, G., Olsson, M., Lovett, S. & Lindahl, T. 1994. Release of 5'-terminal deoxyribose-phosphate residues from incised abasic sites in DNA by the Escherichia coli RecJ protein. *Nucleic Acids Research*, 22, 993-998.
- Dimude, J. U., Midgley-Smith, S. L. & Rudolph, C. J. 2018. Replication-transcription conflicts trigger extensive DNA degradation in Escherichia coli cells lacking RecBCD. *DNA Repair*, 70, 37-48.
- Dombrowski, N., Lee, J. H., Williams, T. A., Offre, P. & Spang, A. 2019. Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *Fems Microbiology Letters*, 366.
- Donnianni, R. a. S., L. S. 2013. Break-induced replication occurs by conservative DNA synthesis. *PNAS*, 110, 13475-13480.
- Douglas, M. E., Ali, F. A., Costa, A. & Diffley, J. F. X. 2018. The mechanism of eukaryotic CMG helicase activation. *Nature*, 555, 265-268.
- Drissi, R., Chauvin, A., Mckenna, A., Levesque, D., Blais-Brochu, S., Jean, D. & Boisvert, F. M. 2018. Destabilization of the MiniChromosome Maintenance (MCM) complex modulates the cellular response to DNA double strand breaks. *Cell Cycle*, 17, 2593-2609.
- Duan, Z. 2009. *Genetic analysis of two structure-specific endonucleases Hef and Fen1 in archaeon Haloferax volcanii*. PhD thesis, University of Nottingham.

Chapter 7: References

- Duderstadt, K. E. & Berger, J. M. 2013. A structural framework for replication origin opening by AAA⁺ initiation factors. *Curr Opin Struct Biol*, 23, 144-53.
- Duggin, I. G., Dubarry, N. & Bell, S. D. 2011. Replication termination and chromosome dimer resolution in the archaeon *Sulfolobus solfataricus*. *EMBO J*, 30, 145-53.
- Duggin, I. G., Wake, R. G., Bell, S. D. & Hill, T. M. 2008. The replication fork trap and termination of chromosome replication. *Molecular Microbiology*, 70, 1323-1333.
- El-Andaloussi, N., Valovka, T., Toueille, M., Steinacher, R., Focke, F., Gehrig, P., Covic, M., Hassa, P. O., *et al.* 2006. Arginine methylation regulates DNA polymerase beta. *Molecular Cell*, 22, 51-62.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., *et al.* 2019. The Pfam protein families database in 2019. *Nucleic Acids Research*, 47, D427-D432.
- Elkins, J. G., Podar, M., Graham, D. E., Makarova, K. S., Wolf, Y., Randau, L., Hedlund, B. P., Brochier-Armanet, C., *et al.* 2008. A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc Natl Acad Sci U S A*, 105, 8102-7.
- Embley, M. & Williams, T. 2016. Only two domains, not three: Changing views on the tree of life. *Microbiology Today*, 43.
- Eme, L., Spang, A., Lombard, J., Stairs, C. W. & Ettema, T. J. G. 2017. Archaea and the origin of eukaryotes. *Nat Rev Microbiol*.
- Erzberger, J. P., Pirruccello, M. M. & Berger, J. M. 2002. The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation. *Embo Journal*, 21, 4763-4773.
- Espeli, O., Lee, C. & Marians, K. J. 2003. A physical and functional interaction between Escherichia coli FtsK and topoisomerase IV. *Journal of Biological Chemistry*, 278, 44639-44644.
- Evrin, C., Clarke, P., Zech, J., Lurz, R., Sun, J. C., Uhle, S., Li, H. L., Stillman, B., *et al.* 2009. A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 20240-20245.
- Fang, L. H., Davey, M. J. & O'donnell, M. 1999. Replisome assembly at oriC, the replication origin of E-coli, reveals an explanation for initiation sites outside an origin. *Molecular Cell*, 4, 541-553.
- Fang, M., Zeisberg, W. M., Condon, C., Ogryzko, V., Danchin, A. & Mechold, U. 2009. Degradation of nanoRNA is performed by multiple redundant RNases in Bacillus subtilis. *Nucleic Acids Research*, 37, 5114-5125.
- Feng, L., Chang, C. C., Song, D., Jiang, C., Song, Y., Wang, C. F., Deng, W., Zou, Y. J., *et al.* 2018. The trimeric Hef-associated nuclease HAN is a 3'-->5' exonuclease and is probably involved in DNA repair. *Nucleic Acids Res*, 46, 9027-9043.
- Fineran, P. C. 2019. CRISPR-Cas impedes archaeal mating. *Nature Microbiology*, 4, 2-3.
- Fischer, S., Maier, L. K., Stoll, B., Brendel, J., Fischer, E., Pfeiffer, F., Dyll-Smith, M. & Marchfelder, A. 2012. An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA. *J Biol Chem*, 287, 33351-63.

- Fitch, W. M. 1970. Distinguishing homologous from analogous proteins. *Systematic Zoology*, 19, 99-&.
- Fletcher, R. J., Shen, J., Gómez-Llorente, Y., Martín, C. S., Carazo, J. M. & Chen, X. S. 2005. Double Hexamer Disruption and Biochemical Activities of *Methanobacterium thermoautotrophicum* MCM. *Journal of Biological Chemistry*, 280, 42405-42410.
- Forouzan, D., Ammelburg, M., Hobel, C. F., Stroh, L. J., Sessler, N., Martin, J. & Lupas, A. N. 2012. The Archaeal Proteasome Is Regulated by a Network of AAA ATPases. *Journal of Biological Chemistry*, 287, 39254-39262.
- Forterre, P., Elie, C. & Kohiyama, M. 1984. Aphidicolin inhibits growth and DNA synthesis in halophilic archaeobacteria. *J Bacteriol*, 159, 800-2.
- Fouqueau, T., Blombach, F. & Werner, F. 2017. Evolutionary Origins of Two-Barrel RNA Polymerases and Site-Specific Transcription Initiation. *Annual Review of Microbiology*, Vol 71, 71, 331-348.
- Frick, D. N. & Richardson, C. C. 2001. DNA primases. *Annual Review of Biochemistry*, 70, 39-80.
- Fu, X., Adams, Z., Liu, R., Hepowit, N. L., Wu, Y. F., Bowmann, C. F., Moskovitz, J. & Maupin-Furlow, J. A. 2017. Methionine Sulfoxide Reductase A (MsrA) and Its Function in Ubiquitin-Like Protein Modification in Archaea. *Mbio*, 8.
- Fujikane, R., Ishino, S., Ishino, Y. & Forterre, P. 2010. Genetic analysis of DNA repair in the hyperthermophilic archaeon, *Thermococcus kodakaraensis*. *Genes Genet Syst*, 85, 243-57.
- Fullbright, G., Rycenga, H. B., Gruber, J. D. & Long, D. T. 2016. p97 Promotes a Conserved Mechanism of Helicase Unloading during DNA Cross-Link Repair. *Molecular and Cellular Biology*, 36, 2983-2994.
- Gabalton, T. & Huynen, M. A. 2004. Prediction of protein function and pathways in the genome era. *Cell Mol Life Sci*, 61, 930-44.
- Galal, W. C., Pan, M., Kelman, Z. & Hurwitz, J. 2012. Characterization of DNA Primase Complex Isolated from the Archaeon, *Thermococcus kodakaraensis*. *Journal of Biological Chemistry*, 287, 16209-16219.
- Gamble-Milner, R. J. 2016. *Genetic analysis of the Hel308 helicase in the archaeon Haloferax volcanii*. University of Nottingham.
- Gambus, A. 2017. Termination of Eukaryotic Replication Forks. *DNA Replication: from Old Principles to New Discoveries*, 1042, 163-187.
- Gambus, A., Jones, R. C., Sanchez-Diaz, A., Kanemaki, M., Van Deursen, F., Edmondson, R. D. & Labib, K. 2006. GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nature Cell Biology*, 8, 358-U41.
- Ganai, R. A., Zhang, X. P., Heyer, W. D. & Johansson, E. 2016. Strand displacement synthesis by yeast DNA polymerase epsilon. *Nucleic Acids Research*, 44, 8229-8240.
- Garcia-Gomez, S., Reyes, A., Martinez-Jimenez, M. I., Chocron, E. S., Mouron, S., Terrados, G., Powel, C., Salido, E., *et al.* 2013. PrimPol, an Archaic Primase/Polymerase Operating in Human Cells. *Molecular Cell*, 52, 541-553.
- Garg, P. & Burgers, P. M. J. 2005. DNA polymerases that propagate the eukaryotic DNA replication fork. *Critical Reviews in Biochemistry and Molecular Biology*, 40, 115-128.

Chapter 7: References

- Garg, P., Stith, C. M., Sabouri, N., Johansson, E. & Burgers, P. M. 2004. Idling by DNA polymerase delta maintains a ligatable nick during lagging-strand DNA replication. *Genes & Development*, 18, 2764-2773.
- Gehring, A. M., Astling, D. P., Matsumi, R., Burkhart, B. W., Kelman, Z., Reeve, J. N., Jones, K. L. & Santangelo, T. J. 2017. Genome Replication in *Thermococcus kodakarensis* Independent of Cdc6 and an Origin of Replication. *Front Microbiol*, 8, 2084.
- Giroux, X. & Macneill, S. A. 2015. Molecular Genetic Methods to Study DNA Replication Protein Function in *Haloferax volcanii*, A Model Archaeal Organism. *Methods Mol Biol*, 1300, 187-218.
- Gogarten, J. P., Doolittle, W. F. & Lawrence, J. G. 2002. Prokaryotic evolution in light of gene transfer. *Mol Biol Evol*, 19, 2226-38.
- Goldfless, S. J., Morag, A. S., Belisle, K. A., Sutera, V. A., Jr. & Lovett, S. T. 2006. DNA repeat rearrangements mediated by DnaK-dependent replication fork repair. *Mol Cell*, 21, 595-604.
- Goldstein, S., Meyerstein, D. & Czapski, G. 1993. The Fenton Reagents. *Free Radical Biology and Medicine*, 15, 435-445.
- Goto, T. & Wang, J. C. 1984. Yeast DNA Topoisomerase II is encoded by a single-copy, essential gene. *Cell*, 36, 1073-1080.
- Greagg, M. A., Fogg, M. J., Panayotou, G., Evans, S. J., Connolly, B. A. & Pearl, L. H. 1999. A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil. *Proc Natl Acad Sci U S A*, 96, 9045-50.
- Greenfeder, S. A. & Newlon, C. S. 1992. Replication forks pause at yeast centromeres. *Molecular and Cellular Biology*, 12, 4056-4066.
- Greenough, L., Kelman, Z. & Gardner, A. F. 2015. The roles of family B and D DNA polymerases in *Thermococcus* species 9 degrees N Okazaki fragment maturation. *J Biol Chem*, 290, 12514-22.
- Gregor, D. & Pfeifer, F. 2005. In vivo analyses of constitutive and regulated promoters in halophilic archaea. *Microbiology-Sgm*, 151, 25-33.
- Griese, M., Lange, C. & Soppa, J. 2011. Ploidy in cyanobacteria. *FEMS Microbiol Lett*, 323, 124-31.
- Gruz, P., Shimizu, M., Pisani, F. M., De Felice, M., Kanke, Y. & Nohmi, T. 2003. Processing of DNA lesions by archaeal DNA polymerases from *Sulfolobus solfataricus*. *Nucleic Acids Research*, 31, 4024-4030.
- Guy, L. & Ettema, T. J. 2011. The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol*, 19, 580-7.
- Haber, J. E. 2013. *Genome Stability: DNA Repair and Recombination*, New York, NY, Garland Science.
- Haldenby, S. 2007. Genetic analysis of RadB, a paralogue of the archaeal Rad51/RecA homologue, RadA. *PhD thesis*, University of Nottingham.
- Han, E. S., Cooper, D. L., Persky, N. S., Sutera, V. A., Whitaker, R. D., Montello, M. L. & Lovett, S. T. 2006. RecJ exonuclease: substrates, products and interaction with SSB. *Nucleic Acids Research*, 34, 1084-1091.
- Hanahan, D. & Weinberg, R. A. 2000. The hallmarks of cancer. *Cell*, 100, 57-70.
- Hartman, A. L., Norais, C., Badger, J. H., Delmas, S., Haldenby, S., Madupu, R., Robinson, J., Khouri, H., *et al.* 2010. The complete genome

- sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS ONE*, 5, e9605.
- Hawkins, M., Malla, S., Blythe, M. J., Nieduszynski, C. A. & Allers, T. 2013a. Accelerated growth in the absence of DNA replication origins. *Nature*, 503, 544-7.
- Hawkins, M., Retkute, R., Muller, C. A., Saner, N., Tanaka, T. U., De Moura, A. P. & Nieduszynski, C. A. 2013b. High-resolution replication profiles define the stochastic nature of genome replication initiation and termination. *Cell Rep*, 5, 1132-41.
- Helleday, T., Eshtad, S. & Nik-Zainal, S. 2014. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet*, 15, 585-98.
- Heller, R. C., Kang, S., Lam, W. M., Chen, S., Chan, C. S. & Bell, S. P. 2011. Eukaryotic Origin-Dependent DNA Replication In Vitro Reveals Sequential Action of DDK and S-CDK Kinases. *Cell*, 146, 80-91.
- Henneke, G. 2012. In vitro reconstitution of RNA primer removal in Archaea reveals the existence of two pathways. *Biochemical Journal*, 447, 271-280.
- Henneke, G., Flament, D., Hubscher, U., Querellou, J. & Raffin, J. P. 2005. The hyperthermophilic euryarchaeota *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication. *Journal of Molecular Biology*, 350, 53-64.
- Hiasa, H. & Marians, K. J. 1994. Topoisomerase III, but not Topoisomerase I, can support nascent chain elongation during Theta-type DNA replication. *Journal of Biological Chemistry*, 269, 32655-32659.
- Hogg, M., Osterman, P., Bylund, G. O., Ganai, R. A., Lundstrom, E. B., Sauer-Eriksson, A. E. & Johansson, E. 2014. Structural basis for processive DNA synthesis by yeast DNA polymerase epsilon. *Nature Structural & Molecular Biology*, 21, 49-+.
- Hogrel, G., Lu, Y., Alexandre, N., Bosse, A., Dulermo, R., Ishino, S., Ishino, Y. & Flament, D. 2020. Role of RadA and DNA Polymerases in Recombination-Associated DNA Synthesis in Hyperthermophilic Archaea. *Biomolecules*, 10.
- Holmes, M. L. & Dyll-Smith, M. L. 2000. Sequence and expression of a halobacterial beta-galactosidase gene. *Mol Microbiol*, 36, 114-122.
- Holmes, M. L., Nuttall, S. D. & Dyll-Smith, M. L. 1991. Construction and use of halobacterial shuttle vectors and further studies on *Haloferax* DNA gyrase. *J Bacteriol*, 173, 3807-13.
- Howley, P. M., Israel, M. A., Law, M. F. & Martin, M. A. 1979. A rapid method for detecting and mapping homology between heterologous DNAs. Evaluation of polyomavirus genomes. *J Biol Chem*, 254, 4876-83.
- Hubscher, U., Maga, G. & Spadari, S. 2002. Eukaryotic DNA polymerases. *Annu Rev Biochem*, 71, 133-63.
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47, D309-D314.

Chapter 7: References

- Huet, J., Schnabel, R., Sentenac, A. & Zillig, W. 1983. Archaeobacteria and eukaryotes possess DNA-dependent RNA polymerases of a common type. *EMBO J*, 2, 1291-4.
- Humbard, M. A., Miranda, H. V., Lim, J. M., Krause, D. J., Pritz, J. R., Zhou, G., Chen, S., Wells, L., *et al.* 2010. Ubiquitin-like small archaeal modifier proteins (SAMPs) in *Haloferax volcanii*. *Nature*, 463, 54-60.
- Hwang, D. S. & Kornberg, A. 1992. Opening of the replication origin of *Escherichia coli* by DnaA protein with protein HU or IHF. *Journal of Biological Chemistry*, 267, 23083-23086.
- Hyjek, M., Figiel, M. & Nowotny, M. 2019. RNases H: Structure and mechanism. *DNA Repair (Amst)*, 102672.
- Ilves, I., Petojevic, T., Pesavento, J. J. & Botchan, M. R. 2010. Activation of the MCM2-7 Helicase by Association with Cdc45 and GINS Proteins. *Molecular Cell*, 37, 247-258.
- Imachi, H., Nobu, M. K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., Takano, Y., Uematsu, K., *et al.* 2020. Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature*, 577, 519-+.
- Indiani, C., Mcinerney, P., Georgescu, R., Goodman, M. F. & O'donnell, M. 2005. A sliding-clamp toolbelt binds high- and low-fidelity DNA polymerases simultaneously. *Mol Cell*, 19, 805-15.
- Indiani, C. & O'donnell, M. 2006. The replication clamp-loading machine at work in the three domains of life. *Nat Rev Mol Cell Biol*, 7, 751-61.
- Inoue, R., Kaito, C., Tanabe, M., Kamura, K., Akimitsu, N. & Sekimizu, K. 2001. Genetic identification of two distinct DNA polymerases, DnaE and PolC, that are essential for chromosomal DNA replication in *Staphylococcus aureus*. *Molecular Genetics and Genomics*, 266, 564-571.
- Ira, G. & Haber, J. E. 2002. Characterization of RAD51-independent break-induced replication that acts preferentially with short homologous sequences. *Molecular and Cellular Biology*, 22, 6384-6392.
- Ishino, S., Fujino, S., Tomita, H., Ogino, H., Takao, K., Daiyasu, H., Kanai, T., Atomi, H., *et al.* 2011. Biochemical and genetical analyses of the three mcm genes from the hyperthermophilic archaeon, *Thermococcus kodakarensis*. *Genes to Cells*, 16, 1176-1189.
- Ishino, S. & Ishino, Y. 2014. DNA polymerases as useful reagents for biotechnology - the history of developmental research in the field. *Frontiers in Microbiology*, 5.
- Ishino, S., Kelman, L. M., Kelman, Z. & Ishino, Y. 2013. The archaeal DNA replication machinery: past, present and future. *Genes Genet Syst*, 88, 315-9.
- Ishino, S., Yamagami, T., Kitamura, M., Kodera, N., Mori, T., Sugiyama, S., Ando, T., Goda, N., *et al.* 2014. Multiple interactions of the intrinsically disordered region between the helicase and nuclease domains of the archaeal Hef protein. *J Biol Chem*, 289, 21627-39.
- Ishino, Y., Iwasaki, H., Fukui, H., Mineno, J., Kato, I. & Shinagawa, H. 1992. Aphidicolin inhibits DNA polymerizing activity but not nucleolytic activity of *Escherichia coli* DNA polymerase II. *Biochimie*, 74, 131-136.
- Ishino, Y., Komori, K., Cann, I. K. & Koga, Y. 1998. A novel DNA polymerase family found in Archaea. *J Bacteriol*, 180, 2232-6.

- Ito, J. & Braithwaite, D. K. 1991. Compilation and alignment of DNA polymerase sequences. *Nucleic Acids Res*, 19, 4045-4057.
- Iyer, L. M., Leipe, D. D., Koonin, E. V. & Aravind, L. 2004. Evolutionary history and higher order classification of AAA plus ATPases. *Journal of Structural Biology*, 146, 11-31.
- Iyer, L. M. K., E. V.; Aravind, L. 2003. Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Structural Biology*, 3.
- Jacob, F., Brenner, S. & Cuzin, F. 1963. On the regulation of DNA replication in bacteria. *Cold Spring Harb Symp Quant Biol*, 28, 329-348.
- Jenkinson, E. R. & Chong, J. P. J. 2006. Minichromosome maintenance helicase activity is controlled by N- and C-terminal motifs and requires the ATPase domain helix-2 insert. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 7613-7618.
- Jeruzalmi, D., O'donnell, M. & Kuriyan, J. 2001. Crystal structure of the processivity clamp loader gamma (gamma) complex of E-coli DNA polymerase III. *Cell*, 106, 429-441.
- Johansson, E. & Dixon, N. 2013. Replicative DNA polymerases. *Cold Spring Harb Perspect Biol*, 5.
- Jokela, M., Eskelinen, A., Pospiech, H., Rouvinen, J. & Syvaaja, J. E. 2004. Characterization of the 3' exonuclease subunit DP1 of Methanococcus jannaschii replicative DNA polymerase D. *Nucleic Acids Res*, 32, 2430-40.
- Jones, N. T. 2019. *The role of homologous recombination in Haloferax volcanii*. PhD, University of Nottingham.
- Kamada, K., Kubota, Y., Arata, T., Shindo, Y. & Hanaoka, F. 2007. Structure of the human GINS complex and its assembly and functional interface in replication initiation. *Nature Structural & Molecular Biology*, 14, 388-396.
- Kanemaki, M., Sanchez-Diaz, A., Gambus, A. & Labib, K. 2003. Functional proteomic identification of DNA replication proteins by induced proteolysis in vivo. *Nature*, 423, 720-724.
- Kazlauskas, D., Krupovic, M., Guglielmini, J., Forterre, P. & Venclovas, C. 2020. Diversity and evolution of B-family DNA polymerases. *Nucleic Acids Res*, 48, 10142-10156.
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10, 845-858.
- Kelman, L. M. & Kelman, Z. 2003. Archaea: an archetype for replication initiation studies? *Mol Microbiol*, 48, 605-15.
- Kelman, L. M. & Kelman, Z. 2004. Multiple origins of replication in archaea. *Trends Microbiol*, 12, 399-401.
- Kelman, L. M. & Kelman, Z. 2014. Archaeal DNA Replication. In: Bassler, B. L. (ed.) *Annual Review of Genetics*, Vol 48.
- Kelman, L. M., O'dell, W. B. & Kelman, Z. 2020. Unwinding 20 Years of the Archaeal Minichromosome Maintenance Helicase. *Journal of Bacteriology*, 202.
- Kelman, Z., Lee, J. K. & Hurwitz, J. 1999a. The single minichromosome maintenance protein of Methanobacterium thermoautotrophicum

Chapter 7: References

- DeltaH contains DNA helicase activity. *Proc Natl Acad Sci U S A*, 96, 14783-8.
- Kelman, Z. & Odonnell, M. 1995. Structural and functional similarities of prokaryotic and eukaryotic DNA polymerase sliding clamps. *Nucleic Acids Research*, 23, 3613-3620.
- Kelman, Z., Pietrokovski, S. & Hurwitz, J. 1999b. Isolation and characterization of a split B-type DNA polymerase from the archaeon *Methanobacterium thermoautotrophicum deltaH*. *J Biol Chem*, 274, 28751-61.
- Kelman, Z. & White, M. F. 2005. Archaeal DNA replication and repair. *Curr Opin Microbiol*, 8, 669-76.
- Kerr, I. D., Wadsworth, R. I. M., Cubeddu, L., Blankenfeldt, W., Naismith, J. H. & White, M. F. 2003. Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of *Sulfolobus* SSB protein. *Embo Journal*, 22, 2561-2570.
- Khelaifia, S. & Raoult, D. 2016. *Haloferax massiliensis* sp. nov., the first human-associated halophilic archaea. *New Microbes New Infect*, 12, 96-8.
- Kogoma, T. 1997. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol Mol Biol Rev*, 61, 212-38.
- Kogoma, T. & Vonmeyerburg, K. 1983. The origin of replication, OriC, and the DnaA protein are dispensable in stable DNA replication (SDRA) mutants of *Escherichia coli* K-12. *Embo Journal*, 2, 463-468.
- Kolodner, R., Fishel, R. A. & Howard, M. 1985. Genetic recombination of bacterial plasmid DNA - Effect of RecF pathway mutations on plasmid recombination in *Escherichia coli*. *Journal of Bacteriology*, 163, 1060-1066.
- Komori, K., Fujikane, R., Shinagawa, H. & Ishino, Y. 2002. Novel endonuclease in Archaea cleaving DNA with various branched structure. *Genes Genet Syst*, 77, 227-41.
- Komori, K., Hidaka, M., Horiuchi, T., Fujikane, R., Shinagawa, H. & Ishino, Y. 2004. Cooperation of the N-terminal Helicase and C-terminal endonuclease activities of Archaeal Hef protein in processing stalled replication forks. *J Biol Chem*, 279, 53175-85.
- Komori, K. & Ishino, Y. 2001. Replication protein A in *Pyrococcus furiosus* is involved in homologous DNA recombination. *J Biol Chem*, 276, 25654-60.
- Koonin, E. V. 2005. Orthologs, paralogs, and evolutionary genomics. *Annual Review of Genetics*, 39, 309-338.
- Koonin, E. V., Krupovic, M., Ishino, S. & Ishino, Y. 2020. The replication machinery of LUCA: common origin of DNA replication and transcription. *BMC Biol*, 18, 61.
- Koonin, E. V. & Wolf, Y. I. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res*, 36, 6688-719.
- Koonin, E. V., Wolf, Y. I. & Aravind, L. 2000. Protein fold recognition using sequence profiles and its application in structural genomics. *Advances in Protein Chemistry*, Vol 54, 54, 245-275.

- Koonin, E. V. & Yutin, N. 2014. The Dispersed Archaeal Eukaryome and the Complex Archaeal Ancestor of Eukaryotes. *Cold Spring Harbor Perspectives in Biology*, 6.
- Korbel, J. O., Jensen, L. J., Von Mering, C. & Bork, P. 2004. Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs. *Nat Biotechnol*, 22, 911-7.
- Kornberg, A., Rao, N. N. & Ault-Riche, D. 1999. Inorganic polyphosphate: A molecule of many functions. *Annual Review of Biochemistry*, 68, 89-125.
- Kowalczykowski, S. C. 2000. Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem Sci*, 25, 156-165.
- Kowalski, D. & Eddy, M. J. 1989. The DNA unwinding element - a novel, cis-acting component that facilitates opening of the Escherichia coli replication origin. *Embo Journal*, 8, 4335-4344.
- Kramara, J., Osia, B. & Malkova, A. 2018. Break-Induced Replication: The Where, The Why, and The How. *Trends Genet*, 34, 518-531.
- Krastanova, I., Sannino, V., Amenitsch, H., Gileadi, O., Pisani, F. M. & Onesti, S. 2012. Structural and Functional Insights into the DNA Replication Factor Cdc45 Reveal an Evolutionary Relationship to the DHH Family of Phosphoesterases. *Journal of Biological Chemistry*, 287, 4121-4128.
- Kreuzer, K. N. & Brister, J. R. 2010. Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives. *Virol J*, 7, 358.
- Kushida, T., Narumi, I., Ishino, S., Ishino, Y., Fujiwara, S., Imanaka, T. & Higashibata, H. 2019. Pol B, a Family B DNA Polymerase, in *Thermococcus kodakarensis* is Important for DNA Repair, but not DNA Replication. *Microbes Environ*, 34, 316-326.
- Labib, K. & Gambus, A. 2007. A key role for the GINS complex at DNA replication forks. *Trends in Cell Biology*, 17, 271-278.
- Lam, W. L., Cohen, A., Tsoulouhas, D. & Doolittle, W. F. 1990. Genes for tryptophan biosynthesis in the archaebacterium *Haloferax volcanii*. *Proceedings of the National Academy of Sciences of the United States of America*, 87, 6614-6618.
- Langston, L. D., Zhang, D., Yurieva, O., Georgescu, R. E., Finkelstein, J., Yao, N. Y., Indiani, C. & O'donnell, M. E. 2014. CMG helicase and DNA polymerase epsilon form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 15390-15395.
- Lao-Sirieix, S. H. & Bell, S. D. 2004. The heterodimeric primase of the hyperthermophilic archaeon *Sulfolobus solfataricus* possesses DNA and RNA primase, polymerase and 3'-terminal nucleotidyl transferase activities. *Journal of Molecular Biology*, 344, 1251-1263.
- Lao-Sirieix, S. H., Pellegrini, L. & Bell, S. D. 2005. The promiscuous primase. *Trends in Genetics*, 21, 568-572.
- Large, A., Stamme, C., Lange, C., Duan, Z., Allers, T., Soppa, J. & Lund, P. A. 2007. Characterization of a tightly controlled promoter of the

Chapter 7: References

- halophilic archaeon *Haloferax volcanii* and its use in the analysis of the essential *cct1* gene. *Mol Microbiol*, 66, 1092-106.
- Lazzaro, F., Novarina, D., Amara, F., Watt, D. L., Stone, J. E., Costanzo, V., Burgers, P. M., Kunkel, T. A., *et al.* 2012. RNase H and postreplication repair protect cells from ribonucleotides incorporated in DNA. *Mol Cell*, 45, 99-110.
- Leigh, J. A., Albers, S. V., Atomi, H. & Allers, T. 2011. Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales. *FEMS Microbiol Rev*, 35, 577-608.
- Leipe, D. D., Aravind, L. & Koonin, E. V. 1999. Did DNA replication evolve twice independently? *Nucleic Acids Research*, 27, 3389-3401.
- Lemon, K. P. & Grossman, A. D. 1998. Localization of bacterial DNA polymerase: Evidence for a factory model of replication. *Science*, 282, 1516-1519.
- Leonard, A. C. & Grimwade, J. E. 2009. Initiating chromosome replication in *E. coli*: it makes sense to recycle. *Genes Dev*, 23, 1145-50.
- Leonard, A. C. & Mechali, M. 2013. DNA replication origins. *Cold Spring Harb Perspect Biol*, 5, a010116.
- Lesterlin, C., Barre, F. X. & Cornet, F. 2004. Genetic recombination and the cell cycle: what we have learned from chromosome dimers. *Molecular Microbiology*, 54, 1151-1160.
- Lestini, R., Duan, Z. & Allers, T. 2010. The archaeal Xpf/Mus81/FANCM homolog Hef and the Holliday junction resolvase Hjc define alternative pathways that are essential for cell viability in *Haloferax volcanii*. *DNA Repair*, 9, 994-1002.
- Lestini, R., Laptinok, S. P., Kuhn, J., Hink, M. A., Schanne-Klein, M. C., Liebl, U. & Myllykallio, H. 2013. Intracellular dynamics of archaeal FANCM homologue Hef in response to halted DNA replication. *Nucleic Acids Res*, 41, 10358-70.
- Lever, R. J. 2019. *Genetic and biochemical analysis of the Hel308 helicase in the archaeon Haloferax volcanii*. University of Nottingham.
- Li, H. & O'donnell, M. E. 2018. The Eukaryotic CMG Helicase at the Replication Fork: Emerging Architecture Reveals an Unexpected Mechanism. *Bioessays*, 40.
- Li, H. & Stillman, B. 2012. The origin recognition complex: a biochemical and structural view. *Subcell Biochem*, 62, 37-58.
- Li, M. J., Yi, G. S., Yu, F., Zhou, H., Chen, J. N., Xu, C. Y., Wang, F. P., Xiao, X., *et al.* 2017. The crystal structure of *Pyrococcus furiosus* RecJ implicates it as an ancestor of eukaryotic Cdc45. *Nucleic Acids Research*, 45, 12551-12564.
- Li, X. & Heyer, W. D. 2008. Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res*, 18, 99-113.
- Li, Z., Pan, M., Santangelo, T. J., Chemnitz, W., Yuan, W., Edwards, J. L., Hurwitz, J., Reeve, J. N., *et al.* 2011. A novel DNA nuclease is stimulated by association with the GINS complex. *Nucleic Acids Research*, 39, 6114-6123.
- Li, Z., Santangelo, T. J., Cubonova, L., Reeve, J. N. & Kelman, Z. 2010. Affinity purification of an archaeal DNA replication protein network. *MBio*, 1.

Chapter 7: References

- Lindås, A. C. & Bernander, R. 2013. The cell cycle of archaea. *Nat Rev Microbiol*, 11, 627-38.
- Liu, X. Q. 2000. Protein-splicing intein: Genetic mobility, origin, and evolution. *Annu Rev Genet*, 34, 61-76.
- Liu, Y., Kao, H. I. & Bambara, R. A. 2004. Flap endonuclease 1: A central component of DNA metabolism. *Annual Review of Biochemistry*, 73, 589-615.
- Llorente, B., Smith, C. E. & Symington, L. S. 2008. Break-induced replication: what is it and what is it for? *Cell Cycle*, 7, 859-64.
- Lloyd, R. G., Porton, M. C. & Buckman, C. 1988. Effect of recF, recJ, recN, recO and ruv mutations on ultraviolet survival and genetic recombination in a recD strain of Escherichia coli K12. *Mol Gen Genet*, 212, 317-24.
- Lodish, H., Berk, A., Zipursky, S. L. & Al., E. 2000. The role of Topoisomerases in DNA replication. In: Freeman, W. H. (ed.) *Molecular Cell Biology 4th Edition*. New York.
- Loeb, L. A. & Monnat, R. J., Jr. 2008. DNA polymerases and human disease. *Nat Rev Genet*, 9, 594-604.
- Lombard, J., Lopez-Garcia, P. & Moreira, D. 2012. The early evolution of lipid membranes and the three domains of life. *Nature Reviews Microbiology*, 10, 507-515.
- Lööke, M., Maloney, M. F. & Bell, S. P. 2017. Mcm10 regulates DNA replication elongation by stimulating the CMG replicative helicase. *Genes Dev*, 31, 291-305.
- Lovett, S. T. & Clark, A. J. 1984. Genetic analysis of the RecJ gene of Escherichia coli K12. *Journal of Bacteriology*, 157, 190-196.
- Lovett, S. T. & Kolodner, R. D. 1989. Identification and purification of a single-stranded DNA-specific exonuclease encoded by the RecJ gene of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 86, 2627-2631.
- Luder, A. & Mosig, G. 1982. Two alternative mechanisms for initiation of DNA replication forks in bacteriophage T4 - Priming by RNA polymerase and by recombination. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, 79, 1101-1105.
- Lujan, S. A., Williams, J. S. & Kunkel, T. A. 2016. DNA Polymerases Divide the Labor of Genome Replication. *Trends Cell Biol*, 26, 640-654.
- Lundgren, M., Andersson, A., Chen, L., Nilsson, P. & Bernander, R. 2004. Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. *Proc Natl Acad Sci U S A*, 101, 7046-51.
- Lundgren, M., Malandrino, L., Eriksson, S., Huber, H. & Bernander, R. 2008. Cell cycle characteristics of crenarchaeota: unity among diversity. *J Bacteriol*, 190, 5362-7.
- Lydeard, J. R. L.-M., Z.; Sheu, Y.; Stillman, B.; Burgers, P. M.; Haber, J. E. 2010. Break-induced replication requires all essential DNA replication factors except those specific for pre-RC assembly. *Genes & Development*, 24, 1133-1144.

Chapter 7: References

- Macheret, M. & Halazonetis, T. D. 2018. Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. *Nature*, 555, 112-116.
- Macleod, F., Kindler, G. S., Wong, H. L., Chen, R. & Burns, B. P. 2019. Asgard archaea: Diversity, function, and evolutionary implications in a range of microbiomes. *Aims Microbiology*, 5, 48-61.
- Macneill, S. A. 2009. The haloarchaeal chromosome replication machinery. *Biochem Soc Trans*, 37, 108-13.
- Macneill, S. A. 2010. Structure and function of the GINS complex, a key component of the eukaryotic replisome. *Biochemical Journal*, 425, 489-500.
- Macneill, S. A. 2011. Protein-protein interactions in the archaeal core replisome. *Biochem Soc Trans*, 39, 163-8.
- Madine, M. a. K., C.; Mills, A. D.; Musahl, C.; Laskey, R. A. 1995. The nuclear envelope prevents reinitiation of replication by regulating the binding of MCM3 to chromatin in *Xenopus* egg extracts. *Current Biology*, 5, 1270-1279.
- Madru, C., Henneke, G., Raia, P., Hugonneau-Beaufet, I., Pehau-Arnaudet, G., England, P., Lindahl, E., Delarue, M., *et al.* 2020. Structural basis for the increased processivity of D-family DNA polymerases in complex with PCNA. *Nature Communications*, 11.
- Maga, G., Villani, G., Tillement, V., Stucki, M., Locatelli, G. A., Frouin, I., Spadari, S. & Hubscher, U. 2001. Okazaki fragment processing: Modulation of the strand displacement activity of DNA polymerase delta by the concerted action of replication protein A, proliferating cell nuclear antigen, and flap endonuclease-1. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 14298-14303.
- Maier, L. K., Alkhnabashi, O. S., Backofen, R. & Marchfelder, A. 2017. CRISPR and Salty: CRISPR-Cas Systems in Haloarchaea. In: B., C.-D. O. (ed.) *RNA Metabolism and Gene Expression in Archaea. Nucleic Acids and Molecular Biology*. Springer, Cham.
- Maier, L. K., Dyll-Smith, M. & Marchfelder, A. 2015a. The Adaptive Immune System of *Haloferax volcanii*. *Life (Basel)*, 5, 521-37.
- Maier, L. K., Lange, S. J., Stoll, B., Haas, K. A., Fischer, S., Fischer, E., Duchardt-Ferner, E., Wohnert, J., *et al.* 2013. Essential requirements for the detection and degradation of invaders by the *Haloferax volcanii* CRISPR/Cas system I-B. *Rna Biology*, 10, 865-874.
- Maier, L. K., Stachler, A. E., Brendel, J., Stoll, B., Fischer, S., Haas, K. A., Schwarz, T. S., Alkhnabashi, O. S., *et al.* 2019. The nuts and bolts of the *Haloferax* CRISPR-Cas system I-B. *Rna Biology*, 16, 469-480.
- Maier, L. K., Stachler, A. E., Saunders, S. J., Backofen, R. & Marchfelder, A. 2015b. An Active Immune Defense with a Minimal CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) RNA and without the Cas6 Protein. *Journal of Biological Chemistry*, 290, 4192-4201.
- Makarova, K. S. & Koonin, E. V. 2005. Evolutionary and functional genomics of the Archaea. *Curr Opin Microbiol*, 8, 586-94.
- Makarova, K. S. & Koonin, E. V. 2013. Archaeology of eukaryotic DNA replication. *Cold Spring Harb Perspect Biol*, 5.

- Makarova, K. S., Koonin, E. V. & Kelman, Z. 2012. The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. *Biol Direct*, 7, 7.
- Makarova, K. S., Krupovic, M. & Koonin, E. V. 2014. Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replication machinery. *Frontiers in Microbiology*, 5, 10.
- Makarova, K. S., Wolf, Y. I. & Koonin, E. V. 2015. Archaeal Clusters of Orthologous Genes (arCOGs): An Update and Application for Analysis of Shared Features between Thermococcales, Methanococcales, and Methanobacteriales. *Life (Basel)*, 5, 818-40.
- Makarova, K. S., Wolf, Y. I. & Koonin, E. V. 2018. Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? *CRISPR J*, 1, 325-336.
- Makkay, A. M., Louyakis, A. S., Ram-Mohan, N., Gophna, U., Gogarten, J. P. & Papke, R. T. 2020. Insights into gene expression changes under conditions that facilitate horizontal gene transfer (mating) of a model archaeon. *Scientific Reports*, 10.
- Malkova, A., Ivanov, E. L. & Haber, J. E. 1996. Double-strand break repair in the absence of RAD51 in yeast: a possible role for break-induced DNA replication. *Proc Natl Acad Sci U S A*, 93, 7131-6.
- Maric, M., Maculins, T., De Piccoli, G. & Labib, K. 2014. Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication. *Science*, 346, 1253596.
- Marinsek, N., Barry, E. R., Makarova, K. S., Dionne, I., Koonin, E. V. & Bell, S. D. 2006. GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep*, 7, 539-45.
- Marriott, H. 2017. Genome architecture and DNA replication in *Haloferax volcanii*. *PhD thesis*, University of Nottingham.
- Masai, H. & Arai, K. 1996. Mechanisms of primer RNA synthesis and D-loop/R-loop-dependent DNA replication in *Escherichia coli*. *Biochimie*, 78, 1109-1117.
- Masai, H., Asai, T., Kubota, Y., Arai, K. & Kogoma, T. 1994. *Escherichia coli* PriA protein is essential for inducible and constitutive stable DNA replication. *Embo Journal*, 13, 5338-5345.
- Maupin-Furlow, J. A. 2013. Archaeal proteasomes and sanylation. *Subcell Biochem*, 66, 297-327.
- Mayle, R. C., I. M.; Beck, C. R.; Yu, Y.; Wilson, M.; Shaw, C. A.; Bjergback, L.; Lupski, J. R.; Ira, G. 2015. Mus81 and converging forks limit the mutagenicity of replication fork breakage. *Science*, 349, 742-747.
- Mcguffee, S. R., Smith, D. J. & Whitehouse, I. 2013. Quantitative, Genome-Wide Analysis of Eukaryotic Replication Initiation and Termination. *Molecular Cell*, 50, 123-135.
- Mcmillan, L. J., Hwang, S. M., Farah, R. E., Koh, J., Chen, S. X. & Maupin-Furlow, J. A. 2018. Multiplex quantitative SILAC for analysis of archaeal proteomes: a case study of oxidative stress responses. *Environmental Microbiology*, 20, 385-401.
- Meagher, M., Epling, L. B. & Enemark, E. J. 2019. DNA translocation mechanism of the MCM complex and implications for replication initiation. *Nature Communications*, 10.

Chapter 7: References

- Meselson, M. & Stahl, F. W. 1958. The replication of DNA in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 44, 671-682.
- Meslet-Cladiere, L., Norais, C., Kuhn, J., Briffotiaux, J., Sloostra, J. W., Ferrari, E., Hubscher, U., Flament, D., *et al.* 2007. A novel proteomic approach identifies new interaction partners for proliferating cell nuclear antigen. *J Mol Biol*, 372, 1137-48.
- Messer, W. 2002. The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiol Rev*, 26, 355-74.
- Mevarech, M. & Werczberger, R. 1985. Genetic transfer in *Halobacterium volcanii*. *J Bacteriol*, 162, 461-2.
- Michel, B. & Bernander, R. 2014. Chromosome replication origins: do we really need them? *Bioessays*, 36, 585-90.
- Miller, E. S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T. & Ruger, W. 2003. Bacteriophage T4 genome. *Microbiology and Molecular Biology Reviews*, 67, 86-+.
- Miller, T. C. R., Locke, J., Greiwe, J. F., Diffley, J. F. X. & Costa, A. 2019. Mechanism of head-to-head MCM double-hexamer formation revealed by cryo-EM. *Nature*, 575, 704-+.
- Miyabe, I., Kunkel, T. A. & Carr, A. M. 2011. The major roles of DNA polymerases epsilon and delta at the eukaryotic replication fork are evolutionarily conserved. *PLoS Genet*, 7, e1002407.
- Miyabe, I., Mizuno, K., Keszthelyi, A., Daigaku, Y., Skouteri, M., Mohebi, S., Kunkel, T. A., Murray, J. M., *et al.* 2015. Polymerase delta replicates both strands after homologous recombination-dependent fork restart. *Nature Structural & Molecular Biology*, 22, 932-938.
- Moalic, Y., Catchpole, R., Leroy, E., Mcteer, L., Cueff-Guachard, V., Aubé, J., Lu, Y., Roussel, E., Oberto, J., Flament, D., Duleromo, R. 2021. Role of Ori in *Thermococcus barophilus*. *bioRxiv*.
- Mojica, F. J. M., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology-Sgm*, 155, 733-740.
- Moon, A. F., Garcia-Diaz, M., Batra, V. K., Beard, W. A., Bebenek, K., Kunkel, T. A., Wilson, S. H. & Pedersen, L. C. 2007. The X family portrait: Structural insights into biological functions of X family polymerases. *DNA Repair*, 6, 1709-1725.
- Morimatsu, K. & Kowalczykowski, S. C. 2014. RecQ helicase and RecJ nuclease provide complementary functions to resect DNA for homologous recombination. *Proceedings of the National Academy of Sciences of the United States of America*, 111, E5133-E5142.
- Mott, M. L., Erzberger, J. P., Coons, M. M. & Berger, J. M. 2008. Structural Synergy and Molecular Crosstalk between Bacterial Helicase Loaders and Replication Initiators. *Cell*, 135, 623-634.
- Moyer, S. E., Lewis, P. W. & Botchan, M. R. 2006. Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 10236-10241.

- Mullakhanbhai, M. F. & Larsen, H. 1975. *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Arch Microbiol*, 104, 207-14.
- Myllykallio, H. & Forterre, P. 2000. Mapping of a chromosome replication origin in an archaeon: response. *Trends Microbiol*, 8, 537-9.
- Nagata, M., Ishino, S., Yamagami, T., Ogino, H., Simons, J.-R., Kanai, T., Atomi, H. & Ishino, Y. 2017a. The Cdc45/RecJ-like protein forms a complex with GINS and MCM, and is important for DNA replication in *Thermococcus kodakarensis*. *Nucleic Acids Res*, gkx740.
- Nagata, M., Ishino, S., Yamagami, T., Simons, J. R., Kanai, T., Atomi, H. & Ishino, Y. 2017b. Possible function of the second RecJ-like protein in stalled replication fork repair by interacting with Hef. *Sci Rep*, 7, 16949.
- Naiman, K., Campillo-Funollet, E., Watson, A. T., Budden, A., Miyabe, I. & Carr, A. M. 2021. Replication dynamics of recombination-dependent replication forks. *Nature Communications*, 12.
- Nakamura, T. M. & Cech, T. R. 1998. Reversing time: Origin of telomerase. *Cell*, 92, 587-590.
- Naor, A., Lazary, R., Barzel, A., Papke, R. T. & Gophna, U. 2011. In Vivo Characterization of the Homing Endonuclease within the polB Gene in the Halophilic Archaeon *Haloferax volcanii*. *Plos One*, 6, 7.
- Naor, A., Thiaville, P. C., Altman-Price, N., Cohen-Or, I., Allers, T., De Crecy-Lagard, V. & Gophna, U. 2012. A genetic investigation of the KEOPS complex in halophilic Archaea. *PLoS One*, 7, e43013.
- Natsuki, T., Sonoko, I., Keisuke, O., Mika, T., Takeshi, Y., Ryotaro, M., Kouta, M. & Yoshizumi, I. 2019. Elucidating Functions of DP1 and DP2 Subunits From the *Thermococcus Kodakarensis* Family D DNA Polymerase. *Extremophiles*, 1, 161-172.
- Noguchi, Y., Yuan, Z. N., Bai, L., Schneider, S., Zhao, G. P., Stillman, B., Speck, C. & Li, H. L. 2017. Cryo-EM structure of Mcm2-7 double hexamer on DNA suggests a lagging-strand DNA extrusion model. *Proceedings of the National Academy of Sciences of the United States of America*, 114, E9529-E9538.
- Norais, C., Hawkins, M., Hartman, A. L., Eisen, J. A., Myllykallio, H. & Allers, T. 2007. Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS Genet*, 3, e77.
- O'donnell, M., Langston, L. & Stillman, B. 2013. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb Perspect Biol*, 5.
- Oberto, J. 2013. SyntTax: a web server linking synteny to prokaryotic taxonomy. *Bmc Bioinformatics*, 14.
- Ogawa, T., Pickett, G. G., Kogoma, T. & Kornberg, A. 1984. RNase H confers specificity in the *dnaA*-dependent initiation of replication at the unique origin of the *Escherichia coli* chromosome *in vivo* and *in vitro*. *Proc Natl Acad Sci U S A*, 81, 1040-4.
- Ogino, H., Ishino, S., Haugland, G. T., Birkeland, N. K., Kohda, D. & Ishino, Y. 2014. Activation of the MCM helicase from the thermophilic archaeon, *Thermoplasma acidophilum* by interactions with GINS and Cdc6-2. *Extremophiles*, 18, 915-24.

Chapter 7: References

- Ogino, H., Ishino, S., Mayanagi, K., Haugland, G. T., Birkeland, N. K., Yamagishi, A. & Ishino, Y. 2011. The GINS complex from the thermophilic archaeon, *Thermoplasma acidophilum* may function as a homotetramer in DNA replication. *Extremophiles*, 15, 529-39.
- Ohbayashi, R., Hirooka, S., Onuma, R., Kanesaki, Y., Hirose, Y., Kobayashi, Y., Fujiwara, T., Furusawa, C., *et al.* 2020. Evolutionary Changes in DnaA-Dependent Chromosomal Replication in Cyanobacteria. *Frontiers in Microbiology*, 11.
- Oki, K., Yamagami, T., Nagata, M., Mayanagi, K., Shirai, T., Adachi, N., Numata, T., Ishino, S., *et al.* 2021. DNA polymerase D temporarily connects primase to the CMG-like helicase before interacting with proliferating cell nuclear antigen. *Nucleic Acids Res*, 49, 4599–4612.
- Olsen, G. J. & Woese, C. R. 1997. Archaeal genomics: an overview. *Cell*, 89, 991-4.
- Oren, A. 2008. Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Systems*, 4, 2.
- Oren, A., Haldal, M., Norland, S. & Galinski, E. A. 2002. Intracellular ion and organic solute concentrations of the extremely halophilic bacterium *Salinibacter ruber*. *Extremophiles*, 6, 491-8.
- Ortenberg, R., Rozenblatt-Rosen, O. & Mevarech, M. 2000. The extremely halophilic archaeon *Haloferax volcanii* has two very different dihydrofolate reductases. *Mol Microbiol*, 35, 1493-1505.
- Oyama, T., Ishino, S., Fujino, S., Ogino, H., Shirai, T., Mayanagi, K., Saito, M., Nagasawa, N., *et al.* 2011. Architectures of archaeal GINS complexes, essential DNA replication initiation factors. *Bmc Biology*, 9, 12.
- Oyama, T., Ishino, S., Shirai, T., Yamagami, T., Nagata, M., Ogino, H., Kusunoki, M. & Ishino, Y. 2016. Atomic structure of an archaeal GAN suggests its dual roles as an exonuclease in DNA repair and a CMG component in DNA replication. *Nucleic Acids Res*, 44, 9505-9517.
- Pace, N. R. 1997. A molecular view of microbial diversity and the biosphere. *Science*, 276, 734-740.
- Pacek, M., Tutter, A. V., Kubota, Y., Takisawa, H. & Walter, J. C. 2006. Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. *Molecular Cell*, 21, 581-587.
- Pan, M., Kelman, L. M. & Kelman, Z. 2011a. The archaeal PCNA proteins. *Biochemical Society Transactions*, 39, 20-24.
- Pan, M., Santangelo, T. J., Cubonova, L., Li, Z., Metangmo, H., Ladner, J., Hurwitz, J., Reeve, J. N., *et al.* 2013. *Thermococcus kodakarensis* has two functional PCNA homologs but only one is required for viability. *Extremophiles*, 17, 453-461.
- Pan, M., Santangelo, T. J., Li, Z., Reeve, J. N. & Kelman, Z. 2011b. *Thermococcus kodakarensis* encodes three MCM homologs but only one is essential. *Nucleic Acids Research*, 39, 9671-9680.
- Park, S. Y. & Asano, M. 2008. The origin recognition complex is dispensable for endoreplication in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 12343-12348.

- Patel, P. H. & Loeb, L. A. 2001. Getting a grip on how DNA polymerases function. *Nature Structural Biology*, 8, 656-659.
- Pavlov, Y. I., Frahm, C., Mcelhinny, S. a. N., Niimi, A., Suzuki, M. & Kunkel, T. A. 2006. Evidence that errors made by DNA polymerase alpha are corrected by DNA polymerase delta. *Current Biology*, 16, 202-207.
- Peeters, E., Driessen, R. P., Werner, F. & Dame, R. T. 2015. The interplay between nucleoid organization and transcription in archaeal genomes. *Nat Rev Microbiol*, 13, 333-41.
- Pellegrini, L. 2017. Structural insights into Cdc45 function: was there a nuclease at the heart of the ancestral replisome? *Biophys Chem*, 225, 10-14.
- Perez-Arnaiz, P., Dattani, A., Smith, V. & Allers, T. 2020. Haloferax volcanii - a model archaeon for studying DNA replication and repair. *Open Biology*, 10.
- Perler, F. B. 2002. InBase: the Intein Database. *Nucleic Acids Res*, 30, 383-4.
- Pfeifer, F., Griffing, J. & Oesterhelt, D. 1993. The fdx gene encoding the [2Fe-2S] ferredoxin of *Halobacterium salinarium* (*H. halobium*). *Mol Gen Genet*, 239, 66-71.
- Pohlschroder, M. & Schulze, S. 2019. Microbe of the Month Haloferax volcanii. *Trends in Microbiology*, 27, 86-87.
- Pursell, Z. F., Isoz, I., Lundstrom, E. B., Johansson, E. & Kunkel, T. A. 2007. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science*, 317, 127-130.
- Raghunathan, S., Ricard, C. S., Lohman, T. M. & Waksman, G. 1997. Crystal structure of the homo-tetrameric DNA binding domain of Escherichia coli single-stranded DNA-binding protein determined by multiwavelength x-ray diffraction on the selenomethionyl protein at 2.9-angstrom resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 6652-6657.
- Raia, P., Carroni, M., Henry, E., Pehau-Arnaudet, G., Brule, S., Beguin, P., Henneke, G., Lindahl, E., *et al.* 2019a. Structure of the DP1-DP2 PolD complex bound with DNA and its implications for the evolutionary history of DNA and RNA polymerases. *Plos Biology*, 17.
- Raia, P., Delarue, M. & Sauguet, L. 2019b. An updated structural classification of replicative DNA polymerases. *Biochemical Society Transactions*, 47, 239-249.
- Rajman, L. A. & Lovett, S. T. 2000. A thermostable single-strand DNase from *Methanococcus jannaschii* related to the RecJ recombination and repair exonuclease from *Escherichia coli*. *J Bacteriol*, 182, 607-12.
- Raymann, K., Forterre, P., Brochier-Armanet, C. & Gribaldo, S. 2014. Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in archaea. *Genome Biol Evol*, 6, 192-212.
- Reeve, J. N. 2003. Archaeal chromatin and transcription. *Mol Microbiol*, 48, 587-98.
- Reeve, J. N., Sandman, K. & Daniels, C. J. 1997. Archaeal histones, nucleosomes, and transcription initiation. *Cell*, 89, 999-1002.
- Reuter, C. J. & Maupin-Furlow, J. 2004. Analysis of proteasome-dependent proteolysis in *Haloferax volcanii* cells, using short-lived green fluorescent proteins. *Appl Environ Microbiol*, 70, 7530-7538.

Chapter 7: References

- Rhodes, M. E., Spear, J. R., Oren, A. & House, C. H. 2011. Differences in lateral gene transfer in hypersaline versus thermal environments. *Bmc Evolutionary Biology*, 11.
- Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J. F., Darling, A., Malfatti, S., *et al.* 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, 499, 431-437.
- Robinson, N. P. & Bell, S. D. 2007. Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proc Natl Acad Sci U S A*, 104, 5806-11.
- Rodriguez, G., Martin, M. T. & De Vega, M. 2019. An array of basic residues is essential for the nucleolytic activity of the PHP domain of bacterial/archaeal PolX DNA polymerases. *Scientific Reports*, 9.
- Roecklein, B., Pelletier, A. & Kuempel, P. 1991. The Tus gene of *Escherichia coli* - Autoregulation, analysis of flanking sequences and identification of a complementary system in *Salmonella typhimurium*. *Research in Microbiology*, 142, 169-175.
- Rogozin, I. B., Makarova, K. S., Pavlov, Y. I. & Koonin, E. V. 2008. A highly conserved family of inactivated archaeal B family DNA polymerases. *Biology Direct*, 3, 5.
- Rosenshine, I., Tchelet, R. & Mevarech, M. 1989. The mechanism of DNA transfer in the mating system of an archaeobacterium. *Science*, 245, 1387-9.
- Rouillon, C., Henneke, G., Flament, D., Querellou, J. & Raffin, J. P. 2007. DNA polymerase switching on homotrimeric PCNA at the replication fork of the euryarchaea *Pyrococcus abyssi*. *Journal of Molecular Biology*, 369, 343-355.
- Rowen, L. & Kornberg, A. 1978. Primase, DnaG protein of *Escherichia coli* - Enzyme which starts DNA chains. *Journal of Biological Chemistry*, 253, 758-764.
- Rudd, S. G., Bianchi, J. & Doherty, A. J. 2014. PrimPol-A new polymerase on the block. *Mol Cell Oncol*, 1, e960754.
- Rudolph, C. J., Upton, A. L., Stockum, A., Nieduszynski, C. A. & Lloyd, R. G. 2013. Avoiding chromosome pathology when replication forks collide. *Nature*, 500, 608-11.
- Ruprich-Robert, G. & Thuriaux, P. 2010. Non-canonical DNA transcription enzymes and the conservation of two-barrel RNA polymerases. *Nucleic Acids Research*, 38, 4559-4569.
- Saini, N., Ramakrishnan, S., Elango, R., Ayyar, S., Zhang, Y., Deem, A., Ira, G., Haber, J. E., *et al.* 2013. Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature*, 502, 389-92.
- Sakakibara, N., Kelman, L. M. & Kelman, Z. 2009a. How is the archaeal MCM helicase assembled at the origin? Possible mechanisms. *Biochem Soc Trans*, 37, 7-11.
- Sakakibara, N., Kelman, L. M. & Kelman, Z. 2009b. Unwinding the structure and function of the archaeal MCM helicase. *Molecular Microbiology*, 72, 286-296.
- Salas-Leiva, D. E., Tromer, E. C., Curtis, B. A., Jerlstrom-Hultqvist, J., Kolisko, M., Yi, Z. Z., Salas-Leiva, J. S., Gallot-Lavallee, L., *et al.* 2021. Genomic analysis finds no evidence of canonical eukaryotic

- DNA processing complexes in a free-living protist. *Nature Communications*, 12.
- Sale, J. E., Lehmann, A. R. & Woodgate, R. 2012. Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat Rev Mol Cell Biol*, 13, 141-52.
- Samson, R. Y., Abeyrathne, P. D. & Bell, S. D. 2016. Mechanism of Archaeal MCM Helicase Recruitment to DNA Replication Origins. *Mol Cell*, 61, 287-96.
- Samson, R. Y. & Bell, S. D. 2016. Archaeal DNA Replication Origins and Recruitment of the MCM Replicative Helicase. *Enzymes*, 39, 169-90.
- Samson, R. Y., Dobro, M. J., Jensen, G. J. & Bell, S. D. 2017. The Structure, Function and Roles of the Archaeal ESCRT Apparatus. *Prokaryotic Cytoskeletons: Filamentous Protein Polymers Active in the Cytoplasm of Bacterial and Archaeal Cells*, 84, 357-377.
- Samson, R. Y., Obita, T., Freund, S. M., Williams, R. L. & Bell, S. D. 2008. A role for the ESCRT system in cell division in archaea. *Science*, 322, 1710-3.
- Samson, R. Y., Xu, Y., Gadelha, C., Stone, T. A., Faqiri, J. N., Li, D., Qin, N., Pu, F., *et al.* 2013. Specificity and function of archaeal DNA replication initiator proteins. *Cell Rep*, 3, 485-96.
- Sancar, G. B. 2000. Enzymatic photoreactivation: 50 years and counting. *Mutat Res*, 451, 25-37.
- Sanchez-Pulido, L. & Ponting, C. P. 2011. Cdc45: the missing RecJ ortholog in eukaryotes? *Bioinformatics*, 27, 1885-8.
- Sanders, G. M., Dallmann, H. G. & Mchenry, C. S. 2010. Reconstitution of the B-subtilis Replisome with 13 Proteins Including Two Distinct Replicases. *Molecular Cell*, 37, 273-281.
- Sanger, F., Nicklen, S. & Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, 74, 5463-7.
- Sarmiento, F., Long, F., Cann, I. & Whitman, W. B. 2014. Diversity of the DNA Replication System in the Archaea Domain. *Archaea-an International Microbiological Journal*.
- Sarmiento, F., Mrazek, J. & Whitman, W. B. 2013. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 4726-4731.
- Sauguet, L. 2019. The Extended 'Two-Barrel' Polymerases Superfamily: Structure, Function and Evolution. *J Mol Biol*.
- Sawa, M. M., H. 2008. Drug design with Cdc7 kinase: a potential novel cancer therapy target. *Drug Design, Development and Therapy*, 2, 255-264.
- Sawitzke, J. & Austin, S. 2001. An analysis of the factory model for chromosome replication and segregation in bacteria. *Molecular Microbiology*, 40, 786-794.
- Schaper, S. & Messer, W. 1995. Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target. *Journal of Biological Chemistry*, 270, 17622-17626.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., *et al.* 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9, 676-682.

Chapter 7: References

- Seitz, K. W., Lazar, C. S., Hinrichs, K. U., Teske, A. P. & Baker, B. J. 2016. Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. *Isme Journal*, 10, 1696-1705.
- Sengupta, S., Van Deursen, F., De Piccoli, G. & Labib, K. 2013. Dpb2 Integrates the Leading-Strand DNA Polymerase into the Eukaryotic Replisome. *Current Biology*, 23, 543-552.
- Shalev, Y., Turgeman-Grott, I., Tamir, A., Eichler, J. & Gophna, U. 2017. Cell Surface Glycosylation Is Required for Efficient Mating of *Haloferax volcanii*. *Frontiers in Microbiology*, 8.
- Sharma, R. & Rao, D. N. 2009. Orchestration of *Haemophilus influenzae* RecJ Exonuclease by Interaction with Single-Stranded DNA-Binding Protein. *Journal of Molecular Biology*, 385, 1375-1396.
- Shen, Y. L., Tang, X. F., Yokoyama, H., Matsui, E. & Matsui, I. 2004. A 21-amino acid peptide from the cysteine cluster II of the family D DNA polymerase from *Pyrococcus horikoshii* stimulates its nuclease activity which is Mre11-like and prefers manganese ion as the cofactor. *Nucleic Acids Research*, 32, 158-168.
- Sherratt, D. J. 2003. Bacterial chromosome dynamics. *Science*, 301, 780-785.
- Shibata, E., Kiran, M., Shibata, Y., Singh, S., Kiran, S. & Dutta, A. 2016. Two subunits of human ORC are dispensable for DNA replication and proliferation. *Elife*, 5.
- Shin, J. H., Grabowski, B., Kasiviswanathan, R., Bell, S. D. & Kelman, Z. 2003. Regulation of minichromosome maintenance helicase activity by Cdc6. *J Biol Chem*, 278, 38059-67.
- Shin, J. H., Heo, G. Y. & Kelman, Z. 2009. The Methanothermobacter thermotrophicus MCM Helicase Is Active as a Hexameric Ring. *J Biol Chem*, 284, 540-6.
- Shin, J. H. & Kelman, Z. 2006. The replicative helicases of bacteria, archaea, and eukarya can unwind RNA-DNA hybrid substrates. *Journal of Biological Chemistry*, 281, 26914-26921.
- Shin, J. H., Santangelo, T. J., Xie, Y. W., Reeve, J. N. & Kelman, Z. 2007. Archaeal minichromosome maintenance (MCM) helicase can unwind DNA bound by archaeal histones and transcription factors. *Journal of Biological Chemistry*, 282, 4908-4915.
- Siglioccolo, A., Paiardini, A., Piscitelli, M. & Pascarella, S. 2011. Structural adaptation of extreme halophilic proteins through decrease of conserved hydrophobic contact surface. *BMC Struct Biol*, 11, 50.
- Simon, A. C., Sannino, V., Costanzo, V. & Pellegrini, L. 2016. Structure of human Cdc45 and implications for CMG helicase function. *Nature Communications*, 7, 15.
- Sinha, R. P. & Hader, D. P. 2002. UV-induced DNA damage and repair: a review. *Photochem Photobiol Sci*, 1, 225-36.
- Sleigh, M. J. 1976. Mechanism of DNA breakage by phleomycin in vitro. *Nucleic Acids Research*, 3, 891-901.
- Slupphaug, G., Kavli, B. & Krokan, H. E. 2003. The interacting pathways for prevention and repair of oxidative DNA damage. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis*, 531, 231-251.

- Sojo, V., Pomiankowski, A. & Lane, N. 2014. A Bioenergetic Basis for Membrane Divergence in Archaea and Bacteria. *Plos Biology*, 12.
- Sonneville, R., Moreno, S. P., Knebel, A., Johnson, C., Hastie, C. J., Gartner, A., Gambus, A. & Labib, K. 2017. CUL-2(LRR-1) and UBXN-3 drive replisome disassembly during DNA replication termination and mitosis. *Nature Cell Biology*, 19, 468-+.
- Spang, A., Caceres, E. F. & Ettema, T. J. G. 2017. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science*, 357.
- Spang, A., Saw, J. H., Jorgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., Van Eijk, R., Schleper, C., *et al.* 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, 521, 173-179.
- Spang, A., Stairs, C. W., Dombrowski, N., Eme, L., Lombard, J., Caceres, E. F., Greening, C., Baker, B., *et al.* 2019. Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nature Microbiology*, 4, 1138-1148.
- Srinivasan, M., Mehta, P., Yu, Y., Prugar, E., Koonin, E. V., Karzai, A. W. & Sternglanz, R. 2011. The highly conserved KEOPS/EKC complex is essential for a universal tRNA modification, t6A. *Embo Journal*, 30, 873-881.
- Stachler, A. E. & Marchfelder, A. 2016. Gene Repression in Haloarchaea Using the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas I-B System. *J Biol Chem*, 291, 15226-42.
- Stachler, A. E., Turgeman-Grott, I., Shtifman-Segal, E., Allers, T., Marchfelder, A. & Gophna, U. 2017. High tolerance to self-targeting of the genome by the endogenous CRISPR-Cas system in an archaeon. *Nucleic Acids Res*, 45, 5208-5216.
- Stanojic, S., Kuk, N., Ullah, I., Sterkers, Y. & Merrick, C. J. 2017. Single-molecule analysis reveals that DNA replication dynamics vary across the course of schizogony in the malaria parasite *Plasmodium falciparum*. *Scientific Reports*, 7.
- Stodola, J. L. & Burgers, P. M. 2016. Resolving individual steps of Okazaki-fragment maturation at a millisecond timescale. *Nature Structural & Molecular Biology*, 23, 402-408.
- Stoll, B., Maier, L. K., Lange, S. J., Brendel, J., Fischer, S., Backofen, R. & Marchfelder, A. 2013. Requirements for a successful defence reaction by the CRISPR-Cas subtype I-B system. *Biochemical Society Transactions*, 41, 1444-1448.
- Stroud, A., Liddell, S. & Allers, T. 2012. Genetic and Biochemical Identification of a Novel Single-Stranded DNA-Binding Complex in *Haloferax volcanii*. *Front Microbiol*, 3, 224.
- Stuckey, R., Garcia-Rodriguez, N., Aguilera, A. & Wellinger, R. E. 2015. Role for RNA:DNA hybrids in origin-independent replication priming in a eukaryotic system. *Proc Natl Acad Sci U S A*, 112, 5779-84.
- Sun, J. C., Shi, Y., Georgescu, R. E., Yuan, Z. N., Chait, B. T., Li, H. L. & O'donnell, M. E. 2015. The architecture of a eukaryotic replisome. *Nature Structural & Molecular Biology*, 22, 976-982.

Chapter 7: References

- Sutera, V. A., Jr., Han, E. S., Rajman, L. A. & Lovett, S. T. 1999. Mutational analysis of the RecJ exonuclease of *Escherichia coli*: identification of phosphoesterase motifs. *J Bacteriol*, 181, 6098-102.
- Sutton, M. D., Narumi, I. & Walker, G. C. 2002. Posttranslational modification of the umuD-encoded subunit of *Escherichia coli* DNA polymerase V regulates its interactions with the beta processivity clamp. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 5307-5312.
- Szambowska, A., Tessmer, I., Kursula, P., Usskilat, C., Prus, P., Pospiech, H. & Grosse, F. 2014. DNA binding properties of human Cdc45 suggest a function as molecular wedge for DNA unwinding. *Nucleic Acids Research*, 42, 2308-2319.
- Takayama, Y., Kamimura, Y., Okawa, M., Muramatsu, S., Sugino, A. & Araki, H. 2003. GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes & Development*, 17, 1153-1165.
- Takeda, D. Y. & Dutta, A. 2005. DNA replication and progression through S phase. *Oncogene*, 24, 2827-43.
- Tanaka, S. & Araki, H. 2013. Helicase activation and establishment of replication forks at chromosomal origins of replication. *Cold Spring Harb Perspect Biol*, 5, a010371.
- Theobald, D. L., Mitton-Fry, R. M. & Wuttke, D. S. 2003. Nucleic acid recognition by OB-fold proteins. *Annual Review of Biophysics and Biomolecular Structure*, 32, 115-133.
- Tomasz, M. 1995. Mitomycin-C - Small, fast and deadly (but very selective). *Chemistry & Biology*, 2, 575-579.
- Tori, K., Kimizu, M., Ishino, S. & Ishino, Y. 2007. DNA polymerases BI and D from the hyperthermophilic archaeon *Pyrococcus furiosus* both bind to proliferating cell nuclear antigen with their C-terminal PIP-box motifs. *J Bacteriol*, 189, 5652-7.
- Trucksis, M. M., J.; Deng, Y. K.; Kaper, J. B. 1998. The *Vibrio cholerae* genome contains two unique circular chromosomes. *Proc Natl Acad Sci USA*, 95, 14464-14469.
- Tyler, J. K. 2002. Chromatin assembly. *European Journal of Biochemistry*, 269, 2268-2274.
- Uemori, T., Sato, Y., Kato, I., Doi, H. & Ishino, Y. 1997. A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus*: gene cloning, expression, and characterization. *Genes Cells*, 2, 499-512.
- Ullsperger, C. J. V., A. V.; Cozzarelli, N. R. 1995. *Unlinking of DNA by Topoisomerases During DNA Replication*. , Springer, Berlin, Heidelberg.
- Vestergaard, G., Garrett, R. A. & Shah, S. A. 2014. CRISPR adaptive immune systems of Archaea. *Rna Biology*, 11, 156-167.
- Volff, J. A., J. 2000. A new beginning with new ends: linearisation of circular chromosomes during bacterial evolution. *FEMS Microbiology Letters*, 186, 143-150.
- Wadsworth, R. I. & White, M. F. 2001. Identification and properties of the crenarchaeal single-stranded DNA binding protein from *Sulfolobus solfataricus*. *Nucleic Acids Res*, 29, 914-20.

Chapter 7: References

- Wan, L. C. K., Pillon, M. C., Thevakumaran, N., Sun, Y. L., Chakrabarty, A., Guarne, A., Kurinov, I., Durocher, D., *et al.* 2016. Structural and functional characterization of KEOPS dimerization by Pcc1 and its role in t(6)A biosynthesis. *Nucleic Acids Research*, 44, 6971-6980.
- Wang, J. C. 1996. DNA topoisomerases. *Annual Review of Biochemistry*, 65, 635-692.
- Wang, W. T., Klein, K. N., Proesmans, K., Yang, H. B., Marchal, C., Zhu, X. P., Borrmann, T., Hastie, A., *et al.* 2021. Genome-wide mapping of human DNA replication by optical replication mapping supports a stochastic model of eukaryotic replication. *Molecular Cell*, 81, 2975-+.
- Warbrick, E., Heatherington, W., Lane, D. P. & Glover, D. N. 1998. PCNA binding proteins in *Drosophila melanogaster*: the analysis of a conserved PCNA binding domain. *Nucleic Acids Research*, 26, 3925-3932.
- Wardell, K. 2013. *Regulation of homologous recombination in the archaeon Haloferax volcanii*. University of Nottingham.
- Wardell, K., Haldenby, S., Jones, N., Liddell, S., Ngo, G. H. P. & Allers, T. 2017. RadB acts in homologous recombination in the archaeon *Haloferax volcanii*, consistent with a role as recombination mediator. *DNA Repair (Amst)*, 55, 7-16.
- Wardle, J., Burgers, P. M. J., Cann, I. K. O., Darley, K., Heslop, P., Johansson, E., Lin, L. J., Mcglynn, P., *et al.* 2008. Uracil recognition by replicative DNA polymerases is limited to the archaea, not occurring with bacteria and eukarya. *Nucleic Acids Research*, 36, 705-711.
- Watson, J. D. C., F. H. C. 1953. Molecular Structure of Nucleic Acids: A structure for deoxyribose nucleic acid. *Nature*, 4356, 737-738.
- Wegrzyn, K. E., Gross, M., Uciechowska, U. & Konieczny, I. 2016. Replisome Assembly at Bacterial Chromosomes and Iteron Plasmids. *Front Mol Biosci*, 3, 39.
- Welch, M. M. & Mchenry, C. S. 1982. Cloning and identification of the product of the DnaE gene of *Escherichia coli*. *Journal of Bacteriology*, 152, 351-356.
- Wendoloski, D., Ferrer, C. & Dyal-Smith, M. L. 2001. A new simvastatin (mevinolin)-resistance marker from *Haloarcula hispanica* and a new *Haloferax volcanii* strain cured of plasmid pHV2. *Microbiology*, 147, 959-64.
- Winker, S. & Woese, C. R. 1991. A definition of the domains Archaea, Bacteria and Eucarya in terms of small subunit ribosomal RNA characteristics. *Syst Appl Microbiol*, 14, 305-10.
- Winter, J. A., Christofi, P., Morroll, S. & Bunting, K. A. 2009. The crystal structure of *Haloferax volcanii* proliferating cell nuclear antigen reveals unique surface charge characteristics due to halophilic adaptation. *BMC Struct Biol*, 9, 55.
- Winter, K., Born, J. & Pfeifer, F. 2018. Interaction of Haloarchaeal Gas Vesicle Proteins Determined by Split-GFP. *Front Microbiol*, 9, 1897.
- Woese, C. R. & Fox, G. E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A*, 74, 5088-90.
- Woese, C. R., Kandler, O. & Wheelis, M. L. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A*, 87, 4576-9.

Chapter 7: References

- Wolf, Y. I., Rogozin, I. B., Kondrashov, A. S. & Koonin, E. V. 2001. Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. *Genome Res*, 11, 356-72.
- Woods, W. G. & Dyall-Smith, M. L. 1997. Construction and analysis of a recombination-deficient (*radA*) mutant of *Haloferax volcanii*. *Mol Microbiol*, 23, 791-7.
- Wörtz, J., Smith, V., Fallman, J., König, S., Thuraisingam, T., Walther, P., Uralaub, H., Stadler, P. F., *et al.* 2022. Cas1 and Fen1 display equivalent functions during archaeal DNA repair. *Front Microbiol*, 13.
- Wu, S., Beard, W. A., Pedersen, L. G. & Wilson, S. H. 2014a. Structural Comparison of DNA Polymerase Architecture Suggests a Nucleotide Gateway to the Polymerase Active Site. *Chemical Reviews*, 114, 2759-2774.
- Wu, Z., Liu, J., Yang, H. & Xiang, H. 2014b. DNA replication origins in archaea. *Front Microbiol*, 5, 179.
- Xia, Y. S., Niu, Y. L., Cui, J. M., Fu, Y., Chen, X. J. S., Lou, H. Q. & Cao, Q. H. 2015. The Helicase Activity of Hyperthermophilic Archaeal MCM is Enhanced at High Temperatures by Lysine Methylation. *Frontiers in Microbiology*, 6.
- Xu, Y., Gristwood, T., Hodgson, B., Trinidad, J. C., Albers, S. V. & Bell, S. D. 2016. Archaeal orthologs of Cdc45 and GINS form a stable complex that stimulates the helicase activity of MCM. *Proc Natl Acad Sci U S A*, 113, 13390-13395.
- Yoshimochi, T., Fujikane, R., Kawanami, M., Matsunaga, F. & Ishino, Y. 2008. The GINS complex from *Pyrococcus furiosus* stimulates the MCM helicase activity. *J Biol Chem*, 283, 1601-9.
- Yuan, H., Liu, X. P., Han, Z., Allers, T., Hou, J. L. & Liu, J. H. 2013. RecJ-like protein from *Pyrococcus furiosus* has 3'-5' exonuclease activity on RNA: implications for proofreading of 3'-mismatched RNA primers in DNA replication. *Nucleic Acids Research*, 41, 5817-5826.
- Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Backstrom, D., Juzokaite, L., Vancaester, E., Seitz, K. W., Anantharaman, K., *et al.* 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*, 541, 353-358.
- Zhang, L., Jiang, D., Shi, H., Wu, M., Gan, Q., Yang, Z. & Oger, P. 2020. Characterization and application of a family B DNA polymerase from the hyperthermophilic and radioresistant euryarchaeon *Thermococcus gammatolerans*. *Int J Biol Macromol*, 156, 217-224.
- Zhao, X.-Q., Hu, J.-F. & Yu, J. 2006. Comparative Analysis of Eubacterial DNA Polymerase III Alpha Subunits. *Genomics, Proteomics & Bioinformatics*, 4, 203-211.

Chapter 8: Appendix

Appendix 1: RecJ2 pulldown mass spectrometry data

Band 2A

Protein accession	Gene name	HVO_number	Predicted MW (Da)	MASCOT score	# of peptides	Peptide sequences
D4GP52	cobN	B0050	141254	1106	25 (25)	23 (23)
D4GV93	<i>Putative secreted glycoprotein</i>	2160	232813	831	13 (13)	11 (11)
D4GT33	nrdJ	2452	114376	782	11 (11)	11 (11)
D4GWB0	recJ2	1147	37450	721	13 (13)	11 (11)
D4GWK9	carB	2361	117,302	699	11 (11)	11 (11)
D4H042	<i>FAD-dependent oxidoreductase</i>	1697	111281	611	13 (13)	12 (12)
D4GRZ3	leuS	0452	109563	601	10 (10)	10 (10)
P25062	csg	2072	85189	557	33 (33)	7 (7)
D4GSM3	dppDF2	0627	97733	407	7 (7)	7 (7)
D4GYH4	aglB	1530	113675	227	3 (3)	3 (3)
D4GT75	pccB2	2471	56772	123	2 (2)	2 (2)
D4GYJ1	ileS	1547	117994	86	2 (2)	2 (2)
D4GTB5	pccA	2486	65346	78	2 (2)	2 (2)
D4GS73	citB2	0541	102210	56	2 (2)	2 (2)
D4GZX3	rpoH	0346	8537	50	1 (1)	1 (1)
D4GUY6	coxA1	0907	65760	45	1 (1)	1 (1)

Band 2B

Protein accession	Gene name	HVO_number	Predicted MW (Da)	MASCOT score	# of peptides	Peptide sequences
D4GWB0	recJ2	1147	37450	1316	85 (85)	16 (16)
D4GS20	pgk	0480	43636	797	12 (12)	12 (12)
D4GUK8	sufB2	0861	44685	710	12 (12)	11 (11)
D4GU92	icd	2588	45810	657	15 (15)	13 (13)
D4GXG2	glyA1	2862	44349	610	10 (10)	9 (9)
D4GYQ3	dppF1	0058	50184	505	8 (8)	8 (8)
D4GUA4	<i>Stomatin family protein</i>	0801	46401	455	8 (8)	7 (7)
D4GX38	sdhA	2808	67641	454	12 (12)	9 (9)
D4GP49	<i>ABC-type transport system periplasmic substrate-binding protein</i>	B0047	44239	429	8 (8)	7 (7)
D4GZ07	ndh1	1578	42816	411	8 (8)	7 (7)
D4GRF0	cbiA	A0487	46883	375	6 (6)	5 (5)
D4GYD4	gnaD	1488	45794	371	9 (9)	7 (7)

Chapter 8: Appendix

D4GW49	rnj	2724	50115	336	7 (7)	7 (7)
D4GYK5	tnaA	0009	48543	328	7 (7)	7 (7)
D4GWX0	eno	2774	42005	309	6 (6)	6 (6)
D4GYJ0	dhaK	1546	34847	306	5 (5)	5 (5)
D4H014	rnr	0388	50358	291	5 (5)	5 (5)
D4GYP4	argG	0049	44483	272	4 (4)	4 (4)
D4GSV7	hisS	1854	47162	272	4 (4)	4 (4)
D4GPJ9	<i>ABC-type transport system periplasmic substrate-binding protein</i>	B0198	43426	258	7 (7)	5 (5)
D4GXII	<i>Pyridoxal phosphate-dependent aminotransferase</i>	2871	49116	254	4 (4)	4 (4)
Q1XBW2	dnaK	1590	67250	239	4 (4)	4 (4)
D4GU24	aroB	0792	43109	234	4 (4)	4 (4)
D4GWA7	metB1	2750	42089	234	4 (4)	4 (4)
D4GZU0	atpI	0311	79504	234	4 (4)	4 (4)
D4GYQ7	dppA1	0062	67307	223	4 (4)	4 (4)
D4GXX4	pmm3	1402	47500	222	5 (5)	5 (5)
D4GVR7	tsgA3	2695	55422	207	5 (5)	5 (5)
D4GT75	pccB2	2471	56772	197	3 (3)	3 (3)
D4GST7	pepB2	1829	39365	179	3 (3)	3 (3)
D4GP72	<i>Pyridoxal phosphate-dependent aminotransferase</i>	B0070	46183	157	3 (3)	3 (3)
D4GZY3	tef2	0356	80441	151	3 (3)	3 (3)
D4GRL8	<i>ABC-type transport system periplasmic substrate-binding protein</i>	A0557	44252	145	3 (3)	3 (3)
D4GW90	metE1	2742	40298	142	3 (3)	3 (3)
D4GTL9	<i>ABC-type transport system periplasmic substrate-binding protein</i>	1991	45174	139	2 (2)	2 (2)
D4GYI4	glpC1	1540	48623	132	3 (3)	3 (3)
D4GX43	<i>Uncharacterised protein</i>	1242	47100	131	3 (3)	3 (3)
D4GY69	gdhA2	1453	45730	126	4 (4)	4 (4)
D4GXX3	tsgA11	1401	39212	124	2 (2)	2 (2)
D4GVN0	gatA	1054	43909	107	2 (2)	2 (2)
D4GYK0	<i>NamA family oxidoreductase</i>	0004	39289	106	1 (1)	1 (1)
D4GS63	tsgA1	0530	48149	102	2 (2)	2 (2)
D4GYS6	hemL	0081	47811	101	2 (2)	2 (2)
D4GW23	rtcB	2712	53434	94	2 (2)	2 (2)
D4GZ58	<i>Uncharacterised protein</i>	1629	56347	87	2 (2)	2 (2)
D4GS58	cca	0521	50492	86	2 (2)	2 (2)
D4GZY6	tef1a1	0359	45759	70	1 (1)	1 (1)
D4GZT3	etfA1	0304	33519	68	1 (1)	1 (1)
D4GTU0	carA	2508	38573	64	1 (1)	1 (1)
O30560	cct2	0455	59298	58	1 (1)	1 (1)
D4GW10	<i>GFO family oxidoreductase</i>	2707	41032	56	1 (1)	1 (1)

Chapter 8: Appendix

D4GSF9	malE	0564	46818	56	1 (1)	1 (1)
D4GTK5	secD	1976	55006	55	1 (1)	1 (1)
D4GPR3	gdhA3	B0266	47309	54	2 (2)	2 (2)
D4GUH5	ywaD1	0836	46938	53	1 (1)	1 (1)
D4GZ33	aspC1	1604	40193	52	1 (1)	1 (1)
D4GZV0	arf1	0321	46582	50	1 (1)	1 (1)
D4GWR	hmgB	2419				
6						
D4GZ66	mtfK2	1637	33784	49	1 (1)	1 (1)

Band 2C

Protein accession	Gene name	HVO_number	Predicted MW (Da)	MASCOT score	# of peptides	Peptide sequences
D4GTB5	pccA	2486	65346	2781	120 (120)	39 (39)
O30560	cct2/thS2	0455	59298	1603	29 (29)	22 (22)
D4GUV5	korA	0888	64026	616	14 (14)	13 (13)
D4GT75	pccB2	2471	56772	293	6 (6)	6 (6)
D4GP84	dppA11	B0082	62804	244	6 (6)	6 (6)
D4GY56	pccB1	1447	63624	230	5 (5)	5 (5)
D4GZ01	gyrB	1572	71213	213	4 (4)	4 (4)
P25062	csg	2072	85189	163	2 (2)	2 (2)
D4GZX5	rpoB1	0348	67762	129	4 (4)	4 (4)
D4GUM1	<i>Zinc-dependent nuclease</i>	0874	71947	103	3 (3)	3 (3)
D4GRM3	hutU	A0562	67982	80	2 (2)	2 (2)
D4GY12	gpdA1	1538	63487	62	2 (2)	2 (2)
D4GW53	gltS	2726	65287	56	1 (1)	1 (1)

Band 2D

Protein accession	Gene name	HVO_number	Predicted MW (Da)	MASCOT score	# of peptides	Peptide sequences
D4GT75	pccB2	2471	56772	2179	184 (184)	31 (31)
D4GTB5	pccA	2486	65346	1867	58 (58)	28 (28)
Q9HHA2	cct3/thS3	0778	55242	1805	35 (35)	21 (21)
D4GVF7	nuoCD	0980	63484	864	17 (17)	15 (15)
D4GT09	aspS	0677	48726	734	17 (17)	11 (11)
D4GTF0	serS	1921	52038	703	12 (12)	11 (11)
D4GT23	gatB	0684	55074	659	11 (11)	10 (10)
D4GWU7	aldH2	1189	56171	645	15 (15)	12 (12)
Q1XBW2	dnaK	1590	67250	643	9 (9)	9 (9)
D4GY36	serA3	2968	55343	533	9 (9)	7 (7)
D4GU83	<i>NPCBM-associated domain-containing protein</i>	2070	58129	527	9 (9)	6 (6)
D4GSM4	dppA2	0628	62819	478	9 (9)	6 (6)
D4GUG6	pyrG	2624	60829	454	8 (8)	8 (8)

Chapter 8: Appendix

O30561	cct1/thS1	0133	58925	403	8 (8)	7 (7)
D4GUL7	proS	0870	54942	390	7 (7)	7 (7)
D4GWJ4	<i>Flavin-dependent pyridine nucleotide oxidoreductase</i>	2345	45796	353	7 (7)	6 (6)
D4GWB0	recJ2	1147	37450	337	6 (6)	6 (6)
O30560	cct2/thS2	0455	59298	317	6 (6)	6 (6)
D4GX92	guaB1	1273	53286	312	6 (6)	6 (6)
P25062	csg	2072	85189	308	11 (11)	3 (3)
D4GST2	<i>Uncharacterised protein</i>	1824	47381	299	4 (4)	4 (4)
D4GZB1	ahcY	0167	46625	277	5 (5)	4 (4)
D4H088	<i>Uncharacterised protein</i>	1749	31044	245	5 (5)	3 (3)
D4GXW9	<i>FAD-dependent oxidoreductase</i>	1396	51068	232	4 (4)	4 (4)
D4GYQ7	dppA1	0062	67307	216	5 (5)	4 (4)
D4GXZ2	pheS	2948	55297	211	4 (4)	3 (3)
D4GP78	hyuA4	B0076	54031	193	4 (4)	4 (4)
D4GSX0	lysS	1867	62118	169	4 (4)	4 (4)
D4GV93	<i>Putative secreted glycoprotein</i>	2160	232813	164	3 (3)	3 (3)
D4H050	<i>ABC-type transport system periplasmic substrate-binding protein</i>	1705	41885	154	3 (3)	3 (3)
D4GY77	pmm4	2989	48528	150	3 (3)	3 (3)
D4GX08	hcpE	1228	83702	113	3 (3)	3 (3)
D4GZX6	rpoA1	0349	108913	108	2 (2)	2 (2)
D4GV94	<i>Putative secreted glycoprotein</i>	2161	36722	98	2 (2)	1 (1)
D4GZY3	tef2	0356	80441	77	2 (2)	2 (2)
D4GUV5	korA	0888	64026	72	2 (2)	2 (2)
D4GW42	btuF	1110	39641	60	1 (1)	1 (1)
D4GRW5	<i>ABCE1 family ribosome recycling factor</i>	0424	67255	53	1 (1)	1 (1)
D4GRV8	cxp	0417	57359	49	2 (2)	2 (2)

Band 2E

Protein accession	Gene name	HVO_number	Predicted MW (Da)	MASCOT score	# of peptides	Peptide sequences
D4GXM0	recJ4	2889	79159	1585	29 (29)	26 (26)
D4GZY3	tef2	0356	80441	1455	34 (34)	27 (27)
D4GYQ7	dppA1	0062	67307	659	19 (19)	10 (10)
D4GXP9	gatE	2902	67597	655	10 (10)	9 (9)
D4GVH7	acdA	1000	74554	648	14 (14)	14 (14)
L9VHM6	sph2	B0118	74332	590	9 (9)	8 (8)
D4GWB0	recJ2	1147	37450	430	8 (8)	7 (7)
D4GZ00	top6B	1571	87438	407	6 (6)	6 (6)
D4GYP9	glyS	0054	64592	369	6 (6)	6 (6)
D4GTI4	citB1	1955	70406	289	5 (5)	5 (5)
Q1XBW2	dnaK	1590	67250	284	5 (5)	5 (5)

Chapter 8: Appendix

P25062	csg	2072	85189	274	4 (4)	3 (3)
D4GX08	hcpE	1228	83702	197	4 (4)	4 (4)
D4GSA2	katG	1778	80004	194	4 (4)	4 (4)
D4GP53	chlID	B0051	73309	146	5 (5)	5 (5)
D4H050	<i>ABC-type transport system periplasmic substrate-binding protein</i>	1705	41885	134	3 (3)	3 (3)
D4GZX6	rpoA1	0349	108913	131	1 (1)	1 (1)
D4GUD0	ppsA	0812	83113	61	2 (2)	2 (2)
D4GYK9	hel308	0014	90355	57	1 (1)	1 (1)

Band 2F

Protein accession	Gene name	HVO_number	Predicted MW (Da)	MASCOT score	# of peptides	Peptide sequences
D4GWM8	cdc48a	2380	81989	2093	43 (43)	35 (35)
Q1XBW2	dnaK	1590	67250	1579	43 (43)	22 (22)
D4GTI4	citB1	1955	70406	1081	27 (27)	18 (18)
D4GXF1	porA	1305	68470	861	20 (20)	13 (13)
D4GUB6	pykA	0806	62042	840	15 (15)	15 (15)
D4GVS7	cdc48b	2700	82817	724	14 (14)	14 (14)
D4GTT9	lon	0783	75482	630	14 (14)	12 (12)
D4GZG5	mcm	0220	78855	600	13 (13)	13 (13)
D4GY19	dsa1	2960	54912	527	10 (1)	9 (9)
D4H037	<i>Putative iron-sulfur protein (4Fe-4S)</i>	1692	80206	520	8 (8)	8 (8)
D4GWB0	recJ2	1147	37450	490	8 (8)	8 (8)
D4GVH7	acdA	1000	74554	477	10 (1)	9 (9)
O30561	cct1/tha1	0133	58925	379	6 (6)	6 (6)
D4GTB5	pccA	2486	65346	330	5 (5)	5 (5)
D4GYQ7	dppA1	0062	67307	279	5 (5)	5 (5)
P25062	csg	2072	85189	249	4 (4)	3 (3)
D4GX08	hcpE	1228	83702	238	5 (5)	5 (5)
D4GXM0	recJ4	2889	79159	198	5 (5)	5 (5)
D4GZY3	tef2	0356	80441	197	5 (5)	5 (5)
D4H050	<i>ABC-type transport periplasmic substrate-binding protein</i>	1705	41885	169	3 (3)	3 (3)
D4GUW1	acs1	0894	74351	131	2 (2)	2 (2)
D4GZU0	atpI	0311	79504	97	2 (2)	2 (2)
D4GXX9	rpap1	1337	64868	95	3 (3)	3 (3)
D4GP53	chlID	B0051	73309	78	2 (2)	2 (2)
D4GT75	pccB2	2471	56772	76	2 (2)	2 (2)
D4GT33	nrdJ	2452	114376	68	1 (1)	1 (1)
D4GSQ3	maeB2	2436	81456	56	1 (1)	1 (1)
D4GZ84	<i>DUF460 domain protein</i>	0140	77555	55	1 (1)	1 (1)

Band 2G

Chapter 8: Appendix

Protein accession	Gene name	HVO_ number	Predicted MW (Da)	MASCOT score	# of peptides	Peptide sequences
D4GZY6	tef1a1	0359	45759	1281	42 (42)	20 (20)
D4GWB0	recJ2	1147	37450	985	23 (23)	13 (13)
D4GVS3	ginS	2698	34301	852	20 (20)	11 (11)
D4GVD7	cetZ1	2204	42064	803	17 (17)	14 (14)
D4GXX3	tsgA11	1401	39212	664	14 (14)	9 (9)
D4GT62	sucC	2465	41005	595	13 (13)	9 (9)
D4GZX7	rpoA2	0350	46140	579	11 (11)	9 (9)
D4GUU7	<i>Uncharacterised protein</i>	0880	36337	536	12 (12)	11 (11)
D4GUN8	<i>Pyridoxal phosphate-dependent aminotransferase</i>	2091	47993	533	10 (10)	9 (9)
D4GVQ6	<i>GFO family oxidoreductase</i>	2690	40028	530	12 (12)	9 (9)
D4GYI3	glpB1	1539	43642	516	7 (7)	7 (7)
D4GY38	thrC3	2969	44404	465	8 (8)	8 (8)
D4GVH6	<i>DRTGG domain protein</i>	0999	39826	463	11 (11)	10 (10)
D4GP49	<i>ABC-type transport system periplasmic substrate-binding protein</i>	B0047	44239	449	10 (10)	8 (8)
D4GWX0	eno	2774	42005	404	8 (8)	6 (6)
D4GU92	icd	2588	45810	393	8 (8)	8 (8)
D4GVK2	acaB2	1025	40732	377	7 (7)	7 (7)
D4GS15	cobT	0590	35605	363	6 (6)	6 (6)
D4GV06	<i>FAD-dependent oxidoreductase</i>	2650	50690	342	6 (6)	6 (6)
D4GYZ9	top6A	1570	41829	293	5 (5)	5 (5)
D4GYD4	gnaD	1488	45794	296	6 (6)	5 (5)
D4GW14	<i>YfhI family protien</i>	2708	35406	250	4 (4)	4 (4)
D4GWT1	<i>Peptidase M42 family protein</i>	2759	37046	245	5 (5)	5 (5)
D4GXW6	<i>M20 family amidohydrolase</i>	1395	45410	223	3 (3)	3 (3)
D4GYN8	argD	0043	39767	216	4 (4)	4 (4)
Q48327	ftsZ1	0717	39805	207	4 (4)	4 (4)
D4GYV5	sufS	0109	46552	205	4 (4)	4 (4)
D4GRM1	hutI	A0560	42409	190	3 (3)	3 (3)
D4GT75	pccB2	2471	56772	177	3 (3)	3 (3)
D4GVU2	trpD2	2226	39646	176	4 (4)	4 (4)
D4GYS6	hemL	0081	47811	155	3 (3)	3 (3)
D4GUJ7	pan1	0850	45802	150	3 (3)	3 (3)
D4GWR0	tef1a2	2413	45628	149	4 (4)	4 (4)
D4GP59	cbiH2	B0057	36370	124	3 (3)	3 (3)
D4GRE8	<i>Pyridoxal phosphate-dependent aminotransferase</i>	A0485	37327	112	2 (2)	2 (2)
D4GZV0	arf1	0321	46582	103	2 (2)	2 (2)
D4GX11	<i>Pyridoxal phosphate-</i>	2871	49116	100	2 (2)	2 (2)

Chapter 8: Appendix

	<i>dependent aminotransferase</i>					
D4GV47	<i>Pyridoxal phosphate- dependent aminotransferase</i>	2671	42412	93	2 (2)	2 (2)
D4GTB8	<i>asd</i>	2487	36612	87	1 (1)	1 (1)
D4GV29	<i>Pyridoxal phosphate- dependent aminotransferase</i>	2661	41877	81	2 (2)	2 (2)
D4GV77	<i>rbcL</i>	0970	43922	80	2 (2)	2 (2)
D4GVS0	<i>graD1</i>	1076	35853	74	2 (2)	2 (2)
Q9HHA2	<i>cct3/tht3</i>	0778	55242	74	2 (2)	2 (2)
D4GYQ7	<i>dppA1</i>	0062	67307	71	1 (1)	1 (1)
D4GP56	<i>cbiX1</i>	B0054	44142	63	1 (1)	1 (1)
D4GTB5	<i>pccA</i>	2486	65346	59	1 (1)	1 (1)
D4GW49	<i>rnj</i>	2724	50115	58	2 (2)	2 (2)
Q977V2	<i>srp54</i>	0123	50918	55	1 (1)	1 (1)
D4GW10	<i>GFO family oxidoreductase</i>	2707	41032	53	1 (1)	1 (1)
D4GXM6	<i>hflX</i>	1346	48212	52	1 (1)	1 (1)
D4GPJ9	<i>ABC-type transport system periplasmic substrate- binding protein</i>	B0198	43426	50	1 (1)	1 (1)

Protein accession code is from the UniProt database, e.g. D4GVS3.

Predicted MW in Daltons (Da) of the protein sequence listed as identified by MASCOT. MASCOT score is the score associated with protein identification. Ions score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event. Individual ions scores > 45 indicate identity or extensive homology ($p < 0.05$), protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits. Number of peptides: number of peptides associated with protein identification by MASCOT. Peptide sequences: the number of distinct peptide sequences associated with the protein identified by MASCOT.

Appendix 2: RecJ1 pulldown mass spectrometry data**H164****Band A**

Protein accession	Gene name	HVO_number	Predicted MW (Da)	MASCOT score	# of peptides	Peptide sequences
Q1XBW2	dnaK	1590	67210	984	21 (21)	17 (17)
D4GZY3	tef2	0356	80448	715	15 (15)	15 (15)
D4GYQ7	dppA1	0062	67380	488	11 (11)	8 (8)
D4GXF1	porA	1305	68541	343	6 (6)	6 (6)
D4GWM8	cdc48a*	2380	81995	164	4 (4)	4 (4)
D4GVH7	acdA/acs	1000	74679	154	4 (4)	4 (4)
D4GTT9	lon	0783	75492	94	3 (3)	3 (3)
D4GTI4	citB1	1955	70819	91	3 (3)	3 (3)
D4GT75	pccB2	2471	56908	48	1 (1)	1 (1)

Band B

Protein accession	Gene name	HVO_number	Predicted MW (Da)	MASCOT score	# of peptides	Peptide sequences
D4GTB5	pccA	2486	65477	1919	35 (35)	28 (28)
THS1	ths1/cct1	0133	58889	974	20 (20)	14 (14)
THS2	ths2/cct2	0455	59319	805	15 (15)	13 (13)
D4GY19	dsa1	2960	54879	538	13 (13)	11 (11)
D4GUV5	korA	0888	64214	307	7 (7)	6 (6)
D4GXY9	pheT	2947	63449	184	4 (4)	4 (4)
D4GZY6	tefla1	0359	45902	99	3 (3)	3 (3)
D4GT75	pccB2	2471	56908	93	1 (1)	1 (1)

Band C

Protein accession	Gene name	HVO_number	Predicted MW (Da)	MASCOT score	# of peptides	Peptide sequences
D4GT75	pccB2	2471	56908	1036	24 (24)	16 (16)
THS3	ths3/ cct3	0778	55265	436	7 (7)	7 (7)
D4GYR9	recJ1*	0073	50864	92	2 (2)	2 (2)
D4H088	<i>Hypothetical protein</i>	1749	31025	50	1 (1)	1 (1)

*predicted to be contaminants from storage

H5199**Band 1A**

Protein accession	Gene name	HVO_number	Predicted MW (Da)	MASCOT score	# of peptides	Peptide sequences
D4GWM8	cdc48a	2380	81995	3809	157 (157)	60 (60)
Q1XBW2	dnaK	1590	67210	1238	22 (22)	17 (17)
D4GXM0	recJ4	2889	79226	913	17 (17)	16 (16)
D4GTI4	citB1	1955	70819	681	12 (12)	11 (11)
D4GYQ7	dppA1	0062	67380	670	18 (18)	10 (10)
D4GYR9	recJ1	0073	50864	498	8 (8)	8 (8)

Chapter 8: Appendix

D4GVS7	<i>cdc48b</i>	2700	82766	469	9 (9)	8 (8)
D4GXF1	<i>porA</i>	1305	68541	369	8 (8)	7 (7)
D4H037	<i>4Fe-4S binding protein</i>	1692	80897	356	7 (7)	7 (7)
D4GTT9	<i>lon</i>	0783	75492	339	7 (7)	6 (6)
D4GVH7	<i>acdA/acs</i>	1000	74679	251	6 (6)	6 (6)
D4GY19	<i>dsa1</i>	2960	54879	197	6 (6)	5 (5)
Q48332	<i>atpA</i>	0316	64702	158	3 (3)	3 (3)
D4GUB6	<i>pykA</i>	0806	62118	145	3 (3)	3 (3)
D4GZU0	<i>atpI</i>	0311	79455	95	3 (3)	3 (3)
D4H050	<i>Iron-III ABC transporter</i>	1705	41916	82	2 (2)	2 (2)
D4GTB5	<i>pccA</i>	2486	65477	57	1 (1)	1 (1)

Band 1B

Protein accession	Gene name	HVO_ number	Predicted MW (Da)	MASCOT score	# of peptides	Peptide sequences
D4GTB5	<i>pccA</i>	2486	65477	2910	118 (118)	39 (39)
THS1	<i>cct1/tht1</i>	0133	58889	1727	33 (33)	21 (21)
THS2	<i>cct2/tht2</i>	0455	59319	1633	28 (28)	23 (23)
D4GY19	<i>dsa1</i>	2960	54879	1427	31 (31)	25 (25)
D4GYR9	<i>recJ1</i>	0073	50864	1076	22 (22)	18 (18)
D4GY56	<i>pccB1</i>	1447	63812	754	11 (11)	11 (11)
D4GT75	<i>pccB2</i>	2471	56908	733	13 (13)	12 (12)
D4GXY9	<i>pheT</i>	2947	63449	647	15 (15)	14 (14)
D4GUV5	<i>korA</i>	0888	64214	586	11 (11)	10 (10)
D4GWM8	<i>cdc48a</i>	2380	81995	568	12 (12)	11 (11)
D4GVJ5	<i>recJ3</i>	1018	69978	201	5 (5)	4 (4)
D4H050	<i>Iron-III ABC transporter</i>	1705	41916	160	3 (3)	3 (3)
D4GUM1	<i>epf1</i>	0874	72130	117	3 (3)	3 (3)
D4GP84	<i>dppA11</i>	B0082	62823	89	2 (2)	2 (2)
D4GZP8	<i>thrS</i>	1684	73555	52	1 (1)	1 (1)

Band 1C

Protein accession	Gene name	HVO_ number	Predicted MW (Da)	MASCOT score	# of peptides	Peptide sequences
D4GYR9	<i>recJ1</i>	0073	50864	2755	227 (227)	38 (38)
D4GT75	<i>pccB2</i>	2471	56908	2151	69 (69)	27 (27)
Q9HHA2	<i>cct3/tht3</i>	0778	55265	1833	33 (33)	22 (22)
D4GVF7	<i>nuoCD</i>	0980	63843	341	8 (8)	7 (7)
D4GSX4	<i>pitA</i>	1871	56054	286	7 (7)	7 (7)
D4GUL7	<i>proS</i>	0870	55137	210	5 (5)	4 (4)
D4GVJ7	<i>HEAT-like repeat containing protein</i>	1020	45579	190	3 (3)	3 (3)
D4GZX4	<i>rpoB2</i>	0347	59202	155	3 (3)	3 (3)

Chapter 8: Appendix

D4GU83	<i>S-layer domain sialidase</i>	2070	58181	135	3 (3)	3 (3)
D4H050	<i>Iron-III ABC transporter</i>	1705	41916	129	2 (2)	2 (2)
D4GXX2	tsgD11	1400	56773	128	2 (2)	2 (2)
D4GXZ2	pheS	2948	55297	109	2 (2)	2 (2)
D4GWM8	cdc48a	2380	81995	84	2 (2)	2 (2)
D4H088	<i>Hypothetical protein</i>	1749	31025	78	1 (1)	1 (1)
D4GR16	dppA7	A0339	61323	55	1 (1)	1 (1)

Protein accession code is from the UniProt database, e.g. D4GVS3.

Predicted MW in Daltons (Da) of the protein sequence listed as identified by MASCOT. MASCOT score is the score associated with protein identification. Ions score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event. Individual ions scores > 45 indicate identity or extensive homology ($p < 0.05$), protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits. Number of peptides: number of peptides associated with protein identification by MASCOT. Peptide sequences: the number of distinct peptide sequences associated with the protein identified by MASCOT.

Appendix 3: RecJ4 pulldown mass spectrometry data**Band 4A**

Protein accession	Gene name	HVO_number	Predicted MW (Da)	MASCOT score	# of peptides	Peptide sequences
D4GTB5	pccA	2486	65346	972	20 (20)	17 (17)
D4GV93	<i>Putative secreted glycoprotein</i>	2160	232813	710	10 (10)	10 (10)
D4GP62	cobN/cbiF	B0060	29688	695	16 (16)	15 (15)
D4GT75	pccB2	2471	56772	575	10 (10)	9 (9)
D4H042	<i>FAD-dependent oxidoreductase</i>	1697	111281	496	10 (10)	10 (10)
P25062	csg	2072	85189	487	35 (35)	5 (5)
D4GT33	nrdJ	2452	114376	429	8 (8)	8 (8)
D4GWK9	carB	2361	117302	427	7 (7)	7 (7)
D4GRZ3	leuS	0452	109563	393	7 (7)	7 (7)
D4GYH4	aglB	1530	113675	392	6 (6)	6 (6)
D4GYJ1	ileS	1547	117994	233	4 (4)	4 (4)
D4H050	<i>ABC-type transport system periplasmic substrate-binding protein</i>	1705	41885	211	4 (4)	4 (4)
D4GWB0	recJ2	1147	37450	182	4 (4)	4 (4)
D4GSM3	dppDF2	0627	97733	161	3 (3)	3 (3)
D4GZU0	atpI	0311	79504	144	2 (2)	2 (2)
D4GZF1	alaS1	0206	102464	104	2 (2)	2 (2)
D4GUI1	petD	0841	28785	89	2 (2)	2 (2)
D4GUI3	petB	0842	30522	82	2 (2)	2 (2)
D4GTK4	sekF	1975	30425	77	2 (2)	2 (2)
D4GUH7	<i>Uncharacterised protein</i>	0838	16190	55	1 (1)	1 (1)
D4GVF5	nuoA	0978	15129	52	1 (1)	1 (1)

Band 4B

Protein accession	Gene name	HVO_number	Predicted MW (Da)	MASCOT score	# of peptides	Peptide sequences
D4GTB5	pccA	2486	65346	3025	205 (205)	38 (38)
O30560	ths2/cct2	0455	59298	1615	33 (33)	23 (23)
D4GY56	pccB1	1447	63624	1037	19 (19)	16 (16)
D4GT75	pccB2	2471	56772	788	11 (11)	11 (11)
D4GVJ5	recJ3	1018	69793	764	13 (13)	12 (12)
D4GYI2	gpdA1	1538	63487	736	13 (13)	13 (13)
D4GUM1	<i>Zinc-dependent nuclease</i>	0874	71947	701	13 (13)	13 (13)
D4GUV5	korA	0888	64026	700	11 (11)	11 (11)
Q1XBW2	dnaK	1590	67250	588	9 (9)	9 (9)
D4GP84	dppA11	B0082	62804	578	12 (12)	11 (11)
O30561	ths1/cct1	0133	58925	456	8 (8)	8 (8)
D4GZX5	rpoB1	0348	67762	454	10 (10)	10 (10)
D4GZ01	gyrB	1572	71213	423	7 (7)	7 (7)
D4GRM3	hutU	A0562	67982	383	7 (7)	7 (7)
D4GUM8	<i>CARDB domain-containing protein</i>	2080	56992	331	6 (6)	6 (6)

Chapter 8: Appendix

D4GW53	gltS	2726	65287	324	6 (6)	6 (6)
D4GWM8	cdc48a	2380	81989	252	5 (5)	5 (5)
D4GXF1	porA	1305	68470	239	3 (3)	3 (3)
P25062	csg	2072	85189	219	5 (5)	3 (3)
D4H050	<i>ABC-type transport system periplasmic substrate-binding protein</i>	1705	41885	215	4 (4)	4 (4)
D4GYQ7	dppA1	0062	67307	214	3 (3)	3 (3)
D4GTJ2	tif5B	1963	65373	199	4 (4)	4 (4)
D4GX08	hcpE	1228	83702	174	4 (4)	4 (4)
D4GXY9	pheT	2947	63317	149	3 (3)	3 (3)
D4GZV3	argS	0324	65034	142	4 (4)	4 (4)
D4GRW2	<i>ArNOG04375 family protein</i>	0421	67691	117	3 (3)	3 (3)
D4GSB2	nirA1	1788	66605	113	2 (2)	2 (2)
D4GTI4	citB1	1955	70406	100	2 (2)	2 (2)
D4GUI1	petD	0841	28785	92	2 (2)	2 (2)
D4GX38	sdhA	2808	67641	91	2 (2)	2 (2)
D4GY19	dsa1	2960	54912	89	2 (2)	2 (2)
D4GZY3	tef2	0356	80441	77	2 (2)	2 (2)
D4H037	<i>Putative iron-sulfur protein (4Fe-4S)</i>	1692	80206	69	1 (1)	1 (1)
D4GYH4	aglB	1530	113675	63	1 (1)	1 (1)
D4GW42	btuF	1110	39641	61	1 (1)	1 (1)
D4GUY6	coxA1	0907	65760	57	1 (1)	1 (1)

Band 4C

Protein accession	Gene name	HVO_ number	Predicted MW (Da)	MASCO T score	# of peptides	Peptide sequences
D4GTB5	pccA	2486	65346	2500	84 (84)	36 (36)
D4GT75	pccB2	2471	56772	2343	183 (183)	34 (34)
Q9HHA2	ths3/cct3	0778	55242	2086	41 (41)	24 (24)
D4GVF7	nuoCD	0980	63484	1333	28 (28)	24 (24)
D4GWU7	aldH2	1189	56171	1046	24 (24)	19 (19)
Q1XBW2	dnaK	1590	67250	1028	16 (16)	16 (16)
D4GT09	aspS	0677	48276	848	19 (19)	14 (14)
D4GWJ4	<i>Flavin-dependent pyridine nucleotide oxidoreductase</i>	2345	45796	824	17 (17)	13 (13)
D4GTF0	serS	1921	52038	795	15 (15)	15 (15)
O30560	ths2/cct2	0455	59298	767	17 (17)	15 (15)

Chapter 8: Appendix

D4GU83	<i>NPCBM assoc domain-containing protein</i>	2070	58129	726	11 (11)	8 (8)
O30561	ths1/cct1	0133	58925	544	10 (10)	9 (9)
D4GXW9	<i>FAD-dependent oxidoreductase</i>	1396	51068	495	8 (8)	8 (8)
D4GT23	gatB	0684	55074	482	11 (11)	10 (10)
D4GSM4	dppA2	0628	62819	458	9 (9)	5 (5)
D4GUG6	pyrG	2624	60829	455	9 (9)	9 (9)
D4GUL7	proS	0870	54942	427	9 (9)	8 (8)
D4GWM8	cdc48a	2380	81989	336	7 (7)	7 (7)
D4GVJ5	recJ3	1018	69793	323	7 (7)	7 (7)
D4GST2	<i>Uncharacterised protein</i>	1824	47381	319	6 (6)	6 (6)
D4GY56	pccB1	1447	63624	293	5 (5)	5 (5)
D4GW42	btuF	1110	39641	292	5 (5)	4 (4)
D4GYQ7	dppA1	0062	67307	271	5 (5)	5 (5)
D4GY19	dsa1	2960	54912	267	7 (7)	6 (6)
D4GZX6	rpoA1	0349	108913	253	5 (5)	5 (5)
D4GX08	hcpE	1228	83702	248	6 (6)	6 (6)
D4GUV5	korA	0888	64026	242	5 (5)	5 (5)
D4H088	<i>Uncharacterised protein</i>	1749	31044	230	4 (4)	3 (3)
D4GX92	guaB1	1273	53286	229	5 (5)	5 (5)
D4H050	<i>ABC-type transport system periplasmic substrate-binding protein</i>	1705	41885	213	4 (4)	4 (4)
D4GZB1	ahcY	0167	46625	210	3 (3)	3 (3)
P25062	csg	2072	85189	195	4 (4)	3 (3)
D4GTK5	secD	1976	55006	170	3 (3)	3 (3)
D4GTT9	lon	0783	75482	168	4 (4)	4 (4)
D4GZY3	tef2	0356	80441	163	4 (4)	4 (4)
D4GYH4	aglB	1530	113675	163	3 (3)	3 (3)
D4GYR4	<i>AlkP-core domain protein</i>	0069	57401	150	4 (4)	4 (4)
D4GUN0	<i>Uncharacterised protein</i>	2082	40885	147	2 (2)	2 (2)
D4H037	<i>Putative iron-sulfur protein (4Fe-4S)</i>	1692	80206	146	3 (3)	3 (3)
D4GXZ2	pheS	2948	55160	145	3 (3)	3 (3)
D4GV94	<i>Putative secreted glycoprotein</i>	2161	36722	124	1 (1)	1 (1)
D4GY36	serA3	2968	55343	116	2 (2)	2 (2)
D4GSX0	lysS	1867	62118	108	3 (3)	3 (3)
D4GTI4	citB1	1955	70406	96	2 (2)	2 (2)
D4GUH7	<i>Uncharacterised protein</i>	0838	16190	88	1 (1)	1 (1)
D4GUI1	petD	0841	28785	85	2 (2)	2 (2)
D4GUM1	<i>Zinc-dependent nuclease</i>	0874	71947	77	2 (2)	2 (2)
D4GP84	dppA11	B0082	62804	72	2 (2)	2 (2)
D4GWT9	<i>DUF1511 family protein</i>	1184	40342	72	2 (2)	2 (2)
D4GUB6	pykA	0806	62042	67	2 (2)	2 (2)
D4GUY6	coxA1	0907	65760	64	1 (1)	1 (1)

Chapter 8: Appendix

D4GUM8	<i>CARDB domain-containing protein</i>	2080	56992	62	1 (1)	1 (1)
D4GRV8	cxp	0417	57359	58	2 (2)	2 (2)
D4GXY9	pheT	2947	63317	50	2 (2)	2 (2)
D4GX49	<i>RimK family protein</i>	2813	49015	46	1 (1)	1 (1)

Band 4D

Protein accession	Gene name	HVO_number	Predicted MW (Da)	MASCOT score	# of peptides	Peptide sequences
D4GZY3	tef2	0356	80441	1930	47 (47)	31 (31)
D4GTB5	pccA	2486	65346	1116	19 (19)	19 (19)
D4GPB9	sph2	B0118	73801	1026	16 (16)	15 (15)
D4GXM0	recJ4	2889	79159	885	18 (18)	18 (18)
D4GYQ7	dppA1	0062	67307	764	27 (27)	11 (11)
D4GUC3	metS	0809	81936	676	11 (11)	11 (11)
D4GX08	hcpE	1228	83702	600	10 (10)	9 (9)
D4GXP9	gatE	2902	67597	588	10 (10)	9 (9)
D4GZ00	top6b	1571	87438	559	10 (10)	10 (10)
Q1XBW2	dnaK	1590	67250	497	7 (7)	7 (7)
D4GSA2	katG	1778	80004	456	9 (9)	9 (9)
D4GT75	pccB2	2471	56772	409	6 (6)	6 (6)
D4GUD0	ppsA	0812	83133	389	7 (7)	7 (7)
D4GVH7	acdA	1000	74554	349	7 (7)	7 (7)
D4GYP9	glyS	0054	64952	339	5 (5)	5 (5)
P25062	csg	2072	85189	318	7 (7)	4 (4)
D4GYH4	aglB	1530	113675	236	5 (5)	5 (5)
D4GUG0	ppc	2621	102399	224	4 (4)	4 (4)
D4H050	<i>ABC-type transport system periplasmic substrate-binding protein</i>	1705	41885	162	3 (3)	3 (3)
D4H037	<i>Putative iron-sulfur protein (4Fe-4S)</i>	1692	80206	128	2 (2)	2 (2)
D4GT20	topA	0681	93686	109	2 (2)	2 (2)
D4GZX6	rpoA1	0349	108913	109	2 (2)	2 (2)
D4GU82	<i>RND superfamily permease</i>	2069	89505	91	2 (2)	2 (2)
D4GUI3	petB	0842	30522	76	2 (2)	2 (2)
D4GUH7	<i>Uncharacterized protein</i>	0838	16190	76	1 (1)	1 (1)
D4GUI1	petD	0841	28785	75	2 (2)	2 (2)
D4GTI4	citB1	1955	70406	72	2 (2)	2 (2)
D4GT33	nrdJ	2452	114376	59	1 (1)	1 (1)
D4GZU0	atpI	0311	79504	59	2 (2)	2 (2)
D4GXX6	rqcH	2883	78933	46	2 (2)	2 (2)

Protein accession code is from the UniProt database, e.g. D4GVS3.

Predicted MW in Daltons (Da) of the protein sequence listed as identified by MASCOT. MASCOT score is the score associated with protein identification. Ions score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event. Individual ions scores > 45 indicate identity or extensive homology ($p < 0.05$), protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits. Number of peptides: number of peptides associated with protein

Chapter 8: Appendix

identification by MASCOT. Peptide sequences: the number of distinct peptide sequences associated with the protein identified by MASCOT.

Appendix 4: GINS pulldown mass spectrometry data**Band A**

Protein accession	Gene name	HVO_ number	Predicted MW (Da)	MASCOT score	# of peptide s	Peptide sequenc es
D4GWM8	cdc48a	2380	81995	4510	361 (195)	69 (48)
Q1XBW2	dnaK	1590	67210	1479	51 (28)	23 (15)
D4GXM0	recJ4	2889	79226	1322	26 (16)	23 (16)
D4GZY3	fusA	0356	80448	1249	38 (17)	25 (13)
D4GTB5	bccA	2486	65477	1077	23 (14)	21 (13)
D4GT75	pccB2	2471	56908	1022	26 (10)	18 (9)
D4GVS7	cdc48b	2700	82766	838	17 (8)	17 (8)
D4GTT9	lon	0783	75492	831	20 (10)	16 (8)
D4GVH7	acdA/a cs	1000	74679	767	16 (10)	13 (9)
D4GXF1	porA	1305	68541	620	17 (8)	10 (6)
D4GYQ7	dppA1	0062	67380	595	20 (8)	9 (6)
D4GXY9	pheT	2947	63449	577	11 (8)	11 (8)
D4GZ84	<i>Possible nuclease of RNase H fold, RuvC/YqgF family</i>	0140	77509	557	11 (4)	11 (4)
Q48332	atpA	0316	64702	488	11 (6)	8 (4)
D4GVS3	ginS	2698	34389	468	9 (4)	7 (3)
D4GXP9	gatE	2902	67726	435	8 (4)	8 (4)
P25062	csg	2072	83341	428	6 (5)	5 (5)
D4GUB6	pykA	0806	62118	413	10 (4)	10 (4)
D4GZG5	mcm	0220	79205	404	9 (3)	9 (3)
D4H037	<i>Putative iron-sulfur protein (4Fe-4S)</i>	1692	80897	381	10 (5)	9 (5)
D4GY19	dsa1	2960	54879	372	9 (3)	7 (3)
D4GPB9	sph2	B0118	74402	360	7 (4)	7 (4)
D4GYP9	glyS	0054	65254	321	7 (4)	7 (4)
D4GZU0	atpI	0311	78700	319	9 (1)	9 (1)
D4GVF7	nuoCD	0980	63843	286	8 (1)	8 (1)
D4GYK9	hel308a	0014	90585	263	6 (4)	6 (4)
D4GTI4	citB1	1955	70819	259	6 (3)	6 (3)
D4GX08	hcpE	1228	82692	193	5 (1)	5 (1)
D4GUV5	korA	0888	64214	171	4 (2)	4 (2)
D4GSQ3	maeB2	2436	81637	153	3 (2)	3 (2)
D4GP53	chlID/h mcA	B0051	73507	145	3 (2)	3 (2)

Chapter 8: Appendix

D4GSA2	katG	1778	79955	136	4 (1)	4 (1)
D4H050	sfuA	1705	41916	129	4 (1)	3 (1)
D4GQG3	<i>Uncharacterised Tat pathway protein</i>	A0133	30988	125	2 (2)	2 (2)
D4GR47	hyuA2	A0379	61993	118	3 (2)	3 (2)
THS2	cct2	0455	58147	98	2 (1)	2 (1)
D4GTU8	gpml	2516	57371	98	2 (1)	2 (1)
THS1	cct1	0133	57717	78	2 (0)	2 (0)
D4GZP8	thrS	1684	73555	70	2 (1)	2 (1)
D4GZY6	tefla1	0359	45902	67	2 (0)	2 (0)
D4GUW1	acs1	0894	74647	56	1 (1)	1 (1)
A0A384KEQ7	rps2P/rps2	2773	28360	46	1 (1)	1 (1)

Band B

Protein accession	Gene name	HVO_number	Predicted MW (Da)	MASCO T score	# of peptides	Peptide sequences
D4GWM8	cdc48a	2380	81995	2152	57 (31)	39 (24)
D4GVS3	ginS	2698	34280	1169	81 (56)	14 (14)
D4GUA4	<i>Stomatin-prohibitin-like protein</i>	0801	46372	1105	27 (20)	18 (15)
D4GUU7	<i>Phosphoserine phosphatase</i>	0880	36315	950	27 (16)	17 (12)
D4GZY6	tefla1	0359	45902	935	34 (24)	15 (12)
D4GT75	pccB2	2471	56908	836	21 (11)	15 (10)
D4GU92	icd	2588	45839	791	23 (15)	15 (11)
D4GWX0	eno	2774	42037	646	17 (7)	12 (7)
D4GT62	sucC	2465	40979	619	13 (10)	10 (8)
D4GTB5	bccA	2486	65477	588	12 (8)	11 (7)
D4GUK8	sufD	0861	44658	556	15 (10)	11 (7)
D4GS20	pgk	0480	43723	552	10 (8)	10 (8)
D4GXG2	glyA1	2862	44493	491	11 (4)	9 (4)
D4GZX7	rpoA2	0350	46112	474	10 (7)	8 (5)
D4GZ07	ndh1	1578	42904	466	9 (5)	8 (5)
D4GYD4	gnaD	1488	45994	463	14 (12)	7 (7)
D4GX38	sdhA	2808	67884	455	14 (9)	9 (6)
D4GW49	rnj	2724	50140	431	11 (6)	9 (6)
D4GVQ6	<i>Oxidoreductase</i>	2690	40060	407	9 (5)	8 (5)
D4GRF0	cbiA/cobB	A048	47139	406	9 (3)	9 (3)

Chapter 8: Appendix

D4GVD7	<i>cetZ1/ftsZ4</i>	2204	42096	396	11 (4)	9 (4)
D4GUJ7	<i>panA</i>	0850	45831	392	7 (4)	7 (4)
D4GP49	<i>Iron ABC transporter substrate-binding protein</i>	B0047	44269	364	6 (5)	6 (5)
D4GVH6	<i>Diamide synthase</i>	0999	39859	346	9 (4)	8 (4)
D4GXX3	<i>tsgA11</i>	1401	39303	336	8 (6)	5 (4)
D4GY69	<i>gdhA2</i>	1453	45816	310	8 (3)	8 (3)
D4GYK9	<i>hel308a</i>	0014	90585	309	7 (4)	7 (4)
D4GST7	<i>pepB2</i>	1829	39398	269	5 (3)	5 (3)
D4GWT1	<i>TET aminopeptidase-like protein</i>	2759	37024	259	6 (2)	6 (2)
D4GZU0	<i>atpI</i>	0311	78700	253	6 (2)	6 (2)
D4GV06	<i>4-hydroxybenzoate 3-monooxygenase</i>	2650	50716	252	5 (3)	5 (3)
P25062	<i>csg</i>	2072	83341	246	5 (3)	4 (2)
D4GYQ3	<i>dppF1</i>	0058	50495	227	5 (3)	5 (3)
D4GWA7	<i>metB1</i>	2750	42234	219	4 (2)	4 (2)
D4GTD4	<i>tif2c/eif2g</i>	1901	44085	203	5 (2)	5 (2)
D4GYK3	<i>lysC</i>	0008	41731	198	4 (2)	4 (2)
D4GYJ0	<i>dhaK</i>	1546	35053	197	4 (3)	4 (3)
D4GYS6	<i>hemL</i>	0081	48066	188	3 (2)	3 (2)
D4GWM3	<i>pstS1</i>	2375	37184	183	4 (4)	3 (3)
D4GRL8	<i>Iron ABC transporter substrate-binding protein</i>	A055 7	44339	179	4 (1)	4 (1)
D4GVU2	<i>trpD2</i>	2226	39679	178	4 (1)	4 (1)
D4GSC3	<i>FAD-dependent oxidoreductase</i>	1799	45649	175	4 (1)	4 (1)
D4GTU3	<i>gatD</i>	2511	44770	172	5 (1)	5 (1)
D4GXW6	<i>Hydrolase</i>	1395	45610	170	3 (3)	3 (3)
D4GP72	<i>gabT1</i>	B0070	46667	163	3 (3)	3 (3)
D4GYP4	<i>argG</i>	0049	44741	163	2 (2)	2 (2)
D4GW90	<i>metE</i>	2742	40558	157	3 (2)	3 (2)
D4GSP2	<i>glnH</i>	2432	32401	143	5 (3)	3 (2)
D4H050	<i>sfuA</i>	1705	41916	140	3 (1)	3 (1)
D4GZV0	<i>erf/arf1/prf1</i>	0321	46724	139	3 (2)	3 (2)
D4GY19	<i>dsa1</i>	2960	54879	135	5 (1)	4 (1)
D4GTU0	<i>carA</i>	2508	38721	133	2 (2)	2 (2)
D4GW10	<i>GFO family oxidoreductase</i>	2707	41121	131	2 (2)	2 (2)
D4H014	<i>vacB</i>	0388	50555	125	3 (1)	3 (1)
D4GY38	<i>thrC3</i>	2969	44775	123	2 (2)	2 (2)
D4GYV5	<i>sufS</i>	0109	46637	119	2 (1)	2 (1)
D4GZT6	<i>Hypothetical protein</i>	0307	49031	117	3 (0)	3 (0)
D4GYQ7	<i>dppA1</i>	0062	67380	113	2 (1)	2 (1)

Chapter 8: Appendix

D4GSF9	malE	0564	46846	113	3 (1)	3 (1)
O30561	cctI	0133	57717	109	3 (1)	2 (1)
Q48327	ftsZ1	0717	39838	108	3 (1)	3 (1)
D4GRM1	hutI	A056	42554	106	3 (0)	3 (0)
		0				
D4GVR2	tsgD3	2692	43496	104	3 (0)	3 (0)
D4GVN0	gatA/aatA	1054	44052	103	3 (1)	3 (1)
D4GYI3	glpB1/gpd B1	1539	43730	102	2 (2)	2 (2)
D4GZA9	mtaD	0165	45315	100	2 (1)	2 (1)
D4GVR7	tsgA3	2695	55559	95	3 (0)	3 (0)
D4GVS1	priS	2697	42648	95	3 (0)	3 (0)
D4GZ82	tyrS	0138	36203	93	2 (1)	2 (1)
D4GZT5	<i>Hypothetical protein</i>	0306	42138	88	2 (0)	2 (0)
D4GYK0	<i>NamA family oxidoreductase</i>	0004	39266	84	2 (1)	2 (1)
D4GZT3	etfA1	0304	33499	84	1 (1)	1 (1)
D4GVT2	gdh	1083	39022	83	1 (1)	1 (1)
D4GPR3	gdhA3	B0266	47393	78	2 (0)	2 (0)
D4GTW0	crtI	2528	56124	77	2 (0)	2 (0)
D4GTJ6	pgi	1967	45691	74	1 (1)	1 (1)
D4GQ81	<i>corC efflux protein</i>	A004	48837	71	2 (0)	2 (0)
		9				
D4GPJ9	<i>Iron ABC transporter substrate- binding protein</i>	B0198	43456	69	2 (0)	2 (0)
D4GSM4	dppA2	0628	62894	68	1 (1)	1 (1)
D4GPB3	<i>Mandelate racemase</i>	B0111	44244	66	2 (0)	2 (0)
D4GXP5	fumC	2900	50119	65	2 (0)	2 (0)
D4GYI4	glpC1/gpd C1	1540	49504	64	2 (0)	2 (0)
D4GYK5	tnaA	0009	48684	64	2 (0)	2 (0)
D4GSV7	hisS	1854	47361	61	2 (0)	2 (0)
A0A384LH W4	rpl3p/rpl3	2564	37156	61	1 (1)	1 (1)
D4GVS0	graD1	1076	35944	56	1 (1)	1 (1)
D4GZ33	aspC1	1604	40339	51	1 (1)	1 (1)
D4GUD8	<i>Flavin- containing amine- oxidoreductase</i>	0817	46249	51	1 (1)	1 (1)
D4GTB7	<i>Sulfatase</i>	0743	50581	50	1 (1)	1 (1)

Protein accession code is from the UniProt database, e.g. D4GVS3.

Predicted MW in Daltons (Da) of the protein sequence listed as identified by MASCOT. MASCOT score is the score associated with protein identification. Ions score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event. Individual ions scores > 45 indicate identity or extensive homology ($p < 0.05$), protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits. Number of peptides: number of peptides associated with protein identification by MASCOT. Peptide sequences: the number of distinct peptide sequences associated with the protein identified by MASCOT.