Genome Architecture and DNA Replication in *Haloferax volcanii*

Hannah Marriott, BSc (Hons)

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

September 2017





Contents

Abstract **xi**

Acknowledgements **xiii**

Abbreviations **xv**

Chapter 1: Introduction **1**

1.1 Archaea 1

1.1.1 Discovery of Archaea 1

1.1.2 Characteristics of Archaea 3

1.1.3 Halophiles 8

1.2 *Haloferax volcanii* 10

1.2.1 Genome Structure 10

1.2.2 Genetic Tools 12

1.2.3 Biochemical Tools 14

1.3 DNA Replication 15

1.3.1 Bacteria 19

1.3.2 Eukaryotes 23

1.3.3 Archaea 26

1.4 DNA Repair 30

1.4.1 DNA Damage 30

1.4.2 Homologous Recombination 31

1.4.2.1 Mechanism 31

1.4.2.2 Regulation 37

1.4.3 Non-Homologous End Joining 39

1.4.3.1 Mechanism 39

1.4.3.2 Conservation 41

1.5 Alternative Forms of DNA Replication 42

1.5.1 Recombination-Dependent Replication 42

1.5.1.1 *Haloferax volcanii* 42

1.5.1.2 T4 Bacteriophage 44

1.5.2 Break-Induced Replication 46

1.5.3 Rolling Circle Replication 46

1.5.4 Mitochondrial DNA Replication 51

1.6 Aims 52

Chapter 2: Materials and Methods 53

2.1 Materials 53

2.1.1 Strains 53

2.1.1.1 *Haloferax volcanii Parental Strains* 56

2.1.2 Plasmids 57

2.1.2.1 Plasmid Construction 59

2.1.3 Oligonucleotides 63

2.1.4 Chemicals and Enzymes 66

2.1.5 Media 66

2.1.6 Other Solutions 69

2.2. Methods 69

2.2.1 General *Escherichia coli* Microbiology 69

2.2.2 General *Haloferax volcanii* Microbiology 71

2.2.3 DNA Extraction from Cells 72

2.2.4 Nucleic Acid Manipulation 73

2.2.5 Genetic Manipulation of *Haloferax volcanii* 78

2.2.5.1 Plasmid Construction 78

2.2.5.2 Strain Construction 80

2.2.6 Genotype Screening 82

2.2.7 Phenotyping of *Haloferax volcanii* 86

2.2.8 Protein Expression and Purification 89

Chapter 3: Comparing Origin-less DNA Replication in the Mega-plasmids pHV3 and pHV4 **93**

3.1 Background 93

3.2 Aims 94

3.3 Results 95

3.3.1 Deletion of *ori-pHV3* and *orc6* from Episomal pHV3 95

3.3.2 Deletion of *ori-pHV4* 101

3.3.3 Integration of pHV3 onto the Main Chromosome 103

Chapter 4: Investigating DNA Replication in the Mega-plasmid pHV1 **121**

4.1 Background 121

4.2 Aims 124

4.3 Results 124

4.3.1 Deletion of *ori-pHV1* and *orc10* 124

4.3.2 Loss of pHV1 130

4.3.3 Pulsed-Field Gel Analysis of pHV1 142

Chapter 5: Exploring Machinery used for Initiation of DNA Replication Including MCM and GINS **149**

5.1 Background 149

5.2 Aims 153

5.3 Results 153

5.3.1 Tryptophan Inducible MCM 153

5.3.2 Epitope-Tagged Replication Enzymes 168

Chapter 6: Discussion **183**

6.1 Genome Architecture 183

6.2 DNA Replication Machinery 191

Chapter 7: Conclusions and Future Work **197**

7.1 Genome Architecture 197

7.1.1 Conclusions 197

7.1.2 Future Work 197

7.2 DNA Replication Machinery 199
7.2.1 Conclusions 199
7.2.2 Future Work 199

References **203**

Appendix **223**
I Synonymous Codon Usage Diagrams 223
II DUE and ORB Alignments 225
III pHV1 226
IV pHV3 234
V pHV4 241
VI Sites of pHV3 Integration 249
VII MCM 250
VIII GINS 253

Microbiology Today Article – Marriott, H. & Allers, T., Archaea and the Meaning of Life. May 2016. **257**

x

Abstract

The archaeon *Haloferax volcanii* is used to study DNA replication and repair, and it is unique amongst cellular organisms as it is able to grow in the absence of DNA replication origins.

There are four DNA replication origins on the main circular chromosome (including the integrated mega-plasmid pHV4) and one on each of the other mega-plasmids pHV1 and pHV3. Replication origins are normally required for the initiation of DNA replication, however *H. volcanii* is able to grow faster when all chromosomal origins have been deleted. Therefore, *H. volcanii* must utilise other methods of DNA replication such as recombination-dependent replication.

The origin found on pHV3 cannot be deleted from the episomal mega-plasmid, whereas the origin can be deleted from episomal pHV4. The pHV3 mega-plasmid can be integrated onto the main chromosome, which allows the pHV3 origin to be deleted from the chromosome.

The pHV1 mega-plasmid origin can be deleted from the episomal mega-plasmid, and the entire mega-plasmid can be lost from the *H. volcanii* cell. This generates a viable, healthy strain, which shows that the pHV1 mega-plasmid is non-essential. It was also found that the pHV1 mega-plasmid exists in *H. volcanii* as a 6x concatemer which is ~510 kb in size, which may explain the reason for being able to delete the origin.

To further investigate the mechanisms that recombination-dependent replication may use, replication machinery (MCM and GINS) were tagged and expressed. Due to time constraints, interactions were not seen. The *mcm* gene was put under the control of a tryptophan inducible promoter. A strain lacking chromosomal origins and therefore primarily using recombination-dependent replication was shown to require more MCM than a wild-type strain.

Acknowledgements

Mostly, I would like to thank Thorsten for letting me work on this project, as well as his guidance and advice throughout. *Haloferax volcanii* is an interesting yet infuriating organism which always seems to come up with plan 'Z'. I would like to think that I can now at least adapt to the unexpected!

Next, I would like to thank everyone past and present in D119 and D100 for all of your help through the years. In particular I thank Becca, Jaime, Nathan and Beccy for your friendship and help around the lab. I would especially like to thank Laura for helping with some of the practical aspects of this project. I also thank Jen and Adam for their friendship and support throughout.

Many thanks have to go to my Mum, who has been behind me every step of the way – even if it is all too complicated! Thanks for all the hugs, advice, supply of food and motivation! I would like to thank Aimée, George and the rest of the DTP cohort for being there along the way. I also thank my best friend Simone for listening to my many rants, for always being there no matter what, and for making me smile.

Finally, I would like to thank my three amazing dogs; Misty, Tag and Darcie for always being there unconditionally whenever I needed a cuddle or some time out from science. I could never be without you.

Abbreviations

5-FOA	5-fluoroorotic Acid
Amp	Ampicillin
ATP	Adenosine 5'–triphosphate
bp	Base pair
BIR	Break induced replication
dHJ	Double Holliday junction
D-loop	Displacement loop
DMSO	Dimethyl sulfoxide
DNA	Deoxyribose nucleic acid
dNTP	Deoxynucleotide
DSB	Double strand break
DSBR	Double strand break repair
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
HJ	Holliday junction
HR	Homologous recombination
kb	Kilobase
LB	Lysogeny broth
Leu	Leucine
Mb	Megabase
NHEJ	Non-homologous end joining
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PMSF	Phenylmethanesulfonylfluoride
RCR	Rolling circle replication
RDR	Recombination-dependent replication
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulphate
SDSA	Synthesis-dependent strand annealing
ssDNA	Single-stranded deoxyribose nucleic acid
SSPE	Saline sodium phosphate EDTA
SW	Salt water
TAE	Tris/Acetic acid/EDTA
TBE	Tris/Borate/EDTA
TE	Tris/EDTA
TEMED	Tetramethylethylenediamine
Thy	Thymidine
Trp	Tryptophan
Ura	Uracil
v/v	Volume per volume
w/v	Weight per volume
WT	Wild-type
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Chapter 1: Introduction

1.1 Archaea

1.1.1 Discovery of Archaea

Methanogens, halophiles and thermophiles had been misclassified as bacteria before Carl Woese and George Fox discovered the domain Archaea in 1977 – in fact, the class comprising of halophilic archaea is still known as Halobacteria. They were originally identified as Archaebacteria by nucleotide sequence analysis of 16S small subunit rRNA (ribosomal RNA), which moved away from more traditional identification methods that relied on visible characteristics including growth conditions, cell morphology or pathology. This showed that the prokaryotes were not one domain but two domains, which were named Eubacteria and Archaebacteria. It was later proposed that because Archaebacteria are more closely related to Eukaryotes rather than Bacteria, they should be renamed Archaea. This led to the proposition that there were three domains of life rather than two; Archaea, Bacteria and Eukaryotes (Figure 1.1) (Woese & Fox 1977; Woese *et al*. 1990). This was initially very controversial in the scientific community; however, the new classification was supported by biochemical analysis of RNA polymerases by Wolfram Zillig, which prompted others to investigate Archaea more closely (Albers *et al*. 2013).

From this it was then discovered that the domain Archaea is split into two distinct groups based on further analysis of rRNA nucleotide sequences: the Crenarchaeota and the Euryarchaeota (Winker & Woese 1991). As new DNA sequencing methods were developed and became more widely accessible, several more groups of archaea were discovered. The Thaumarchaeota, which were originally classified as mesophilic Crenarchaeota (Pester *et al*. 2011), the Korarchaeota, thought to have diverged very early away from Crenarchaeota and Euryarchaeota (Elkins *et al*. 2008; Auchtung *et al*. 2006), and the Aigarchaeota, a sister lineage to the Thaumarchaeota (Hedlund *et al*. 2015). These groups are normally combined with the Crenarchaeota and are generally referred to as the 'TACK' superphylum.

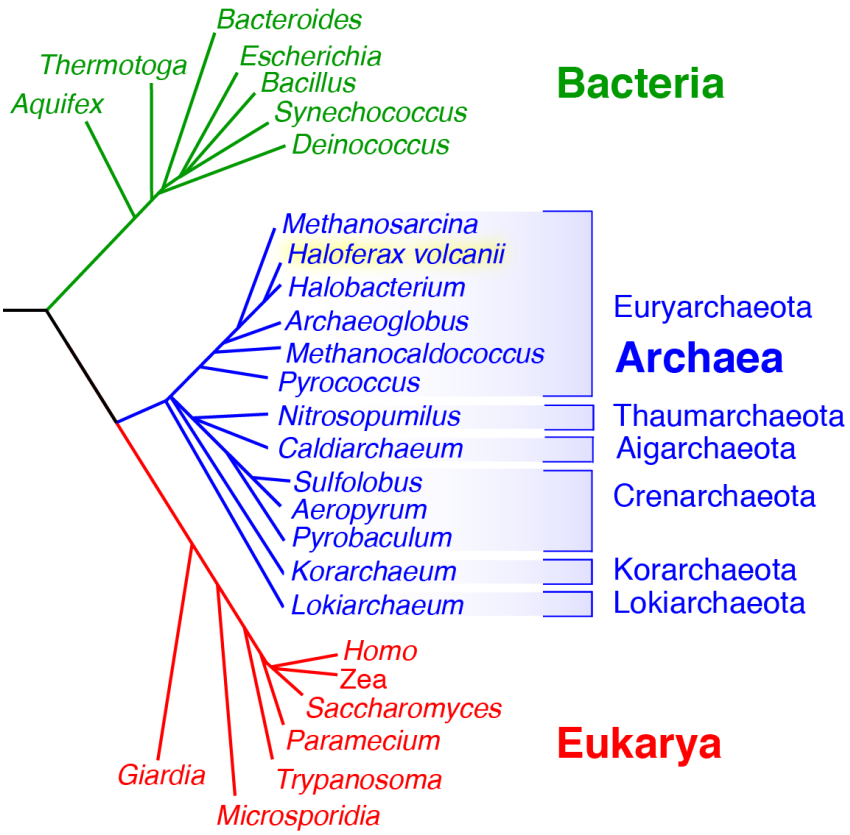


Figure 1.1 **Carl Woese's three domain tree.** Bacteria, Archaea and Eukaryotes are separate domains

Other groups include Bathyarchaeota and Geoarchaeota, about which little is known (Koonin 2015), and Nanoarchaeota which are renowned for having very few genes and a small cell size (Huber *et al*. 2003; Munsen-McGee *et al*. 2015). More recently another group of archaea has been discovered near a hydrothermal vent called Loki's Castle in the Arctic Ocean. This group are called the Lokiarchaeota and are believed to be the closest related to eukaryotes. Lokiarchaeota are part of a group of uncultured archaea called the Asgard superphylum which also includes Thorarchaeota, Odinarchaeota and Heimdallarchaeota. The genomes of these archaea all have high similarities to eukaryotes including homologues of some genes previously thought to be only found in eukaryotes (Zaremba-Niedzwiedzka *et al*. 2017). Phylogenetically this now places eukaryotes as direct descendants of archaea. This changes the tree of life once again into two domains (Figure 1.2); Bacteria and Archaea, with Eukaryotes descending from Archaea (Williams *et al*. 2013; Spang *et al*. 2015; Embley & Williams 2015). It has also been found that biological methane, most probably produced by methanogenic Euryarchaea, has been found in rocks that

are 3.4 billion years old. This predates the earliest known Eukaryotic microfossils which are around 1.8 billion years old (Jaakkola *et al*. 2016), providing further evidence that archaea are an ancient form of life that predates eukaryotes.

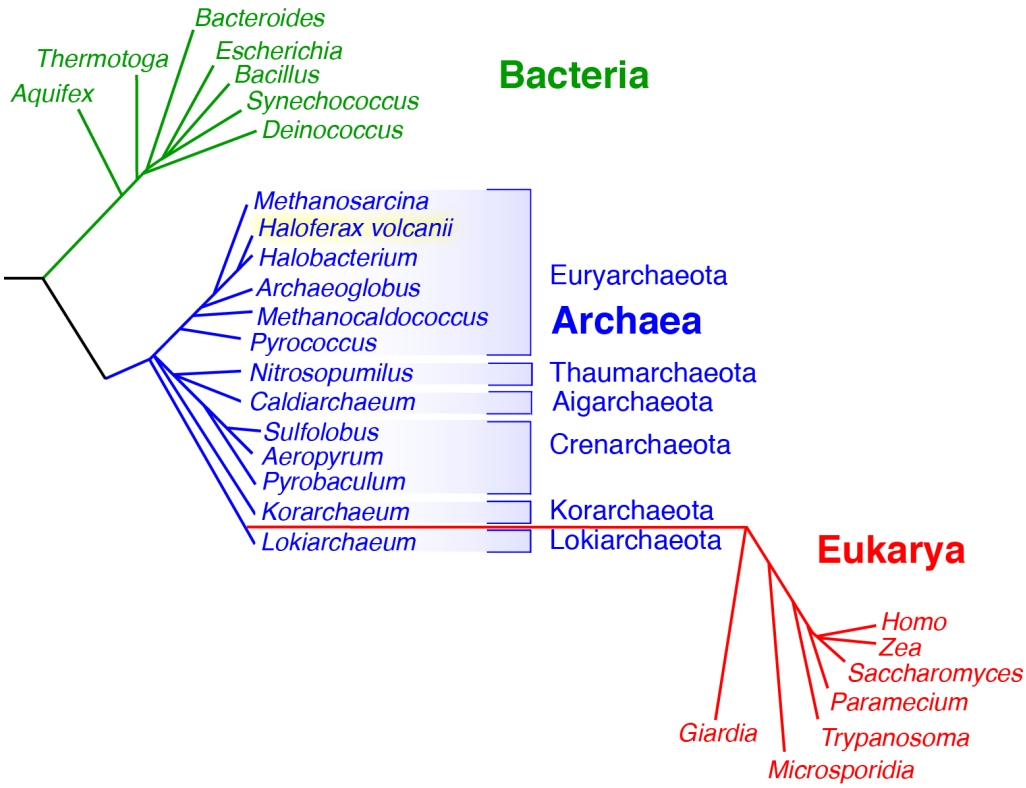


Figure 1.2 **New two domain tree.** Eukaryotes descend from Archaea and Bacteria are a separate domain.

1.1.2 Characteristics of Archaea

Habitats

Archaea are often thought of as extremophiles that live in harsh environments including extremes in temperature, salinity, acidity and alkalinity. However, this is not always the case and archaea can be often found in non-extreme environments such as soil, oceans or the human gut, where they often cohabit with bacteria (Chaban *et al*. 2006; Makarova & Koonin 2003; Khelaifia & Raoult 2016). Halophiles have even been found colonising wall mural paintings (Figure 1.3) causing discolouration (Ettenauer *et al*. 2014).



Figure 1.3 **Halophiles give a pink discolouration to wall paintings.** Halophiles give discolouration to walls in a variety of places including **(A, B)** The Johannes Chapel in Pürgg, **(C)** the castle Rappottenstein and **(D)** the Saint Rupert Chapel in Weißpriach. Adapted from Ettenauer *et al*. (2014).

Cellular and Genome Organisation

Archaea have many characteristics in common with both bacteria and eukaryotes. With regards to bacteria, archaea are also single celled organisms with circular chromosomes and have no cellular organelles – neither are their genomes contained within nuclei. Archaeal and bacterial genomes are also similar in size (normally around 2-4 Mb), use operonic organisation of genes and have a bidirectional origin of replication (Koonin & Wolf 2008); some archaea have more than one origin (Wu *et al*. 2014). Archaea also differ from bacteria as regards their cell envelope. Most archaea have a single membrane that is coated with a proteinaceous surface (S-) layer, which can be glycosylated, whereas bacteria have a cell membrane which is protected by a cell wall consisting of peptidoglycan as well as other components. The proteinaceous S-layer is a barrier to the uptake of DNA, although it has been seen in *Thermococcus* that membrane vesicles can be produced that can carry DNA, and is suggested that this could be a mechanism for the transfer of genetic material. The S-layer can be removed by chemicals or detergents (Rodrigues-Oliveira *et al*. 2017) for example EDTA is used during transformation of *Haloferax volcanii* to remove the S-layer and therefore aid DNA uptake. Archaea use isoprenoid side chains in their cell membranes rather than the fatty acid side chains used in bacteria and eukaryotes (Albers & Meyer 2011). However, archaea are more often seen as simpler models for more complex eukaryotic systems, especially in relation to DNA replication, transcription and translation, where the proteins

share many features (Makarova & Koonin 2003; Makarova & Koonin 2013; Olsen & Woese 1997).

DNA Replication and Repair

Archaeal enzymes and mechanisms of DNA replication and repair are more similar to those found in eukaryotes than those found in bacteria, although archaeal mechanisms are simpler and in some cases do have some bacterial homologues (Ishino *et al*. 2014). This is further discussed in *Section 1.3 DNA Replication*.

Transcription and Translation

Archaeal transcription has been characterised as a combination of eukaryotic and bacterial systems. The proteins involved, namely RNA polymerase (RNAP), share many similarities with eukaryotic proteins; however, regulation of transcription in archaea is similar to bacterial regulation systems (Jun *et al*. 2011). Neither archaea nor bacteria have nuclear membranes and it has been shown that like bacteria, transcription and translation in archaea are coupled, with translation initiated before the full-length mRNA has been produced (French *et al*. 2007). The similarity of the RNAP between archaea and eukaryotes can be seen in Figure 1.4 Archaeal RNAP contains at least 10 subunits and requires TATA-box-binding protein (TBP) and transcription factor B (TFB), both of which have eukaryotic homologues, to initiate transcription (Bell & Jackson 2001). Transcription in archaea is commonly controlled by the Leucine-responsive Regulatory Protein (Lrp) family which is more well known as a regulator in bacteria (Peeters & Charlier 2010).

Translation in archaea is also more similar to eukaryotes than to bacteria. Archaea have at least five initiation factors (IFs) which share homology with eukaryotic IFs (Benelli & Londei 2011).

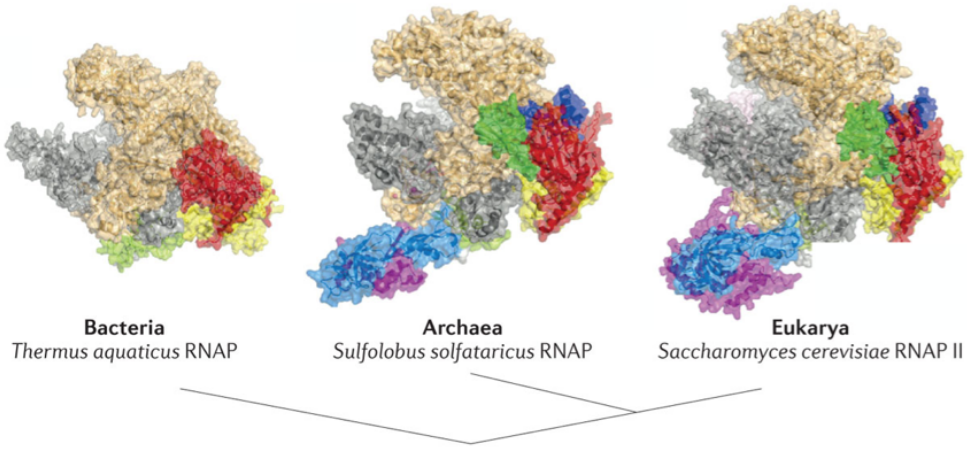


Figure 1.4 **Structure of RNA polymerase across the three domains of life.** Archaeal RNA polymerase (RNAP) is more closely related to eukaryotic RNAP than bacterial RNAP. Adapted from Albers *et al*. (2013).

Genome Packaging

Genome packaging varies between different groups of archaea (Figure 1.5), some use histones similar to eukaryotes, while others use alternative chromatin packaging proteins. Histones are found in the majority of euryarchaeal, thaumarchaeal, aigarchaeal, korarchaeal, and nanoarchaeal genomes. They are also found in limited numbers of crenarchaeal genomes. Eukaryotic histones form two heterodimers of H2A-H2B and H3-H4 which come together as octamers. These octamers then wrap the DNA which both protects the DNA from damage and compacts it, as well as affecting transcription. Archaeal histones are homologues of H3 and H4 proteins, which can form either homodimers or heterodimers, but bind to DNA as tetramers. They are generally smaller than eukaryotic histones due to the lack of N- and C-terminal tails. However, it has recently been found that some archaeal histone sequences have tails, which in eukaryotes contribute to nucleosome stability and gene regulation. These histone tetramers wrap around DNA in a horseshoe structure and protect ~60 bp of DNA, which is about half that of eukaryotic nucleosomes. Archaeal chromatin is not limited to one discrete structure like eukaryotic chromatin; archaeal histones are able to form a wide range of histone-DNA complexes, since archaeal genomes can code up to 6 histone homologues. Some archaea like *Haloferax volcanii* have uniform tetrameric histones that are evenly spaced along the genome, however some archaea such as *Thermococcus kodakarensis*

can form a wide range of histone multimers that can protect up to 500 bp of DNA. *Methanothermobacter thermautotrophicus* also forms histone multimers, but may use a dimer rather than a tetramer to bind DNA. All of these characteristics confer a lot of flexibility to the archaeal nucleosome but may result in less stability compared to eukaryotes (White & Bell 2002; Sandman & Reeve 2005; Ammar *et al*. 2012; Peeters *et al*. 2015; Mattiroli *et al*. 2017).

Alba is a small protein around 10 kDa in size that is known to bind both DNA and RNA, although its primary role is chromatin organisation. It is commonly found in the crenarchaea, most of which do not have histone proteins, but can also be found in some euryarchaea, thaumarchaea, aigarchaea, korarchaea and even eukaryotes, where it is always found alongside a histone. Alba normally acts as a homodimer, however there are two paralogues of Alba in archaea; Alba1 and Alba2. Alba1 is the most common form with Alba2 only ever found as a heterodimer with Alba1. There is some evidence that Alba2 may help DNA compaction. Alongside genome compaction, it is thought that Alba has another function to form filaments that coat the DNA and help protect it from nuclease damage (Sandman & Reeve 2005; White & Bell 2002; Peeters *et al*. 2015; Laurens *et al*. 2012; Crnigoj *et al*. 2013).

There are many other chromatin proteins found in archaea. These include MC1 which has been found in some methanogenic archaea, a homologue of the bacterial protein HU is found in *Thermoplasma acidphilum* and Sul7 is found in *Sulfolobus* (White & Bell 2002; Makarova & Koonin 2003; Sandman & Reeve 2005; Peeters *et al*. 2015).

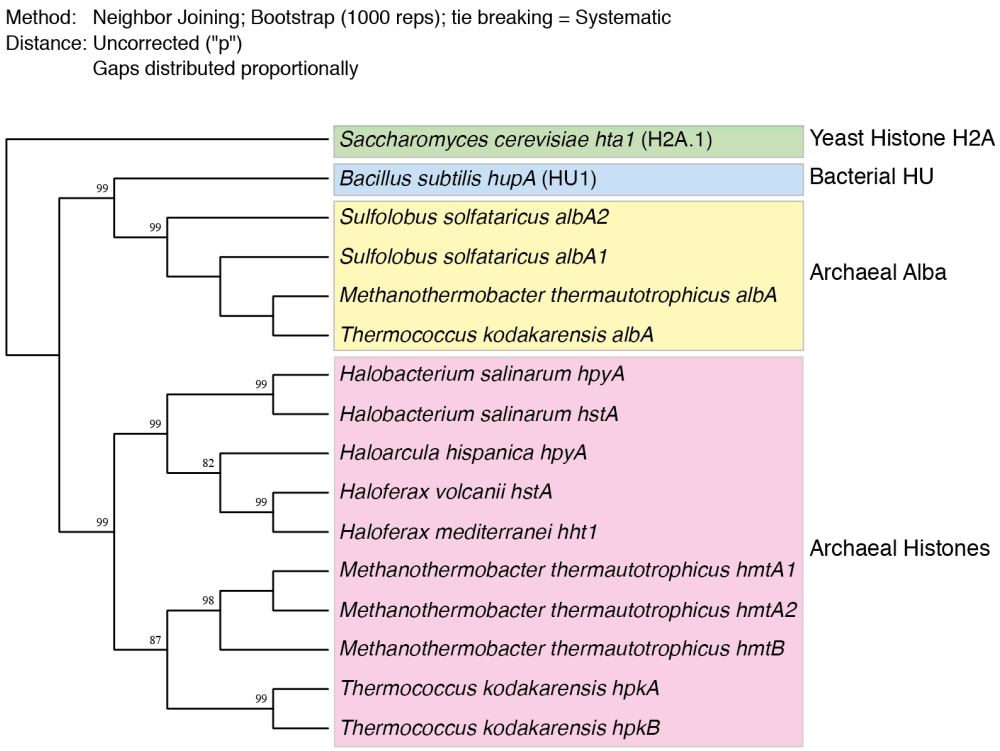


Figure 1.5 **Alignment of archaeal histone and alba proteins.** Tree is rooted to *Saccharomyces cerevisiae* H2A.1 (green box), bacterial HU (blue box) is related to the archaeal alba proteins (yellow box), while archaeal histone proteins are in a separate clade (pink box). Protein sequences were obtained from UniProt and alignment was created using MacVector software.

1.1.3 Halophiles

Halophiles are extremophiles that live in environments where there are high concentrations of salt. Most halophilic archaea use a method called 'salt in' to maintain osmotic balance between intracellular and extracellular salt concentrations. This is where the cells take up salts including potassium chloride to maintain the same salt concentration in the cytoplasm as in the environment. For this adaptation to be successful, the intracellular proteins have adapted to function in high salt conditions. Halophilic proteins are known to have an excess of acidic amino acid residues compared to basic amino acids. This makes the surfaces of halophilic proteins highly negatively charged (Figure 1.6) which is thought to be important for solubility, but also for making them likely to denature in low salt conditions. However, there are not always obvious differences between halophilic proteins and their non-halophilic counterparts in other

organisms, for example dihydrofolate reductase from *Haloferax volcanii* (Oren 2008; Madern *et al*. 2000; Winter *et al*. 2009; Allers *et al*. 2010).

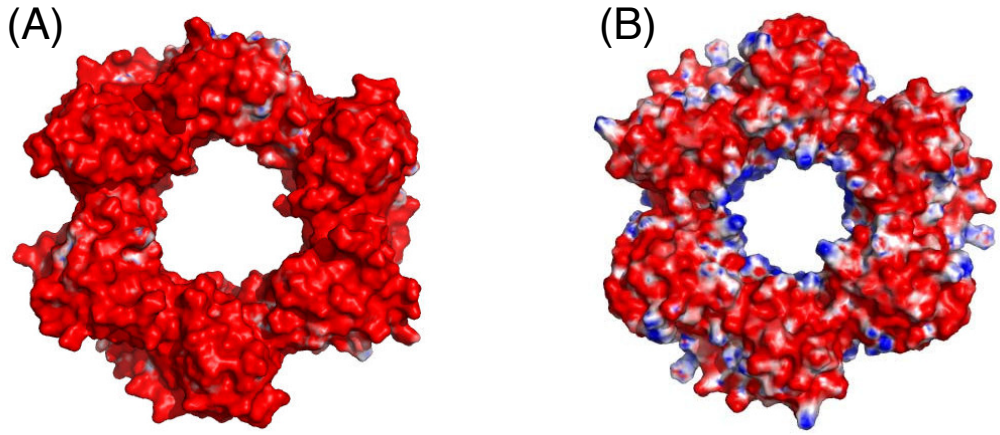


Figure 1.6 **Differences in surface charge between halophilic and non-halophilic PCNA sliding clamp complexes. (A)** *Haloferax volcanii* PCNA. Halophilic proteins have highly negatively charged, acidic surfaces **(B)** *Archaeoglobus fulgidus* PCNA. Non-halophilic proteins have positively charged, basic surfaces. Red indicates acidic/negative and blue indicates positive/basic. Adapted from Winter *et al*. (2009).

The 'salt in' mechanism used by halophilic archaea is very different from the much more widespread alternative, the 'salt-out' mechanism used by most halophilic bacteria. The 'salt out' mechanism is where cells accumulate high levels of organic solutes, such as glycine betaine, sugars, amino acid derivatives and sugar alcohols, within their cytoplasm to maintain an osmotic balance with the environment. The benefit of this system is that intracellular proteins do not need to be modified for a high salt environment, as these organic solutes generally do not interfere with normal enzyme activity. Also, this system allows organisms to adapt quickly to changes in salt concentrations in their environment. On the other hand, this method is energetically costly to the cells, since they must either make or take up organic solutes, and also prevent the entry of high levels of salts into the cytoplasm (Siglioccolo *et al*. 2011; Oren *et al*. 2008).

1.2 *Haloferax volcanii*

Haloferax volcanii is a member of the Euryarchaeota, it is halophilic and can be used to study DNA replication, recombination and repair (Allers & Mevarech 2005; Ammar *et al*. 2012). It is coccoid, pigmented red due to carotenoids, and has a glycoprotein S layer on its outer cell surface. There is a wide range of genetic tools available for *H. volcanii* and it is relatively easy to culture in a laboratory environment, with an optimum growth temperature of 45°C in either a rich or minimal agar or broth-based media. It has a doubling time of around 2.5 hours, grows moderately fast on plates and is able to be grown in overnight liquid cultures (Mullakhanbhai & Larsen 1975; Hartman *et al*. 2010). In the natural environment, it is normally found in hypersaline environments like the Dead Sea, where it was first isolated. *H. volcanii* is able to grow in media containing between 1.7 and 2.5 M NaCl, for which it uses a 'salt in' mechanism (Siglioccolo *et al*. 2011).

1.2.1 Genome Structure

The *Haloferax volcanii* genome consists of a main circular chromosome (2.85 Mb) and three mega-plasmids, pHV1 (85 kb), pHV3 (438 kb) and pHV4 (636 kb). In the wild type strain *H. volcanii* DS2 there is also a small plasmid pHV2 (6.4 kb), however this is not present in the laboratory strains (Figure 1.7) (Charlebois *et al*., 1987; Hartman *et al*., 2010; Wendoloski *et al*., 2001). It was recently found that in the laboratory strain, the mega-plasmid pHV4 has been integrated onto the main chromosome via recombination between identical insertion sequence (IS) elements (Hawkins *et al*., 2013). The genome has 65% GC content and is highly polyploid with up to 20 genome copies in each cell (Breuert *et al*. 2006). There is a full genome sequence available for *Haloferax volcanii* (Hartman *et al*., 2010).

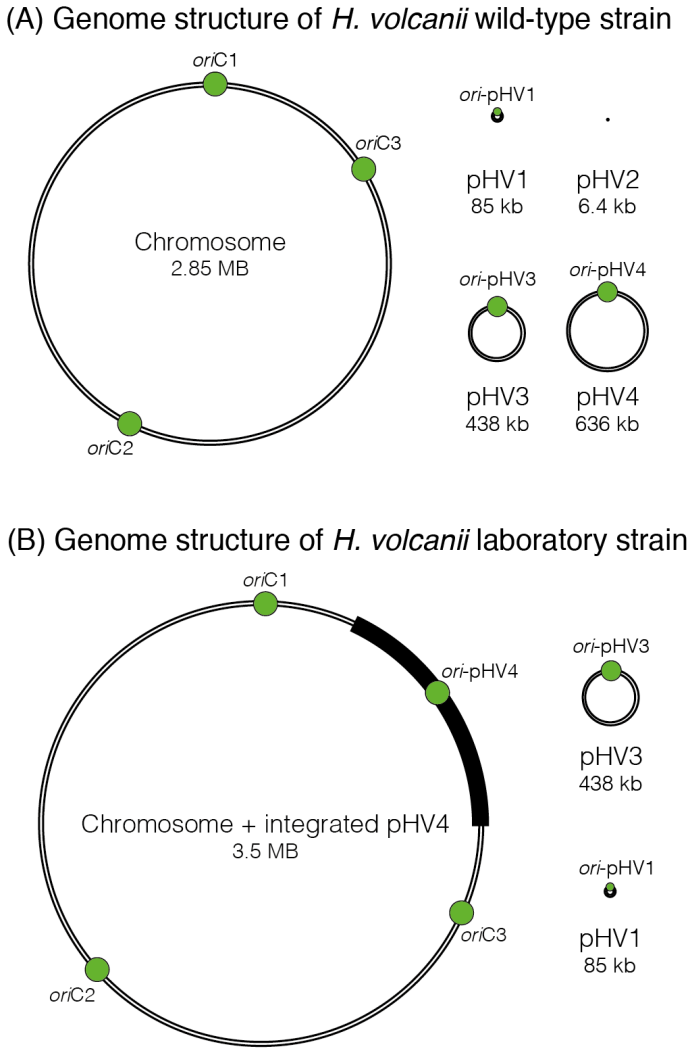


Figure 1.7 **Genome structure of *Haloferax volcanii* showing the locations of DNA replication origins. (A)** Genome structure of wild-type strain DS2 with a single circular chromosome, three mega-plasmids (pHV1, pHV3 and pHV4), and a small plasmid (pHV2). **(B)** Genome structure of laboratory strain H26 consisting of a single circular chromosome with the integrated pHV4 mega-plasmid (solid black line), and the mega-plasmids pHV1 and pHV3.

Origin Deletion

Normal DNA replication of the main chromosome and mega-plasmids requires the firing of origins which then produce a bi-directional replication fork. Similar to eukaryotes, the main chromosome of wild-type *H. volcanii* has more than one origin of replication – three in total (Figure 1.5). However due to the integration of pHV4 onto the main chromosome, there are now four chromosomal origins of replication in the laboratory strain. Each of the mega-plasmids pHV1 and

pHV3 also have one replication origin each (Hartman *et al*. 2010; Hawkins *et al*. 2013).

It has been shown previously by Hawkins *et al*. (2013) that all four of the replication origins found on the main chromosome (*oriC1,2,3,pHV4*) can be easily deleted and in fact the resulting strain grows 7.5% faster than the wild-type laboratory strain. This is surprising as in other organisms, for example *E. coli,* deletion of origins of DNA replication leads to cell death or impaired growth (Ogawa *et* al. 1984).

1.2.2 Genetic Tools

Selectable Markers

There are multiple selectable markers all based on essential amino acid or nucleotide biosynthesis pathways. The corresponding gene can be deleted from the required *H. volcanii* strain and then complemented on a plasmid. This can be useful when transforming *H. volcanii* or studying gene knockouts. As Table 1.1 shows, these genetic markers include *pyrE2* (uracil biosynthesis) (Bitan-Banin *et al*. 2003), *trpA* (tryptophan biosynthesis), *leuB* (leucine biosynthesis) (Allers *et al*. 2004), *hdrB* (thymidine biosynthesis) (Ortenberg *et al*. 2000), *hisC* (histidine biosynthesis), *metX* (methionine biosynthesis) (Leigh *et al*. 2011), *argH* (arginine biosynthesis), and *lysA* (lysine biosynthesis) (Julie Maupin-Furlow, personal communication). The *pyrE2, trpA* and *hdrB* genes are most commonly used.

Table 1.1 **Auxotrophic selection methods available for *Haloferax volcanii*.**

Gene	Encodes	Pathway
pyrE2	Oronate phosphoribosyl transferase	Uracil biosynthesis
trpA	Tryptophan synthase	Tryptophan biosynthesis
leuB	3-isoprpylmalate dehydrogenase	Leucine biosynthesis
hdrB	Dihydrofolate reductase	Thymidine biosynthesis
hisC	Histidinol-phosphate aminotransferase	Histidine biosynthesis
metX	Homoserine O-acetyltransferase	Methionine biosynthesis
argH	Argininosuccinate lyase	Arginine biosynthesis
lysA	Diaminopimelate decarboxylase	Lysine biosynthesis

Antibiotics can also be used as selectable markers in *H. volcanii*, which include a *gyrB* mutant that gives resistance to novobiocin (inhibits DNA gyrase) and an *hmgA* mutant that gives resistance to mevinolin (inhibits HMG-CoA reductase) (Holmes *et al*. 1991; Lam & Doolittle 1989). However, they are not commonly used due to their homology to the native *gyrB* and *hmgA* genes in *H. volcanii*. These genes therefore have a high possibility of recombining with the chromosomal copy of these genes, resulting in an unstable marker and acquisition of antibiotic resistance by the host strain. However, an alternative mevinolin resistance marker has been isolated from *Haloarcula hispanica* which can be used as it has a lower level of identity to the *H. volcanii* gene (Wendoloski *et al*. 2001).

Reporter Genes

The reporter gene *bgaH* (β –galactosidase) allows blue colonies to be detected after cells are treated with X-gal (5-bromo-4chloro-3-indolyl ß-D-galactopyranoside). This is because wild type *Haloferax volcanii* does not have a functioning β-galactosidase gene. An active version, which has been isolated from the close relative *Haloferax alicantei*, can be introduced into *H. volcanii* and then used successfully in blue/white screening of transformed strains (Holmes & Dyall-Smith 2000; Gregor & Pfeifer 2001).

Another reporter gene that can be used is green fluorescent protein (GFP), which has been modified with three separate point mutations so that it will fluoresce in a halophilic environment with improved solubility (Reuter & Maupin-Furlow 2004; Leigh *et al*. 2011).

Transformation

The ease of transformation of *H. volcanii* makes it an ideal candidate for genetic studies. Briefly, EDTA is used to remove the S-layer (by chelation of magnesium) followed by the addition of polyethylene glycol 600, which aids the uptake of DNA into the cells (Cline *et al*. 1989). For further details see *Chapter 2: Materials and Methods, Section 2.2.2 General Haloferax volcanii Microbiology*.

Gene Deletion/ Replacement

There is a successful gene knock-out/ replacement system that can be easily employed in *H. volcanii*, which involves the use of selectable markers including *pyrE2* and the pop in/pop out method of gene deletion (Bitan-Banin *et al*. 2003). This method is further described in *Chapter 2: Materials and Methods, Section 2.2.5 Genetic Manipulation of Haloferax volcanii*.

Due to *Haloferax volcanii* being highly polyploid with 20 genome copies, if a gene is difficult to delete or essential to the cell *H. volcanii* has the possibility of forming a mero-diploid genome where some copies have the original gene and some copies have the gene deletion or replacement construct; thus, allowing the cell to maintain the essential gene and the possible selectable marker which will provide a growth advantage. This requires *H. volcanii* to be checked through Southern blot and not by colony PCR to ensure all copies of the genome have the correct genetic information.

Gene Expression

(Over)expression of genes in *Haloferax volcanii* can be achieved by using the tightly controlled tryptophan-inducible promoters *p.tnaA* (Large *et al*. 2007), or *p.tnaAM3* (Stuart MacNeill, University of St. Andrews, personal communication), which both originate from the tryptophanase gene in *H. volcanii*, *tnaA*. The *p.tnaA* promoter has been used successfully for the overexpression of proteins from plasmids, and for expression of proteins at native levels from the chromosome (Allers *et al*. 2010; Hawkins *et al*. 2013). The use of these promoters is discussed further in *Chapter 5: Exploring Machinery used for Initiation of DNA Replication Including MCM and GINS*.

1.2.3 Biochemical Tools

Special methods must be utilised for protein purification and biochemical assays in *Haloferax volcanii*. Proteins found in *H. volcanii* are halophilic and therefore are normally insoluble and prone to misfolding in low salt conditions.

Consequently, it is not advisable to express halophilic proteins in other expression systems such as *E. coli* or insect cells.

Biochemical tools have been developed to aid the (over)expression and purification of halophilic proteins from *H. volcanii*, which have included the generation of protein expression strains. These strains allow for the efficient transformation of *H. volcanii* with protein-overexpression plasmids by deleting the restriction enzyme Mrr, which degrades foreign GATC-methylated DNA. The deletion of Mrr allows the direct transformation without the need to isolate plasmid DNA from *dam- E. coli* strains first. The host strains also have been improved to reduce the co-purification of histidine-rich proteins with His-tagged proteins, which was accomplished by removing a histidine-rich linker from PitA and removing a histidine-rich C-terminal domain from Cdc48d. As well as hexa-histidine (His_6) tags, streptavidin (StrepII) tags have been also been developed to be used in *Haloferax volcanii* (Allers *et al*. 2010; Gamble- Milner, 2016).

1.3 DNA Replication

Deoxyribonucleic acid (DNA) is a molecule in our cells that stores all our genetic information, it is found in all domains of life; Bacteria, Eukarya and Archaea. It is critical for the survival of the cell, and therefore to the survival of the organism, that DNA is copied accurately and that upon cell division, both daughter cells contain a complete genome. Mistakes made during DNA replication can lead to genetic disorders and diseases such as cancer.

The Replicon Model

Another requirement of DNA replication is that it must occur only when the cell needs to divide. This requires some level of control to ensure that DNA replication initiation is spatially and temporally coordinated with the cell cycle. The replicon model of DNA replication was proposed in 1963 by Jacob, Brenner and Cuzin (Jacob *et al*. 1963), which suggested that replication of DNA begins from *cis*-acting 'replicators', which were only active after being bound by a *trans*-acting 'initiator' (Figure 1.8).

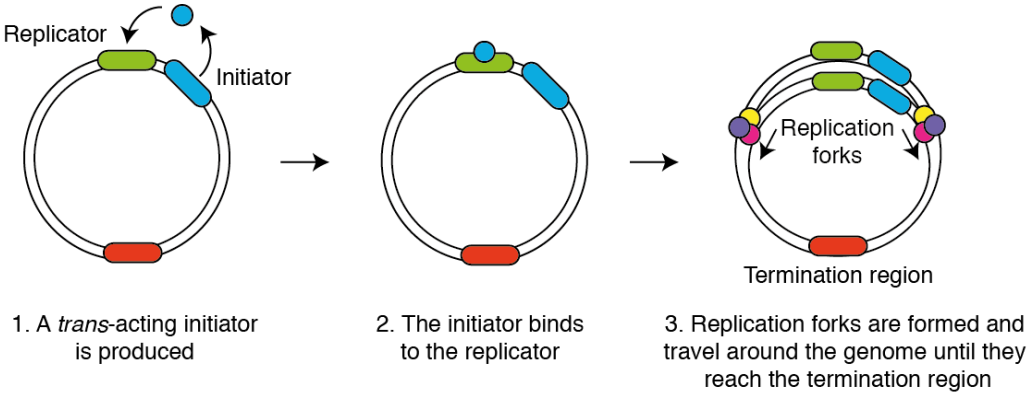


Figure 1.8 **DNA replication model.** DNA replication in a circular bacterial genome. Adapted from Jacob *et al*. (1963).

The binding of the initiator to the replicator then causes DNA replication to initiate from this point, which is also known as an 'origin of replication'. After the initiator protein has bound to an origin, the DNA is unwound slightly allowing the loading of replication machinery, normally starting with the replicative helicase. The replicative helicase plays a vital part and is rate-limiting, as the replication fork cannot progress without the DNA being unwound into two single-stranded DNA (ssDNA) strands (DePamphilis & Bell 2011).

The replication fork proteins vary across Bacteria, Eukarya and Archaea, with some conservation. In addition to the replicative helicase, these include sliding clamps (PCNA), polymerases, primase and ssDNA binding proteins, which all play important roles in successfully and faithful replicating DNA (Leman & Noguchi 2013). Figure 1.9 shows an archaeal replication fork with bacterial and eukaryotic comparisons shown in Table 1.2. Two replication forks are formed travelling in opposite directions along the genome, each with its own replisome, which synthesise the new DNA strands. The forks are eventually terminated when two daughter chromosomes have been produced. Unlike archaea and eukaryotes, bacteria have specific termination (Ter) sites where progression of replication forks is terminated (Gowrishankar 2015).

Table 1.2 **Comparison of DNA replication machinery in Bacteria, Eukarya and Archaea.**

Feature	Role	Bacteria	Eukarya	Archaea
Genome	Genetic information	Small, circular, single chromosome	Large, linear, multiple chromosomes	Small, circular chromosome(s)
Replication Origin	Where DNA replication starts	Single	Multiple	Single or multiple
Initiator Protein	Initially binds to origin to begin replication	DnaA	ORC	Orc/Cdc6
Helicase Loader	Loads the replicative helicase	DnaC/DnaI	ORC, Cdt1, Cdc6	Orc/Cdc6
Replicative Helicase	Unwinds DNA ready for replication	DnaB (Lagging strand)	MCM(2-7) (Leading strand)	MCM (Leading strand)
ssDNA-binding Protein	Binds to ssDNA to protect it from degradation and from forming secondary structures	SSB	RPA	SSB/RPA
Primase	Synthesises short RNA primers for the DNA polymerase to prime from	DnaG	Polα, PriSL	PriSL, possibly DnaG
DNA Polymerase	Synthesises DNA 5'-3' continuously on the leading strand and discontinuously on the lagging strand	PolIII	Polε, Polδ	PolB, sometimes PolD (leading strand only)
Clamp Loader	Loads and unloads the sliding clamp	ɣ-complex	RFC	RFC
Sliding Clamp	Stops DNA polymerase dissociating from the DNA	β-sliding clamp	PCNA	PCNA
Primer Removal	Removes DNA primers	PolI, RNaseH	Fen1, RNaseH	Fen1, RNaseH
Lagging Strand Maturation	Ligates Okazaki fragments	DNA ligase	DNA ligase I	DNA ligase
Topoisomerase	Changes DNA supercoiling near the replication fork	Type I, Type II, Reverse gyrase	Type I, Type II	Type I, Type II, Topo VI

Adapted from DePamphilis & Bell (2011); Hawkins (2009); Kelman & White (2005); Wardell (2013).

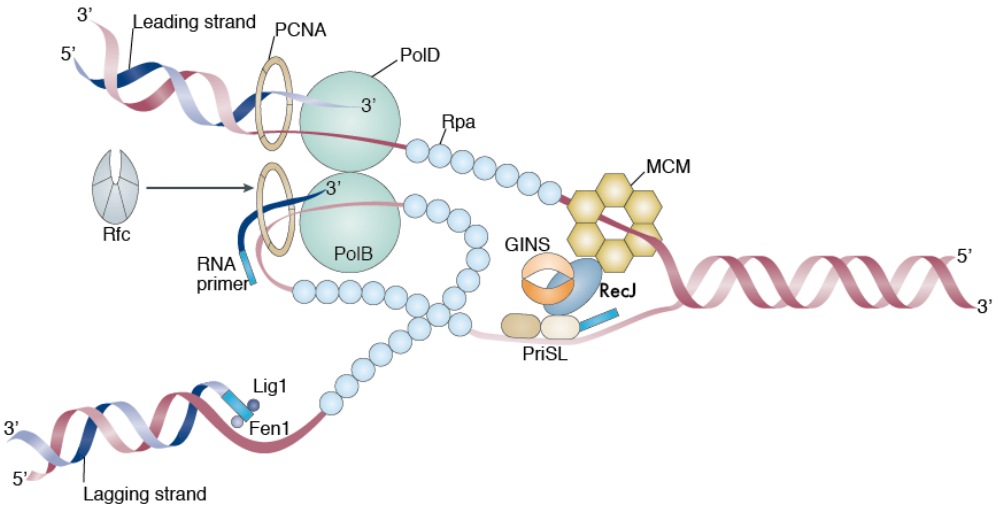


Figure 1.9 **Archaeal replication fork.** The replicative helicase complex (MCM-GINS-RecJ) unwinds DNA so it can be replicated by DNA polymerase held in place by sliding clamps PCNA. PriSL adds RNA primers and RPA coats single-stranded DNA. Adapted from Lindås & Bernander (2013).

Origins of Replication

Replication origins are *cis*-acting elements of DNA where replication is initiated. Core features of origins include an origin-recognition element where the initiator protein binds, and a DNA unwinding element (DUE) where the initial DNA unwinding occurs. These features are usually found in an AT-rich region of DNA, which is easier to melt than a GC-rich region due to A-T having two hydrogen bonds whereas G-C base pairs have three (DePamphilis & Bell 2011).

Origins of replication perform an important role of controlling DNA replication. For example, it is critical in eukaryotes that the genome is replicated once and only once per cell cycle, otherwise this leads to aneuploidy, genome instability and the possibility of chromosome rearrangements or mutations. In bacteria over-replication of the genome can cause increased formation of double strand breaks, a serious form of DNA damage, which are detrimental to growth and could lead to cell death (Michel & Bernander 2014).

Ploidy refers to the number of genome copies present in each cell, which origins can help to control. Maintenance of ploidy in eukaryotic cells is vital for genome integrity and avoiding increased mutation rates which can lead to cancer and other genetic diseases (O'Donnell *et al*. 2013). Single-celled eukaryotes, e.g.

yeast, can be either diploid (two copies) or haploid (one copy), depending on the stage of their cell cycle. Higher eukaryotes are generally diploid in somatic cells and haploid in gametes (eggs/sperm) (DePamphilis & Bell 2011). Bacteria are normally haploid but archaea can vary in copy number (O'Donnell *et al*. 2013). For example, *Sulfolobus acidocaldarius* has 1-2 copies of its genome but *Halobacterium salinarum* is thought to have up to 25 copies of its genome in exponential growth and *Haloferax volcanii* up to 20 copies (Samson & Bell 2011; Breuert *et al*. 2006). This means that one of two modes of replication may occur; either all the copies of genomes are replicated simultaneously or one copy is replicated multiple times and the other copies lay dormant. If only one copy of the genome were replicated multiple times, gene deletions in *H. volcanii* would be very difficult as it would take a long time for gene conversion to occur. Conversely, if all the copies of the genome were replicated simultaneously, gene conversion would act quickly to maintain homozygosity and genes could easily be replaced/ deleted. It is currently unknown which method is used in *Haloferax volcanii*, however it is thought that the all the genome copies are replicated simultaneously due to the speed of gene conversion in the organism when undergoing genetic manipulation, and the ease of gene deletion/ replacement.

1.3.1 Bacteria

The majority of bacteria replicate their circular chromosome using a single origin of replication, *oriC*, which is a well-defined sequence where the initiator DnaA binds (O'Donnell *et al*. 2013). After origin firing, two replication forks are formed that travel in opposite directions around the genome until they reach the termination region, where DNA replication ends with two complete copies of the genome. To avoid replication forks stalling due to colliding with the transcription apparatus, genes are mostly transcribed in the same direction as DNA replication (DePamphilis & Bell 2011).

Initiation

In the model bacterium *E. coli,* DnaA is the initiator that binds to *oriC* causing it to fire. It binds to the DnaA box in the DnaA assembly region, which is located

next to the AT rich DUE region (Ozaki & Katayama 2012; DePamphilis & Bell 2011). There are normally one or two DnaA box clusters in the origins which have a consensus sequence of 5'-TTATNCACA-3', although this varies between organisms (Donczew *et al*. 2014; Messer 2002; Mackiewicz *et al*. 2004). The DnaA filament initially binds ATP and unwinds the DUE forming a ssDNA bubble, thus allowing loading of the replicative helicase DnaB by the helicase loader DnaC. The helicase DnaB is a homohexameric ring that forms around the lagging strand and uses ATP hydrolysis to translocate along the DNA in a 5'-3' direction, unwinding it as it goes. DnaC binds to inactive DnaB using ATP and facilitates the binding of DnaB to the ssDNA bubble created by DnaA. DnaB becomes an active helicase when ATP hydrolysis occurs and DnaC is ejected (O'Donnell *et al*. 2013). Other parts of the replication machinery such as ssDNA binding protein (SSB) and polymerases are then loaded and the replication forks are formed (DePamphilis & Bell 2011).

oriC firing is controlled by SeqA, which binds to hemi-methylated DNA at GATC sites (Rotman *et al*. 2014). Hemi-methylated DNA is the newly synthesised DNA strand has not yet been methylated by Dam methylase (the complementary template strand is still methylated). SeqA has been found to bind more frequently near newly-formed replication forks compared to older, more advanced forks (Waldminghaus *et al*. 2012), this may stop immediate re-firing of the origin. However, this does not completely stop origin firing before the genome is fully replicated, as some bacteria can re-initiate origin firing before the previous round of replication has finished. This is the case in rapidly-growing *E. coli*, where cells can divide every 20 minutes but the genome takes around 40 minutes to replicate (Leonard & Grimwade 2009). Another method for control of origin firing is by the sequestration of DnaA. There are many DnaA boxes termed *datA* that are not located at *oriC*. *dat*A is a region of DNA that contains five DnaA boxes and it thought to bind DnaA with a higher affinity than *oriC*. This stops DnaA from binding to *oriC* and so acts as a negative regulator of origin firing (Kasho & Katayama 2013).

Rep proteins can act as a control mechanism for origin firing in plasmids. They can either be a monomer or a dimer, and can be changed from a dimer to

monomer by other DNA replication associated proteins DnaK, DnaJ and GrpE. The monomer form of Rep proteins is able to act as initiator of origins promoting DNA replication (replication-active), whereas the dimeric form prevents the unwinding of the DUE inhibiting origin activity (replication-inactive) (DePamphilis & Bell 2011). It is thought that the binding of Rep proteins to ssDNA at the DUE region may help to stabilise the formation of the open region of DNA ready for helicase loading (Wegrzyn *et al*. 2014).

Elongation

After the replisome has been assembled at the origin, two DNA replication forks travel around the chromosome, each replicating half of the genome: one leading strand that is synthesised in a continuous fashion, and one lagging strand that is discontinuous in the form of Okazki fragments. The replisome contains all the enzymes required for DNA synthesis including the replicative DNA polymerase Pol III in *E. coli*. Pol III replicates both the leading and lagging stands, and has three subunits including Pol IIIα which is a catalytic subunit, Pol IIIε is a proofreading exonuclease, and Pol IIIθ is a poorly conserved subunit with unknown function that is thought to stabilise and/or stimulate the exonuclease activity of Pol IIIε. Pol III is held by β-sliding clamp that is assembled by the clamp loader, ɣ-complex. DnaB and Pol III both interact with clamp loader, which couples template unwinding with DNA synthesis. The primase DnaG also interacts with DnaB. On the lagging strand, initiation of Okazaki fragments requires synthesis of an RNA primer that is synthesised by DnaG and extended by Pol III, which is then replaced by Pol I that removes the RNA primer and fills the remaining gap. *Bacillus subtilis*, a Gram-positive bacterium, uses two DNA polymerases: DnaE and PolC. DnaE is related to Pol III in *E. coli*, however PolC is not related to Pol III and although it is known to have exonuclease activity, PolC does not have proofreading ability. PolC synthesises the leading strand and DnaE extends RNA primers on the lagging strand a short distance, before full extension by PolC. PolC linked to replicative helicase by the clamp loader, and DnaE interacts directly with helicase and DnaG. These replisome components are conserved across a range of bacteria including *Staphylococcus aureus*,

Streptococcus pyogenes and *Pseudomonas aeruginosa* (Beattie & Reyes-Lamothe 2015; Fijalkowska *et al*. 2012; McHenry 2011).

Termination

In *E. coli*, DNA replication forks converge in the terminus region, which is located on the opposite side of the genome to the origin. The terminus region is flanked by Ter sites: TerA, TerB and TerC. Ter sites are DNA sequences that act as direction-specific replication fork blocks, therefore replisomes can enter but cannot leave the terminus region. This is achieved by deactivation of the DnaB replicative helicase. The Tus protein binds to Ter sites to halt the progression of the replisome; this specific interaction allows termination to be coordinated with other process such as chromosome dimer resolution by XerCD. The Tus protein is not essential and can be deleted, and the Tus-Ter system is poorly conserved across bacteria, although it also found in *Vibrio cholerae*. *Bacillus subtilis* has a functional homologue of Tus, RTP, but RTP lacks both sequence and structural homology to Tus. Positive supercoiling ahead of the replication fork is normally relieved by DNA gyrase, however it is thought that the 3'-5'RecQ helicase and Topoisomerase III may relieve topological stress ahead of the converging replication forks. In *E. coli,* the RecG translocase and helicase-nuclease RecBCD resolve the overlapping DNA sequences at the terminus into single daughter strands (Beattie & Reyes-Lamothe 2015; Gowrishankar 2015; Mulugu *et al*. 2001; Xie *et al*. 2017). Spirochetes such as *Borrelia* have linear genomes that are terminated through covalently closed hairpin ends. Replication starts in the centre of the replicon and travels to the ends, essentially forming a circular DNA molecule. This is resolved through a telomere resolvase ResT, which recognises a *rTel* site located at the joined DNA ends and cleaves the DNA. There is then a conformational change which forms a hairpin at the DNA ends, splitting them into two separate molecules (Chaconas & Kobryn 2010).

1.3.2 Eukaryotes

Eukaryotic genomes are much larger than either bacterial or archaeal genomes. For example, the human genome is around 3000 Mb whereas a typical bacterial genome is around 5 Mb in size (O'Donnell *et al*. 2013). Eukaryotic genomes are split up into several chromosomes, which are linear in contrast to the circular chromosomes found in bacteria and archaea. Each of these chromosomes has many origins where DNA replication may be initiated. To fit inside the nucleus of eukaryotic cells, the genome is tightly compacted by supercoiling, which is achieved by wrapping the DNA around histone proteins that further wrap and coil the genome. This affects which genes and origins can be accessed at any one time, and complicates DNA replication as the DNA strands around the histones must be unwound and then the two daughter strands must be repackaged and supercoiled. There are proteins called histone chaperones and histone remodellers, which remove histones ahead of the replication forks and bind histones to the nascent strands after the replication forks (Leman & Noguchi 2013).

Initiation

The eukaryotic process of origin licensing and origin firing is very tightly regulated. ORC (origin recognition complex) is the initiator protein in eukaryotes, which is a complex made up of six proteins. ORC binds to origins in G1-phase, however the origins do not fire until S phase. ORC then recruits other proteins; Cdc6 and Cdt1. This complex of ORC-Cdc6-Cdt1 allows the loading of the replicative helicase, called the minichromosome maintenance complex (MCM) (Makarova & Koonin 2013), thereby forming the pre-replicative complex (preRC). The process is called origin licensing. Origin firing occurs in S-phase of the cell cycle. This is where the replicative helicase MCM is activated and the replisome is formed. This in turn forms two replication forks that move away from the origin bi-directionally and replicate the DNA (Coster *et al*. 2014).

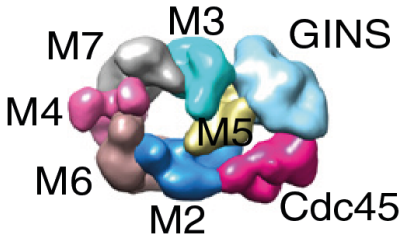


Figure 1.10 **MCM2-7 in complex with GINS and Cdc45.** Eukaryotic MCM2-7 is a hexameric complex which forms into the active CMG helicase complex when associated with GINS and Cdc45. Adapted from Yuan *et al*. (2016).

MCM is a hexameric complex made up of six different proteins, MCM2-7 (Makarova & Koonin 2013). It is initially loaded onto dsDNA and is inactive. Upon loading of MCM, ORC, Cdc6 and Cdt1 are no longer required (Coster *et al*. 2014). Origin firing depends upon the activation of MCM by switching it from dsDNA binding to ssDNA binding. When the cell enters S-phase, the proteins GINS and Cdc45 associate with MCM, forming the CMG complex (Figure 1.10). This process is dependent on protein kinases, CDK and DDK, which modify MCM causing it to switch to binding ssDNA, although the full process is not known. Other replication proteins such as DNA polymerases are recruited and the CMG complex then translocates along the leading strand of DNA of the replication fork in a 3'-5' direction, unwinding DNA ready for replication (Costa *et al*. 2014). MCM is important in DNA replication and if there is decreased expression of MCM, such as in aged haematopoietic stem cells (self-renewing blood cells), this can lead to cell cycle defects and incorrect replication of chromosomes (Flach *et al*. 2014).

The multiple origins present in eukaryotic genomes are not all fired simultaneously, some are fired early in S phase, some in late S phase and some only at certain developmental stages. Unlike in bacteria, binding of the initiator protein complex ORC is not thought to dictate when an origin fires. This is because some origins with very high affinity for ORC do not undergo efficient early S-phase firing. In the model eukaryote *Saccharomyces cerevisiae*, origins are defined by specific sequences, generally those with a high affinity for ORC binding. Chromatin structure plays a major part in ORC binding and therefore

has a regulatory effect. This means that origins can be sequence- dependent or chromatin-dependent, although both specific sequences and chromatin may play a part in origin licensing. It has been found in yeast that chromatin-dependent origins generally fire in early S-phase, whereas sequence-dependent origins are normally found to fire in late S-phase. It was also noted that early-firing origins, on the whole, did not have high affinity for ORC binding suggesting that chromatin structure plays an important part in eukaryotic origin firing and may promote origin activation (Hoggard *et al*. 2013).

Elongation

In eukaryotes, there are three polymerases involved in the elongation of nascent DNA strands: Polε, Polδ and Polα. Polε replicates the leading strand and has proofreading ability due to 3' exonuclease activity. The Dpb2 subunit of Polε interacts with the Psf1 subunit of GINS, linking DNA synthesis with the CMG complex. Polα initially synthesises the lagging strand and does not have proofreading ability. However, mismatch repair is more active on the lagging strand because it uses RNA primers generated by the primase PriSL to distinguish the nascent from the template strand. Polα mediated synthesis is terminated after ~20-30 nucleotides, which allows Polδ-dependent replication to take over (Polδ has proofreading ability). When Polδ reaches 5'end of the preceding Okazaki fragment, it initiates strand displacement synthesis and generates small 5' flaps; they are no longer than ~1 nucleotide due to a molecular brake. Flap endonuclease FEN1 cuts 3' flaps that are produced by remodelling the 5' flap, and the nick is ligated. In yeast, Dna2 cuts longer 5' flaps, and the 3' exonuclease activity of Polδ can trim ends left by Dna2 into ligatable nick. The Ctf4 protein thought to coordinate replisome and interacts with Polα on lagging strand, and GINS and Polε on leading strand (Burgers & Kunkel 2017; Miyabe *et al*. 2011).

Termination

Eukaryotic replication forks terminate when replisomes from adjacent origins converge – it is important that DNA replication is completed accurately, to avoid any genomic rearrangements. There are four steps to eukaryotic termination: dissolution, where the last stretch of parental DNA is unwound between forks; ligation, where remaining gaps are filled in nascent strands and are ligated; decatenation, where dsDNA catenanes are removed; and replisome disassembly. The order of events is not known, however the CMG complex dissociates from dsDNA late during termination, after the ligation step, which avoids premature replisome disassembly. This is achieved by poly-ubiquitination of the MCM subunit MCM7. Converging forks do not stall or slow down during termination events, and leading strands pass each other unhindered before ligation to the downstream lagging strand. Termination regions are called TERs and unlike bacteria, they are not sequence specific. Rrn3 is required for fork progression across TERs and Pfh1 is also involved in termination. Overwinding of DNA caused by converging replisomes may stop progression of replication forks and need to be avoided. Forks can remove positive supercoiling by rotating in a clockwise direction which counteracts the overwinding of DNA. This causes the two DNA strands to cross over and form pre-catenaries, which are resolved by type II topoisomerases (Dewar *et al*. 2015; Dewar & Walter 2017; Bailey *et al*. 2015; Fachinetti *et al*. 2010).

1.3.3 Archaea

Archaea have genomes that are similar to bacteria in that they have circular dsDNA chromosomes. However, unlike bacteria they can have one or more origins of replication, more like eukaryotes. For example, *Pyrococcus* sp. only have one origin of replication whereas *Sulfolobus* sp. have three (Ausiannikava & Allers 2017). Organisms in the genus *Sulfolobus* are known to have a defined cell cycle which can be compared to the eukaryotic cell cycle with defined G1 (pre-replicative), S (DNA synthesis), G2 (growth), M (genome segregation) and D (cell division) phases. This defined cell cycle is common across the crenarchaea but different phases may vary in length between organisms.

Haloferax is a member of the euryarchaea and due to its high levels of polyploidy it has not yet been possible to define a cell cycle (Lindås & Bernander 2013).

Initiation

The archaeal initiator protein shows homology to the eukaryotic initiator proteins Orc1 and Cdc6 and so in archaea it is denoted either Orc, Cdc6 or Orc1/Cdc6. Here it will be called Orc. Orc is used in the vast majority of cases although other initiator proteins have been found such as the WhiP protein in *Sulfolobus*. Orc binds to origins and is also thought to load the replicative helicase MCM, which is homologous to eukaryotic MCM (Wu *et al*., 2014). Archaeal MCM forms a homohexameric ring around the leading strand of replication and translocates 3'-5' along the DNA (DePamphilis & Bell, 2011). Archaea have also been found to contain homologs of GINS and Cdc45/RecJ, which combined with MCM could form a complex with a role similar to the CMG complex seen in eukaryotes (Makarova *et al*. 2012).

Archaeal origins of replication are similar to those found in bacteria (and *S. cerevisiae*) in that they are sequence-specific. There are normally AT rich regions surrounded by origin recognition boxes (ORBs). However, archaeal origins are not conserved across different species. It is notable that archaeal origins and *orc* initiator genes are physically co-located on the genome, suggesting that each origin is individually controlled by its own Orc protein. In fact, the three replication origins found in *Sulfolobus* differ, and each origin is recognised by its own specific initiator protein; no ORBs have been observed at *oriC3*, whereas ORBs are present at *oriC1* and *oriC2*. This is also the case in *Haloarcula hispanica*, where each origin is recognised by a specific initiator protein (Wu *et al*. 2014). There may be some redundancy in other archaea such as *Haloferax volcanii*, where it has been shown that there are nine *orc* initiator genes present on the main chromosome but only three origins (Hartman *et al*. 2010).

Far less is understood about control of replication origins in archaea than in bacteria or eukaryotes. It is generally thought that recognition of origins by Orc

is DnaA- like. For example, in *Haloarcula hispanica*, G-rich inverted repeats have been found which are thought to enhance the binding of initiators; similar repeats are present in bacteria. A model has been proposed for negative regulation of origin firing that is comparable to the *dat*A sequestration of initiators in *E. coli*. Since archaea use a replication machinery that is eukaryotic-like, it has been proposed that there are regulatory factors in *Sulfolobus* that possibly correspond to CDK in eukaryotes. This suggests that there may be further control imposed on origin firing, perhaps similar to eukaryotic origin licensing. There is also an ATP hydrolysis-dependent conformational change in the Orc of *Sulfolobus islandicus*, which inhibits further recruiting of MCM to the origin (Wu *et al*. 2014).

It has been shown that unlike bacteria and eukaryotes, archaea do not necessarily require origins to replicate their genome. The halophile *Haloferax volcanii* has been shown to be capable of faster growth when all the chromosomal origins have been deleted. This indicates that another form of DNA replication is taking place, which was suggested to be initiated by homologous recombination, using the multiple copies of the chromosome in the polyploid cell (Hawkins *et al*. 2013).

Elongation

Euryarchaea use two DNA polymerases for replication; Pol D family and Pol B family polymerases. Crenarchaea do not have D family polymerase, but have multiple B family polymerases. D family polymerases have two subunits, a smaller DP1 3'-5' exonuclease proofreading subunit and a larger DP2 polymerase catalytic subunit, which can both potentially interact with the sliding clamp PCNA. B family polymerases are also found in eukaryotes, they have 3'-5' exonuclease proofreading ability and the archaeal enzymes can detect deaminated bases such as uracil. In *Pyrococcus*, it is thought that Pol B synthesises the leading strand and Pol D synthesises the lagging strand. Archaea also have homologues of the eukaryotic primase subunits PriS and PriL, but not Polα. PriS can synthesise DNA and RNA, although the presence of PriL is required for efficient RNA synthesis. Archaeal Okazaki fragments have been

found to have RNA at the 5' end and it is possible that the archaeal primase extends RNA primers in a manner analogous to Polα in eukaryotes; however, there is currently no evidence for this. Archaeal PriS has similarities to the Pol X family of DNA polymerases, which in eukaryotes function in DNA replication, repair and recombination. However, most archaea do not have PolX family proteins. Therefore, the archaeal primase PriSL may have role in DNA repair; this could explain why it has DNA polymerase and transferase activities (PrimPol) that are not normally associated with a primase. Some archaea also have a DnaG homologue but it is not thought to play a role in DNA replication. Similar to eukaryotes, archaea use FEN1 for Okazaki fragment processing (Barry & Bell 2006; Čuboňová *et al*. 2013; Johansson & Dixon 2013).

Termination

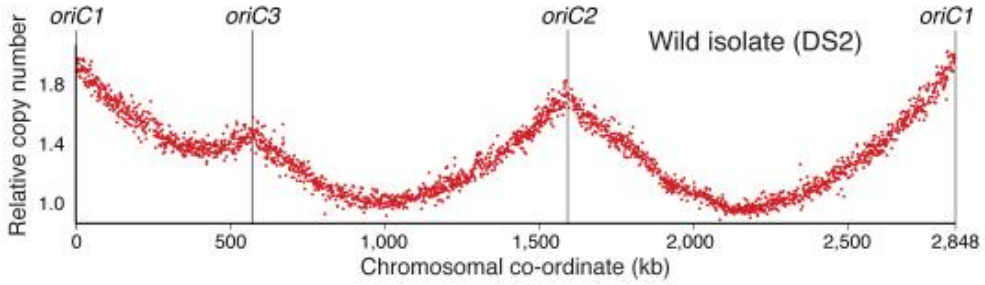


Figure 1.11 **Marker frequency analysis of *Haloferax volcanii* DS2.** Sharp peaks represent discrete origins of replication and smooth valleys represent broad zones of termination. Adapted from Hawkins *et al*. (2013).

Termination in the archaea has yet to be studied in detail, however some initial studies have been carried out into termination in *Sulfolobus*, a member of the Crenarchaea that has three replication origins. It was found that termination in *Sulfolobus* is asynchronous and it is unknown if there are specific fork arrest sites, like the terminus region in bacteria, or whether forks collide at random sites midway between origins like in eukaryotes. It is thought that archaeal termination is in zones rather than specific sites, on account of the broad valleys seen in marker frequency analysis of DNA replication (Figure 1.11). *Sulfolobus* does encode a Xer homologue for chromosome dimer resolution, however unlike

in bacteria, termination and dimer resolution are temporally and spatially distinct processes (Hawkins *et al*. 2013; Duggin *et al*. 2011; Lundgren *et al*. 2004).

1.4 DNA Repair

1.4.1 DNA Damage

DNA damage is a common occurrence that can arise from endogenous (spontaneous) and exogenous (environmental) sources (Friedberg *et al*. 2004). As can be seen in Figure 1.12, there are many different types of DNA damage that can occur.

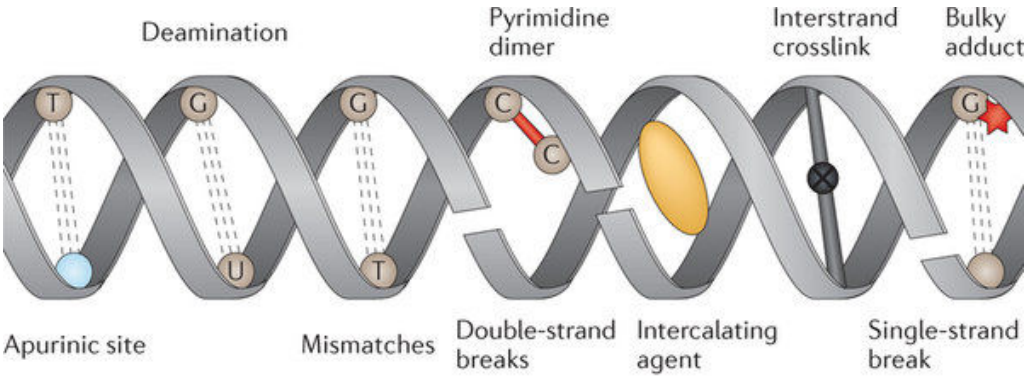


Figure 1.12 **Types of DNA damage.** Common types of DNA damage that can be caused by endogenous and exogenous sources. Adapted from Helleday *et al*. (2014).

Endogenous DNA damage can arise from spontaneous or enzymatic conversions of bases, but can also be caused by deamination, replication errors, and free radial species. DNA damage can be extremely harmful to the cell and organism, for example, enzyme-catalysed deamination of cytosine to uracil in DNA has been linked to cancer. Exogenous sources of DNA damage can be physical or chemical (Helleday *et al*. 2014). Non-ionising radiation such as UV can cause pyrimidine dimers that distort the DNA backbone and can block DNA polymerases (Sinha & Häder 2002). Chemical compounds can intercalate or covalently bond to DNA. These include compounds such as cisplatin (chemotherapy drug) and aflatoxin (compound found in mouldy food), which can cause bulky adducts as well as inter- and intrastrand crosslinks (Hu *et al*. 2016; Poirier 2004), which can also be caused by intercalating agents such as

actinomycin-D. DNA alkylating agents can be caused by endogenous sources from products produced in the cell such as betaine and choline, but can also be caused by exogenous agents e.g. nitrogen mustards, methyl methanesulphonate (MMS). These can cause inter- and intrastrand crosslinks, and methylation of bases that can cause disruption in DNA replication. Free radial species such as reactive oxygen or nitrogen oxide are produced either by ionising radiation (exogenous source) and by normal cellular metabolism (endogenous source) and can cause single and double strand breaks in the DNA backbone, and many base modifications including oxidation and deamination (Helleday *et al*. 2014; De Bont & van Larebeke 2004; Cooke *et al*. 2003).

1.4.2 Homologous Recombination

Homologous recombination (HR) is a form of double strand break (DSB) repair that uses homologous DNA sequences to repair the damaged DNA. It is an error-free repair system, unlike non-homologous end joining that can result in mutations in the DNA; however, if HR is not controlled it can lead to rearrangements in the DNA. HR not only repairs double strand breaks but it also important in the repair of single strand DNA gaps, interstrand DNA crosslinks, as well as replication fork restart. It is a vital repair process that helps limit mutations in the genome, but when the proteins involved in this process are mutated it can result in many forms of cancer, which shows the importance of HR in normal cellular life (Krejci *et al*. 2012; Lui *et al*. 2011).

1.4.2.1 Mechanism

There are three stages of repair by homologous recombination; pre-synapsis, synapsis and post-synapsis (Figure 1.13). This process is dependent on Rad51 recombinase in eukaryotes. RecA and RadA are the homologues of Rad51 in bacteria and archaea, respectively. Rad51 is loaded onto single-stranded DNA (ssDNA) and forms the pre-synaptic filament. In the synapsis stage, the ssDNA coated in Rad51 undergoes a homology search to find intact homologous DNA. Rad51 then helps the formation of the D-loop structure by invading homologous DNA. Rad51 dissociates in the post-synapsis stage so DNA synthesis can occur,

repairing the DNA damage or replication fork. This leads to either crossover, where DNA strands exchange genetic material, or non-crossover products, where no exchange occurs (Krejci *et al*. 2012).

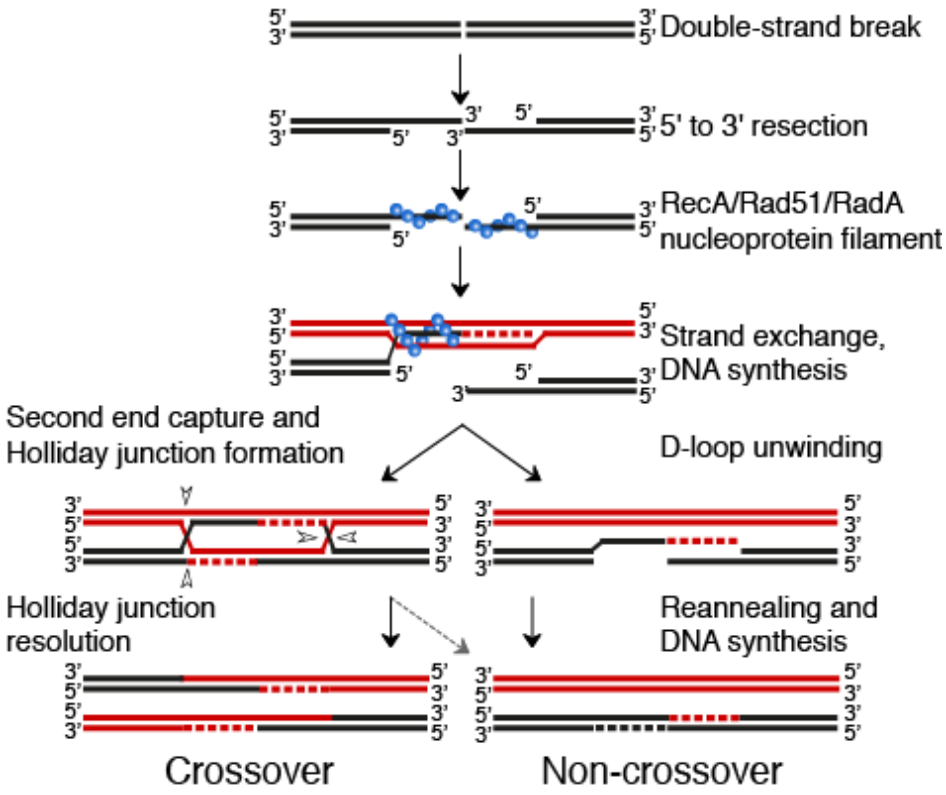


Figure 1.13 **Homologous recombination mechanism.** DNA ends are resected and nucleoprotein filament of recombinase forms on the ssDNA (pre-synapsis). Pre-synaptic filament invades homologous DNA and initiates strand exchange and DNA synthesis (synapsis). Resolution can end with crossover or non-crossover events (post-synapsis). Adapted from Haldenby *et al*. (2009).

Pre-synapsis

In pre-synapsis, dsDNA ends are resected to form single stranded 3' overhangs that are initially coated with single stranded DNA binding protein, which removes inhibitory secondary structures, before being displaced by the loading of the RecA-family recombinase to form the pre-synaptic filament; recombinases are highly conserved across viruses, bacteria, eukaryotes and archaea. Assembly of the pre- synaptic filament is a two-step process. The first step is nucleation, which is initial protein-DNA binding, followed by propagation where additional protein monomers are recruited to the nucleation

sites and extend in both directions covering the ssDNA region. This pre-synaptic filament is not only used for HR, but also features in the SOS response of bacteria and in meiosis in eukaryotes, which also uses the RecA homologue Dmc1 (Liu *et al*. 2011).

In bacteria, specifically *E. coli*, HR can be initiated by the RecBCD pathway or the RecFOR pathway. Both these pathways produce ssDNA onto which the bacterial recombinase, RecA, can be loaded. The RecBCD complex is able to repair DSBs by binding to blunt dsDNA ends and processing them so RecA can be loaded. RecB is a 3'-5' helicase and nuclease, RecC is a subunit that splits duplex DNA and recognises Chi sites, and RecD is a 5'-3' helicase. Together they unwind and resect dsDNA until the complex reaches a Chi site, which is a regulatory sequence (5'-GCTGGTGG-3') that is present throughout the *E. coli* genome. When RecBCD reaches this sequence, it switches function to generate 3' ssDNA ends and facilitates the loading of RecA (Dillingham & Kowalczykowski 2008; Smith 2012). In the RecFOR pathway, the RecFOR complex binds to gapped ssDNA and facilitates the loading of RecA by recognising the 5'- end of a dsDNA-ssDNA junction. Elongation of the filament displaces single stranded DNA binding protein, SSB. The exonuclease RecJ and the helicase RecQ are also involved in the RecFOR pathway (Morimatsu *et al*. 2012; Bentchikou *et al*. 2010). RecA has a natural bias towards binding ssDNA and forms an extended right-handed filament that covers the whole region of ssDNA. It is able to slide along DNA to aid its search for homology during synapsis (Bell & Kowalczykowski 2016).

End resection is initiated by the Mre11-Rad50-Xrs2 (MRX) complex in yeast and the Mre11-Rad50-Nbs1 (MRN) complex in mammals. MRX/MRN is one of the first protein complexes that are recruited to DSBs and controls resection and the tethering of DNA ends until they can be repaired by HR or NHEJ. The MRX/MRN complex is aided by the protein Sae2 in yeast and CtlP in mammals to resect DNA ends. It has been suggested that the MRX complex initiates end resection by creating a nick in the DNA, which allows 5'-3' nucleases, Exo1 and Dna2, to degrade DNA generating a 3' overhang. Dna2 requires Sgs1 helicase (BLM in mammals) to unwind the DNA (Gobbini *et al*. 2016; Cejka 2015).

Rad51 recombinase must then be loaded onto ssDNA that has already been bound by the single stranded DNA binding protein RPA, which removes secondary structures that inhibit Rad51 binding. There are multiple nucleation sites and this results in the formation of discontinuous filaments that have small gaps between them, differently from bacteria where RecA forms continuous filaments. Rad51 requires mediator proteins to be loaded onto RPA-coated ssDNA. In yeast the primary mediator is Rad52, forming Rad52-RPA-ssDNA complexes that recruit Rad51, although Rad54 and Swi5-Sfr2 also play a role. In mammals, BRCA2 directly interacts with around 6 Rad51 molecules for nucleation and loads Rad51 onto RPA-coated ssDNA. BRCA1 and the Rad51 paralogues, Rad51B, Rad51C, Rad51D, XRCC2 and XRCC3, also have a role in recruitment of Rad51. In mammals, Rad52 is involved in second end capture and not in recombinase mediation (Liu *et al*. 2011; Kolinjivadi *et al*. 2017).

Viruses such as T4 bacteriophage also have homologous recombination systems, however their primary function is normally genome replication rather than DNA repair. T4 encodes a single stranded DNA binding protein Gp32, a recombinase UvsX, and a mediator protein UvsY. An exonuclease is also encoded, Gp46-Gp47 complex, which is homologous to the Mre11-Rad50 complex and resects DNA ends to generate ssDNA tails or gaps. Gp32 initially coats the ssDNA and UvsY loads UvsX onto a UvsY-Gp32-ssDNA intermediate. UvsY stimulates UvsX to displace Gp32 by altering the structure of the ssDNA, and also promotes UvsX to bind to ssDNA rather than its preferred dsDNA. Gp32 coverage is not always required prior to UvsX loading, as there is evidence that UvsY and another T4 encoded protein, MR, can load UvsX directly onto ssDNA (Liu & Morrical 2010; Liu *et al*. 2011).

Mre11 and Rad50 are conserved in archaea (Gobbini *et al*. 2016). In thermophilic archaea, for example *Sulfolobus* and *Thermococcus*, the Mre11-Rad50 complex is joined by HerA and NurA for end resection, which are thought to be the functional homologues of Sgs1/BLM and Exo1/Dna2 proteins, respectively, in eukaryotes. HerA is an ATPase with helicase activity and NurA is a 5'-3' exonuclease as well as a ssDNA endonuclease. It is thought that HerA acts as a translocase that feeds DNA into the nuclease sites of NurA (Rzechorzek

et al. 2014; Huang *et al*. 2015; Byrne *et al*. 2014; De Falco *et al*. 2015). *Haloferax volcanii* is highly polyploid and therefore can place more reliance on HR for DNA repair. The Mre11-Rad50 complex is present in *H. volcanii*, however it is thought that DSBs are not repaired primarily by HR and that Mre11- Rad50 acts instead in micro-homology end joining. Mre11-Rad50 may temporarily block HR by stopping formation of RadA filaments (Delmas *et al*. 2009). In archaea, the recombinase involved in HR is RadA, which is a homologue of RecA and Rad51. RadA has DNA-dependent ATPase activity that enables it to form nucleoprotein pre-synaptic filaments, as well as catalysing homology searching and strand exchange. RadA has a paralogue called RadB that directly interacts with RadA. RadB acts a recombination mediator by changing the conformation of RadA promoting polymerisation of RadA on DNA (Figure 1.14) (Haldenby *et al*. 2009; Wardell *et al*. 2017; Komori *et al*. 2000).

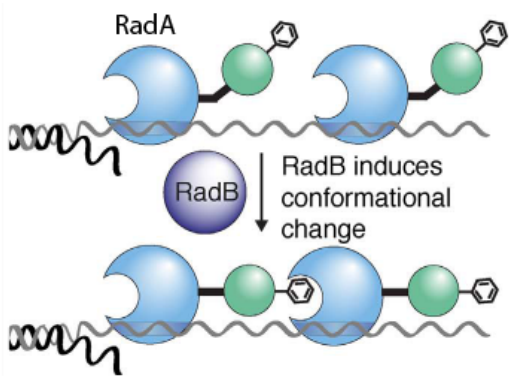


Figure 1.14 **RadB causes a conformational change in RadA**. RadB promotes polymerisation of RadA on DNA by causing a conformational change. RadA monomers polymerise when a phenylalanine is inserted into the hydrophobic pocket of the adjacent monomer. Adapted from Wardell *et al*. (2017).

Synapsis

Synapsis occurs after the formation of the pre-synaptic filament, and is the stage in which the recombinase mediates a homology search for intact homologous DNA. The pre-synaptic filament then invades the homologous DNA and undergoes strand exchange forming a displacement-loop (D-loop) structure with a 3' end from which DNA synthesis can be initiated (Northall *et al*. 2016; Li & Heyer 2008).

Post-synapsis

There are several pathways that can resolve the recombination intermediates generated during synapsis, which can lead to crossover and non-crossover products. Recombination intermediates can also be channelled into other repair pathways other than double strand break repair (DSBR), such as synthesis-dependent strand annealing (SDSA) (Figure 1.15) or break-induced replication (BIR) (Northall *et al*. 2016).

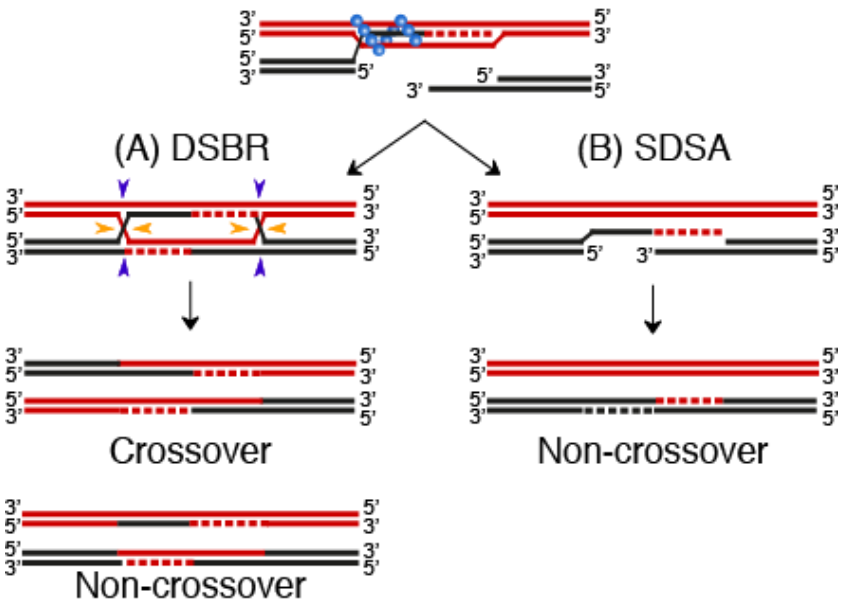


Figure 1.15 **Resolution of homologous recombination intermediates. (A)** DSBR forms a double Holliday junction which can be resolved to form crossover or non-crossover products. **(B)** Synthesis-dependent strand annealing dissolves the D-loop and forms a non-crossover product. Adapted from Sung & Klein (2006).

Synthesis-Dependent Strand Annealing

In synthesis-dependent strand annealing (SDSA)n a small amount of DNA synthesis occurs extending the D-loop, which is then unwound leading to the disengagement of the invading strand. The disengaged strand then reanneals with the resected second end of the DSB. This mechanism of resolution of HR intermediates always leads to non-crossover products, avoiding genomic rearrangements, and in eukaryotes is thought to involve Rad52 and BRCA2. It also requires a motor protein which is Srs2 and Irc20 in yeast or BLM in mammals (Li & Heyer 2008; Heyer *et al*. 2010; Miura *et al*. 2012).

Double Strand Break Repair

Double strand break repair (DSBR) is an alternative mechanism of HR resolution to SDSA, but can either result in a crossover or a non-crossover event, with a bias towards crossover events caused by MutL and Exo1. Crossover events increase the possibility for genomic instability through deletions, inversions, or translocations in repeated sequences. In this pathway of resolution, the second end of the DSB is captured to form a double Holliday junction, or is involved in a second invasion event.

Annealing of the second end is catalysed by Rad52 and forms a double Holliday junction (dHJ). The dHJ is then resolved by resolvases, either endonucleases or a dsDNA nuclease. In bacteria, the dHJ is cleaved by the endonuclease RuvC. In eukaryotes, resolvases include Gen1, SLX1/SLX4 and Mus81/EME1. dHJs can also be processed by a RecQ family helicase and topoisomerase to form non-crossover products, BLM and type 1 topoisomerase in mammals and Sgs1-Top3-Rmi1 in yeast (Li & Heyer 2008; Heyer et al 2010; Jasin & Rothstein 2013). This is known as dHJ dissolution.

Break-Induced Replication

Break-induced replication (BIR) is a form of DNA repair that uses migrating D-loops to form bubble-like replication forks (Saini *et al*. 2013). This process is described further in *Section 1.5.2 Break-Induced Replication*.

1.4.2.2 Regulation

Regulation of HR is essential to maintain genome integrity and avoid DNA rearrangements (Heyer *et al*. 2010). Regulation is mainly based around Rad51/RecA/RadA and its regulators, due to its central role in HR, and is focussed on the relationship between Rad51, single- stranded DNA binding protein RPA and ssDNA. RPA has higher affinity for ssDNA than Rad51, and mediators allow RPA to be replaced by Rad51 and thus to enable HR. RPA also promotes recombination by removing secondary structures that could hinder formation of Rad51 filament (Krejci *et al*. 2012).

Positive regulators promote HR. Rad52 in yeast is a recombination mediator that allows Rad51 to replace RPA. It facilitates loading of Rad51 by increasing the stability of the Rad51 presynaptic filament and protects it from being removed by helicases. BRCA2 in humans shares no homology with yeast Rad52 but has a similar function of aiding the assembly of Rad51 filaments. BRCA2 stabilises filaments and prevents Mre11-mediated degradation of replication forks, providing a link between HR and DNA replication. In archaea, the mediators RadB (e.g. *Haloferax volcanii*) and SSoRal1 (*Sulfolobus solfataricus*) aid RadA binding to ssDNA (Wardell *et al*. 2017). Rad51 paralogues in eukaryotes include Rad51B, Rad51C, Rad51D, XRCC2 and XRCC3, which form complexes that are thought to mediate, promote and/or stabilise Rad51 filaments. They are also involved in Holliday junction resolution and may act with BRCA2 (Krejci *et al*. 2012; Sung & Klein 2006). The Shu complex in *Schizosaccharomyces pombe* contains Rad51 paralogues and is thought to play a role in HR during replication stress, as well as inhibiting the Srs2 anti-recombinase (Bernstein *et al*. 2011). Rad54 is an example of the Snf2/Swi2 family of DNA-dependent ATPases that stabilise Rad51 filaments. They stimulate Rad51-mediated strand invasion and promote branch migration of D-loops and Holliday junctions. These ATPases also act as negative regulators in later stages of HR to prevent non-specific binding of Rad51 to dsDNA and to remove Rad51 to expose the free 3' end for initiation of DNA synthesis (Krejci *et al*. 2012; Sung & Klein 2006; Heyer *et al*. 2010).

Negative regulators either act to remove Rad51 or block its action. This is important as sometimes recombination can be harmful to the cell. For example, translesion DNA synthesis can be a safer way of restoring stalled replication forks. Binding of Rad51 can interfere with replication progression and needs to be removed to allow subsequent DNA synthesis following HR (Krejci *et al*. 2012). In yeast, Srs2 is an anti-recombinase with homology to *E. coli* UvrD and Rep helicases. Srs2 unwinds DNA 3'-5' and removes Rad51 from DNA allowing RPA to bind in its place, including from D-loop structures. It also has translocase activity which aids the dismantling of Rad51 filaments (Marini & Krejci 2010). Translocases, for example FANCM in humans, can unwind D-loop intermediates, catalyse branch migration and help to supresses crossover

events. RecQ family helicases can dismantle Rad51 filaments, thereby allowing the resolution of recombination intermediates (including double Holliday junctions) in a non-crossover configuration (Krejci *et al*. 2012). Rad51 and its regulators can be regulated by post-translational modifications e.g. phosphorylation, ubiquitination and SUMOylation (Heyer *et al*. 2010).

1.4.3 Non-Homologous End Joining

Non-homologous end joining (NHEJ) is another method of repairing DNA double strand breaks. NHEJ does not require a homologous DNA template like HR and therefore is able to be implemented at any point in the cell cycle, not only when there are multiple copies of DNA. NHEJ plays a role in maintaining genome stability, with deficiencies in NHEJ being implicated in various cancers (Davis & Chen 2013). However, NHEJ can be error-prone and introduce insertions or deletions into the DNA sequence, although this is not always the case (Chang *et al*. 2017; Bétermier *et al*. 2014).

1.4.3.1 Mechanism

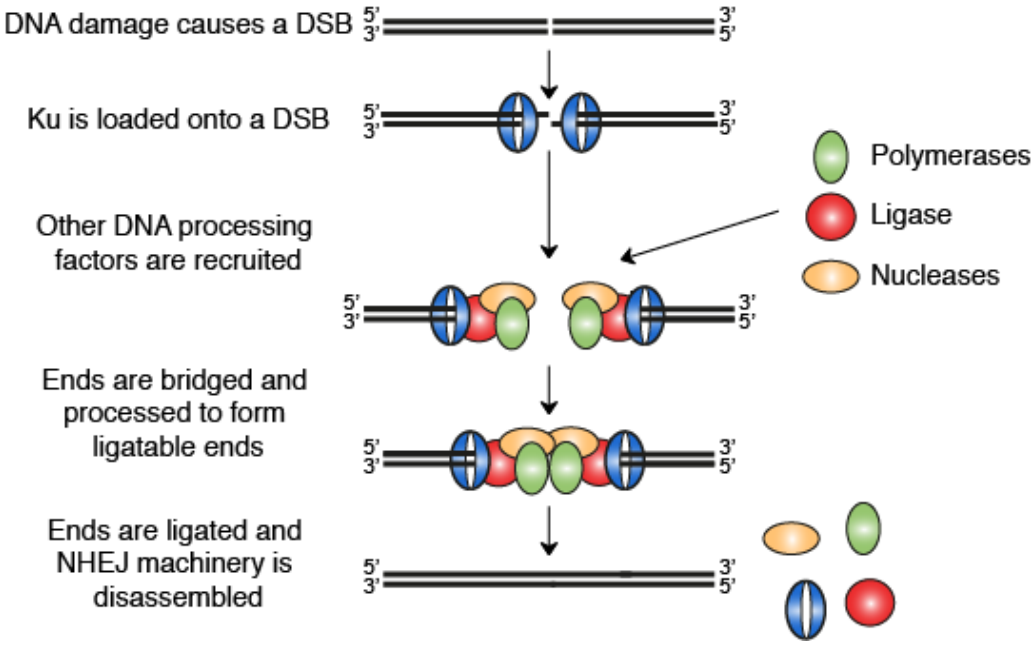


Figure 1.16 **Non-Homologous End Joining pathway.** Ku loads onto ends of double strand breaks and other factors are recruited including ligase, nucleases and polymerases Ends are then processed so they can be ligated and NHEJ machinery is disassembled when repair is complete. Adapted from Bowater & Doherty (2006).

There are four steps of NHEJ (Figure 1.16). Firstly, DNA ends are recognised and NHEJ machinery is assembled at the DSB, which are then bridged, holding them together. This is followed by DNA end processing, ligation of the broken ends and dissolution of the NHEJ complex (Davis & Chen 2013).

The first step of NHEJ is the recognition of the DSB ends and in higher eukaryotes the binding of the Ku heterodimer to DSB, which is made up of Ku70 and Ku80 subunits. Ku is abundant in the cell and can quickly localise to DSBs in sequence independent manner by binding to the DNA backbone. Ku makes a ring structure that slides onto the ends of DSBs and acts as scaffold to recruit other NHEJ factors, including DNA-PKcs and the XRCC4-DNA ligase IV complex. XRCC4 also acts as scaffold to recruit other NHEJ factors. Ku also helps maintain the stability of the DNA ends by protecting them from non-specific processing that may increase genomic instability. The next step is to bridge the two DNA ends, holding them together while they are repaired. This step is also called synapsis. This is mainly executed by DNA-PKcs, although the XRCC2-XLF complex also plays a role in bridging by forming a filament; it is unknown if they work together or independently. Ends can have some homology allowing them to base pair, which is called micro-homology end joining, although this is not essential in NHEJ. The third step of NHEJ is to process the DNA ends to create ligatable ends. This can be achieved through end resection, filling in and removing 5' hydroxyls and 3' phosphates that block ligation. These DNA processing enzymes are most likely recruited by the Ku-XRCC4 scaffold. Artemis is a 5' endonuclease that nicks 5' overhangs to leave a blunt duplex end, but also has 5'-3' exonuclease activity on ssDNA and can remove 3' phosphoglycolate groups. PNKP is a kinase and a phosphatase that removes 3' phosphates and adds phosphate to 5' OH groups. APLF has both endonuclease and 3'-5' exonuclease activity which give it's the ability to resect 3' overhangs. WRN is a RecQ family helicase that has 3'-5' exonuclease in presence of Ku and XRCC4, and Ku itself is able to excise abasic sites near DSBs. There are also polymerases recruited, both in the PolX family of polymerases; polymerase μ has template-dependent synthesis and can polymerase across discontinuous template strand, and polymerase λ which has template-independent activity with

β lyase activity that allows it to remove individual damaged bases. Both polymerases can either incorporate dNTPs or rNTPs, with the rNTPs being removed later by base excision repair. The final step is to ligate ends by DNA ligase IV, stabilised by XRCC4. XLF stimulates activity of DNA ligase IV towards mismatched and non-cohesive DNA ends. APLF also stimulates ligation of XRCC4-DNA ligase IV in the presence of Ku. Ku is required for the ligation of non-homologous ends, but is not required if there is micro-homology. After ligation and completion of the repair, the NHEJ complex is dissolved (Lieber *et al*. 2003; Davis & Chen 2013; Chang *et al*. 2017).

1.4.3.2 Conservation

In yeast, the NHEJ mechanism of action is very similar to that found in higher eukaryotes. *S. cerevisiae* has functional homologues of Ku, XRCC4 and DNA ligase IV, called Hdf1/Hdf2, Lif1 and Dnl4, respectively. Dnl4 and Lif1 also form complex similar to the XRCC4-DNA ligase IV complex. However, there is no known DNA-PKcs homologue and instead yeast has the Mre11-Rad50-Xrs2 complex as the functional equivalent of DNA-PKcs/Artemis complex (Hefferin & Tomkinson 2005). Unlike higher eukaryotes in which HR is limited by the cell cycle and therefore NHEJ is utilised as a major DNA repair pathway, NHEJ seems to be the minor pathway in yeast to repair DSBs (Bowater & Doherty 2006).

Many bacteria been identified that have a protein homologous to Ku, however it forms homodimers rather than heterodimers as in eukaryotes. Ligase D is the bacterial homologue of DNA-ligase IV, however in many bacteria it also often contains polymerase and nuclease domains. *Bacillus subtilis* has YkoV, which is a Ku homologue, and YkoU which has ligase, primase and nuclease domains. *Mycobacterium tuberculosis* has Mt-Ku and an ATP-dependent ligase, Mt-Lig, which also has primase and nuclease domains. In this case, Mt-Ku homodimer is thought to also act as the bridging factor in place of DNA-PKcs and Mt-Lig is a 3'-5' single-stranded exonuclease that is able to process DNA ends. Together Mt-Lig and Mt-Ku form a two component NHEJ system. Most bacteria with a two component NHEJ system either sporulate (*B. subtilis*) or have long

stationary phases (*M. tuberculosis*); in both cases, a prolonged haploid phase exists during which HR is impossible. A functional NHEJ system is also found in *Pseudomonas*, however is not found in the common laboratory strain *E. coli* K12. The polymerases found in bacterial systems prefer to incorporate NTPs, which differs from eukaryotic systems (Hefferin & Tomkinson 2005; Pitcher *et al*. 2007; Alonso *et al*. 2013; Bowater & Doherty 2006).

Bacteriophage Mu encodes a protein Gam with homology to Ku, but it is not known to be involved in NHEJ; however it is thought to be the functional counterpart of Ku. NHEJ helps to circularise Mycobacteriophage genomes, which is required for rolling circle recombination and also protects the linear genome from exonucleolytic degradation (Pitcher et al 2007; di Fagagna *et al*. 2003).

NHEJ is also found in a few archaea and is homologous to bacterial systems. *Methanocella paludicola* has a Ku homodimer, DNA ligase, DNA polymerase and phosphoesterase (PE). PE regulates polymerase activity which preferentially incorporates rNTPs. Both bacterial and archaeal polymerases use strand displacement synthesis (Bartlett et al 2016). The archaeon *Haloferax volcanii* has a Mre11-Rad50 complex and like most archaea, does not encode a Ku homologue. The system in *H. volcanii* requires RadA (Rad51 homologue) and resembles micro-homology end joining (Delmas *et al*. 2009).

1.5 Alternative Forms of DNA Replication

1.5.1 Recombination-Dependent Replication

1.5.1.1 *Haloferax volcanii*

Homologous recombination can be associated with origin-independent replication, where recombination intermediates known as D-loops (Figure 1.17A) are remodelled into replication forks to initiate DNA replication (Kogoma 1997). Recombination-dependent replication (RDR) was proposed by Hawkins *et al*. (2013), as a mechanism for *H. volcanii* to replicate its chromosome when all origins are deleted. It was found that RadA (homologue

of bacterial RecA and eukaryotic Rad51) becomes essential in a strain lacking all chromosomal origins, but is not essential in a wild-type strain (Figure 1.17B). This indicates that homologous recombination is now an essential process in an origin-deleted strain. Deletion of RadA has previously been shown to make the strain deficient in homologous recombination (Woods & Dyall-Smith 1997). RDR may be more efficient at replicating the genome due to the fact that it can theoretically be initiated from anywhere in the genome, rather than origin-dependent replication which is limited to specific sites on the genome for initiation (Figure 1.17C). However, this also creates the possibility of many recombination intermediates within the genome, which may then be difficult to separate into individual chromosomes (Michel & Bernander 2014).

It is thought that RDR initiates from recombination intermediates such as D-loops, possibly formed by the collapse of replication forks or from other causes of DNA damage. However, it is also possible that RDR initiates from R-loops, which are RNA:DNA hybrids that are structurally similar to D-loops but are due to abortive transcription. R-loops are known to initiate DNA replication in mitochondria (Lee & Clayton 1998). RNaseH mutants of *E. coli* can undergo origin-independent replication, since R-loops are not destroyed following the removal of the RNA polymerase and can be converted into DNA replication forks (Kogoma 1997). This process is dependent on the polymerase PriA and is prone to mutagenesis of the genome. R-loops have also been observed to be involved in unscheduled origin-independent replication in yeast (Gowrishankar 2015; Michel & Bernander 2014; Stuckey *et al*. 2015). Novel replication origins have been observed in human cancer cells at highly transcribed oncogenes, which indicates initiation through the use of R-loops (Macheret & Halazonetis 2018). While it is possible that DNA replication in *H. volcanii* may initiate from R-loops, it has been shown that none of the RNaseH genes are mutated and levels of transcription are unchanged in origin-deleted strains (Hawkins *et* al. 2013; Thorsten Allers, personal communication). It was suggested by Michel and Bernander (2014) that *H. volcanii* may be using dormant origins when all its chromosomal origins are deleted, similar to observations made in the closely related *Haloferax mediterranei* (Yang *et al*. 2015), however this was not seen in the study by Hawkins *et al*. (2013).

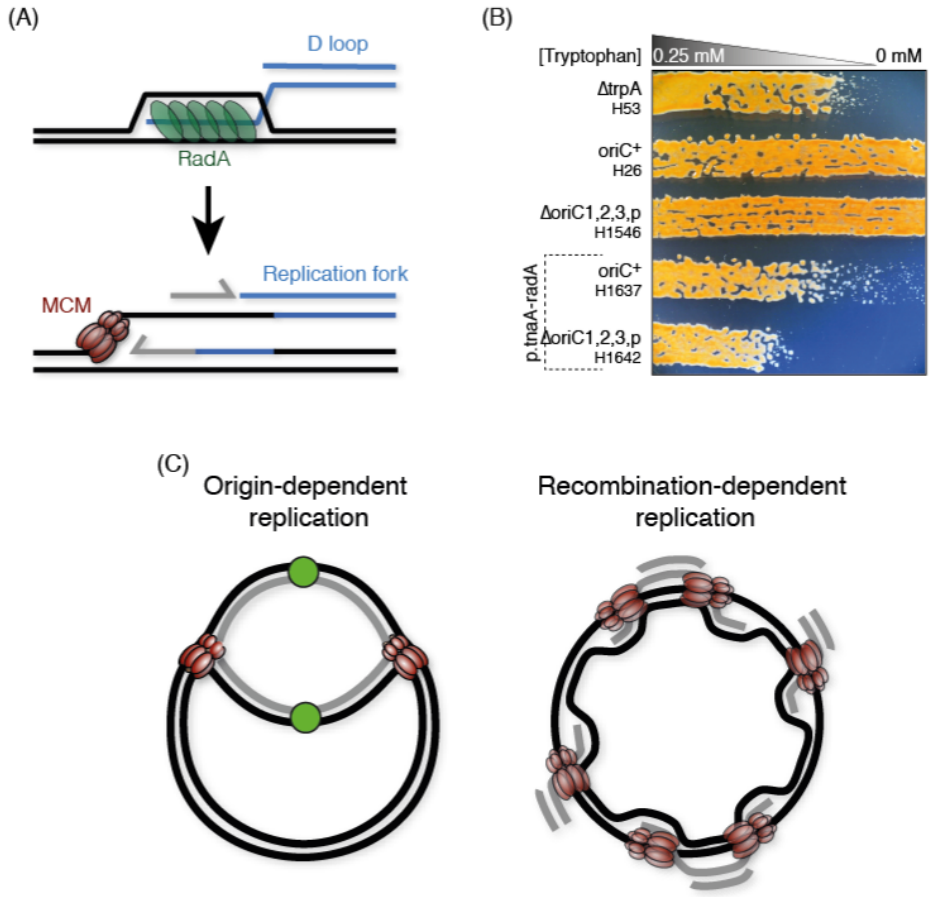


Figure 1.17 **Recombination dependent replication in *Haloferax volcanii*. (A)** Homologous recombination is able to form a D-loop catalysed by RadA, which can be used to form a replication fork. **(B)** Tryptophan gradient of wild-type and origin-deleted strains with and without *radA* under the control of a tryptophan inducible promoter showing RadA is essential in origin-deleted strains. Adapted from Hawkins *et al.* (2013). **(C)** Recombination-dependent replication can be initiated from anywhere on the genome, potentially allowing faster DNA replication than origin-dependent replication.

1.5.1.2 T4 Bacteriophage

The bacteriophage T4 is able to replicate in an origin-dependent manner using R-loops, as well as by recombination-dependent replication using D-loops for initiation. There are many common characteristics between replication initiation, DNA repair, recombination and replication fork restart, for example recombination proteins are used in origin-based replication and DSB repair mechanism in T4 bacteriophage, which is closely related to RDR. The T4 bacteriophage genome is 169 kb in size with five replication origins. At some T4 replication origins, persistent R-loops are formed in the DUE that are processed by viral RNaseH to generate free 3' ends, that are used to initiate

leading strand synthesis. Primase is thought to be involved in initiation at other origins and is used to prime discontinuous lagging strand synthesis. T4 bacteriophage encodes many of its own proteins that are required for DNA replication from R-loops, these include DNA polymerase, polymerase clamp, clamp loader, single-stranded DNA binding protein Gp32 and replicative helicase Gp41. A helicase loading protein, type II topoisomerase, and two other helicases, UvsW and Dda are also encoded in the T4 genome. Some of these proteins are also used in recombination-dependent replication.

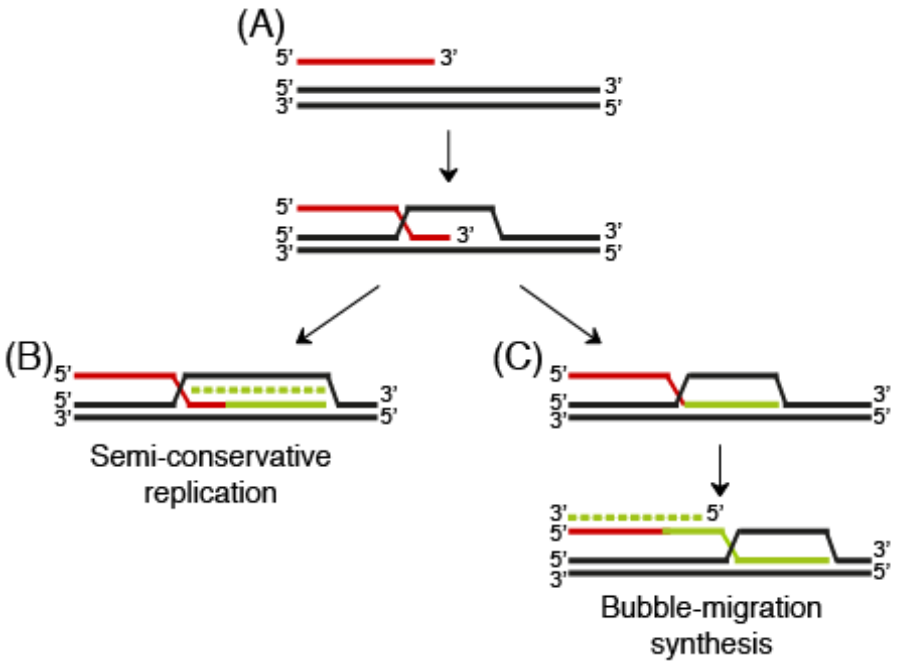


Figure 1.18 **Two types of RDR in T4 Bacteriophage. (A)** Strand invasion occurs and a D-loop is formed. **(B)** Semi-conservative replication. Lagging strand synthesis occurs on the displaced D-loop strand. **(C)** Bubble-migration synthesis. The nascent leading strand is released from the D-loop structure as it is synthesised where lagging strand synthesis can the occur. Red indicates invading DNA. Solid green indicates new leading strand replication and new lagging strand replication is shown in dashed green. Adapted from Kreuzer & Brister (2010).

RDR in T4 bacteriophage is initiated from D-loop structures that are generated by recombination machinery, they are formed via invading 3' ends of DNA. Infecting T4 DNA is linear and its ends are preferential sites for homologous recombination, but single-strand annealing intermediates may also be used to initiate RDR. Helicases unwind DNA ends to form ssDNA regions that are coated with single-stranded DNA binding protein Gp32. UvsY loads UvsX, a

Rad51/RecA homologue, onto Gp32-coated DNA, and the UvsW helicase stabilises the D-loop structure by promoting branch migration. Conservative replication is achieved using bubble-migration synthesis, where a single-stranded copy is generated and lagging strand synthesis occurs outside of D-loop (Figure 1.18). A conventional semi-conservative replication fork can also be generated where lagging strand synthesis occurs on the displaced DNA strand within the D-loop (Kreuzer & Brister 2010).

1.5.2 Break-Induced Replication

Break-induced replication (BIR) is a form of recombination-dependent replication that is found in yeast. Similar to homologous recombination, end resection of DNA occurs and the 3' ssDNA tails formed are coated with Rad51, which then undergo strand invasion with a homologous DNA strand to form a D-loop. In BIR, a uni-directional replication fork is then established and DNA synthesis occurs (Llorente *et al*. 2008; Malkova & Ira 2013). Unlike normal DNA replication, BIR proceeds via a bubble-like replication fork that is dependent on the Pif1 helicase and Polδ. BIR is highly mutagenic to the cells resulting in high numbers of mutations including genome rearrangements, copy number variations, and accumulation of ssDNA, therefore BIR is considered very costly to the cell (Saini *et al*. 2013; Wilson *et al*. 2013; Byrd & Raney 2017). On the other hand, RDR in *H. volcanii* is believed to maintain genome integrity (Michel & Bernander 2014), suggesting that the mechanism used is not BIR.

1.5.3 Rolling Circle Replication

Rolling circle replication (RCR) is a form of unidirectional DNA replication where leading and lagging strand synthesis are uncoupled (del Solar *et al*. 1998). It is commonly used by prokaryotic plasmids and viruses to replicate their DNA, however there are slight differences between them (Novick 1998).

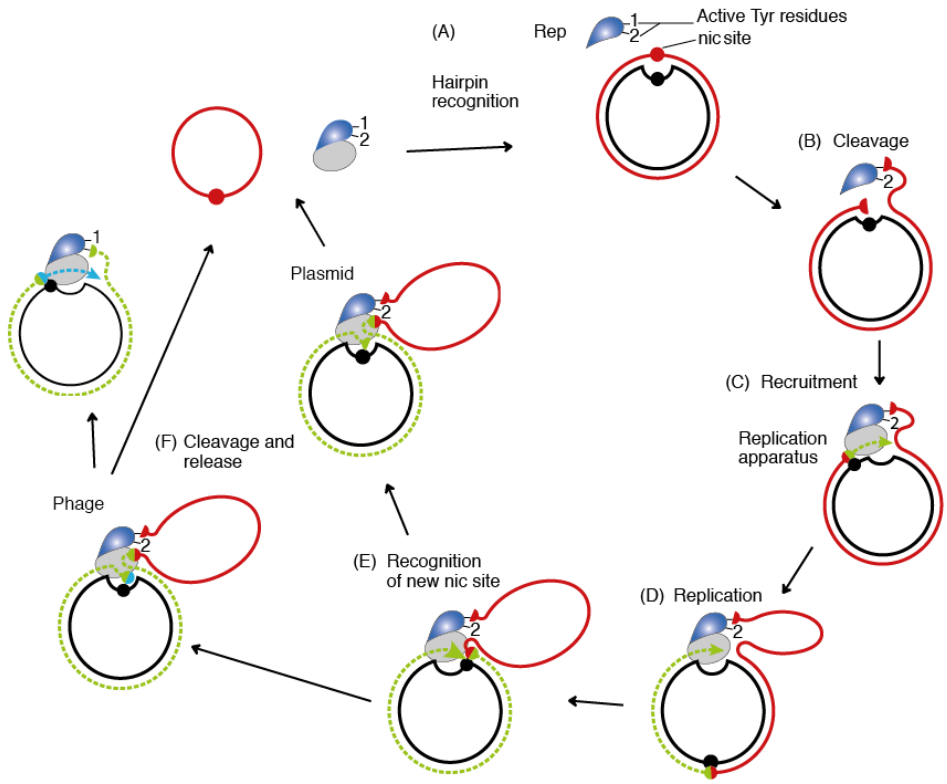


Figure 1.19 **Rolling circle replication. (A)** Initiator protein (Rep) binds to the double stranded origin region and **(B)** makes a single-stranded nick in the DNA. **(C)** DNA Replication apparatus is recruited and **(D)** the replication fork travels around the DNA producing a dsDNA molecule and a ssDNA molecule. **(E)** New nic site is recognised **(F)** causing cleavage and release. Phage/ virus replication continues to create more DNA molecules whereas plasmid replication is terminated. The ssDNA molecule produced can be made into dsDNA and the process can be repeated. Adapted from Chandler *et al*. (2013).

As shown in Figure 1.19, rolling circle replication is initiated by the binding of the initiator protein to a specific site (origin), which differs in plasmids and viruses. This only occurs if the DNA is supercoiled, which helps prevent unwanted binding of the initiator protein and therefore re-replication. The nick site is another region contained within the origin where the initiator protein nicks the DNA, and a Rep helicase as well as other components of the replication fork are recruited from host machinery. The replication fork advances using the 3' end generated at the nick site, displacing the parental leading strand of DNA whilst synthesising a new nascent leading strand. The initiator protein stays bound to the DNA through one of the two tyrosine residues located in its active site and travels with the replication fork, producing a looped displaced leading DNA strand.

This ssDNA displaced leading strand is released when the replication fork reaches the origin region; replication continues for around 10 bp past the nick site, which fully displaces the DNA strand (Novick 1998; Khan 2005).

Plasmids

Plasmids that use rolling circle replication were originally discovered in Gram-positive bacteria such as *Bacillus subtilis* and *Staphylococcus aureus*, however have since also been found in Gram-negative bacteria, for example *Thermus aquaticus, Helicobacter pylori* and *E. coli,* and in archaea (Guglielmetti & Parini 2007; Meijer *et al*. 1998; Chu *et al*. 2006; Bruand & Ehrlich 2000; Khan 1997; Khan 2005). Archaeal RCR plasmids have been isolated from a variety of species including *Pyrococcus abyssi*, *Thermococcus nautilus* and other hyperthermophiles (Erauso *et al*. 1996; Soler *et al*. 2007), and the haloalkaliphilic archaeon *Natronobacterium* (Zhou *et al*. 2004). These plasmids are normally small in size (<10 kb) and present in multiple copies within their hosts. They can carry a variety of genes including antibiotic and heavy metal resistance genes, mobilisation and transfer genes, and IS-elements and transposons (Khan 2005). Some of these genes allow them to be mobilised during horizontal gene transfer, allowing them to spread throughout a population (Fernández-López *et al*. 2014). To proliferate successfully, RCR plasmids must be able maintain a stable hereditary relationship with the host and therefore their replication is tightly controlled (Novick 1998).

The initiator protein in plasmids is known as Rep and binds to a sequence-specific site in the double-stranded origin (DSO) on the plasmid; both of these have homology to ssDNA bacteriophages in *E. coli*. The binding generates a sharp bend or hairpin in the DNA where the nick site is located, which is also within the DSO region. The nick site is highly conserved, however the sequence-specific binding site is less conserved and therefore serves as a source of specificity for plasmid replication. Rep nicks the DNA, becomes covalently attached though a tyrosine residue, and recruits a DNA helicase. This method of initiation produces a 3' DNA end that host polymerases can use to initiate DNA replication, and therefore does not require an RNA primer. The DNA is unwound

and in bacteria, coated in single-stranded binding protein (SSB) to protect the ssDNA strand. Leading strand synthesis continues until the old leading strand has been fully displaced. A combination of cleavage and ligation events occur and a relaxed, closed circular dsDNA molecule is produced alongside a ssDNA molecule that is the displaced leading strand. The Rep protein is then released and the ssDNA is converted to a dsDNA molecule via initiation of lagging strand synthesis at a single-stranded origin (SSO). SSO are non-coding regions that are strand specific and are subject to extensive secondary structures. Synthesis of the lagging strand is achieved by the recruitment of host RNA polymerase to produce an RNA primer, which is then extended by host DNA polymerase. The DNA is then ligated by host DNA ligase and supercoiled by host DNA gyrase, along with the dsDNA molecule already produced (Khan 2005; del Solar 1998; Ruiz-Masó *et al*. 2015). This process results in two supercoiled plasmids, each with one old and one nascent DNA strand.

Archaeal RCR plasmids use the same mechanism and have similar binding sequences in their DSO regions to bacteria. Their Rep proteins are also related to bacterial IS-elements that use RCR for transposition (Soler *et al*. 2007; Soler *et al*. 2011). The plasmid found in *Haloferax volcanii*, pHV2, contains a homologue of the RepH gene, suggesting pHV2 can replicate using RCR (Norais *et al*. 2007).

Viruses

Viral rolling circle replication mechanisms are similar to the plasmid RCR mechanisms describes above, however they do not necessarily use a Rep protein as an initiator, and may have ssDNA genomes rather than dsDNA. There is also a difference in their termination, since plasmids control their copy number within their host cells but viruses do not, since the latter aim to produce as many progeny as possible before cell lysis. Viruses, including bacteriophage, do not fully terminate RCR. Instead, the DNA polymerase continues to travel around the original DNA strand, producing (and displacing) many copies of the leading strand It is thought that this is achieved via the initiator protein switching

between the two tyrosine residues in the active site, so the ssDNA molecule can be released while the initiator stays bound (Novick 1998). The long multimer of displaced leading strand DNA is then either packaged into viral particles as ssDNA, or is converted into dsDNA and then packaged as either ssDNA or dsDNA, dependent on the virus genome. One of the most well-characterised bacteriophages, lambda from *E. coli*, can use rolling circle replication (sigma mode) as well as the alternative (origin-dependent) theta mode (Taylor & Wegrzyn 1995; Konopa *et al*. 2000). Eukaryotic geminivirus is known to use recombination-dependent replication as well as RCR (Rizvi *et al*. 2015). The halophilic archaeal *Natrinema* virus SNJ1, a member of the family *Sphaerolipoviridae*, has been found to use RCR to replicate its genome. The *repA* gene it encodes was found to be essential and most closely related to bacterial transposases, indicating evolutionary connections between archaeal virus replication and other mobile-genetic elements (Wang *et al.* 2018).

Transposons

Transposons normally use a 'cut-and-paste' mechanism, however the eukaryotic transposon category known as *Helitron* have been found to use rolling circle replication to transpose. *Helitron* transposons were first discovered in *Arabidopsis thaliana* (Thale cress), *Oryza sativa* (Asian rice) and *Caenorhabditis elegans* (nematode worm), and have since been discovered in a wide variety of eukaryotic organisms. These transposons encode their own Rep-like protein and can also encode their own single-stranded DNA binding protein, but depend upon the host for all other proteins required for DNA replication (Kapitonov & Jurka 2001). A model for their replication has been proposed, which is based on the model for bacterial transposons that also use RCR. Both *Helitron* transposons and the bacterial transposon IS91 were found to produce a dsDNA molecule and a circular ssDNA molecule, although the mechanisms are not yet fully understood (Kapitonov & Jurka 2007; del Pilar Garcillán-Barcia *et al*. 2001).

1.5.4 Mitochondrial DNA Replication

Mitochondria are known to have bacterial ancestors, namely *Alphaproteobacteria*, and are most closely related to *Rickettsiales* which are obligate intracellular parasitic bacteria (Gray 2012). However, mitochondria use a different method of replicating their small circular genome (~16.5 kb in mammals) and are known not to use recombination-based methods. Many mitochondrial DNA (mtDNA) molecules are known to contain triple stranded DNA regions, which presumably serve as origins of replication. These regions can form D-loops to initiate replication. In mammals, replication then occurs through the strand displacement mechanism, which has uncoupled leading and lagging strand synthesis. Around two thirds of the leading strand is synthesised before the lagging strand begins. It was assumed that the long region of ssDNA produced by this method would be coated in RPA single-strand binding protein, similar to a standard replication fork, to protect it from potential DNA damage. However, it has since been discovered that RNA anneals to the lagging strand until it is displaced by DNA synthesis. This is called the bootlace model, which helps mitochondria avoid the use of repair mechanisms such as NHEJ, which could introduce mutations into mtDNA. Synthesis of mtDNA is very slow compared to normal DNA synthesis but this is thought to improve fidelity, avoid collisions between replication and transcription (therefore preventing fork collapse) and protect against viruses due to less abundant dNTPs. Conventional DNA synthesis, where leading and lagging strand synthesis is coupled, has been noted in *Drosophila* mitochondria. Origin-independent mechanisms also exist in yeast and plant cell mitochondria, which use recombination-dependent replication and rolling circle replication to synthesise new mtDNA strands (Holt & Reyes 2012; Holt & Jacobs 2014; Bailey & Doherty 2017; Chen & Clark-Walker 2017). Rolling circle replication has also been seen in *C. elegans* mtDNA replication (Lewis *et al*. 2015).

1.6 Aims

The aims of this study are to further understand recombination-dependent replication in *Haloferax volcanii* and to investigate the role of DNA replication machinery for this purpose. Firstly, it is necessary to delete the remaining known replication origins from the cell, which are found on the mega-plasmids pHV1 and pHV3. This will reveal whether the cell can survive with no origins present throughout the genome (rather than just on the main chromosome), and whether this leads to any change in growth rates. Secondly, the DNA replication machinery of *H. volcanii*, namely the proteins MCM and GINS, will be analysed for their possible roles in origin-independent DNA replication versus origin-dependent replication. Together these approaches will enable greater understanding of DNA replication in *Haloferax volcanii* of both origin-dependent and recombination-dependent replication, as well as investigating possible alternative methods of DNA replication.

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	Parent	Genotype	Notes
H26	H18	*ΔpyrE2*	Constructed by TA (Allers *et al*. 2004)
H53	(H26)	*ΔpyrE2, ΔtrpA*	Constructed by TA (Allers *et al*. 2004)
H112	(H26)	*ΔradA*	Constructed by GN (Norais *et al*. 2007)
H678 (DS2)	n/a	*Wild-type – pHV2+, Episomal pHV4*	NCIMB
H2047	H1611	*ΔpyrE2, Nph-pitA, Δmrr, cdc48-ct, ΔtrpA*	Constructed by TA (Wardell *et al*. 2017)
H100	(H26)	*ΔpyrE2, ΔleuB, ΔhdrB*	Constructed by TA (Allers *et al*. 2004)

Parent:H53 (*ΔpyrE2, ΔtrpA*)			
Strain	**Parent**	**Genotype**	**Notes**
H99	(H53)	*ΔhdrB*	Constructed by TA (Allers *et al*. 2004)
H1539	(H53)	*Δori-pHV4::trpA+*	Constructed by TA (Hawkins *et al*. 2013)
H1802	(H53)	*ΔoriC1, ΔoriC2, ΔoriC3, Δori-pHV4, ΔhdrB, trpA+, p.tnaA-mcm+::hdrB+*	Constructed by TA (unpublished data)
H1804	(H53)	*ΔoriC1, ΔoriC2, ΔoriC3, Δori-pHV4*	Constructed by KP (unpublished data)
H1911	(H1608)	*trpA+, ΔoriC1, ΔoriC2, ΔoriC3, Δori-pHV4, ΔhdrB, p.tnaAM3-mcm+::hdrB+*	Chapter 5.3.1
H1941	(H53)	*Δorc10::trpA+*	Constructed by TA (unpublished data)
H1945	(H53)	*Δori-pHV1::trpA+*	Constructed by TA (unpublished data)
H1948	(H53)	*Δori-pHV1 Δorc10::trpA+*	Constructed by TA (unpublished data)
H2184	(H1948)	*ΔpHV1*	Chapter 3.3.3
H2475	(H1945)	*Δori-pHV1*	Chapter 4.3.2
H2477	(H1941)	*Δorc10*	Chapter 4.3.2
H2479	(H1948)	*Δori-pHV1 Δorc10*	Chapter 4.3.2

H2931	(H1911)	*trpA+, ΔoriC1, ΔoriC2, ΔoriC3, Δori-pHV4, ΔhdrB, p.tnaAM3-mcm+::hdrB+, tnaAΔEcoN*	Chapter 5.3.1
H3222	(H53)	*His_6 tag-ginS+-StrepII tag*	Chapter 5.3.2
H3640	H53	*pHV1::pyrE2+*	Chapter 4.3.2
H3644	H2475	*Δori-pHV1, pHV1::pyrE2+*	Chapter 4.3.2

Parent:H99 (*ΔhdrB*)			
Strain	**Parent**	**Genotype**	**Notes**
H1530	H99	*trpA+*	Constructed by TA (unpublished data)
H1794	(H1530)	*trpA+, p.tnaA-mcm+::hdrB+*	Constructed by TA (unpublished data)
H1904	(H1530)	*trpA+, ΔoriC1, ΔoriC2, ΔoriC3, Δori-pHV4, p.tnaAM3-mcm+::hdrB+*	Chapter 5.3.1
H2928	(H1904)	*trpA+, ΔoriC1, ΔoriC2, ΔoriC3, Δori-pHV4, p.tnaAM3-mcm+::hdrB+, tnaAΔEcoN*	Chapter 5.3.1

Parent:H678 (DS2 *wild-type – pHV2+, Episomal pHV4*)			
Strain	**Parent**	**Genotype**	**Notes**
H2218	(H678)	*ΔpyrE2, ΔtrpA*	Constructed by TA (unpublished data)
H2589	(H2218)	*ΔpyrE2, ΔtrpA, Δori-pHV4::trpA+*	Chapter 3.3.2

Parent:H1804 (*ΔpyrE2, ΔtrpA, ΔoriC1, ΔoriC2, ΔoriC3, Δori-pHV4*)			
Strain	**Parent**	**Genotype**	**Notes**
H1939	(H1804)	*Δori-pHV1::trpA+*	Constructed by TA (unpublished data)
H1943	(H1804)	*Δori-pHV1 Δorc10::trpA+*	Constructed by TA (unpublished data)
H1949	(H1804)	*Δorc10::trpA+*	Constructed by TA (unpublished data)
H2175	(H1939)	*Δori-pHV1*	Chapter 4.3.2
H2183	(H1949)	*ΔpHV1*	Chapter 4.3.2
H2481	(H1939)	*Δori-pHV1*	Chapter 4.3.2
H2483	(H1949)	*Δorc10*	Chapter 4.3.2
H2485	(H1943)	*Δori-pHV1 Δorc10*	Chapter 4.3.2
H3225	(H2175)	*His_6 tag-ginS+-StrepII tag*	Chapter 5.3.2

Parent:H2047 (*ΔpyrE2, Nph-pitA, Δmrr, cdc48-ct, ΔtrpA*)			
Strain	**Parent**	**Genotype**	**Notes**
H2594	H2047	*mcm+::[His_6 tag-mcm+-StrepII tag pyrE2+]*	Chapter 5.3.2
H2660	H2594	*His_6 tag-mcm+*	Chapter 5.3.2
H2839	H2047	*ginS+::[His_6 tag-ginS+-StrepII tag pyrE2+]*	Chapter 5.3.2
H2840	H2047	*ginS+::[His_6 tag-ginS+-StrepII tag pyrE2+]*	Chapter 5.3.2

H2962	H2839	*ginS+-StrepII tag*	Chapter 5.3.2
H2963	H2839	*His_6 tag-ginS+*	Chapter 5.3.2
H2964	H2840	*His_6 tag-ginS+-StrepII tag*	Chapter 5.3.2
H3628	(H2962)	*ginS+-StrepII tag, His_6 tag-mcm+*	Chapter 5.3.2

Integrated pHV3 strains				
Strain	**Parent 1**	**Parent 2**	**Genotype**	**Notes**
H2461	(H53)	(H2240)	*ΔpyrE2, leuB+::[pHV3]*	Chapter 3.3.3
H3512	(H2461)	n/a	*ΔpyrE2, leuB+::[pHV3], Δori-pHV3 Δorc6*	Chapter 3.3.3
H3718	(H2461)	n/a	*ΔpyrE2, leuB+::[pHV3], ΔtrpA*	Chapter 3.3.3
H3794	(H1804)	(H3726)	*ΔpyrE2, ΔtrpA, ΔoriC1, ΔoriC2, ΔoriC3, Δori-pHV4, leuB+::[pHV3]*	Chapter 3.3.3
H3796	(H2175)	(H3728)	*ΔpyrE2, ΔtrpA, ΔoriC1, ΔoriC2, ΔoriC3, Δori-pHV4, Δori-pHV1, leuB+::[pHV3]*	Chapter 3.3.3
H3798	(H2183)	(H3730)	*ΔpyrE2, ΔtrpA, ΔoriC1, ΔoriC2, ΔoriC3, Δori-pHV4, ΔpHV1, leuB+::[pHV3]*	Chapter 3.3.3
H3800	(H2184)	(H3733)	*ΔpyrE2, ΔtrpA, ΔpHV1, leuB+::[pHV3]*	Chapter 3.3.3
H3954	(H3800)	n/a	*ΔpyrE2, ΔtrpA, ΔpHV1, leuB+::[pHV3], Δori-pHV3::trpA+*	Chapter 3.3.3

TA = Thorsten Allers, KP = Katarzyna Ptasinska, GN = Greg Ngo. Parental strains in brackets indicate intermediate strains were generated during construction but these are not listed. [ ] indicated integrated plasmid.

Escherichia coli strains

Table 2.2 ***E. coli* strains used in this study**

Strain	**Genotype**	**Notes**
XL1-Blue MRF'	*endA1, gyrA96* (NalR), *lac* [F' proAB *lacIqZΔM15 tn10* (TetR)], *Δ(mcrA)183, Δ(mcrCB-hsdSMR-mrr)173, recA1, relA1, supE44, thi-1*	Standard cloning strain for blue/white screening using pBluescript based plasmids. Tetracycline resistant. Restriction endonuclease and recombination deficient. *dam+*. From Stratagene.
N2338 (GM121)	*F-, ara-14, dam-3, dcm-6, fhuA31, galK2, galT22, hsdR3, lacY1, leu-6, thi-1, thr-1, tsx-78*	*dam- dcm-* mutant for preparing DNA for *Haloferax volcanii* transformations (Allers *et al*. 2004). From RG Lloyd.

2.1.1.1 *Haloferax volcanii* Parental Strains

Several parental strains of *Haloferax volcanii* were used for the construction of strains in this study. They are described below:

H26 (*ΔpyrE2*)

H26 is the wild-type laboratory strain, derived from the wild-type isolate DS2. H26 was made by curing DS2 of the mega-plasmid pHV2, to allow for the use of its origin in episomal vectors, and deletion of the uracil biosynthesis gene *pyrE2* (Allers *et al*. 2004, Wendoloski *et al*. 2001).

H53 (*ΔpyrE2 ΔtrpA*)

H53 is deleted for the auxotrophic selection markers *ΔpyrE2* (uracil) and *ΔtrpA* (tryptophan), derived from H26. The *trpA* deletion is used in the selection of integrated *trpA* marked gene replacements and deletions (Allers *et al*. 2004).

H99 (*ΔpryE2 ΔtrpA ΔhdrB*)

H99 contains the auxotrophic selection markers *ΔpyrE2* (uracil), *ΔtrpA* (tryptophan) and *ΔhdrB* (thymidine). The *ΔhdrB* is an alternative to *ΔtrpA* which is utilised in the selection of integrated *hdrB* marked gene replacement or deletion constructs for genes that are difficult to delete (Allers *et al*. 2004).

H678 (*wild-type, pHV2+, Episomal pHV4*)

H678 is also known as *Haloferax volcanii* DS2 and NCIMB 2012. Isolated by Mullakhanbhai, and Larsen (1975) from the Dead Sea. H678 contains the pHV2 plasmid and the mega-plasmid pHV4 is episomal.

H2047 (*ΔpyrE2 Nph-pitA Δmrr cdc48d-ct ΔtrpA*)

H2047 was generated from the strain H1424 via H1611. H1424 (*ΔpyrE2, Nph-pitA, Δmrr, cdc48d-ct,Δ hdrB*) was derived from H1209 (*ΔpyrE2, Nph-pitA, Δmrr, ΔhdrB*) which is a strain produced for improved protein overexpression

and purification (Allers *et al*. 2010, Stroud *et al*. 2012). The *pitA* gene of *Haloferax volcanii,* which encodes a histidine rich linker region causing it to co-purify with His_6 tagged recombinant proteins, has been previously replaced by a non-histidine rich ortholog from *Natronomonas pharaonis* (Allers et al 2010) *Nph-pitA*. The *mrr* gene encodes a restriction enzyme which cuts methylated 5'-GATC-3' sequences. This causes a problem as most plasmids are passaged through *E. coli* which contains a Dam methylase which methylates 5'-GATC-3' sequences and so the DNA is mostly destroyed by the Mrr enzyme in *H. volcanii* making transformations inefficient. Deleting the *mrr* gene allows the strain to be transformed without the need to passage the DNA through an *E. coli dam*$^-$ mutant first, which reduces the risk of mutations due to impaired DNA mismatch repair in *E. coli dam*$^-$ strains. This strain also contains a truncated version of the *cdc48d* gene, *cdc48d-ct*, which is missing the 3' end which encodes its histidine rich C-terminus that causes it to co-purify with His_6 tagged recombinant proteins. H1424 was used to make H1611 by deleting the *trpA* gene to allow for selection of integrated *trpA* marked gene replacements and deletions. H1611 was then transformed with a linear fragment containing *hdrB* to make the strain *hdrB*+, generating H2047 (Wardell *et al*. 2017).

2.1.2 Plasmids

Table 2.3 **Plasmids used in this study**

Name	Use	Notes
pBluescript II SK+	Standard *E. coli* cloning vector with functional blue/white screen capability	From Stratagene
pTA44	Genomic clone of *leuB*	Constructed by TA (Allers *et al*. 2004)
pTA52	*pyrE2* under the control of the ferredoxin promoter *p.fdx*	Constructed by TA (Bitan-Banin *et al*. 2003)
pTA95	*trpA* deletion construct	Constructed by TA (Allers *et al*. 2004)
pTA131	For making deletions in *ΔpyrE2* backgrounds	Constructed by TA (Allers *et al*. 2004)
pTA298	Contains *trpA* under control of the ferredoxin promoter *p.fdx*	Constructed by TA (Lestini *et al*. 2010)
pTA700	Contains inactive *leuB* allele *leuBΔ3'Bb*	Constructed by SH (unpublished data)

pTA1331	Deletion construct for *ori-pHV4::trpA+*	Constructed by TA (Hawkins *et al*. 2013)
pTA1369	Vector for tryptophan inducible gene expression under the control of *p.tnaA* with N-terminal His_6 tag and C-terminal StrepII tag, and *hdrB* selection	Constructed by TA (unpublished data)
pTA1392	Overexpression vector for tryptophan inducible gene expression under the control of *p.tnaA* with an N-terminal His_6 tag and a C-terminal StrepII tag	Constructed by RGM (Gamble-Milner, 2016)
pTA1404	Genomic clone of *mcm* including upstream and downstream flanking regions	Constructed by TA (unpublished data)
pTA1411	Gene replacement construct of *mcm* under the control of tryptophan inducible promoter *p.tnaA* with *hdrB* selection	Constructed by TA (unpublished data)
pTA1451	pTA1369 with tryptophan inducible promoter *p.tnaA* replaced with low activity tryptophan inducible promoter *p.tnaAM3*	Chapter 5.3.1
pTA1460	*p.tnaAM3-mcm::hdrB+* gene replacement construct	Chapter 5.3.1
pTA1484	Genomic clone of *ori-pHV1* and *orc10*	Constructed by TA (unpublished data)
pTA1494	*ori-pHV1* deletion construct with *trpA* selection	Constructed by TA (unpublished data)
pTA1495	*orc10* deletion construct with *trpA* selection	Constructed by TA (unpublished data)
pTA1496	*ori-pHV1 orc10* deletion construct with *trpA* selection	Constructed by TA (unpublished data)
pTA1509	Genomic clone of *ori-pHV3* and *orc6*	Chapter 3.3.1
pTA1520	*ori-pHV3* deletion construct	Chapter 3.3.1
pTA1521	*orc6* deletion construct	Chapter 3.3.1
pTA1522	*ori-pHV3 orc6* deletion construct	Chapter 3.3.1
pTA1553	pTA131 with *PciI* site removed	Chapter 5.3.2
pTA1557	*ori-pHV1* deletion construct	Chapter 4.3.1
pTA1558	*orc10* deletion construct	Chapter 4.3.1
pTA1559	*ori-pHV1 orc10* deletion construct	Chapter 4.3.1
pTA1560	For the expression of N-terminally His_6 tagged and C-terminally StrepII tagged genes under the control of their native promoter on the chromosome	Chapter 5.3.2
pTA1621	*Δadh2::leuBΔ5'Dd* gene replacement construct	Chapter 3.3.3
pTA1627	Intermediate construct for creating pTA1663	Chapter 5.3.2
pTA1655	Intermediate construct for creating pTA1663	Chapter 5.3.2
pTA1663	N-terminally His_6 tagged, C-terminally StrepII tagged *mcm* gene replacement construct	Chapter 5.3.2
pTA1667	pTA1560 with unique *NspI* site	Constructed by TA (unpublished data)
pTA1677	*pyrE2* under the control of the ferredoxin promoter *p.fdx* flanked by *BamHI* sites	Chapter 2.1.2.1
pTA1716	*ginS* genomic clone including *priS*	Chapter 5.3.2
pTA1730	*tnaA* deletion construct (partial)	Constructed by TA (unpublished data)

pTA1776	Intermediate construct for creating pTA1791	Chapter 5.3.2
pTA1790	Intermediate construct for creating pTA1791	Chapter 5.3.2
pTA1791	N-terminally His_6 tagged, C-terminally StrepII tagged *ginS* gene replacement construct	Chapter 5.3.2
pTA1798	*ori-pHV3* deletion construct with *trpA* selection	Chapter 3.3.1
pTA1799	*orc6* deletion construct with *trpA* selection	Chapter 3.3.1
pTA1800	*ori-pHV3 orc6* deletion construct with *trpA* selection	Chapter 3.3.1
pTA1878	Genomic clone of pHV1 region to insert marker onto pHV1	Chapter 4.3.2
pTA1914	pHV1 region with *pyrE2* marker	Chapter 4.3.2
pCN26	Genomic clone of *ori-pHV3*	Constructed by CN (Norais *et al*. 2007)
pNPM-tnaAM3-HfxMCM	Tryptophan inducible promoter *p.tnaAM3* with 50% reduced activity	Courtesy of SMN (unpublished data)

TA = Thorsten Allers, SH = Sam Haldenby, CN = Cédric Norais, RGM = Rebecca Gamble-Milner, SMN = Stuart MacNeill (University of St. Andrews).

2.1.2.1 Plasmid Construction

All plasmids were verified by restriction digest and sequencing, and all plasmid numbers given throughout this study refer to plasmids grown in *E. coli* XL1-Blue. The following abbreviations apply in all figures: *AmpR* (Ampicillin resistance gene, *E. coli*), *ColE1* (*E. coli* origin of replication), *f1(+) ori* (*E. coli* origin of replication), *lacZ* (ß-galactosidase used for blue/white selection in *E. coli*), MCS (multiple cloning site), *p.lac* (promoter for *lacZ*, *E. coli*), *p.fdx* (ferredoxin promoter, *H. volcanii*), *p.tnaA* (tryptophan-inducible promoter, *H. volcanii*), *p.tnaAM3* (tryptophan-inducible promoter with 50% reduced activity), *hdrB* (thymidine biosynthesis, *H. volcanii*), *trpA* (tryptophan biosynthesis, *H. volcanii*), *pyrE2* (uracil biosynthesis, *H. volcanii*), His_6 tag (hexahistidine tag), StrepII tag (streptavidin-binding tag), *t.L11e* (L11e rRNA terminator, *H. volcanii*), *t.syn* (synthetic terminator, *H. volcanii*).

Genomic Clone Construction

Genomic clones allow easier genetic manipulation compared to the whole genome and reduce the possibility of PCR errors in downstream cloning. Genomic clones were constructed using the standard *E. coli* cloning vector pBluescript II SK+ (Stratagene) (Figure 2.1) and grown in *E. coli* XL1-Blue to

allow for blue/white screening. pBluescript II SK+ contains the multiple cloning site (MCS) within *lacZ* (ß-galactosidase gene). If *lacZ* is disrupted colonies will be white when cells are plated onto media containing X-gal, whereas if there has been no insertion of DNA into the plasmid the *lacZ* gene will remain intact and therefore the colonies will be blue. Genomic clones are constructed by cloning a region of interest either by restriction digest or PCR, including ~1 kb of upstream and downstream regions, into the MCS of pBluescript II SK+.

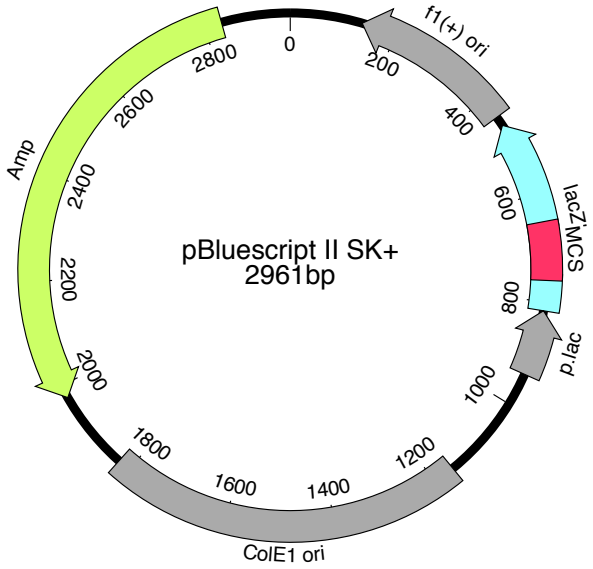


Figure 2.1 **pBluescript II SK+.** Standard *E. coli* cloning vector with the multiple cloning site (MCS) located within *lacZ* allowing for blue/white screening in *E. coli* (from Stratagene).

Gene Deletion and Replacement Plasmid Construction

For gene deletions and replacements constructs were cloned into the vector pTA131 (Figure 2.2) which is a derivative of pBluescript II SK+ that also contains a *pyrE2* marker. This encodes the final enzyme used in the uracil biosynthesis pathway, orotate phosphoribosyl transferase (Allers *et al*. 2004).

For deletion constructs, upstream and downstream flanking regions of the region/gene to be deleted were inserted into the MCS of pTA131 within the *lacZ* gene. For gene replacements, the replacement region is inserted alongside upstream and downstream flanking regions. The *pyrE2* marker gene is used to

select for *H. volcanii* transformants (previously *ΔpyrE2*) that have taken up the plasmid when plated on selective media that is lacking in uracil as *ΔpyrE2* strains cannot grow without the addition of uracil to Hv-Ca media.

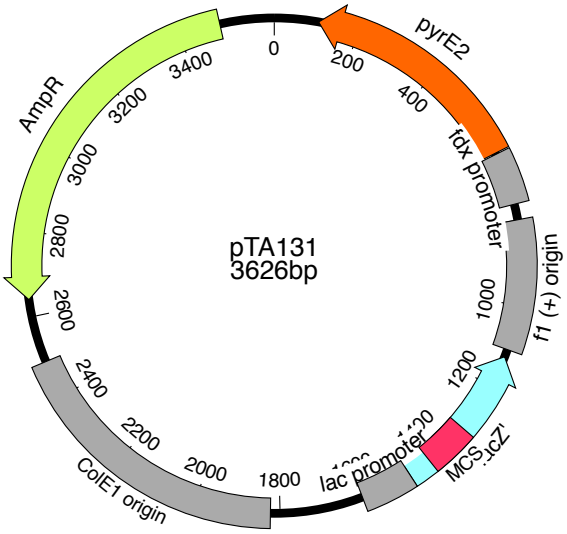


Figure 2.2 **pTA131**. Vector used for making gene deletions and replacements constructs for genetic manipulation of *H. volcanii*. A pBluescript II SK+ derivative containing the *pyrE2* marker for the uracil biosynthesis pathway allowing selection in *ΔpyrE2* strains (Allers *et al*. 2004).

Selectable Markers

Plasmids containing selectable markers for *Haloferax volcanii* were used to insert the selectable markers for *trpA* and *pyrE2* into deletion and gene replacement constructs.

The plasmid pTA298 (*p.fdx-trpA*) construct was constructed by Thorsten Allers in 2004 (Figure 2.3). It contains the *trpA* selectable marker which is flanked by two *BamHI* sites (Lestini *et al*. 2010).

To construct the plasmid pTA1677 (*p.fdx-pyrE2*), the plasmid pTA52 (Bitan-Banin *et al*. 2003) was digested with *XbaI* and ligated with a linker consisting of the annealed oligonucleotides BamLinkXba and XbaLinkBam using the *XbaI* sites. The linker contains a *BamHI* site which means that the *pyrE2* gene is flanked by two *BamHI* sites (Figure 2.4).

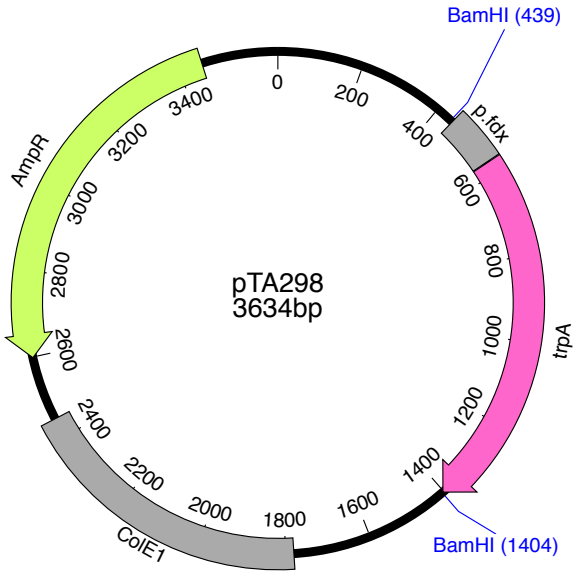


Figure 2.3 **pTA298**. *p.fdx-trpA* construct flanked by two *BamHI* sites (Lestini *et al*, 2010).

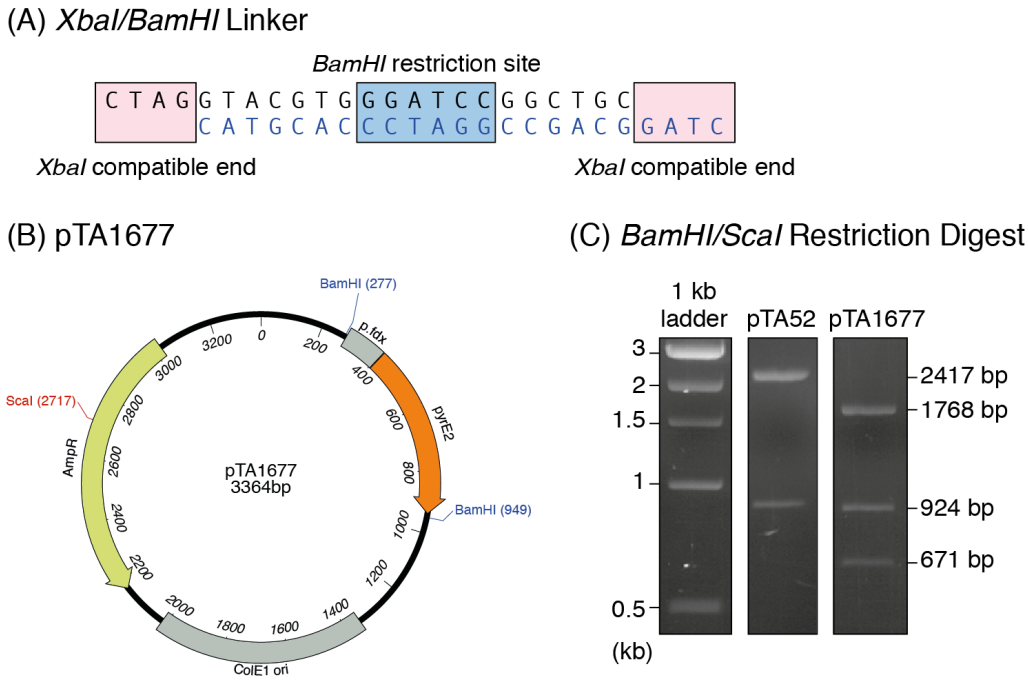


Figure 2.4 **pTA1677**. **(A)** *XbaI/BamHI* linker with *XbaI* compatible ends and internal *BamHI* site. Constructed by annealing BamLinkXba and XbaLinkBam oligos. **(B)** pTA1677 *p.fdx-pyrE2* construct flanked by two *BamHI* sites, also showing *ScaI* restriction site. **(C)** Restriction digest to confirm insertion of linker using *BamHI* and *ScaI*. Parental plasmid pTA52 is predicted to give two bands of 2417 bp and 924 bp, and pTA1677 is predicted to give three bands of 1768 bp, 924 bp and 671 bp. When digested both plasmids produced the correct band sizes.

2.1.3 Oligonucleotides

Table 2.4 **PCR primers used in this study**

Name	Sequence (5'-3')	Description	Used to construct
BamLinkXba	CTAGGTACGTGGGATCCGGCTGC	Annealed with XbaLinkBam to produce linker	pTA1677
XbaLinkBam	CTAGGCAGCCGGATCCCACGTAC	Annealed with BamLinkXba to produce linker	pTA1677
dAdh2UF	CCGGGgtACCTCCAGAAGACC	Amplification of *adh2* with the introduction of a *KpnI* site	pTA1688
dAdh2DR	CTCGTCTaGaTCGACCAGCAC	Amplification of *adh2* with the introduction of a *XbaI* site	pTA1688
doripHV3USBglR	CTTCCAGaTCtCAACCTCCCAACCACGCC	Amplification of *ori-pHV3* upstream flanking region introducing a *BglII* site	pTA1520 pTA1522
pBSR3	ACCCCAGGCTTTACACTTTATGC	Amplification of *ori-pHV3* upstream flanking region	pTA1520 pTA1522
doripHV3DSBglF	TCGTaGATCtTCGAATCCAAAACCGGG	Amplification of *ori-pHV3* downstream flanking region introducing a *BglII* site	pTA1520
doripHV3DSR	GACAGAGACGGACGTGTGTTAAGGG	Amplification of *ori-pHV3* downstream flanking region	pTA1520
dorc6USBglR	TGAGAGATcTTGAGGATATCCTCGTTCGC	Amplification of upstream flanking region of *orc6* introducing a *BglII* site	pTA1521
dorc6USEcoF	CACCGAAtTCGGAATCCCGATTCCCGACG	Amplification of upstream flanking region of *orc6* introducing an *EcoRI* site	pTA1521
dorc6DSBglF	CCGGAgATCtTCAAGCGAGTGCTCCTCCG	Amplification of downstream flanking region of *orc6* introducing a *BglII* site	pTA1521 pTA1522
pBSF2	TTAAGTTGGGTAACGCCAGGG	Amplification of downstream flanking region of *orc6*	pTA1521 pTA1522
mcmtagUSKpnF	AGAGGTAcCCGGCGGTCGGGCCGACGAGC	Amplification of upstream flanking region of *mcm* introducing a *KpnI* site	pTA1655
mcmtagUSNdeR	CTGCGCCATatgACAGGGGTATGTCTCCAATCGG	Amplification of upstream flanking region of *mcm* introducing a *NdeI* site	pTA1655
mcmtagGNcoF	CTGTGCcATGGCGCAGGCCCCCCAGAACC	Amplification of *mcm* gene introducing a *NcoI* site	pTA1627

mcmtagGNheR	GAGGACGctagc AGTCGCGCGG TACACGTCGC	Amplification of *mcm* gene removing stop codon and introducing a *NheI* site	pTA1627
mcmtagDSBamF	TGAGggaTCCT CGGCTGCTGC GCCCGCCC	Amplification of downstream region of *mcm* introducing a *BamHI* site	pTA1663
mcmtagDSXbaR	TGAGtCtAGAC TCGTCTTCTCC GGCGGCG	Amplification of downstream region of *mcm* introducing a *XbaI* site	pTA1663
oripHV3probeF	TCGCGCAAGT AAGACCGTCA GACACCACC	Amplification of internal region of *ori-pHV3*	-
oripHV3probeR	CGGCAGACCG GTCACGGATT CGTTGTTCG	Amplification of internal region of *ori-pHV3*	-
pHV1probeF	GAGGAATCGG GAATGGACTG G	Amplification of pHV1 region	-
pHV1probeR	GACACTCTGA GTATCCACGTC CG	Amplification of pHV1 region	-
FleuBpHV3chrF	CGGCTACTAC AAGAACGTCA TGCGCTTCC	For colony PCR to check integration of pHV3 onto the chromosome via presence of a functional *leuB* gene	-
FleuBpHV3chrR	GAGAACGACG AACCCAAGGA CCACCCG	For colony PCR to check integration of pHV3 onto the chromosome via presence of a functional *leuB* gene	-
NFleuBpHV3F	ACGGTCTTCG AGTTCGCGCA GGTCACG	For colony PCR to check integration of pHV3 onto the chromosome via presence of a non-functional *leuB* gene	-
NFleuBpHV3R	CGCGACAGAC TACGGCCGAG GAGCTGG	For colony PCR to check integration of pHV3 onto the chromosome via presence of a non-functional *leuB* gene	-
GinSF	CGGACGAAAT CGAGACCGCG GAGGAGG	Amplification of internal region of *ginS*	-
GinSR	CGCAGAGCCG GAGTTCGCAG CGTCGG	Amplification of internal region of *ginS*	-
mcmHisF	GCTCGTCGCG CTCGTCGTCGG CTTCTTCG	For colony PCR to check N-terminal His_6 tagged *mcm*	-
mcmHisR	TGGCCGGGCT TCGTGATGTA GTCGTCCGC	For colony PCR to check N-terminal His_6 tagged *mcm*	-
mcmStrepF	CGGCATGGAC CCAGAGACCG GCGAGTTCG	For colony PCR to check C-terminal StrepII tagged *mcm*	-

mcmStrepR	GATGGCGTCTTCGATGTTGCGGATGGCGG	For colony PCR to check C-terminal StrepII tagged *mcm*	-
gintagUSKpnF	TCGGggTACCTCGCCCGCAGCGCCCCTCG	Amplification of upstream region of *ginS* introducing a *KpnI* site	pTA1776
gintagUSNdeR	GAGGTCaTatgCGTTCATTCTGTCACCTTCTCGGC	Amplification of upstream region of *ginS* introducing a *NdeI* site	pTA1776
gintagGBspF	TGACAtcATGAACGTGGACGACCTCAGGAGCG	Amplification of *ginS* gene introducing a *BspHI* site	pTA1790
gintagGNheR	AAGGGTCTgctagcGTCGAGTCGCTCGGCCGCGCC	Amplification of *ginS* gene removing stop codon and introducing a *NheI* site	pTA1790
gintagDSBamF	AGAAgGAtCCTTCTTTCCGCCGCCTGAACAGG	Amplification of downstream region of *ginS* introducing a *BamHI* site	pTA1791
gintagDSXbaR	ACGCtCTaGACGAAATCGGCGCGAGCGTCACCG	Amplification of downstream region of *ginS* introducing a *XbaI* site	pTA1791
priSintF	GAGATGGAAGAAGACGCCGCGCTCTCG	For colony PCR to check N-terminal His_6 tagged *ginS*	-
ginSintR	TCCGCGGTCTCGATTTCGTCCGTGAGCCG	For colony PCR to check N-terminal His_6 tagged *ginS*	-
ginSintF	ACTCCGACGCTGCGAACTCCGGCTCTGC	For colony PCR to check C-terminal StrepII tagged *ginS*	-
bcpintR	CCGACGAGGTAGGAGTTGCGGAACACG	For colony PCR to check C-terminal StrepII tagged *ginS*	-
orc6probeF	GATGTCCGGGTCCTCGAGCAGGCGGTGGG	Amplification of on internal region of *orc6*	-
orc6probeR	GAAGGCCGTCCTCTACTCGCTGGTCCTCC	Amplification of on internal region of *orc6*	-
pHV1mkrprobeF	GACACACTCGGCTCGAATCGGGTCAGCG	To check insertion of marker into pHV1 region	-
pHV1mkrprobeR	CATGGTCAGTTGGTCGCGGGGTCATCGGG	To check insertion of marker into pHV1 region	-

ori_pHV42_F	GTCACAGAG AACCACGGA GAACG	Amplification of region internal to *ori-pHV4*
ori_pHV42_R	ATACTTAGG CGGGAACAC ACCCC	Amplification of region internal to *ori-pHV4*
pHV3probe2F	CGGCCGCGA CTACATGTAC TGGAACAAC G	Amplification of region internal to *narH* gene for a pHV3 specific probe
pHV3probe2R	GGTGTCGGT CAGCATATC CAGCAGTTC GC	Amplification of region internal to *narH* gene for a pHV3 specific probe

Lowercase letters indicate mismatches with the wild-type gene. Novel restriction sites are underlined.

2.1.4 Chemicals and Enzymes

All enzymes were purchased from New England BioLabs (NEB) and all chemicals from Sigma, unless otherwise stated. Appropriate methods give details of specific chemicals, buffers and solutions.

2.1.5 Media

Haloferax Media

Media are sterilised by autoclaving for 1 minute at 121ºC. Liquid media are stored at room temperature in the dark and solid media in the form of plates are stored at 4ºC in sealed bags as to prevent desiccation. All plates were dried for at least 30 mins before use.

30% salt water (SW): 4 M NaCl, 148 mM $MgCl_2.6H_2O$, 122 mM $MgSO_4.7H_2O$, 94 mM KCl, 20 mM Tris.HCl pH7.5.

18% salt water (SW): Made with 30% SW, 3 mM $CaCl_2$.

$CaCl_2$ added after autoclaving.

Trace elements: 1.82 mM $MnCl_2.4H_2O$, 1.53 mM $ZnSO_4.7H_2O$, 8.3 mM $FeSO_4.7H_2O$, 200 μM $CuSO_4.5H_2O$.

Filter sterilised and stored at 4ºC.

Hv-Min Salts: 0.4 M NH_4Cl, 0.25 M $CaCl_2$, 8% v/v of trace element solution.

Stored at 4ºC.

Hv-Min carbon source: 10% DL-lactic acid Na_2 salt, 8% succinic acid Na_2 salt·$6H_2O$, 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 $CaCl_2$, 8.3% v/v of trace elements, 615 μg/ml thiamine 77 μg/ml biotin.

KPO_4 Buffer: 308 mM K_2HPO_4, 192 mM KH_2PO_4 pH7.0.

Hv-YPC agar: 1.6% agar (Bacto), 18% SW, 1 x YPC, 3mM $CaCl_2$.

Microwaved without 10 x YPC to dissolve agar. 10 x YPC added, then autoclaved. $CaCl_2$ added prior to pouring.

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 KPO_4 buffer (pH 7.0).

Microwaved without 10 x Ca to dissolve agar. 10 x Ca added, then autoclaved. Hv-Ca salts and KPO_4 buffer added prior to pouring.

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 KPO_4 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, KPO_4 buffer, biotin and thiamine added prior to pouring.

Hv-YPC broth: 18% SW, 1 x YPC, 3 mM $CaCl_2$.

$CaCl_2$ 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 KPO_4 buffer (pH 7.0), 444 nM biotin, 2.5 μM thiamine.

30% SW, dH_2O and Tris-HCl pH 7.0 autoclaved. All other components added when cool.

Haloferax volcanii Media Supplements

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

Table 2.5 **Media supplements used 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

Table 2.6 **Growth of auxotrophic *Haloferax volcanii* mutants on different media**

Genotype	Hv-YPC	Hv-Ca	Hv-Min
ΔpyrE2	+	-	-
ΔleuB	+	+	-
ΔtrpA	+	-	-
ΔhdrB	-	-*	-*

+ indicates growth, - indicates no growth (supplement must be added to media). *Supplemented with hypoxanthine in Hv-Ca and hypoxanthine, methionine, glycine, and pantothenic acid in Hv-Min.

Escherichia coli Media

Sterilised by autoclaving at 121ºC for 15 mins. Stored at room temperature until supplements added, then stored at 4ºC.

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

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

Escherichia coli Media Supplements

Table 2.7 **Media supplements used with *E. coli***

Supplement	Abbreviation	Final Concentration
Ampicillin	Amp	50 µg/ml
Tetracycline	Tet	3.5 µg/ml
5-Bromo-4chloro-3-indolyl ß-D-galactopyranoside	X-Gal	40 µg/ml
Isopropyl ß-D-1-thiogalactopyranoside	IPTG	0.2 mM

2.1.6 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

Cultures on solid media were grown in a static incubator (LEEC) at 37ºC overnight. Small liquid cultures (1-10 ml) were grown overnight at 37ºC in the same incubator with 8 rpm rotation. Large scale liquid cultures (300 ml) were grown overnight at 37ºC with 110 rpm shaking in an Innova 4330 floor-standing incubator (New Brunswick Scientific). For short term storage, all cultures were

stored at 4ºC. For long term storage, glycerol was added (20% v/v) to liquid cultures, snap frozen on dry ice and stored at -80ºC.

Preparation of Electrocompetent Cells

The 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 37ºC with 8 rpm rotation. Cells were diluted 1/100 in LB broth supplemented with appropriate antibiotics. These were grown at 37ºC to A_{650} = 0.5-0.8. Cells were pelleted at 6000 x *g* for 12 minutes at 4ºC. 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 -80ºC.

Transformation of *Escherichia coli* by Electroporation

1-2 µg of DNA in 5 µl of sterile dH_2O 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 37ºC with 8 rpm rotation for 1 hour, to allow for recovery of the cells. Cells were plated onto LB+2xAmp plates and incubated at 37ºC overnight.

Buffers and solutions:

SOC Broth: 2% tryptone (Bacto), 0.5% yeast extract (Difco), 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose.

2.2.2 General *Haloferax volcanii* Microbiology

Growth and Storage

Cultures on solid media were grown in a static incubator (LEEC) at 45ºC for 5-14 days in a plastic bag to prevent drying of plates. Small liquid cultures (1-10 ml) were grown overnight at 45ºC in the same incubator with 8 rpm rotation. Large scale liquid cultures (50-600 ml) were grown overnight at 45ºC with 110 rpm shaking in an Innova 4330 floor-standing incubator (New Brunswick Scientific). For short term storage, all cultures were stored at room temperature. For long term storage, glycerol (80% glycerol in 6% SW) was added (20% v/v) to liquid cultures, snap frozen on dry ice and stored at -80ºC.

Transformation of *Haloferax volcanii* Using PEG600

Haloferax volcanii can be easily and efficiently transformed by using PEG600 (Cline *et al*. 1989). DNA must first be passaged through a *dam- E. coli* host prior to transformation as *H. volcanii* encodes a restriction endonuclease (Mrr) that targets methylated DNA (Holmes *et al*. 1991). *Δmrr* strains of *H. volcanii* (deficient in Mrr) can be transformed directly with *dam+* plasmid DNA.

A 5-10 ml culture in Hv-YPC broth (+Thy) of *H. volcanii* was grown overnight at 45ºC with 8 rpm rotation until A_{650} = 0.6-0.8. Cells were pelleted by centrifugation at 3300 x *g* for 8 minutes in a 15 ml round-bottomed tube. The supernatant was removed and the cells gently resuspended in 2 ml buffered spheroplasting solution. Cells were transferred to a fresh 2 ml round-bottomed tube, pelleted again and supernatant removed. Cells were gently resuspended in 400-800 µl buffered spheroplasting solution. A 200 µl aliquot per transformation was transferred to a fresh 2 ml round-bottomed tube and a 20 µl drop of EDTA pH 8.0 was pipetted onto the side of the tube, gently inverted and incubated for 10 minutes at room temperature. Transforming DNA was added in the same manner as EDTA and incubated at room temperature for 5 minutes. 250 µl of 60% PEG600 was added to the side of the tube and mixed by gentle rocking followed by incubation at room temperature for 30 minutes. 1.5 ml of

spheroplast dilution solution was added, mixed by gentle inversion and incubated at room temperature for 2 minutes. It was then centrifuged at 3300 x *g* for 8 minutes at 25ºC. The cell pellet was then transferred whole to a sterile 4 ml tube containing 1 ml regeneration solution (+ 40 µg/ml thymidine for *ΔhdrB* strains). The cells were left to recover undisturbed for 90 minutes at 45ºC after which the pellet was resuspended by gently tapping the side of the tube and incubated for a further 3-4 hours at 45ºC with 8 rpm rotation. Cells were transferred to a fresh 2ml round-bottomed tube and centrifuged at 3300 x *g* for 8 minutes at 25ºC. The supernatant was removed and the cell pellet gently resuspended in 1 ml of transformation dilution solution. Appropriate dilutions were made and 100 µl of each dilution plated on appropriate media. Plates were incubated for at least 5 days at 45ºC.

Buffers and Solutions:

All buffers and solutions sterilised by filtration through a 0.2 µm Minisart filter (Sartorius).

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 µl 0.5 M EDTA pH 8.0, 15 µl unbuffered spheroplasting solution, 10 µl DNA (~1-2 µg).
60% Polyethylene Glycol 600 (PEG600): 150 µl PEG600, 100 µl unbuffered spheroplasting solution.
Spheroplast Dilution Solution: 23% SW, 15% sucrose, 37.5 mM $CaCl_2$.
Regeneration Solution: 18% SW, 1 x YPC, 15% sucrose, 30 mM $CaCl_2$.
Transformation Dilution Solution: 18% SW, 15% sucrose, 30 mM $CaCl_2$.

2.2.3 DNA Extraction from Cells

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 1-2 ml *E. coli* cell culture (LB broth+Amp) was used and eluted in 30 µl 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 µl of TE, and stored at -20ºC.

Genomic DNA Extraction from *Haloferax volcanii* by Spooling

A 5 ml culture of *H. volcanii* was grown overnight in Hv-YPC (+Thy) at 45ºC until A_{650} = 0.6-0.8. 2 ml of culture was transferred to a round-bottomed 2 ml tube and centrifuged at 3300 x *g* for 8 minutes at 25ºC. The supernatant was removed and the pellet resuspended in 200 µl of ST buffer. 200 µl of lysis solution was added, the tube mixed by inversion, and the cell lysate overlaid with 1 ml of 100% EtOH. DNA was spooled at the interface with a capillary tip until the liquid was homogenous and clear. The spool of DNA was washed twice in 100% EtOH, and excess EtOH allowed to drain from the DNA. The DNA was air-dried, resuspended in 500 µl of TE, and precipitated with 400 µl isopropanol and 50 µl sodium acetate. The samples were centrifuged at 11,000 x *g* for 5 minutes, washed in 1 ml 70% EtOH, and dried thoroughly. The pellet was resuspended in 100-500 µl of TE and stored at 4ºC.

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.

2.2.4 Nucleic Acid Manipulation

PCR Amplification

Amplification of DNA was carried out using Phusion, DyNAzyme EXT (Finnzymes), Q5 Hotstart or OneTaq. These enzymes are suitable for amplifying templates with high GC content. Phusion and Q5 Hotstart were used for amplifications that required high fidelity. DyNAzyme EXT and One*Taq* were

used for diagnostic amplifications. Reaction conditions are shown below. All reactions were carried out using a Techne TC-512 thermocycler.

Table 2.8 **PCR Components**

Phusion	**Q5 Hotstart**	**DyNAzyme EXT**	**One*Taq***
200 µM of each dNTP	200 µM of each dNTP	200 µM of each dNTP	200 µM of each dNTP
0.5 µM of each primer	0.5 µM of each primer	0.5 µM of each primer	0.5 µM of each primer
10 ng of template DNA	1 ng – 1 µg genomic DNA or 1 pg – 1 ng plasmid DNA template	10 ng of template DNA	10 ng of template DNA
1 x Phusion GC Buffer	1 x Q5 Reaction Buffer	1 x Optimised DyNAzyme EXT Buffer	1 x OneTaq GC Buffer
3% DMSO	1 x Q5 High GC Enhancer	5% DMSO	-
1 U of Phusion	0.02 U/µl Q5 Hotstart	1 U of DyNazyme EXT	0.025 U/µl One*Taq*

Table 2.9 **PCR Conditions for Phusion and Q5 Hotstart**

Step	**Phusion**	**Q5 Hotstart**	
Initial Denaturation	98ºC, 30 seconds	98ºC, 30 seconds	
Denaturation	98ºC, 10 seconds	98ºC, 5-10 seconds	30 cycles
Annealing	TmºC, 20 seconds	TmºC, 10-30 seconds	
Extension	72ºC, 20 seconds/kb	72ºC, 20-30 seconds/kb	
Final Extension	72ºC, 5 minutes	72ºC, 5 minutes	

Table 2.10 **PCR Conditions for DyNAzyme EXT and One*Taq***

Step	**DyNAzyme EXT**	**One*Taq***	
Initial Denaturation	94ºC, 2 minutes	94ºC, 30 seconds	
Denaturation	94ºC, 15 seconds	94ºC, 30 seconds	30 cycles
Annealing	TmºC, 20 seconds	TmºC, 30 seconds	
Extension	72ºC, 40 seconds/kb	68ºC, 60 seconds/kb	
Final Extension	72ºC, 5 minutes	68ºC, 5 minutes	

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

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

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.

Touchdown PCR

When primers did not have 100% homology to the template DNA (e.g. when introducing restriction sites or mutations), two annealing temperatures were calculated. Tm_S was based on the original percentage of homology between primer and template, Tm_E was based on 100% homology. The reaction temperature started at Tm_S and increased linearly to Tm_E across 10 cycles. The remaining 20 cycles used Tm_E as the annealing temperature.

Colony PCR

Colony PCRs were used for screening large numbers of colonies for a desired plasmid or chromosomal gene. *H. volcanii* or *E. coli* colonies were grown on solid media and touched lightly with a sterile yellow tip, ensuring only a small number of cells were collected and the colony was not disturbed. The yellow tip was pipetted up and down in 100 µl of sterile dH_2O to wash cells off the yellow tip. This was then boiled at 100°C to lyse cells for 10 minutes and immediately cooled on ice. 1 µl of the preparation was then used in a PCR with DyNAzyme EXT polymerase or One*Taq* polymerase.

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 Enzyme

When overhangs generated by restriction digests were 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 x NEB buffer 4 and 1 mM dNTPs for 30 minutes at 25ºC. The reaction was stopped by heat inactivation at 75ºC for 20 minutes.

Dephosphorylation of Vector DNA

To prevent self-ligation of vector DNA, Shrimp alkaline phosphatase was used to remove 5' phosphate groups. Samples were incubated with 5 units of Shrimp alkaline phosphatase / μg of DNA and 1 x Antarctic phosphatase buffer for 30 minutes at 37ºC. Phosphatase was heat inactivated at 65ºC for 10 minutes.

Ligation of DNA

Ligations were performed using T4 DNA ligase. For each μg of DNA, 5 units of ligase were used in a reaction with 1 x T4 ligase buffer. For vector:insert ligations, reactions contained a molar ratio of >3:1 insert to vector DNA. Ligations were carried out at 15ºC overnight or 4ºC for 36 hours. Ligations were ethanol precipitated, resuspended in 5 μl sterile dH_2O and transformed 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 (pH5.2) were added to DNA and incubated at -20ºC for at least 1 hour. Samples were centrifuged at 20,000 x *g* for 30 minutes at 4ºC and the supernatant removed. Pellets were washed in 400 μl 70% EtOH followed by centrifugation at 20,000 x *g* for 10 minutes at 4ºC. The supernatant was removed and the pellets air-dried thoroughly before resuspension in sterile dH_2O.

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 by washing with ethanol. DNA was eluted in 30 µl of the provided elution buffer.

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

All oligonucleotides were synthesised by Eurofins MWG, Germany.

Nucleic Acid Quantification

To determine the concentration and purity of plasmid preparations the absorbance at 260 nm and the 260:280 nm absorbance ration, respectively, were measured by spectrometer (Beckman Coulter DU 530).

Agarose Gel Electrophoresis

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 1 x. All samples and molecular markers, either 1 kb (NEB) or 100 bp (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 0.5 x final concentration. Gels used for DNA extraction were stained with ethidium bromide after the removal of sample lanes.

Buffers and Solutions:

TBE (Tris/Borate/EDTA): 89 mM Tris, 89 mM boric acid, 2 mM EDTA pH 8.0.
TAE (Tris/Acetic acid/EDTA): 40 mM Tris, 20 mM acetic acid, 1 mM EDTA pH 8.0.
Gel loading dye (5x): 50 mM Tris.HCl, 100 mM EDTA pH 8.0, 15% Ficoll (w/v), 0.25% Bromophenol Blue (w/v), 0.25% Xylene Cyanol FF (w/v).

Agarose Gel Extraction and Purification of DNA

To purify DNA from agarose gels without UV exposure, sample lanes were protected using foil while the appropriate band was excised. DNA was visualised by UV light using a UV transilluminator (UVP inc.). DNA was purified using the Macherey-Nagel DNA purification kit following manufacturer's guidelines.

2.2.5 Genetic Manipulation of *Haloferax volcanii*

2.2.5.1 Plasmid Construction

Generating a Genomic Clone

Genomic clones of genes were obtained by restriction digest of high quality genomic DNA. Restriction sites were chosen so that around 1 kb of upstream and downstream regions of the gene of interest were included in the digest. Restriction sites were also chosen to be compatible for the multiple cloning site in pBluescript II SK+ (Stratagene). Genomic DNA was isolated from H26 and digested overnight with appropriate enzymes. The digested genomic DNA was run on a 0.75% TAE agarose gel for 16 hours at 50 V. The band of the expected size was then extracted (± 0.5 kb). DNA was purified and ligated into compatible sites in either pBluescript II SK+ or a pBluescript II SK+ derivative, that had been digested with appropriate enzymes, dephosphoylated (if required) and gel

extracted. The resulting plasmid library was then transformed into *E. coli* XL1-Blue. Transformants were selected by plating on LB+Amp+Xgal+IPTG and using blue/white screening. White colonies were patched and screened by colony hybridisation using a gene specific probe. Presence of the desired gene was confirmed by DNA sequencing.

Generating a Deletion Construct by PCR

Deletion constructs were generating from the genomic clone by inserting the upstream and downstream flanking regions of the gene of interest into pTA131 or a derivative. Upstream and downstream flanking regions were generated by PCR (Figure 2.5).

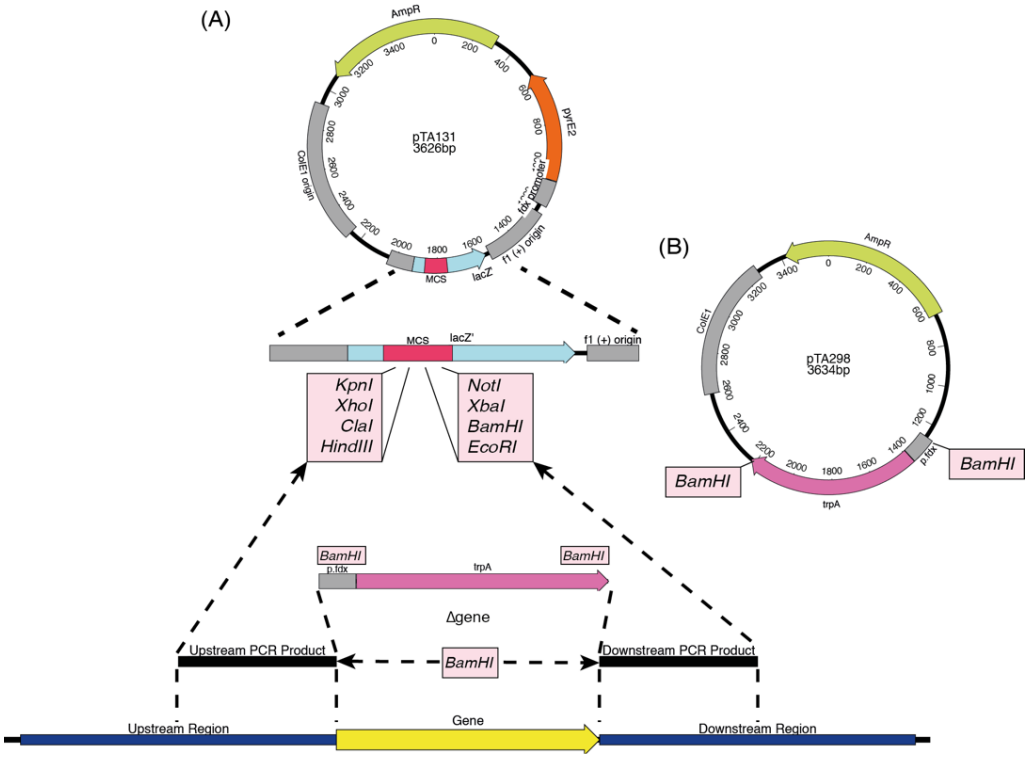


Figure 2.5 **Schematic for generating deletion constructs by PCR. (A)** Upstream and downstream flanking regions of the gene to be deleted are amplified by PCR. Primers can introduce novel restriction sites that are compatible with the multiple cloning site (MCS) in the *lacZ* gene allowing the deletion construct to be cloned into pTA131. The PCR products are ligated internally using *BamHI* sites introduced by the primers. **(B)** Optional addition of *trpA* marker gene excised from pTA298 using *BamHI* and ligated between the compatible upstream and downstream PCR products.

Two separate PCRs were carried out; one to amplify the upstream flanking region (US) and one to amplify the downstream flanking region (DS) of the gene of interest. External primers were designed to include restriction sites (or to be outside of existing restriction sites) that were compatible with the multiple cloning site in pTA131 or a derivative. Internal primers were designed to delete the gene of interest and introduce either a *BamHI* or a *BglII* restriction site at the site of the deletion. This is to allow the insertion of markers e.g. *trpA,* that are flanked with *BamHI*, into this region of the deletion construct if required. The PCR products were then digested with either *BamHI* or a *BglII* as appropriate and ligated together. This product was then digested using the appropriate external restriction enzymes and ligated into pTA131 or a derivative. Plasmids were transformed into *E. coli* XL-1 Blue cells and plated onto LB+Amp. Selected colonies were grown for miniprep using Macherey-Nagel NucleoSpin Plasmid (Mini) according to manufacturer's instructions. The plasmid DNA was then subjected to a restriction digest to confirm the presence of the insert. The final plasmid was sequenced to ensure the correct insert was present without point mutations.

Generating a Gene Replacement Construct

Gene replacement constructs were made by inserting the gene of interest (usually with the addition of an inducible promotor 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.

2.2.5.2 Strain Construction

Gene Deletion and Replacement in *Haloferax volcanii*

Gene deletion and replacement constructs were integrated at the gene locus by transformation of *H. volcanii ΔpyrE2* strains. The strains were plated on Hv-Ca (+ additives as required) to select for the integrated *pyrE2* marked plasmid (pop-in). A pop-in colony was picked and grown without selection in a 5 ml Hv-YPC (+Thy) overnight culture until A_{650} = 1.0. This culture was then diluted 1/500 into fresh 5 ml Hv-YPC (+Thy) broth. The growth and the dilution was repeated.

By relieving selection for uracil, it allows for the integrated plasmid and native gene to be lost from the chromosome. To select for *pyrE2*- pop-outs, the final culture was plated onto Hv-Ca +5FOA (+ additives as required). Depending on the location of the recombination event either wild-type or deletion/replacement mutants can be recovered. Replacing the gene with a selectable marker e.g. *trpA* enables direct selection of deletion mutants. Colonies were then restreaked onto selective media and tested for the desired genotype, normally colony hybridisation to a membrane and probed with a radiolabelled probe specific for the gene of interest, or by colony PCR. The pop-in/pop-out method was developed by Bitan-Banin *et al*. (2003) and is shown in Figure 2.6.

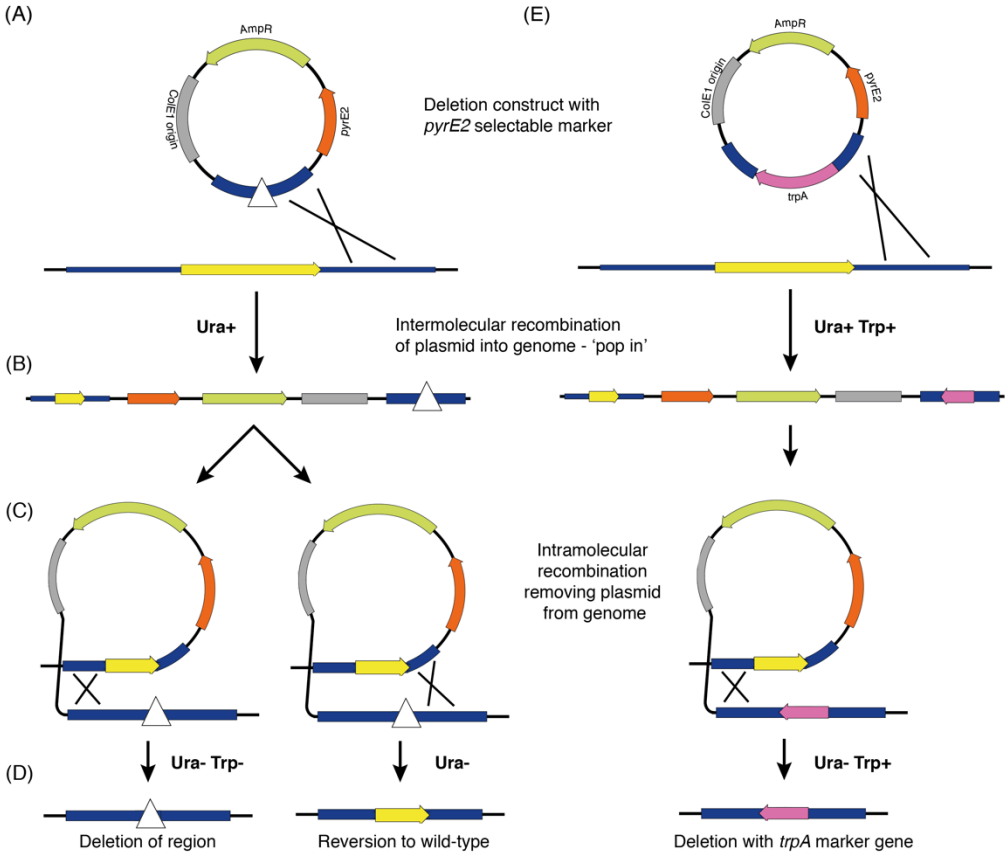


Figure 2.6 **Gene deletion by pop-in/pop-out method. (A)** A *ΔpyrE2* strain is transformed with *pyrE2+* deletion construct. **(B)** Pop-in colonies selected using *ura+* phenotype. Cells undergo pop-out when uracil selection is relieved. This is selected for by plating on 5-FOA. **(C)** Recombination can be in the downstream region (left) or the upstream region (right). **(D)** The gene is deleted (left) or it reverts to wild-type (right). **(E)** A *trpA* marker can be used in the deletion construct to allow for direct selection of the deletion.

Mating of *Haloferax volcanii*

To mate *Haloferax volcanii* requires two parent strains. Cultures of these strains were set up in 5 ml Hv-YPC and incubated overnight at 45ºC until A_{650} = 0.6-0.8. These cultures were then diluted and grown until they reached A_{650} = 0.9. Meanwhile, Millipore Swinnex 25 filter units with 0.45 µm HA 25 mm filters (Millipore) were prepared by autoclaving. 1 ml of cell culture of each of the parent strains at an equal OD A_{650} were mixed in a 2 ml tube. 2 ml of unmixed parent strains were also aliquoted into separate 2 ml tubes to act as negative controls. The aliquots were then transferred individually to a 5 ml syringe and the cell suspension was gently filtered through the Swinnex filter unit. The filter unit was then disassembled and the filter disc removed using sterile forceps. The filter disc was then placed cell side up on an Hv-YPC +Thy plate. These were incubated overnight at 45ºC. Sterile forceps were then used to transfer the filter disc into a 5 ml tube containing 1 ml 18% SW, which were vortexed gently to resuspend the cells and incubated at 45ºC with 8 rpm rotation for 1 hour. The cell suspensions were then diluted in 18% and 100 µl plated on appropriate selective media. 20 µl spots were plated on Hv-YPC (+Thy) for viable counts.

2.2.6 Genotype Screening

There are variety of methods that can be used to screen for the presence or absence of a specific gene in *Haloferax volcanii*. If the strain has a selectable phenotype e.g. *ΔtrpA* colonies can be plated onto selectable media. However, due to the polyploid nature of *H. volcanii* strains can be mero-diploid meaning that they have a mixture of mutant and wild-type alleles present on different chromosome copies. To screen for colonies in which all chromosome copies carry the same allele colony hybridisation and Southern blotting are used. For both of these methods DNA is denatured and transferred to a positively charged membrane by either colony lift or vacuum transfer, respectively.

Colony Lift

A colony lift was used when there was a large number of colonies to be screened at once. Candidate colonies were patched onto Hv-YPC (+Thy) plates using sterile wooden toothpicks and incubated at 45ºC for ~3 days until grown. The patched colonies were lifted from the plate by placing a circle of positively charged membrane (BioRad Zeta-Probe GT or Amersham Hybond-N+) onto the surface of the plate for 1 minute. The membrane was transferred colony side up to Whatman paper soaked in 10% SDS and left 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 <30 seconds in 2 x SSPE before being thoroughly air-dried. DNA was crosslinked to the membrane with 120 mJ/cm^2 UV.

Buffers and Solutions:

20 x SSPE: 3 M NaCl, 230 mM NaH_2PO_4, 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.

Southern Blot Vacuum Transfer

Purified *H. volcanii* 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 gel for 16 hours at 50 V, with buffer circulation. The gel was post-stained with 0.5 µg/ml ethidium bromide 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 dH_2O. The DNA was then denatured in denaturing solution for 45 minutes. Membrane (BioRad Zeta-Probe GT or Amersham Hybond-XL) was soaked in dH_2O 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/cm^2 UV.

Buffers and Solutions:

20 x SSPE: 3 M NaCl, 230 mM NaH_2PO_4, 32 mM EDTA, pH 7.4.
Denaturing Solution: 1.5 M NaCl, 0.5 M NaOH.

Hybridisation

Membranes from either colony lifts or Southern blots were pre-hybridised for >3 hours at 65ºC in 40 ml pre-hybridisation solution. Radiolabelled DNA probes were made with 50 ng of DNA and 0.74 MBq of [α-^{32}P] dCTP (Perkin Elmer). DNA was denatured at 100ºC for 5 minutes then incubated with the radioisotope and HiPrime random priming mix (Roche) for 15-20 minutes at 37ºC. 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 100ºC 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 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 65ºC. 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).

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 NaH_2PO_4, 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 100ºC prior to addition).
Hybridisation Solution: 6 x SSPE, 1% SDS, 5% dextran sulfate.
Low Stringency Wash Solution: 2 x SSPE, 0.5% SDS.
High Stringency Wash Solution: 0.2 x SSPE, 0.5% SDS.

Pulsed-Field Gel Electrophoresis

A 5 ml Hv-YPC (+Thy) *H. volcanii* culture was incubated overnight at 45ºC until A_{650} = 0.8-1.0. 1 ml of cells were pelleted at 3300 x *g* for 10 minutes in a round-bottomed 2 ml tube. The supernatant was removed and the pellet resuspended in 100 µl 18% SW. 80 µl of resuspended cells were transferred to a fresh tube, incubated at 42ºC in a water bath for 10 minutes and gently mixed with 320 µl 1% low-melt agarose (SeaPlaque GTG, Lonza) in 18% SW also at 42ºC. 80 µl of the mixture was quickly pipetted into plug molds (BioRad) avoiding air bubbles and left at 4°C for 30 minutes. Once set the plugs were transferred to a 2 ml tube containing 2 ml lysis solution and 1 mg/ml proteinase K, and incubated with gentle shaking at 45ºC for 3 hours. The lysis solution was replaced with 2 ml fresh lysis solution containing 1 mg/ml proteinase K and 10 µg/ml RNase A, and incubated overnight with gentle shaking at 45ºC. Lysis solution was removed, replaced with 2 ml wash solution and incubated with gentle shaking at 37ºC for 30 minutes. If plugs were gamma irradiated, plugs were transferred to a fresh tube containing 500 µl wash solution and exposed to approximately 35 Gy (Gammacell 1000 Elite Model ^{137}Cs source). Regardless of gamma irradiation, plugs were then transferred to a fresh 2 ml tube containing 2 ml wash solution and 1 mM PMSF, and incubated with gentle shaking for 1 hour at 37ºC, which was then repeated. If plugs were to be subjected to restriction digest, they were washed twice in 2 ml wash solution diluted 1/10 for 30 minutes with gentle shaking at 37ºC, washed once in 2 ml wash solution diluted 1/100 for 30 minutes with gentle shaking at 37ºC, and then washed once in 2 ml wash solution diluted 1/1000 for 30 minutes with gentle shaking at 37ºC. The plugs were then transferred to a fresh tube containing 1 ml 1 x restriction buffer as appropriate for the required enzyme, and incubated for 1 hour at 37ºC with gentle shaking. Restriction buffer was then replaced with 250 µl fresh 1 x

restriction buffer and 50 U of restriction enzyme and incubated overnight at 37ºC with gentle shaking. Plugs were then washed twice in 2 ml 0.5 x TBE at 37ºC for 30 minutes with gentle shaking before loading onto the gel. For short term storage plugs were stored at 4ºC in 1 ml wash solution and for long term storage plugs were stored at -20ºC in 1 ml wash solution containing 50% glycerol.

A 1.2% agarose (SeaKem Gold, Lonza) gel was prepared with 0.5 x TBE and allowed to set for >1 hour. Plugs were then loaded alongside Lambda PFG ladder and sealed with molten agarose. The gel was run in a CHEF Mapper (BioRad) using 2.2 L 0.5 x TBE at 14ºC for 20 hours 46 minutes with a two-state included angle of 120º, voltage gradient of 6 V/cm (linear ramping), and switch times of 0.64 seconds (initial) to 1 minute 13.22 seconds (final). The gel was post-stained with 0.5 μg/ml ethidium bromide, visualised, and Southern blotted if required.

Buffers and Solutions:

Lysis Solution: 20 mM Tris.HCl pH 8.8, 500 mM EDTA, 1% N-lauroylsarcosine, 1 mg/ml proteinase K (if required), 10 μg/ml RNase A (Ribonuclease A from bovine pancreas, if required).
Wash Solution: 25 mM Tris.HCl pH 7.5, 100 mM EDTA, 1 mM PMSF (Phenylmethylsulfonyl fluoride, if required), 50% glycerol (if required).
TBE (Tris/Borate/EDTA): 89 mM Tris, 89 mM boric acid, 2 mM EDTA pH 8.0.

2.2.7 Phenotyping of *Haloferax volcanii*

Tryptophan Gradient Plates

Tryptophan gradient plates were poured with 10 ml Hv-Ca +Trp (in appropriate concentration) on a slant forming a tapered wedge of agar (Figure 2.7). Once set, this was covered with Hv-Ca agar (without tryptophan) with the plate flat, therefore forming a gradient across the plate, with the desired tryptophan concentration at one side of the plate reducing to no tryptophan on the other side (Adapted from Bryson & Szybalski 1952).

Cultures of strains were set up in 5 ml Hv-YPC and incubated overnight at 45ºC with 8 rpm rotation until A_{650} = 0.6-0.8. These were then diluted into fresh Hv-YPC and incubated at 45ºC until A_{650} = 1.0. Cultures were then diluted appropriately in 18% SW. Autoclaved paintbrushes were first dipped in 18% SW, before being dipped into the diluted culture and, then painted in one direction across the gradient plate. The same paintbrush was then dipped into the diluted culture again and painted over the top of the first streak in the other direction. The plates were then incubated at 45ºC for 4-5 days.

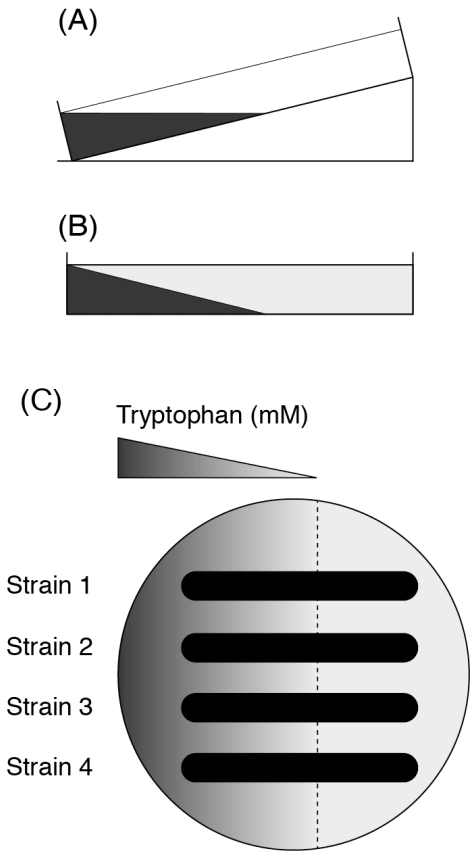


Figure 2.7 **Tryptophan gradient plates. (A)** A wedge of Hv-Ca media with the desired concentration of tryptophan is poured while the plate is slanted. **(B)** The plate is then laid flat and Hv-Ca media (without tryptophan) is poured over the rest of the plate to form a flat surface. **(C)** Strains are painted across the tryptophan gradient.

Growth Curves

For testing growth rates in liquid media, cultures were set up in 5 ml Hv-YPA broth and grown at 45ºC to an A_{650} = 0.2-0.8 (exponential phase). Cells were diluted and grown again to A_{650} = 0.2-0.8. Cells were diluted again and grown

overnight at 45ºC until they reached A_{650} = 0.4. 250 µl of culture and appropriate blanks were added to the wells of a 48 well microtiter plate (Corning). The plate was sealed with 1.25 cm microporous tape (Boots) and incubated at 45ºC with double orbital shaking at 425 rpm for between 48-72 hours in an Epoch 2 Microplate Spectrophotometer (BioTek). Readings were taken every 15 minutes at A_{600}. The generation time was calculated by plotting the growth on a $\log_2$ scale and using the equation below (Equation 2.2).

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

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

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

B = start OD
b = end OD

Equation 2.2 **Calculating the generation time of a liquid culture using the OD A_{600}.**

It had previously been noted that generation times of strains varied between experiments when measuring growth in this method (Gamble-Milner, 2016). Therefore, in this study, comparisons are made only between duplicates of strains within the same experiment i.e. strains that have been grown on the same microtiter plate and the A_{600} measured simultaneously during a single run on the Epoch2 Microplate Spectrophotometer (BioTek). As generation times are seen to vary between experiments, generation times stated are not absolute. However, the relationship between different strains is observed to be consistent, meaning that reliable comparisons are able to be made between different strains in the same experiment. This means that the generation times and growth curves generated by this method are used to illustrate the differences between different strains rather than an exact generation time. It is also worth noting that this method is not sensitive enough to detect very small differences in growth that can be observed with growth competition methods (Delmas *et al*. 2009).

2.2.8 Protein Expression and Purification

Protein Expression

A 5 ml overnight culture in Hv-YPC was set up and incubated at 45ºC with 8 rpm rotation. This was grown to A_{650} = 1.0 and then diluted into 50 ml Hv-YPC broth and grown at 45ºC with 150 rpm orbital shaking until an A_{650} = 0.5. This was then diluted into 333 ml Hv-YPC broth and incubated overnight at 45ºC with 150 rpm orbital shaking until A_{650} = 0.4. The culture was centrifuged at 3300 x *g* for 8 minutes at 4ºC. Cells were resuspended in 7 ml ice-cold Buffer A (20 mM imidazole, 1 x Sigma*FAST* protease inhibitor). Cells were then lysed by sonication (4-5 x 30 seconds at ≤ 8 μm amplitude) on ice. Cell lysate was transferred to a 15 ml round bottomed centrifuge tube and centrifuged at 20,000 x *g* for 15 minutes at 4ºC. The sample was filtered sequentially through 0.8 μm, 0.45 μm and 0.2 μm Minisart syringe filters (Sartorius) to remove DNA.

Buffers and Solutions:

Buffer A: 20 mM HEPES pH 7.5, 2 M NaCl, 1 mM PMSF (phenylmethanesuphonyl fluoride), imidazole to desired concentration.

Charging of Ni^{2+} Beads

To purify His_6 tagged proteins, IMAC Sepharose 6 Fast Flow beads (GE Healthcare) were charged with Ni^{2+}. Beads (0.5 ml per column) were washed twice with ≥2 volumes of dH_2O, and then equilibrated for 30 minutes with 2 volumes of 0.2 M $NiSO_4$ with rotation at 4ºC. Beads were then washed twice with 5 ml dH_2O, washed once with 5 ml Buffer A (500 mM imidazole) and a further 3 times with 5 ml Buffer A (20 mM imidazole). The beads were then resuspended in 0.5 ml Buffer A (20 mM imidazole) per sample and stored at 4ºC until used (<24 hours).

Ni^{2+} Gravity Column

Cell lysates were incubated for 1 hour at 4ºC with rotation with 0.5 ml Ni^{2+} charged beads. The slurry was applied to a Poly-Prep gravity column (BioRad). The flow through was collected and reapplied to the column. The column was washed with 5 ml ice-cold Buffer A (20 mM imidazole), and then again with 10 ml ice-cold Buffer A (20 mM imidazole). Bound protein was step-eluted with 1 ml Buffer A containing 100, 200 and 500 mM imidazole. All purification steps were performed at 4ºC.

Protein Storage

For short term storage, elutions were stored at 4ºC.

SDS-Polyacrylamide Gel Electrophoresis

Protein samples were analysed using SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis). Gels were cast in Novex 1 mm cassettes (Invitrogen). A 10% resolving gel was poured and covered with a layer of isopropanol to leave a flat surface. Once set, the isopropanol was washed off, a 3% stacking gel poured and a comb inserted. Protein samples were mixed ¼ by final volume of 4x Laemmli buffer and denatured by boiling at 94ºC for 5 minutes. Samples were run alongside a Precision Plus Protein Kaleidoscope Protein Standard (BioRad). If the gel was to be used for Western blot, samples were run alongside a Biotinylated Protein Ladder Detection Pack (Cell Signalling Technologies) instead. Gels were run for ~1 hour 20 minutes in 1 x SDS-PAGE running buffer (200 V, 36 mA per gel) and stained with PageBlue Protein Staining Solution (Thermo Scientific) for visualisation of proteins.

Buffers and Solutions:

10% SDS-PAGE gel (resolving): 10% acrylamide/bisacrylamide Protogel (National Diagnostics), 0.37 M Tris pH 8.8, 0.1% SDS, 0.05% AMPS (ammonium persulfate), 0.05% TEMED (tetramethylethylenediamine).

3% SDS-PAGE gel (stacking): 3% acrylamide/bisacrylamide Protogel, 0.37 M Tris pH 6.8, 0.1% SDS, 0.05% AMPS, 0.05% TEMED.
SDS-PAGE running buffer: 0.25 M Tris pH 6.8, 1.92 M glycine, 1% SDS.
Laemmli buffer (4x): 50 mM Tris pH 6.8, 100 mM DTT, 2% SDS, 0.1% bromophenol blue, 10% glycerol.

Protein Identification by Western Blot

Freshly run, unstained SDS-PAGE gels were removed from cassettes and the stacking gel and foot removed. Sponges, Whatman paper and, Amersham Protan 0.45 μm NC membrane (GE Heathcare) were soaked in transfer buffer for 5 minutes before the Western blot apparatus was assembled, ensuring there were no air bubbles. Transfer was run at 25 V for 2 hours in NovexTM XCell IITM blotting apparatus. When transfer was complete, the membrane was blocked with TBST with 5% w/v BSA for 1 hour at room temperature, with agitation. The membrane was then incubated with primary antibody at the appropriate dilution in TBST (1% w/v BSA) at 4ºC overnight. The membranes were then washed 4 times with TBST and then incubated with the appropriate secondary antibody at the appropriate dilution in TBST at room temperature for 2 hours. The membrane was then washed 4 times in TBST. The membrane was then placed in Saran wrap and incubated with Amersham ECL prime reagent (GE Healthcare). Excess was removed after 1 minute. The membrane was imaged using Amersham Hyperfilm ECL (GE Healthcare).

To detect His_6 tagged MCM, the primary antibody used was 6x-His Tag Monoclonal Antibody HIS.H8 (Thermo Fisher) raised in mouse, which was in turn detected by Anti-Mouse IgG peroxidase antibody produced in Goat (Sigma) secondary antibody for detection of mouse antibodies using HRP.

StrepII tagged GINS was detected by the use of the primary antibody Anti-Strep-tag II antibody (Abcam) which is a rabbit monoclonal antibody. The secondary antibody used was Anti-Rabbit IgG peroxidase antibody produced in Goat (Sigma) for detection of rabbit antibodies using HRP.

To detect the protein ladder Biotinylated Protein Ladder Detection Pack (Cell Signaling Technologies) was used. This pack includes both the biotinylated protein ladder and a polyclonal conjugate antibody raised in guinea-pig for its detection. This was added during incubation with the secondary antibody.

Buffers and Solutions:

Transfer buffer: 12 mM Tris base, 96 mM Glycine, 20% v/v Methanol. Stored at 4ºC
10 x TBS: 198 mM Tris base, 936 mM $NaCl_2$. pH adjusted to 7.6 with 1 M HCl.
TBST: 1 x TBS, 0.1% Tween20.

Chapter 3: Comparing Origin-less DNA Replication in the Mega-plasmids pHV3 and pHV4

3.1 Background

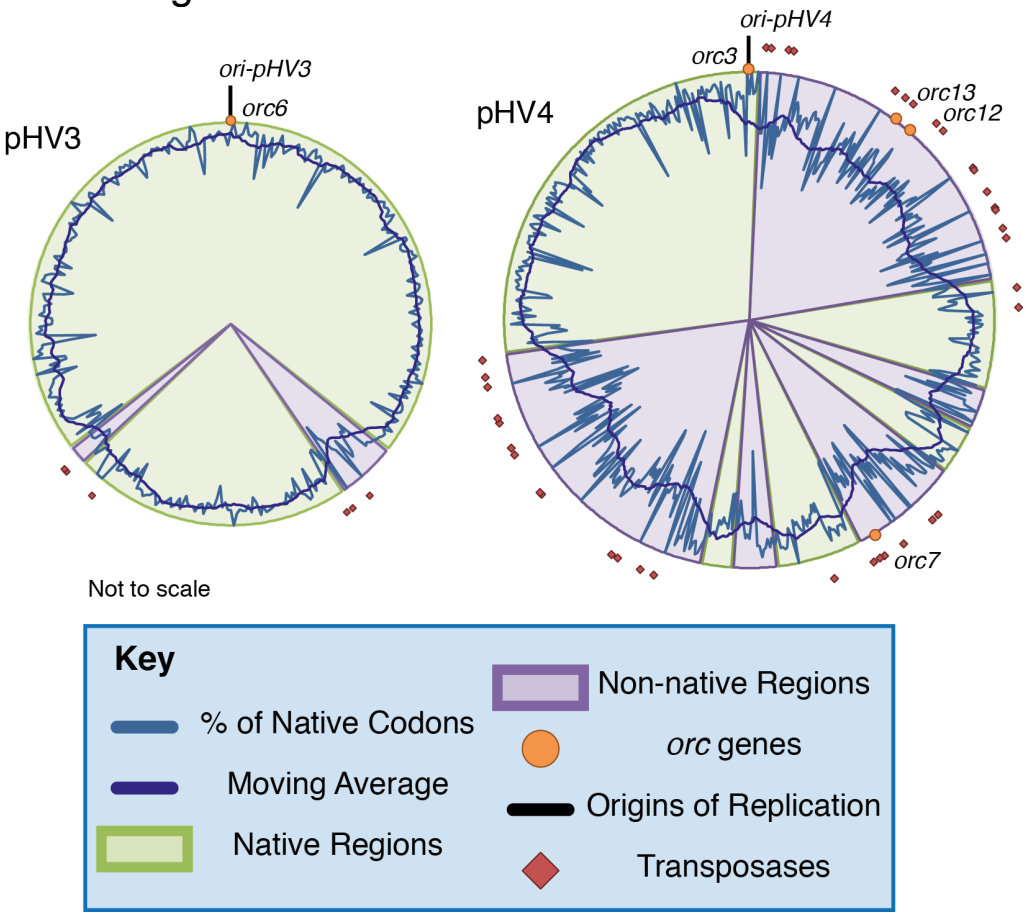


Figure 3.1 **Synonymous codon usage of pHV3 and pHV4**. Percentage of native codons by synonymous codon usage has been plotted for each gene on each replicon (light blue line) and a moving average has been taken for each replicon (dark blue line). For pHV3 this is 10 genes and for pHV4 this is 20 genes. Regions that contain 10 or more genes with >85% native codons have been designated native regions (green shading) and regions that contain 10 or more genes with <85% native codons were designated non-native regions (purple shading). Origins of replication (black line), *orc* genes (orange circle) and genes annotated as transposases (red diamond) have also been indicated. Nearly all the genes on pHV3 are native to *H. volcanii*, therefore pHV3 is considered native to *Haloferax volcanii*. Around half of the genes on pHV4 are native, and half are non-native. Adapted from Hartman *et al.* (2010). Full figure shown in Appendix *I Synonymous Codon Usage Diagrams*.

The mega-plasmid pHV3 has one replication origin, *ori-pHV3*, and one associated *orc* gene, *orc6*. It is 438 kb in size and contains many essential genes.

Almost all genes on pHV3 have a synonymous codon usage that matches the main chromosome (Figure 3.1), therefore it is a native component of the *H. volcanii* genome. The mega-plasmid pHV4 is 636 kb in size and in the laboratory strain has integrated onto the main chromosome via homologous recombination between IS-elements. There is one origin *ori-pHV4*, which has previously been deleted from integrated pHV4, and an associated *orc* gene *orc3*. There are also 3 other *orc* genes on pHV4 that are not considered in this study; *orc7, orc12* and *orc13*. When genes on pHV4 are assessed by synonymous codon usage (Figure 3.1), it is found that approximately half are native (match the main chromosome) and half are non-native (do not match the main chromosome), therefore there may be some non-essential regions. However, it is important to note that the replication origin and associated *orc* gene reside in a native region, and that the majority of the non-native regions are rich with IS-elements indicating that they could have been acquired (Hartman *et al*. 2010; Hawkins *et al*. 2013; Norais *et al*. 2007).

3.2 Aims

All the chromosomal origins, including the origin from the integrated pHV4 mega-plasmid, can be deleted from *H. volcanii* with no adverse effects on the cell (Hawkins *et al*. 2013). However, neither the pHV3 origin *ori-pHV3*, its associated *orc* gene *orc6* nor the origin from the episomal pHV4 have been deleted yet. This study will further investigate the replication of pHV3, including the possibility of a size threshold that may limit RDR for small DNA molecules.

The aims of this chapter are to:

- Delete *ori-pHV3* and *orc6* singularly and in combination with each other, to determine whether they are essential for the replication of pHV3.
- Compare the ability to delete *ori-pHV4* from episomal (and integrated) pHV4 with deletion of *ori-pHV3*.

3.3 Results

3.3.1 Deletion of *ori-pHV3* and *orc6* from Episomal pHV3

In order to construct deletion plasmids for *ori-pHV3* (HVO_B0001ori) and *orc6* (HVO_B0001) a genomic clone (pTA1509) was first constructed. Plasmid deletion constructs for the deletion of *ori-pHV3* and/or *orc6* were then constructed both with and without tryptophan selection. The pHV3 origin region is shown in Figure 3.2 (for full alignments of ORBs and DUEs see Appendix *II DUE and ORB Alignments*).

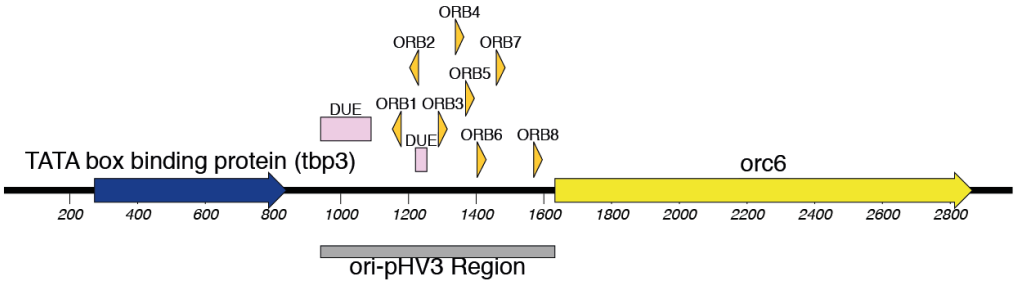


Figure 3.2 **pHV3 origin region and associated *orc* gene**. 656 bp *ori-pHV3* region is highlighted (grey box) with duplex unwinding elements (DUEs) (pink box) and origin recognition boxes (ORBs) (orange arrows) shown. Orientation of ORBs is indicated by arrows. Origin is flanked by TATA box binding protein (*tbp3*) upstream and *orc6* downstream. Adapted from Norais *et al*. (2007).

pTA1509 *ori-pHV3* (HVO_B0001ori), *orc6* (HVO_B0001) Genomic clone

Genomic DNA from H26 (wild-type *H. volcanii*) was digested with *BstBI* and *XmaI* (Figure 3.3A) resulting in a 4.5 kb fragment. Digested DNA was then electrophoresed on an agarose gel and DNA between 4-5kb was extracted and ligated into pBluescript II SK+ at *ClaI* and *XmaI* sites (Figure 3.3B, C). These ligations were then transformed into *E. coli* XL1-blue cells and transformants plated onto LB+Ampicillin+X-gal+IPTG agar for blue/white screening. White colonies were patched onto LB+Amp agar and transferred to a membrane via a colony lift. Colony lifts were then probed using a radiolabelled 699 bp fragment digested using *EcoRI* and *BamHI* sites from pCN26 (Norais *et al*. 2007), specific for the region of interest (Figure 3.3D, E). Patches that hybridised with the probe

were restreaked, and sequencing was used to confirm the presence of the correct insert.

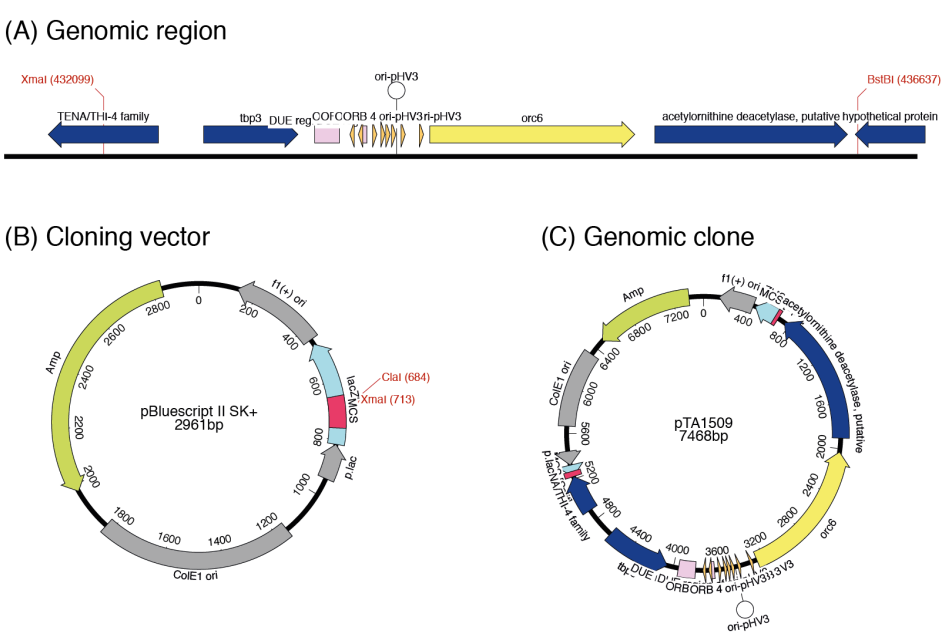


Figure 3.3 **pTA1509. (A)** The location of *ori-pHV3* and *orc6* on the *H. volcanii* genome, genomic DNA was digested with *XmaI* and *BstBI*. **(B)** pBluescript II SK+ plasmid where the *XmaI* to *BstBI* fragment was inserted into *XmaI* and *ClaI* sites. **(C)** Genomic clone pTA1509 constructed by inserting a *XmaI* to *BstBI* fragment containing *ori-pHV3* and *orc6* into pBluescript II SK+. **(D)** 699 bp *ori-pHV3* probe was constructed by digesting pCN26 using *BamHI* and *EcoRI* sites, was radiolabelled and used to probe colonies. **(E)** Colony hybridisation showing a patch which the probe hybridised to (containing *ori-pHV3* and *orc6* genomic clone) and a patch to which the probe did not hybridise to (does not contain *ori-pHV3* and *orc6* genomic clone).

pTA1520 (*ori-pHV3* deletion construct) and pTA1798 (*ori-pHV3::trpA+* deletion construct)

From the genomic clone pTA1509 the upstream region of *ori-pHV3* was PCR amplified using doripHV3USBglR and pBSF3 primers introducing a novel *BglII* restriction site. The downstream region amplified using doripHV3DSBglF and doripHV3DSR which introduces a *BglII* site. Both PCR products were digested with *BglII* and ligated. The resulting product was ligated into pTA131 at *XbaI* and *XhoI* sites to create pTA1520 *ori-pHV3* deletion construct. To introduce a *trpA* marker to the deletion construct pTA1520 was digested using *BglII* and the compatible *BamHI* digested fragment from pTA298 containing *p.fdx-trpA* was ligated to create pTA1798 *ori-pHV3::trpA+* deletion construct (Figure 3.4).

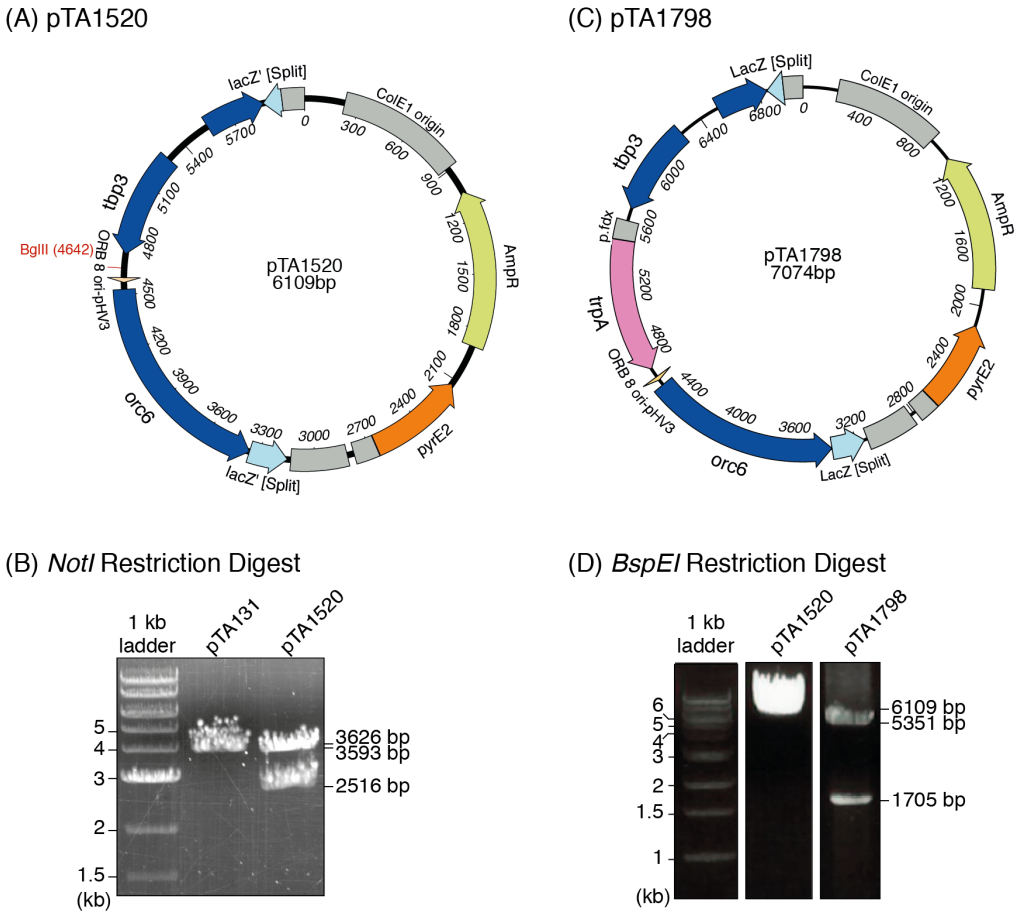


Figure 3.4 ***ori-pHV3* deletion constructs**. **(A)** pTA1520 showing internal *BglII* site. **(B)** *NotI* restriction digest of pTA131 and pTA1520 giving band sizes as predicted at 3626 bp, and 3593 bp and 2516 bp, respectively. **(C)** pTA1798 *ori-pHV3* deletion construct with *trpA* marker gene. **(D)** *BspEI* restriction digest of pTA1520 producing a bands size of 6109 bp and pTA1798 producing band sizes of 5351 bp and 1705 bp, as predicted.

pTA1521 (*orc6* deletion construct) and pTA1799 (*orc6::trpA+* deletion construct)

Using the genomic clone pTA1509 as a template a dorc6USEcoF-dorc6USBglR upstream region was PCR amplified introducing *EcoRI* and *BglII* sites respectively. A dorc6DSBglF-pBSF2 downstream region was also PCR amplified introducing a novel *BglII* site. The two PCR products were digested with *BglII* and ligated together. This product was then ligated into pTA131 at *EcoRI* and *KpnI* sites to create the *orc6* deletion construct pTA1521. To create pTA1799 *orc6::trpA+* deletion construct *BamHI* digested *p.fdx-trpA* from pTA298 was ligated into *BglII* in pTA1521 (Figure 3.5).

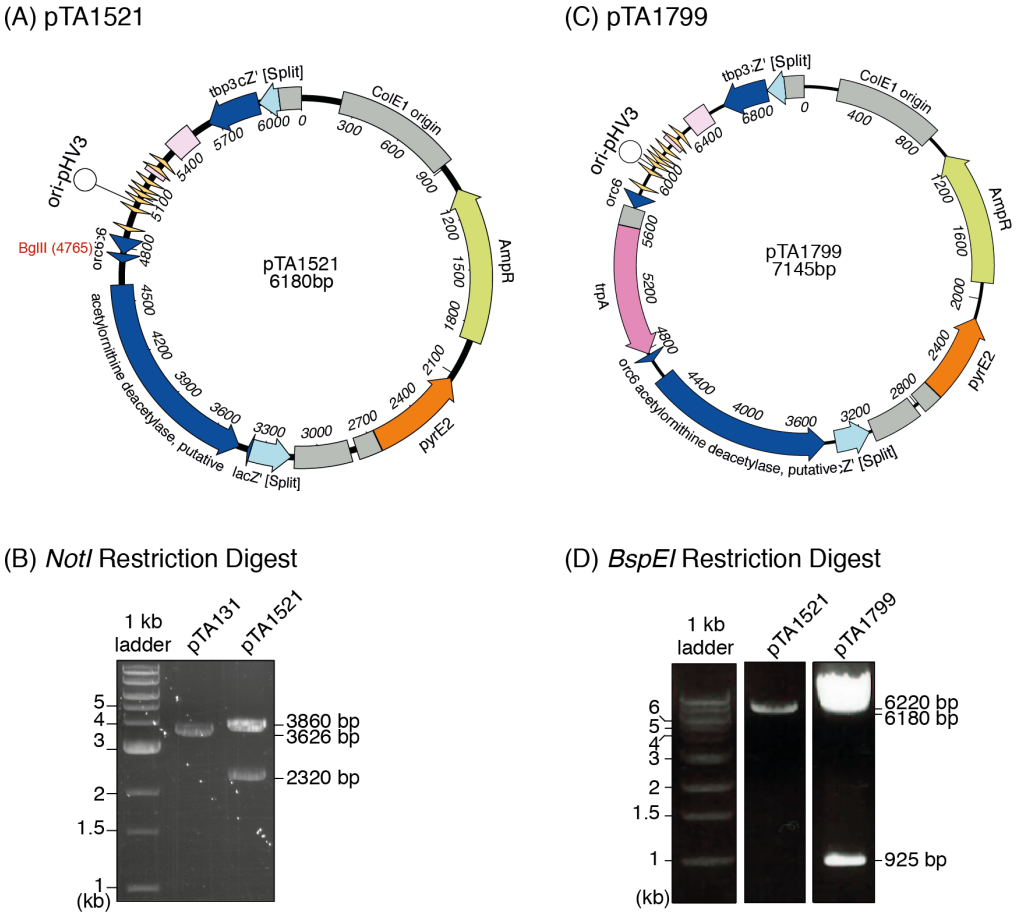


Figure 3.5 ***orc6* deletion constructs**. **(A)** pTA1521 showing internal *BglII* site. **(B)** *NotI* restriction digest of pTA131 giving a band of 3626 bp and pTA1521 giving bands of 3860 bp and 2320 bp, as predicted. **(C)** pTA1799 *orc6* deletion construct with *trpA* marker gene. **(D)** *BspEI* restriction digest of pTA1521 giving a single band of 6220 bp and pTA1799 giving bands of 6180 bp and 925 bp, as predicted.

pTA1522 (*ori-pHV3 orc6* deletion construct) and pTA1800 (*ori-pHV3 orc6::trpA+* deletion construct)

From the genomic clone pTA1509, the primers doripHV3USBglR and pBSR3 were used to PCR amplify the upstream region, introducing a *BglII* site. The downstream was PCR amplified using the primers dorc6DSBglF and pBSF2 introducing a *BglII* site. The two PCR products were digested with *BglII* and ligated. The resulting product was then ligated into pTA131 at *XbaI* and *KpnI* sites. This created the *ori-pHV3 orc6* deletion construct pTA1522. The plasmid pTA298 was digested with *BamHI* to excise the *p.fdx-trpA* fragment which was ligated into compatible the *BglII* site in pTA1522 to create the *ori-pHV3 orc6::trpA+* deletion construct pTA1800 (Figure 3.6).

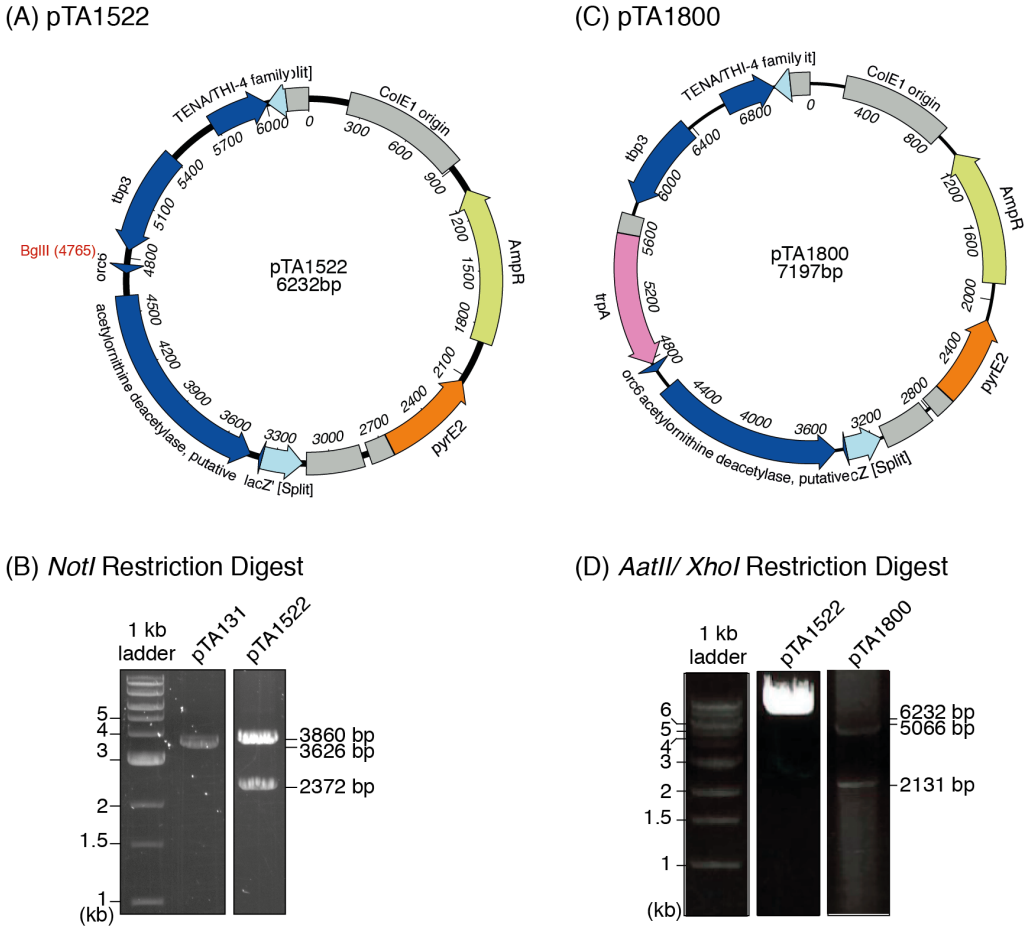


Figure 3.6 ***ori-pHV3* and *orc6* double deletion constructs**. **(A)** pTA1522 showing internal *BglII* site. **(B)** *NotI* restriction digest of pTA131 giving a band of 3626 bp and pTA1521 giving bands of 3860 bp and 2372 bp, as predicted. **(C)** pTA1800 double deletion construct with *trpA* marker gene. **(D)** *AatII/ XhoI restriction* digest of pTA1522 giving a band of 6232 bp and pTA1800 giving bands of 5066 bp and 2131 bp, as predicted.

Plasmids with tryptophan selection, pTA1798 (*Δori-pHV3::trpA+*), pTA1799 (*Δorc6::trpA+*) and pTA1800 (*Δori-pHV3 Δorc6::trpA+*), were transformed into strains H53 (wild-type) and H1804 (origin deleted). Pop-in strains were successfully generated, however it proved impossible to derive deletion strains; instead, all pop-outs reverted back to the original wild-type genotype. As the example in Figure 3.7B shows, when colonies were patched and subjected to colony hybridisation with either an *ori-pHV3* or *orc6*-specific probe, all colonies hybridised with the probe. The *ori-pHV3*-specific probe was generated by PCR using the primers oripHV3probeF and oripHV3probeR, and the *orc6* specific probe was also generated by PCR using the primers orc6probeF and orc6probeR (Figure 3.7A). This process was repeated several times with different pop-in strains and the same result was obtained each time. This indicates that neither *ori-pHV3* or *orc6* can be deleted from pHV3 when it is episomal, meaning that both *ori-pHV3* and *orc6* are essential for the maintenance for pHV3 in the cell. This could be due to the fact there are essential genes located on pHV3 (Hartman *et al*. 2010) and therefore it is vital that the mega-plasmid is maintained, however the question remains as to why *ori-pHV3* is essential when the origins found on the main chromosome are not, and why pHV3 cannot be replicated via recombination-dependent replication instead?

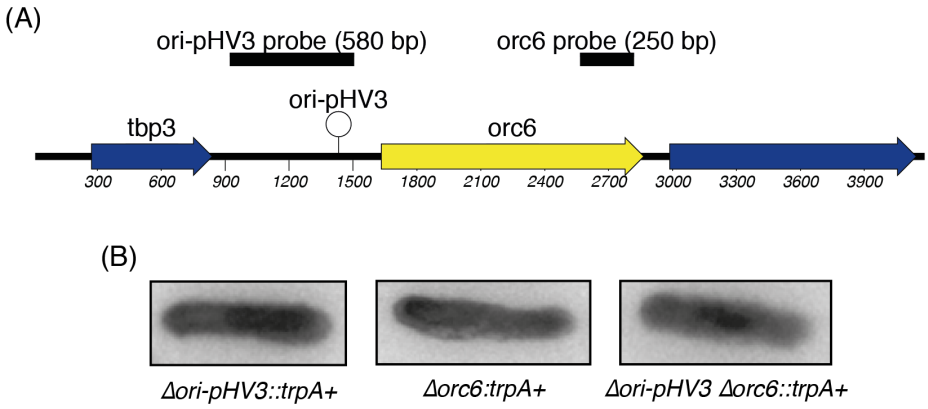


Figure 3.7 ***ori-pHV3* and *orc6* cannot be deleted from the episomal pHV3 mega-plasmid**. **(A)** *ori-pHV3* and *orc6* specific probes were generated by PCR using the primers oripHV3probeF and oripHV3probeR (580 bp), and orc6probeF and orc6probeR (250 bp), respectively. **(B)** Examples of colony hybridisations in each of the attempted deletions *Δori-pHV3::trpA+*, *Δorc6::trpA+* and *Δori-pHV3 Δorc6::trpA+* in H53 (wild-type). In each case the probe hybridises to the colony, indicating that the region has not been deleted. For the *Δori-pHV3::trpA+* and *Δori-pHV3 Δorc6::trpA+* the *ori-pHV3* specific probe was used and for the *Δorc6::trpA+* the *orc6* specific probe was used.

3.3.2 Deletion of *ori-pHV4*

The pHV4 origin region can be seen in Figure 3.8 (for full alignments of ORBs and DUEs see Appendix *II DUE and ORB Alignments*).

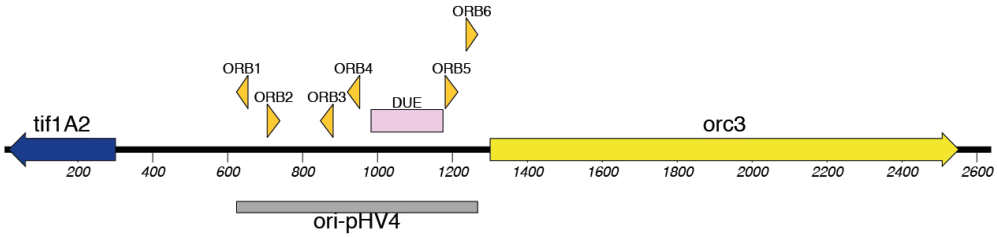


Figure 3.8 **pHV4 origin region and associated *orc* gene**. 631 bp *ori-pHV4* region is highlighted (grey box) with duplex unwinding elements (DUEs) (pink box) and origin recognition boxes (ORBs) (orange arrows) shown. Orientation of ORBs is indicated by arrows. Origin is flanked by a translation initiation factor (*tif1A2*) upstream and *orc3* downstream. Adapted from Norais *et al*. (2007).

pTA1331 (*ori-pHV4::trpA+* deletion construct)

pTA1331 *ori-pHV4::trpA+* deletion construct was constructed by Thorsten Allers in 2011 (unpublished data) (Figure 3.9). The origin has been replaced by a *trpA* marker gene and also contains upstream and downstream flanking regions including *orc3*.

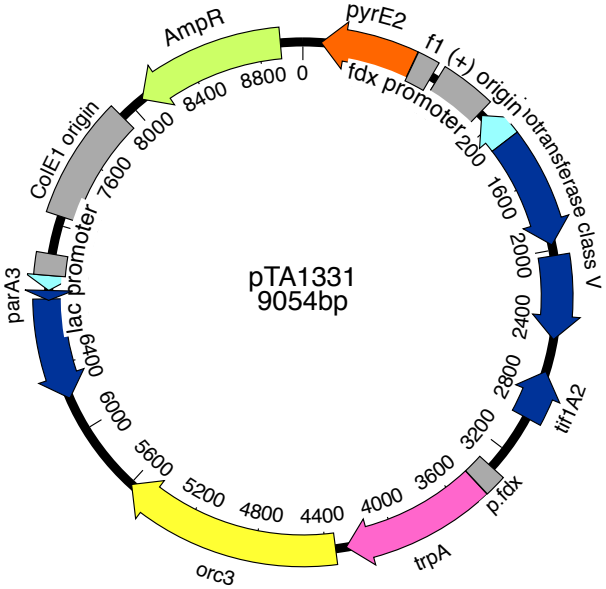


Figure 3.9 **pTA1331**. *ori-pHV4::trpA+* deletion construct. Constructed by Thorsten Allers in 2011 (unpublished data).

Integrated pHV4

In the laboratory strain H26 and its derivatives, the mega-plasmid pHV4 has integrated onto the main chromosome; this was an unintended consequence of curing the plasmid pHV2 from the wild-type strain *H. volcanii* DS2 (Wendolski *et al*. 2001; Hawkins *et al*. 2013). The origin on pHV4, *ori-pHV4*, has been previously been deleted from the integrated pHV4 mega-plasmid by transforming H53 with the plasmid pTA1331 (*Δori-pHV4::trpA+*) and was characterised by Hawkins *et al*. (2013); this strain, H1539, was used in this study.

Episomal pHV4

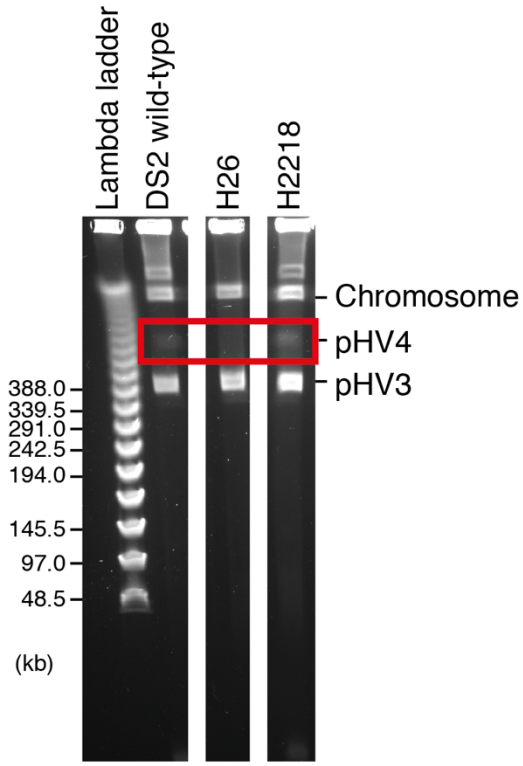


Figure 3.10 **In H2218, pHV4 mega-plasmid is episomal**. H2218 and its parent DS2 maintain pHV4 as an episomal mega-plasmid. In H26, pHV4 has been integrated onto the main chromosome (red box indicates pHV4 band). Adapted from Thorsten Allers (personal communication).

To test whether pHV4 exceeds the proposed size threshold required for efficient recombination-dependent replication, *ori-pHV4* was deleted from a strain containing an episomal pHV4 (H2218). H2218 is a *ΔpyrE2 ΔtrpA* derivative of the wild-type strain *H. volcanii* DS2, it was confirmed by pulsed field gel

(Thorsten Allers, personal communication) that pHV4 is episomal in H2218 (Figure 3.10), rather than integrated onto the main chromosome. Therefore, it can be tested as an individual plasmid rather than as part of the main chromosome.

In this study, H2218 was transformed with the plasmid pTA1331 to produce the *Δori-pHV4::trpA+* strain H2589. This was confirmed by colony hybridisation with a PCR generated probe using the primers ori_pHV42_F and ori_pHV4_R (Figure 3.11). Therefore, *ori-pHV4* can be deleted from an episomal pHV4 mega-plasmid (as well as the integrated pHV4). This demonstrates that pHV4 (636 kb) can be replicated using recombination-dependent replication, unlike pHV3 (438 kb), providing further evidence that there is a size dependent threshold of between 438 and 636 kb.

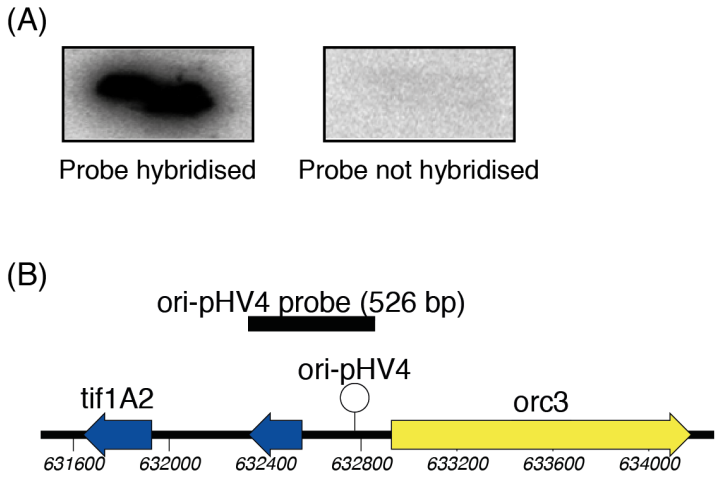


Figure 3.11 **Deletion of *ori-pHV4::trpA+* in H2218**. **(A)** Examples of colony hybridisations of *Δori-pHV4::trpA+* in H2218 with episomal pHV4. In most cases the probe does not hybridise, indicating that *ori-pHV4* has been deleted generating the strain H2589 (shown above). In some cases, the probe hybridised to the colonies indicating the region has not been deleted. **(B)** A specific probe for the *ori-pHV4* (526 bp) was generated by PCR using the primers ori_pHV42_F and ori_pHV42_R.

3.3.3 Integration of pHV3 onto the Main Chromosome

The pHV4 origin *ori-pHV4* can be deleted from both episomal and chromosomally integrated pHV4, whereas *ori-pHV3* cannot be deleted from episomal pHV3. In contrast to pHV4, there are no known naturally occurring strains where pHV3 has been integrated into the main chromosome. In order to determine whether the pHV3 origin itself is essential (regardless of where it is

located in the genome), or whether pHV3 is unable to utilise recombination-dependent replication in the absence of an origin, pHV3 must be integrated into the main chromosome. Briefly, two strains were generated with different mutant *leuB* alleles for leucine biosynthesis, one on the chromosome and one on pHV3. When these strains are mated, the *leuB* alleles recombine to generate a Leu$^+$ cell, thereby integrating pHV3 into the chromosome at the *leuB* locus (Figure 3.12).

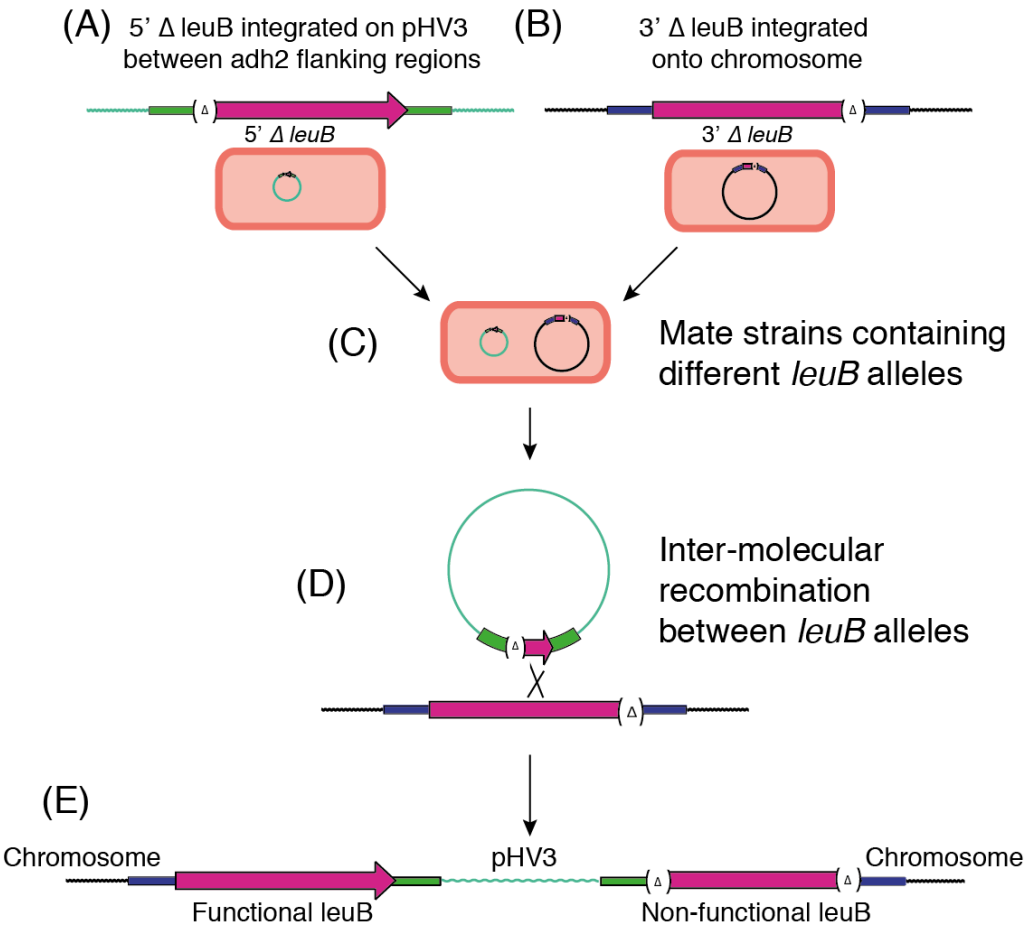


Figure 3.12 **Scheme for the integration of pHV3 onto the main chromosome**. **(A)** A strain containing the mutant *leuBΔ5'Dd* allele on pHV3 at the *adh2* locus **(B)** and a strain containing the mutant *leuBΔ3'Bb* allele on the main chromosome **(C)** are mated. **(D)** Intermolecular recombination occurs between the pHV3 mega-plasmid and the main chromosome **(E)** to produce a strain with pHV3 integrated on the chromosome flanked by a functional *leuB* and non-functional *leuBΔ5'DdΔ3'Bb* allele.

Leucine Alleles

Two leucine alleles were constructed (prior to this study), *leuBΔ3'Bb* and *leuBΔ5'Dd*, which are missing the 3' end of the gene and the 5' end of the gene, respectively. Both these alleles inactivate the *leuB* gene, therefore if the functional *leuB* gene is replaced by one of these alleles, cells will be deficient in the production of leucine. Leucine was chosen as a selectable marker so that later gene deletions could still utilise tryptophan and uracil selection (Allers *et al*. 2004).

The plasmid pTA700 replaces *leuB* (on the chromosome) with the *leuBΔ3'Bb* allele and the plasmid pTA1621 replaces the gene *adh2* (on pHV3) with *leuBΔ5'Dd*. The *adh2* gene was chosen as a locus for the gene replacement as it is located away from the origin (and therefore will not affect DNA replication), and has previously been successfully deleted with no adverse effects on cell growth (Timpson *et al*. 2013).

pTA700 (*leuBΔ3'Bb* gene replacement construct)

The plasmid pTA700 was constructed by Sam Haldenby in 2006 (unpublished data) (Figure 3.13). It contains the inactive *leuBΔ3'Bb* allele of the *leuB* gene which is missing the 3' end of the gene.

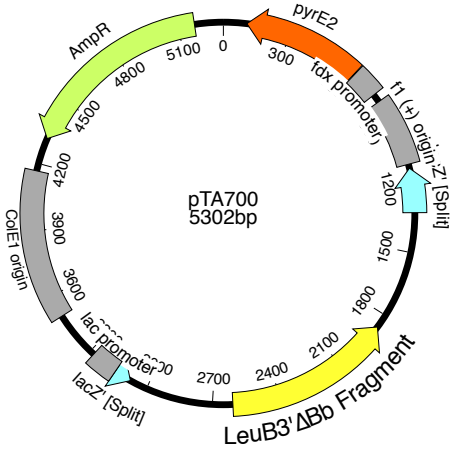


Figure 3.13 **pTA700**. *leuBΔ3'Bb* fragment. Constructed by Sam Haldenby in 2006 (unpublished data).

pTA1621 (*Δadh2::leuBΔ5'Dd* gene replacement construct)

The plasmid pTA1621 was constructed by Thorsten Allers in 2015 (unpublished data). It contains a *DdeI* fragment of *leuB* which removes the first 8bp from the 5' end of the *leuB* gene inactivating it forming the *leuBΔ5'Dd* allele. This has then been inserted in the *adh2* deletion construct meaning the *leuBΔ5'Dd* allele will be replacing the *adh2* gene, hence *Δadh2::leuBΔ5'Dd* (Figure 3.14).

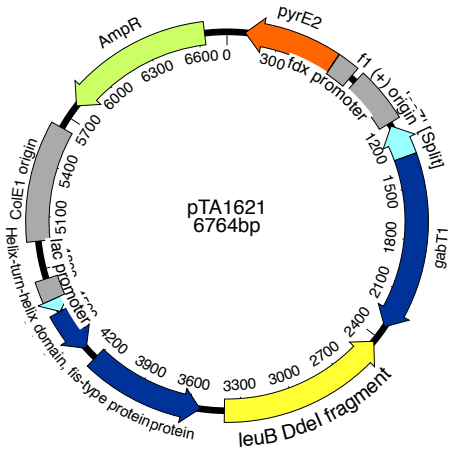


Figure 3.14 **pTA1621**. *leuBΔ5'Dd* fragment inserted into the *adh2* deletion construct to make the *Δadh2::leuBΔ5'Dd* gene replacement construct. Constructed by Thorsten Allers in 2015 (unpublished data).

Strain Construction

Strains H53, H1804, H2175, H2183 and, H2184 were transformed with pTA700 (*leuBΔ3'Bb*) to produce the strains H2163, H3210, H3211, H3216 and H3217, respectively. The strains H2240, H3726, H3728, H3730 and H3733 were generated by transforming H53, H1804, H2175, H2183 and H2184, respectively, with the plasmid pTA1621 (*leuBΔ5'Dd*). These strains were then mated together to produce the strains H2461, H3794, H3796, H3798 and, H3800, as can be seen in Table 3.1 (for details see *Chapter 2: Materials and Methods, Section 2.2.5 Genetic Manipulation of Haloferax volcanii*.

It should be noted that the two alleles of *leuB* were transformed separately into the different strain backgrounds, which were then subsequently mated. During mating two *Haloferax volcanii* cells form cytoplasmic bridges which join them together, followed by fusion of the cell membranes and S-layer. The

cytoplasms to not mix, however, genetic material can be exchanged (Rosenshine *et al* 1989; Zhao *et al*. 2006). This does not necessarily have to occur between two *H. volcanii* cells as it can also be achieved with closely related strains e.g. *H. volcanii* and *H. mediterranei* (Naor *et al*. 2012). This was undertaken to avoid the possibility of recombination between the two *leuB* alleles (on the plasmids) in a manner that would result in a Leu^+ cell without forcing the integration of pHV3 onto the chromosome. For example, transformation of a strain carrying the chromosomal *leuBΔ3'Bb* allele with pTA1621 (carrying *leuBΔ5'Dd*) could result in recombination of the two *leuB* alleles, rather than integrating the latter plasmid on pHV3 by recombination at the *adh2* locus. The new genome structure of the strain H2461 can be seen in Figure 3.15).

Table 3.1 **Strains mated to integrate pHV3 onto the chromosome.**

Parent 1 (*leuBΔ3'Bb*)	**Parent 2 (*leuBΔ5'Dd*)**	**Integrated pHV3 Strain**	**Background Genotype**
H2163	H2240	H2461	Wild-type (H53)
H3210	H3726	H3794	*ΔoriC1,2,3,pHV4* (H1804)
H3211	H3728	H3796	*ΔoriC1,2,3,pHV4,pHV1* (H2175)
H3216	H3730	H3798	*ΔoriC1,2,3,pHV4 ΔpHV1* (H2183)
H3217	H3733	H3800	*ΔpHV1* (H2184)

As shown in Figure 3.15, the integrated pHV3 mega-plasmid is now located close to the origin, *oriC2*, which means that both *oriC2* and *ori-pHV3* are now in close proximity to one another. This may have an effect on replication and transcription, especially as there is an rRNA locus located next to *oriC2*. In both processes, the DNA must be unwound and forks may collide more frequently causing fork stalling, which must then be repaired before transcription or DNA replication can continue. Another issue with the proximity of the two origins is that the origins may fire at different times or with different frequencies, which may also interfere with DNA replication and/or transcription. A way to test if this is an issue would be to integrate pHV3 into other loci on the main chromosome, however this was not possible within the time frame of this study.

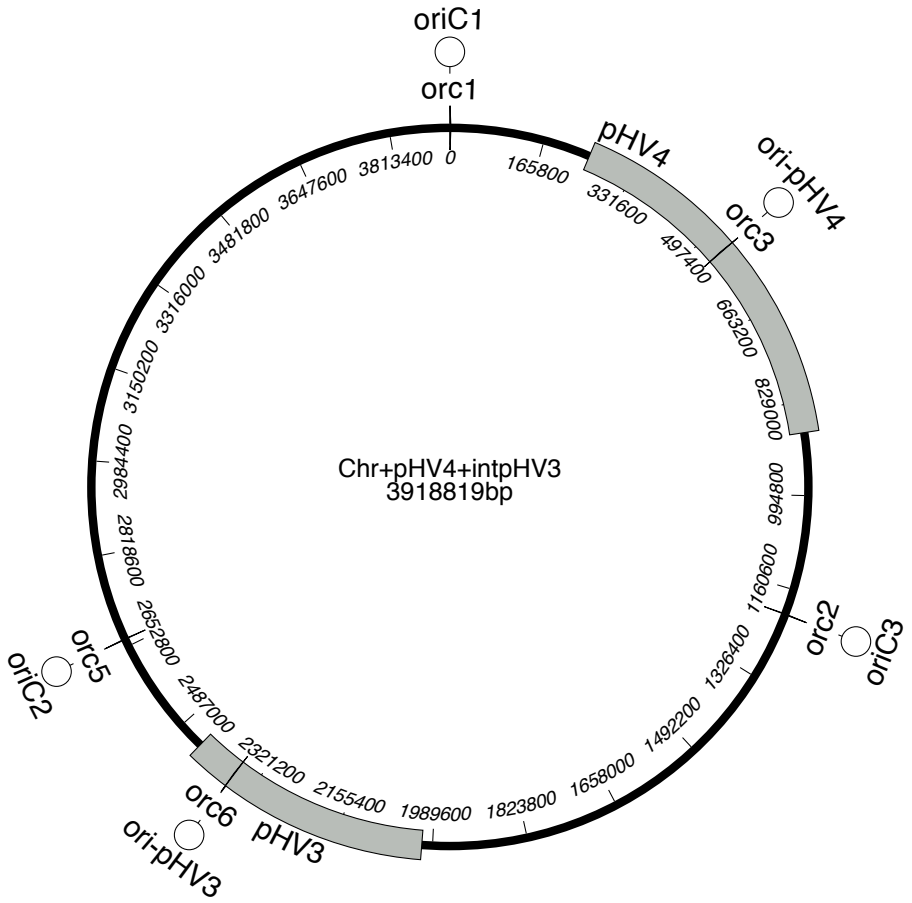


Figure 3.15 **Locations of chromosomal origins and their associated *orc* genes, including integrated pHV4 and pHV3 mega-plasmids**. The main chromosome with integrated mega-plasmids pHV3 and pHV4 of *Haloferax volcanii* has five origins of replication.

Exact sites of the integration of pHV3 are shown in Appendix *VI Sites of pHV3 Integration*. Strains were verified by colony PCR (Figure 3.16), which looked at the ends where pHV3 and the chromosome joined. Colony PCR used the primer pairs FleuBpHV3chrF and FleuBpHV3chrR (product size 1630 bp) and, NFleuBpHV3chrF and NFleuBpHV3chrR (product size 1320 bp).

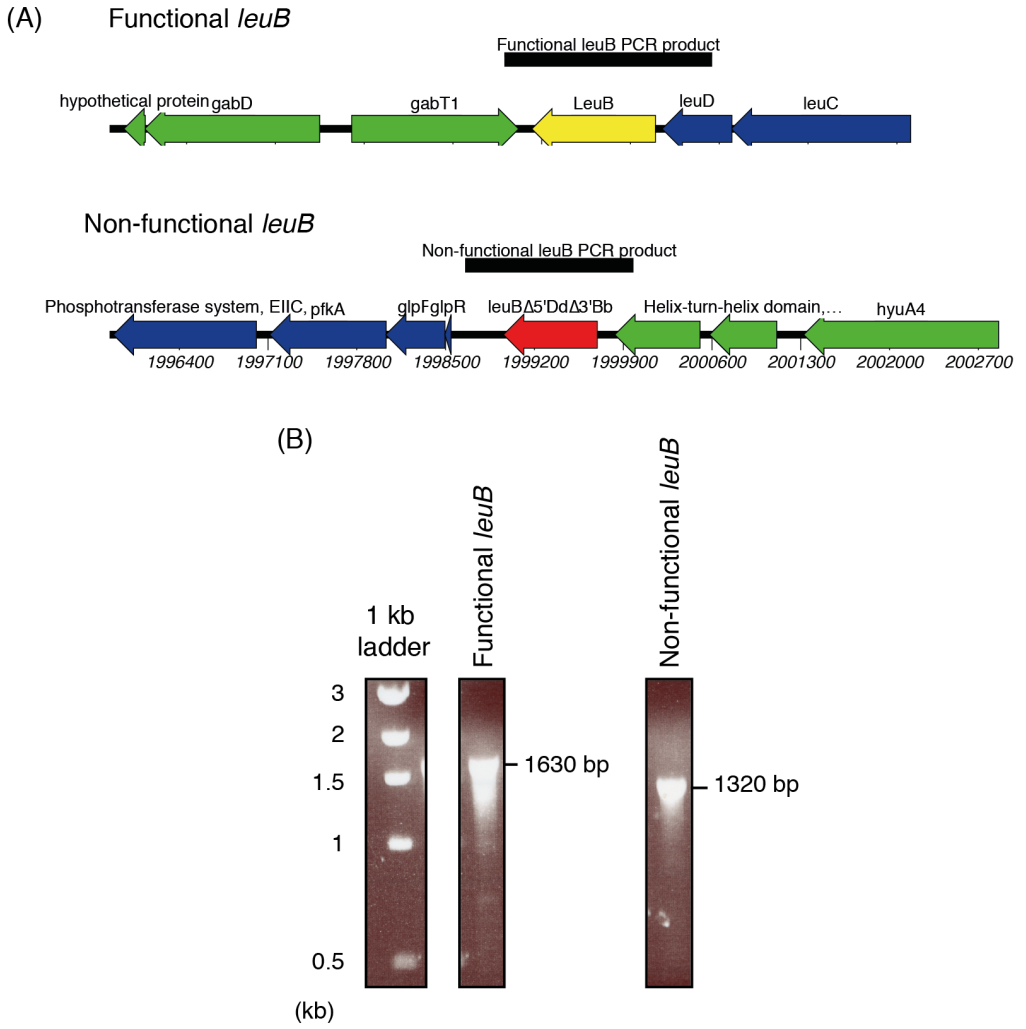


Figure 3.16 **Integrated pHV3 strains were verified by PCR**. **(A)** PCR products produced by FleuBpHV3chrF and FleuBpHV3chrR (functional *leuB*) and, NFleuBpHV3chrF and NFleuBpHV3chrR (non-functional *leuB*). Green indicates genes on pHV3, blue indicates genes found on the chromosome, yellow indicates functional *leuB* gene and red indicates non-functional *leuBΔ5'DdΔ3'Bb* allele. **(B)** PCR of H2461. Correct bands sizes of 1630 bp and 1320 bp for functional and non-functional *leuB*, respectively, showing pHV3 has integrated onto the chromosome.

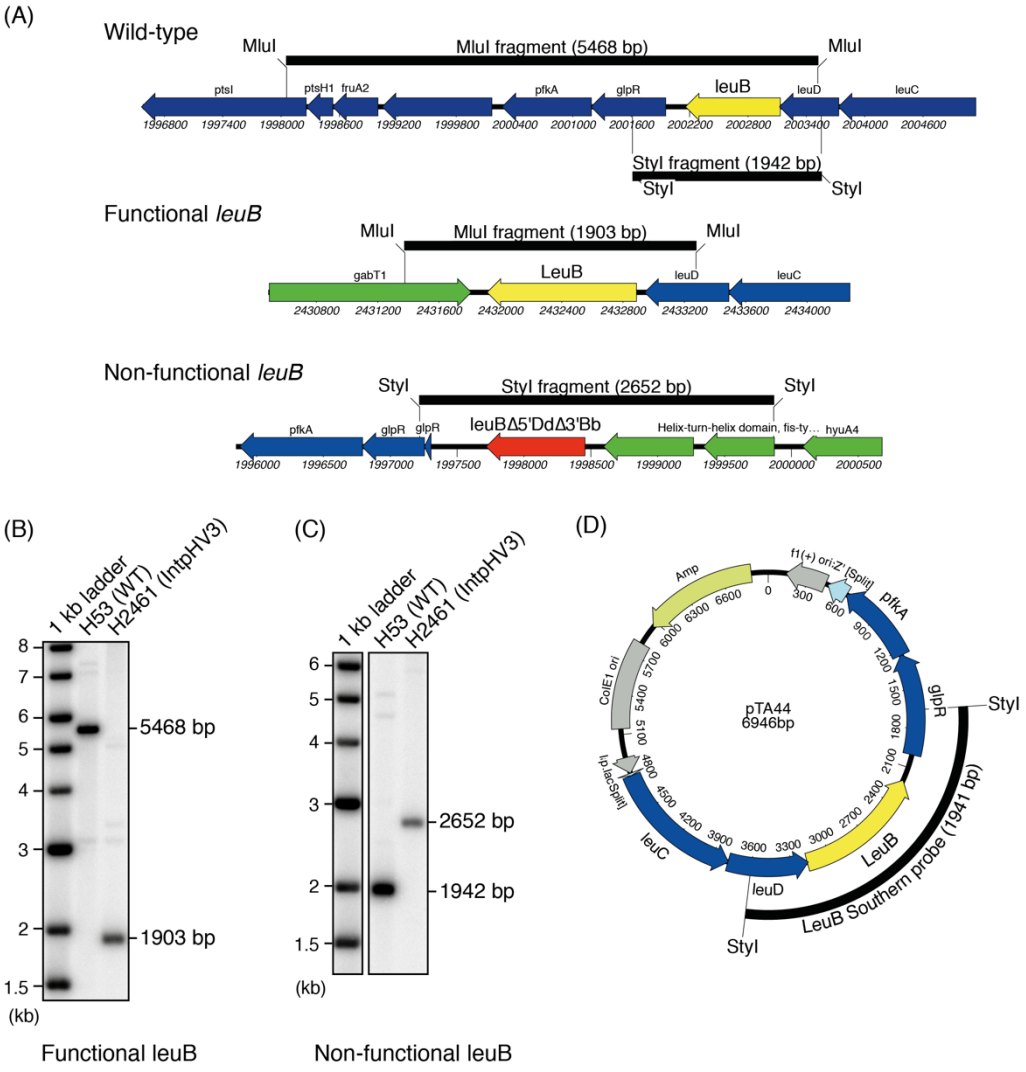


Figure 3.17 **Integrated pHV3 strains were verified by Southern blot**. **(A)** Predicted Southern blot fragment sizes. Functional *leuB* (yellow) and non-functional *leuB* (red) flank the integration of pHV3 (green) onto the main chromosome (dark blue). Functional *leuB* fragments were digested with *MluI* giving a 5468 bp fragment for wild-type and a 1903 bp fragment for integrated pHV3. Non-functional *leuB* fragments were digested with *StyI* giving a 1942 bp fragment for wild-type and a 2652 bp fragment for integrated pHV3. **(B)** H53 (wild-type) and H2461 (intpHV3) were digested with *MluI* and Southern blotted. When analysed they gave the correct predicted fragment sizes for the functional *leuB*. **(C)** H53 (wild-type) and H2461 (intpHV3) were digested with *StyI* and Southern blotted. When analysed they gave the correct predicted fragment sizes for the non-functional *leuB*. **(D)** Probe used for Southern blots of *LeuB* region. pTA44 was digested with *StyI* (1941 bp).

Alongside colony PCR, integration of the pHV3 mega-plasmid was confirmed by Southern blot. H2461 (IntpHV3, wild-type) is shown as an example in Figure 3.17. For this the sites of integration between pHV3 and the chromosome were tested, one looking at the end where there is a functional *leuB* gene present and one where the non-functional *leuB* is present (Figure 3.17A). The integration of

pHV3 onto the chromosome causes a change in band sizes compared to the wild-type strain (Figure 3.17B). The Southern blot was probed with the *leuB* region digested using *StyI* from the genomic clone of *leuB* pTA44 (Figure 3.17C).

Parental strains and pHV3 integrated strains were also verified by pulsed-field gel. Whole genomic DNA was treated with approximately 35 Gy (Gammacell 1000 Elite Model ^{137}Cs source) to cause single, double-strand breaks in the circular DNA therefore linearising the DNA and allowing it to enter the gel (circular DNA cannot enter agarose gels). The DNA was then analysed by pulsed-field gel electrophoresis, which was subsequently subjected to Southern blot (Figure 3.18). The Southern blot was probed with pHV3probe2 that targets the *narH* gene (HVO_B0165), which is in a different region to the origin/ *orc* and to *adh2*. (Figure 3.18B) The probe was made via PCR amplifying genomic DNA using the primers pHV3probe2F and pHV3probe2R.

All parental strains H53 (wild-type), H1804 (*ΔoriC1,2,3,pHV4*), H2175 (*ΔoriC1,2,3,pHV4,pHV1*), H2183 (*ΔoriC1,2,3,pHV4, ΔpHV1*), and H2184 (*ΔpHV1*) have bands corresponding to episomal pHV3 (438 kb), showing that pHV3 has not naturally integrated onto the main chromosome. H2461 (IntpHV3, wild-type), H3796 (IntpHV3, *ΔoriC1,2,3,pHV4,pHV1*), H3798 (IntpHV3, *ΔoriC1,2,3,pHV4, ΔpHV1*), H3800 (IntpHV3, *ΔpHV1*), H3512 (IntpHV3, *Δori-pHV3 Δorc6*) and H3954 (IntpHV3, *ΔpHV1, Δori-pHV3::trpA+*) only have bands corresponding to pHV3 being integrated onto the main chromosome, meaning that there is no episomal pHV3 present and all copies have integrated onto the main chromosome. H3794 (*ΔoriC1,2,3,pHV4,* IntpHV3) has bands corresponding to both integrated and episomal pHV3, which means that it is mero-diploid and either not all the copies of pHV3 were integrated onto the main chromosome or the integration process has subsequently reversed via recombination between the two *leuB* alleles still present in the genome at either end of the integration. H3974 requires re-making before further experiments can be done. The presence of bands in all samples shows that all of the strains have retained the pHV3 mega-plasmid despite genetic manipulation.

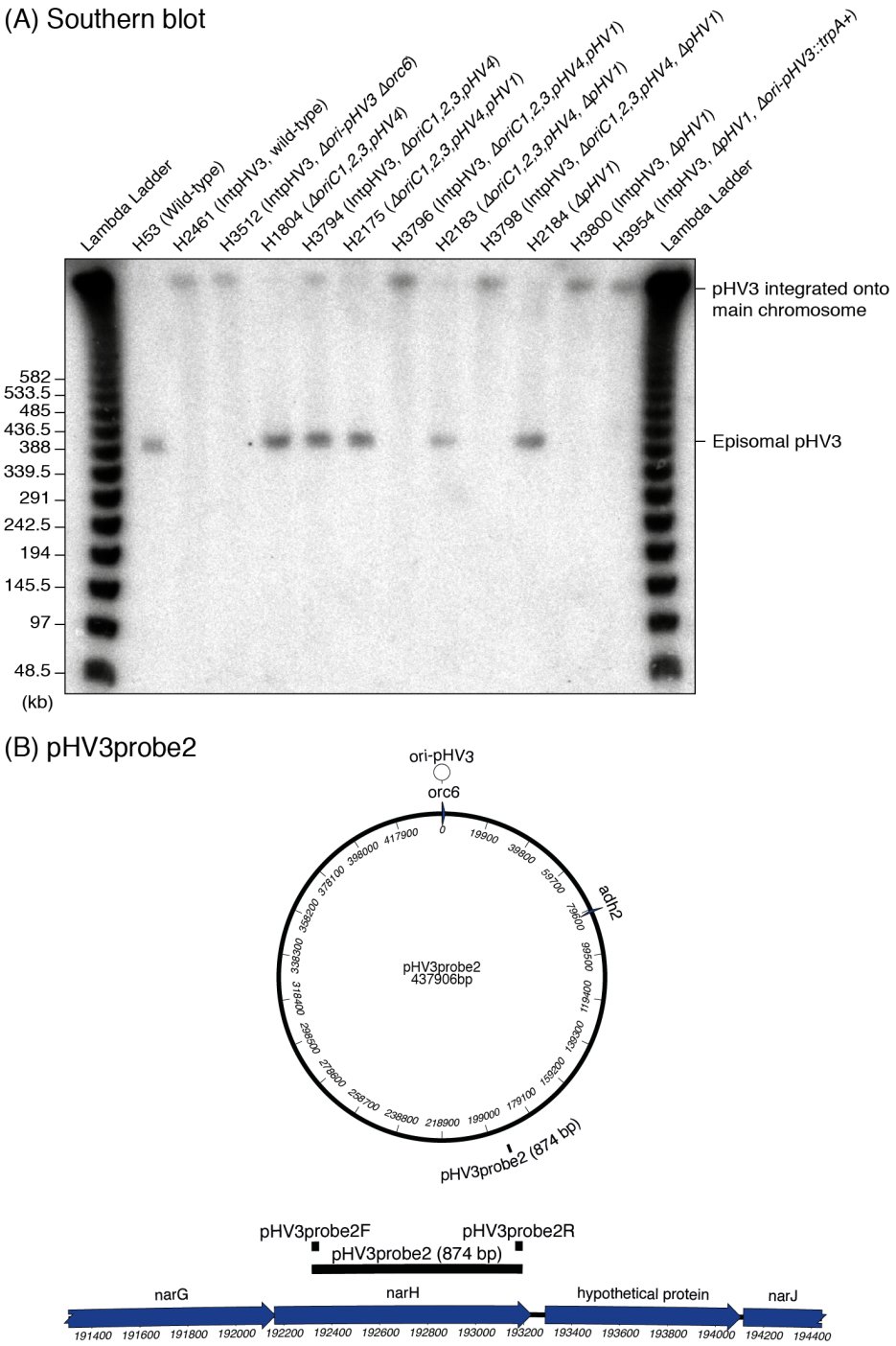


Figure 3.18 **Integrated pHV3 strains were verified by Pulsed-Field Gel**. **(A)** Southern blot of a Pulsed-field gel of whole genomic DNA. IntpHV3 denotes pHV3 integrated onto the main chromosome. All parental strains H53 (wild-type), H1804 (*ΔoriC1,2,3,pHV4*), H2175 (*ΔoriC1,2,3,pHV4,pHV1*), H2183 (*ΔoriC1,2,3,pHV4, ΔpHV1*), and H2184 (*ΔpHV1*) have bands corresponding to episomal pHV3 (438 kb). H2461 (wild-type, IntpHV3), H3796 (*ΔoriC1,2,3,pHV4,pHV1,* IntpHV3), H3798 (*ΔoriC1,2,3,pHV4, ΔpHV1,* IntpHV3), H3800 (*ΔpHV1,* IntpHV3), H3512 (IntpHV3, *Δori-pHV3 Δorc6*) and H3954 (IntpHV3, *ΔpHV1, Δori-pHV3::trpA+*) only have bands corresponding to IntpHV3. H3794 (*ΔoriC1,2,3,pHV4,* IntpHV3) has bands corresponding to both integrated and episomal pHV3. Blot was probed using pHV3probe2. **(B)** pHV3probe2 (874 bp) was produced by amplifying genomic DNA using primers pHV3probe2F and pHV3probe2R. Diagram shows location of probe on pHV3, specifically targeting the *narH* gene (HVO_B0165)

Stability of pHV3 Integration

As there are two alleles of *leuB* still present in the strains where pHV3 has been integrated onto the chromosome, it is possible that when selection for the *leuB* gene pressure is relieved the integration could be reversed, making pHV3 episomal again. Another possibility is that some copies of pHV3 may become episomal and that some may stay integrated on the chromosome (mero-diploid), as there are 20 copies of the chromosome and pHV3. It could also be the case that the integrated pHV3 is stable and does not become episomal.

To test this, H53 (wild-type) and H2461 (integrated pHV3) were grown for 28 consecutive overnights, diluting each day, and subjected to pulsed-field gel electrophoresis to look at the bands present (Figure 3.19). As shown in Figure 3.19, there is a band thought to be corresponding to the episomal pHV3 mega-plasmid in the wild-type lane (H53). In H2461 which has been grown for only one overnight, there is no band at that location indicating that all of the copies of pHV3 are integrated onto the main chromosome, however in H2461 which has been grown for 28 consecutive overnights, there is a faint band seen. As the band is not as strong as what is seen in the wild-type, this suggests that some of the copies of pHV3 have become episomal and the integration has been partially reversed. This would be beneficial for the strain as the integration may have disrupted gene clusters/ operons on both pHV3 and the chromosome which may not be functioning properly as a result of the integration, and if pHV3 is episomal these gene neighbourhoods will be intact. Nonetheless, keeping some of the copies of pHV3 integrated would confer the benefit of the functional *leuB* gene as well. Ideally the location of pHV3 should be confirmed via Southern blot of the PFG, however there was not sufficient time as part of this project to complete a Southern blot of the PFG to definitively confirm the location of pHV3 in the genome.

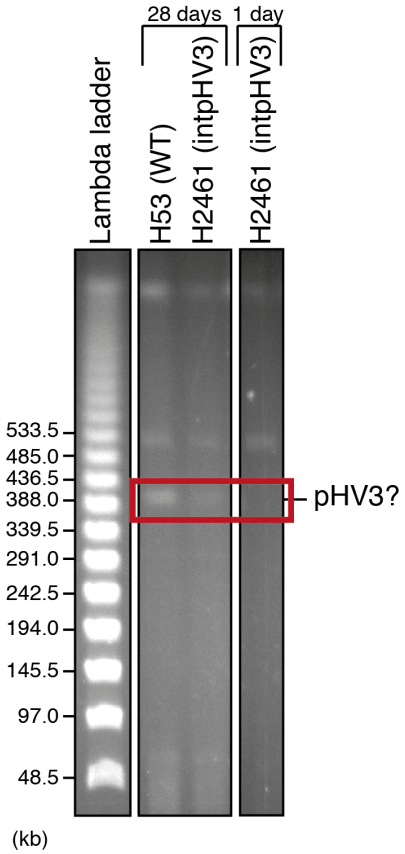


Figure 3.19 **The integration of pHV3 is not stable**. Pulsed-field gel of H53 (wild-type) and H2461 (integrated pHV3) grown for 28 consecutive overnights and H2461 grown for one overnight prior to agarose plugs being made. In H2461 (one overnight) there is no band corresponding to what is thought to be episomal pHV3, showing all copies of pHV3 are integrated onto the main chromosome. In H53 (28 overnights) there is a band thought to be episomal pHV3 and in H2461 (28 overnights) there is a faint band thought to be pHV3 therefore indicating that some copies of pHV3 are no longer integrated onto the main chromosome. Red box indicates band thought to be pHV3.

Deletion from Integrated pHV3

The wild-type strain with pHV3 integrated onto the main chromosome, H2461, was then transformed with the plasmids pTA1520, pTA1521 and, pTA1522 to delete *ori-pHV3*, *orc6* and *ori-pHV3 orc6*, respectively. This was to test whether it is the origin and/or associated *orc* gene that are essential, or whether it is essential that the rest of the mega-plasmid is maintained. This also examines the hypothesis that there is a size threshold for efficient recombination-dependent replication. Plasmids without tryptophan selection were used as the strain H2461 is not deleted for tryptophan despite being derived from H53 (*ΔpyrE2, ΔtrpA*).

This is because the other parent strain used in the generation of H2461, H2240, was not *ΔtrpA*, meaning that during the transfer of genetic material the daughter (H2461) gained the *trpA* gene and became Trp+.

It was found to be possible to delete *ori-pHV3* and *orc6* together in the strain H2461, generating the strain H3512. However, it was not possible to delete *ori-pHV3* or *orc6* singularly. These were confirmed by colony hybridisation with a probe generated by PCR using the primers oripHV3probeF and oripHV3probeR, and orc6probeF and orc6probeR, for *ori-pHV3* and *orc6* respectively. The *ori-pHV3*-specific probe was used for the *Δori-pHV3* and *Δori-pHV3 Δorc6* deletions, and the *orc6* specific probe was used for the *Δorc6* deletion (Figure 3.20). This shows that neither the origin or the associated *orc* gene are essential for the survival of the cell. The fact they could not be deleted individually is most likely to be due to the lack of additional tryptophan selection pressure in the deletion constructs, although it is possible that leaving either the origin or the associated *orc* gene may disrupt the balance of DNA replication in the cell, for example if *ori-pHV3* is deleted this means that Orc6 will no longer be able to bind there meaning there is an excess of Orc6 in the cell, which could in turn interfere with other origin/Orc binding. Deletion of *orc6* may mean that *ori-pHV3* is no longer functional as an origin or that it is then able the use other Orc proteins to fire, sequestering them away from other origins. It is not known at this point if Orc6 exclusively binds to *ori-pHV3* (could be binding to other origins) or whether *ori-pHV3* exclusively binds *orc6* (other Orcs may be required for firing). As pHV3 is integrated onto the main chromosome and the origin region is relatively close to the integration site with the chromosome and to the chromosomal origin *oriC2* and its associated *orc, orc5* (Figure 3.15) deleting *ori-pHV3* or *orc6* individually may also have an effect on the replication of the genes in that region. The integration itself may disrupt genes as those that were near to *oriC2* are now further away and therefore may have a reduced gene dosage, which may affect the cell as a whole.

The deletion of *ori-pHV3* and *orc6,* both individually and together, needs to be repeated in the future to ensure that this is the case and to further investigate the relationship of *ori-pHV3* and *orc6* with other DNA replication mechanisms in

Haloferax volcanii. It is highly probable that *ori-pHV3* and *orc6* can be deleted individually, however they may require the additional selective pressure of a *trpA* marker in the deletion construct.

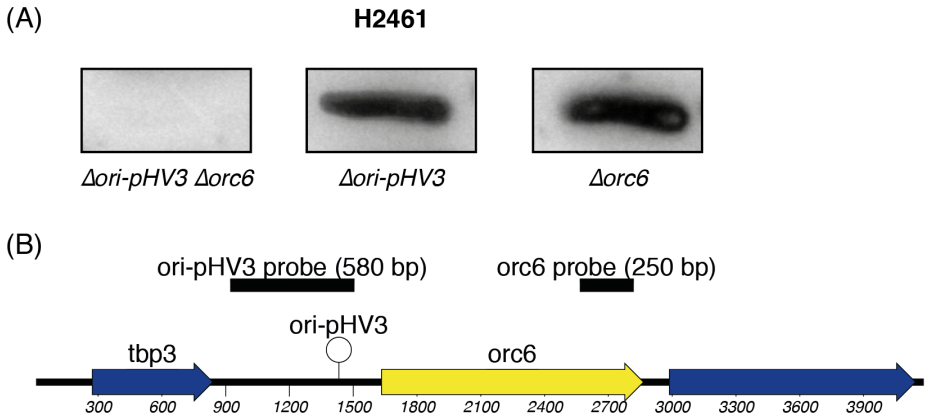


Figure 3.20 ***ori-pHV3* and *orc6* can be deleted together but not individually from H2461**. **(A)** Examples of colony hybridisations in each of the attempted deletions *Δori-pHV3*, *Δorc6* and *Δori-pHV3 Δorc6* in H2461 (integrated pHV3). For the double deletion, the probe does not hybridise, indicating that *ori-pHV3* and *orc6* have been deleted. For the single deletions, the probes hybridise to the colonies indicating the region has not been deleted. **(B)** Specific probes for the *ori-pHV3* and *orc6* regions were generated by PCR using the primers oripHV3probeF and oripHV3probeR (580 bp), and orc6probeF and orc6probeR (250 bp), respectively. For the *Δori-pHV3* and *Δori-pHV3 Δorc6*, the *ori-pHV3*-specific probe was used and for the *Δorc6*, the *orc6* specific probe was used.

Deletion of *trpA* from H2461

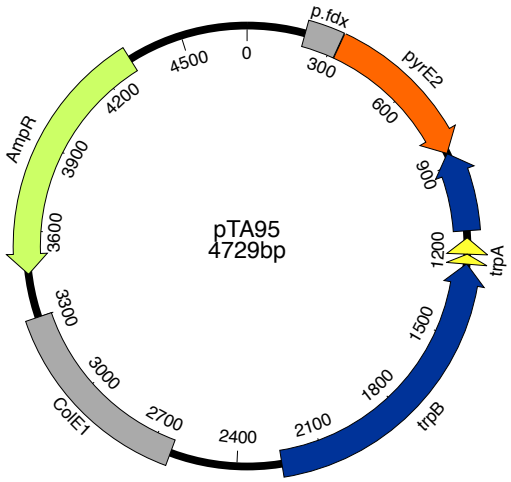


Figure 3.21 **pTA95**. *trpA* deletion construct. Constructed by Thorsten Allers in 2002 (Allers *et al*. 2004).

The plasmid pTA95 (*trpA* deletion construct) was constructed by Thorsten Allers in 2002 (Allers *et al*. 2004). It contains an asymmetric deletion construct (downstream flanking region is smaller than upstream flanking region) for the deletion of *trpA* (Figure 3.22).

In order to utilise deletion constructs that have the tryptophan selectable marker, H2461 was transformed with pTA95 to delete *trpA*. This was verified by selecting colonies unable to grow on tryptophan deficient media (Figure 3.23). This was successful and generated the strain H3718. However, due to time constraints, it was not possible to test whether *ori-pHV3* and *orc6* could be deleted individually from this strain using the *trpA* selectable marker. Since the conclusion of this study is has been shown that it is possible to delete *ori-pHV3* and *orc6* both individually and together.

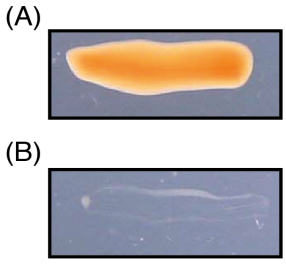


Figure 3.21 **Deletion of *trpA* gene verified by *trp-* phenotype. (A)** Colony not deleted for *trpA* is able to grow on media lacking tryptophan. **(B)** Colony deleted for *trpA* is unable to grow on media lacking tryptophan. Colonies were patched in duplicate onto media with tryptophan.

Deletion of *ori-pHV3::trpA+*

The other mated strains that have an integrated pHV3 mega-plasmid, namely H3794 (*ΔoriC1,2,3,pHV4* background), H3796 (*ΔoriC1,2,3,pHV4,pHV1* background), H3798 (*ΔoriC1,2,3,pHV4 ΔpHV1* background) and, H3800 (*ΔpHV1* background), could be transformed with the plasmid pTA1798 (*Δori-pHV3::trpA+*) since they had already been deleted for *trpA*. It was found that it was only possible to delete *ori-pHV3* from the strain H3800 (39/40 colonies deleted for *ori-pHV3::trpA+)* and not the strains H3794, H3796 and H3798 (0/40 colonies deleted) (Figure 3.22).

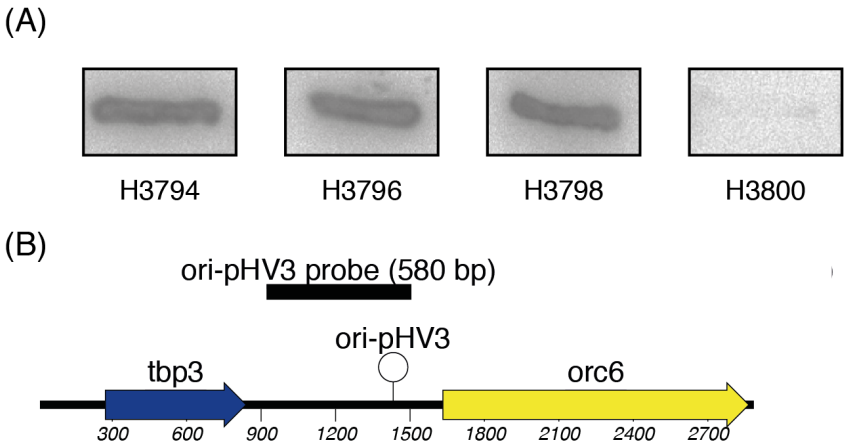


Figure 3.22 ***ori-pHV3* can be deleted from the strain H3800 but not H3794, H3796 or H3798**. **(A)** Examples of colony hybridisations in each of the attempted deletion of *Δori-pHV3::trpA+* in H3794, H3796, H3798 and H3800. In all cases except H3800, the probe hybridised to the colonies, indicating the region has not been deleted. In H3800 the probe did not hybridise, this indicating that *ori-pHV3* has been deleted. **(B)** A specific probe for the *ori-pHV3* region was generated by PCR using the primers oripHV3probeF and oripHV3probeR (580 bp).

The strains where *ori-pHV3* could not be deleted (H3794, H3798 and H3796) had been deleted for all chromosomal origins, including the integrated pHV4; by contrast, H3800 still has all the chromosomal origins. There is also the possibility that *ori-pHV3* and *orc6* are easier to delete together, but further investigation is required. It may also be possible that *ori-pHV3* is difficult to delete even with the added selection provided by the *trpA* marker gene and that more colonies need to be tested before conclusions can be drawn. In this study only 40 colonies were tested which is a low number and not enough to be statistically relevant.

The fact that so far *ori-pHV3* has only been able to be deleted in strain backgrounds that still contain the chromosomal origins may be significant. It could be the case that *ori-pHV3* is essential in the absence of the chromosomal origins and it has been seen that it is essential when pHV3 is episomal. It is possible that there are regions on pHV3 that must be replicated using origin-dependent replication rather than recombination-dependent replication. This means that *ori-pHV3* can be deleted when the mega-plasmid is integrated onto the chromosome as those regions are relatively close to *oriC2* and therefore would be replicated by that origin. In strain backgrounds where the chromosomal

origins have been deleted, *oriC2* is no longer able to replicate those regions and so *ori-pHV3* becomes essential once again.

It could also be the case that *ori-pHV3* can actually be deleted in any strain background once it has been integrated onto the main chromosome, making the size of the DNA molecule larger and above a possible threshold required for recombination-dependent replication, which the episomal pHV3 is too small to be able to accomplish efficiently, therefore making *ori-pHV3* essential.

Since the conclusion of this study it has been shown that it is possible to delete *ori-pHV3* in both of the strains H3796 (*ΔoriC1,2,3,pHV4,pHV1* background) and H3798 (*ΔoriC1,2,3,pHV4 ΔpHV1* background), showing that it is possible (but difficult) to delete. Therefore, all replication origins can be deleted from *H. volcanii* when pHV3 is integrated onto the main chromosome. This supports the proposal of a size threshold for recombination-dependent replication.

Chapter 4: Investigating DNA Replication in the Mega-plasmid pHV1

4.1 Background

pHV1 is one of three mega-plasmids in *Haloferax volcanii* and is 85 kb in size (Hartman *et al*. 2010). The mega-plasmid pHV1 is known to contain one replication origin that was characterised in a study by Norais *et al*. (2007), and two *orc* genes. One *orc* gene, *orc10*, is located next to the pHV1 origin but the other gene, *orc8*, is not origin-associated. When analysed by synonymous codon usage (Figure 4.1), the whole of pHV1 is designated non-native. When the nucleotide sequence of pHV1 is subjected to a BLASTn search (Appendix *III pHV1*), it shows homology to other closely-related archaeal species, for example *Haloferax mediterranei, Haloferax gibbonsii, Haloarcula hispanica* and *Halobacterium salinarum*, as well as to IS-elements. This suggests that pHV1 was acquired, possibly by lateral gene transfer, from another (closely-related) species.

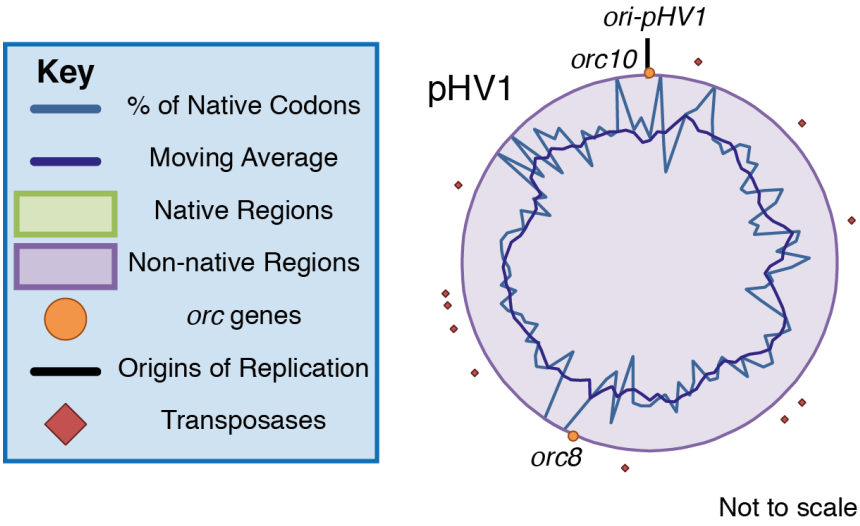


Figure 4.1 **Synonymous codon usage of pHV1**. Percentage of native codons by synonymous codon usage plotted for each gene (light blue line) and a moving average of 5 genes (dark blue line). Regions that contain 10 or more genes with >85% native codons were designated native (green shading) and regions that contain 10 or more genes with <85% native codons were designated non-native (purple shading). Origins of replication (black line), *orc* genes (orange circle) and genes annotated as transposases (red diamond) are indicated. All of the genes on pHV1 are designated as non-native. Full figure shown in Appendix *I Synonymous Codon Usage Diagrams*.

A previous study by Hawkins *et al*. 2013 found that marker frequency analysis of pHV1 gave very noisy data with no clear origin seen. This could mean that pHV1 does not ordinarily use origin-dependent replication, but uses an alternative method such as phage derived methods of DNA replication. Whether this is recombination-dependent replication is yet to be determined.

The pHV1 origin was found in an assay by Norais *et al*. (2007) looking for autonomous replicating sequence (ARS) elements. Briefly, this involves a restriction digest of whole genomic DNA that was then cloned into a non-replicating plasmid, which was marked with the *pyrE2* uracil biosynthesis pathway gene. This plasmid library was then transformed into *ΔpyrE2 ΔradA* strain of *H. volcanii* that is deficient in homologous recombination; this was to avoid plasmid integration into the genome by recombination. Cells could only grow on media without uracil if the plasmid is replicated by an ARS element (a replication origin). Initially only one ARS element was recovered from the screen, which was found to be from pHV1. This region was found to have two divergently transcribed genes (one of which is *orc10*), separated by an intergenic region with direct repeats (Figure 4.2) and 68% AT putative DUE (duplex unwinding element), which is similar to almost all characterised archaeal origins (for full alignments of ORBs and DUEs see Appendix *II DUE and ORB Alignments*).

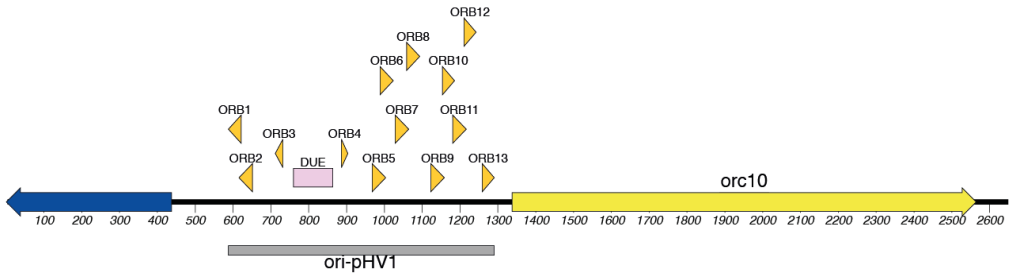


Figure 4.2 **The *ori-pHV1* region and associated *orc* gene**. The origin *ori-pHV1* (671 bp) is flanked by two divergently transcribed genes including its associated *orc10* gene and contains a duplex unwinding element similar to other archaeal origins. The origin region is highlighted (grey box) with duplex unwinding elements (DUEs) (pink box) and origin recognition boxes (ORBs) (orange arrows) shown. Orientation of ORBs is indicated by arrows. Adapted from Norais *et al*. (2007).

In the same study by Norais *et al*. (2007), the origin was deleted from pHV1 using a *trpA* marker and the pHV1 mega-plasmid was found to be still present in the genome. Because this origin was isolated several times in the ARS screen, it was concluded that it is dominant over other origins in *Haloferax volcanii* as no other origins where found unless *ori-pHV1* had been deleted first. The copy number of pHV1 is at a ratio of 2:1 compared to the chromosome and the other mega-plasmids, meaning that it is highly abundant in the cell, however this does not account for only *ori-pHV1* being seen in the ARS assay performed by Norais *et al*. (2007) as it was expected that other origins should have been found as well, although perhaps in lower numbers compared to *ori-pHV1*. The origin *ori-pHV1* could be dominant in the cell due to sequestering of replication machinery or more efficient binding of Orc.

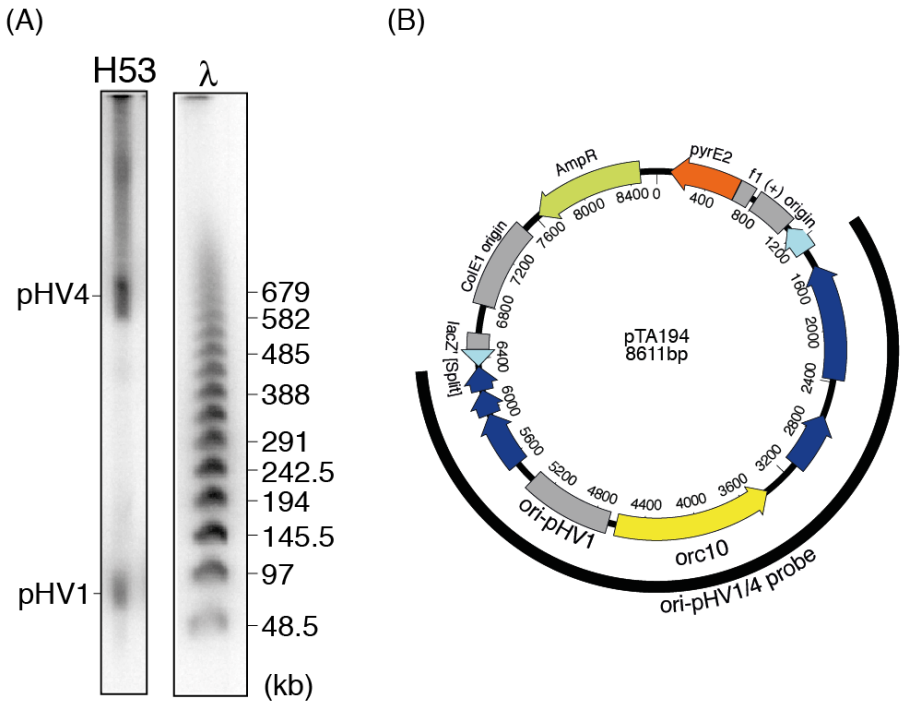


Figure 4.3 **Two bands hybridise to a probe for *ori-pHV1*. (A)** In a study by Norais *et al*. (2007) two bands are seen at ~500-700 kb and at ~85 kb when a Southern blot of a PFG of wild-type whole genomic DNA (H53) was probed with **(B)** the isolated ARS element (*ori-pHV1/4*, designated *ori-pHV1* in this study) which was cloned into the plasmid pTA194. Norais *et* al. (2007) designated the higher band as the pHV4 mega-plasmid and the lower band as the pHV1 mega-plasmid according to their sizes, 636 kb and 85 kb, respectively. Adapted from Norais *et al*. (2007).

The putative origin region isolated in the ARS assay was used to probe a Southern blot of a pulsed-field gel of whole genomic DNA. Two bands were

seen: one at ~500-700 kb and one at ~85 kb, which were thought by Norais *et al*. (2007) to correspond to pHV4 and pHV1, respectively (Figure 4.3). For this reason, the ARS assay element was named by Norais *et al*. (2007) as *ori-pHV1/4*. However, in this study it will be referred to as *ori-pHV1* because (as will be shown) it is present only on pHV1. Furthermore, it has been found by Hawkins *et al*. (2013) that in laboratory strains of *Haloferax volcanii*, the pHV4 mega-plasmid had integrated on the chromosome and therefore does not represent the ~500-700 kb band seen on the pulsed field gel.

4.2 Aims

Further investigation of pHV1, its origin and *orc* genes is required. The aims of this chapter are:

- To independently delete *ori-pHV1* and its associated *orc* gene, *orc10*, either individually or together, both in wild-type and origin-deleted backgrounds. These deletions will initially be *trpA*-marked, before removing the *trpA* marker so it can be used in further deletions.
- To analyse the requirement for a size threshold for recombination-dependent replication, as suggested by findings with pHV3 and pHV4.
- Expand on the discovery made by Norais *et al*. 2007 of two bands of ~500-700 kb and ~85 kb seen on a pulsed field gel Southern blot, that both hybridise with an *ori-pHV1* probe.

4.3 Results

4.3.1 Deletion of *ori-pHV1* and *orc10*

Plasmids pTA1494 (*Δori-pHV1::trpA+*), pTA1495 (*Δorc10::trpA+*) and pTA1496 (*Δori-pHV1 Δorc10::trpA+*) were used to transform strains H53 (wild-type) and H1804 (origin-deleted) to determine if the origin and the associated *orc10* gene present on pHV1 could be deleted in these strain backgrounds. Plasmids with a *trpA* selection marker were used in the knowledge that *ori-pHV1* and *orc10* would be difficult to delete and would require extra selection pressure (Norais *et al*. 2007). In this study plasmids

were constructed from the genomic clone of the *ori-pHV1* region for the deletion of *ori-pHV1* (pTA1557), *orc10* (pTA1558) and, *ori-pHV1* and *orc10* together (pTA1559).

However, unlike Norais *et al*. (2007), this study will also delete *ori-pHV1* and *orc10* in an origin-deleted background. There is another *orc* gene *orc8* present on pHV1, however this is not located next to the origin and will not be investigated in this study.

pTA1484 *ori-pHV1* (HVO_C0001ori) *orc10* (HVO_C0001)

The genomic clone of *ori-pHV1* and *orc10* (pTA1484) was constructed by Thorsten Allers prior to this study in 2014 (unpublished data) (Figure 4.4).

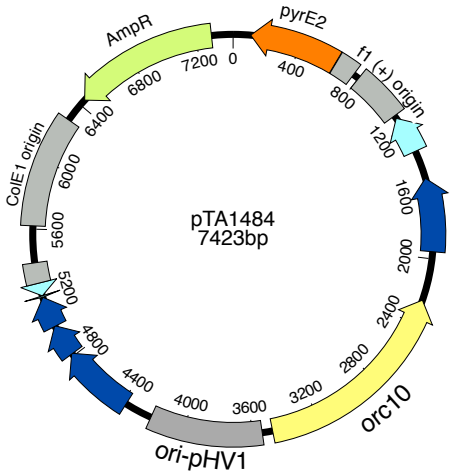


Figure 4.4 **pTA1484.** Genomic clone of *ori-pHV1* and *orc10*. Constructed by Thorsten Allers in 2014 (unpublished data).

pTA1494 (*ori-pHV1::trpA+* deletion construct) and pTA1557 (*ori-pHV1* deletion construct

The plasmid pTA1494 (*ori-pHV1::trpA+* deletion construct) was constructed by Thorsten Allers in 2014 (unpublished data). To create the *ori-pHV1* deletion construct the genomic clone pTA1484 was digested with *BsmI* and *XbaI* and blunt-ended using Klenow. It was then self-ligated to make pTA1557 (Figure 4.5).

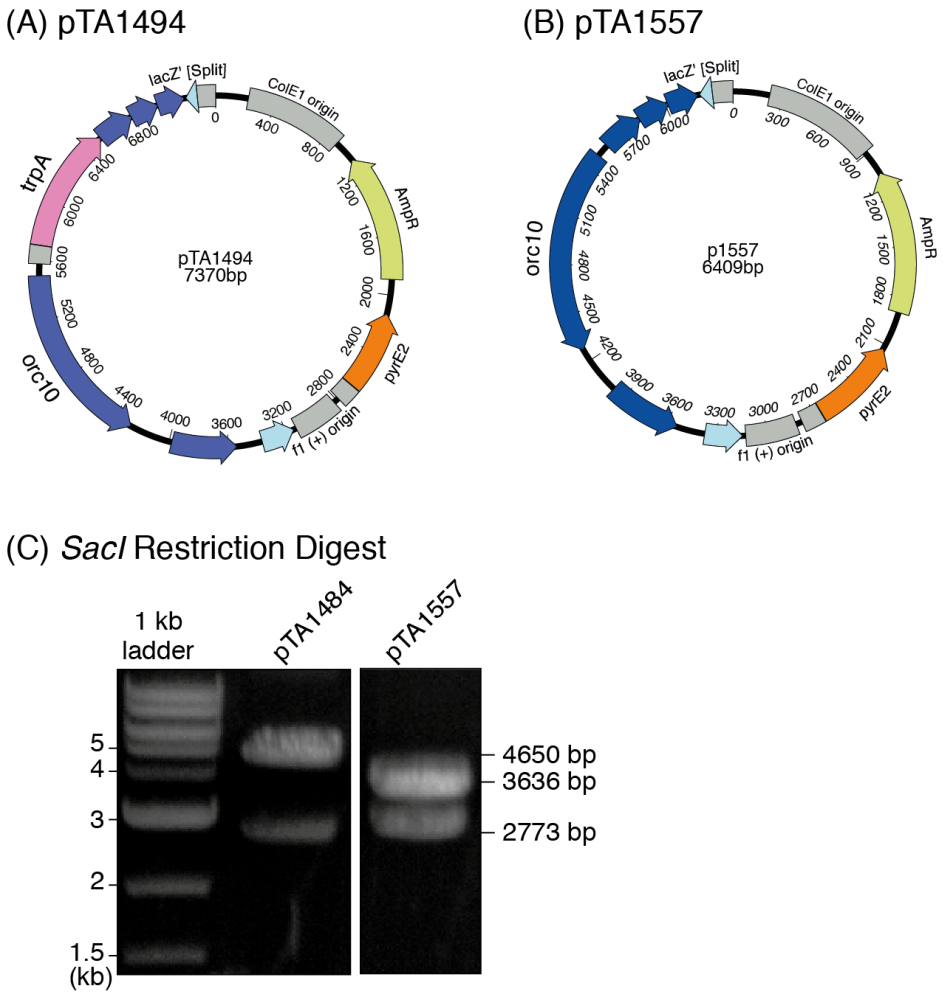


Figure 4.5 ***ori-pHV1* deletion constructs**. **(A)** pTA1494 *ori-pHV1* deletion construct with *trpA* marker. Constructed by Thorsten Allers in 2014 (unpublished data). **(B)** pTA1557 *ori-pHV1* deletion construct. **(C)** *SacI* restriction digest of pTA1484 producing band sizes of 4650 bp and 2773 bp and, pTA1557 producing band sizes of 3636 bp and 2773 bp, as predicted.

pTA1495 (*orc10::trpA+* deletion construct) and pTA1558 (*orc10* deletion construct)

The plasmid pTA1495 (*orc10::trpA+* deletion construct) was constructed by Thorsten Allers in 2014 (unpublished data). The genomic clone pTA1484 was digested with *HindIII* and *StuI* and blunt-ended with Klenow. The product was then self-ligated to create pTA1558 *orc10* deletion construct (Figure 4.6).

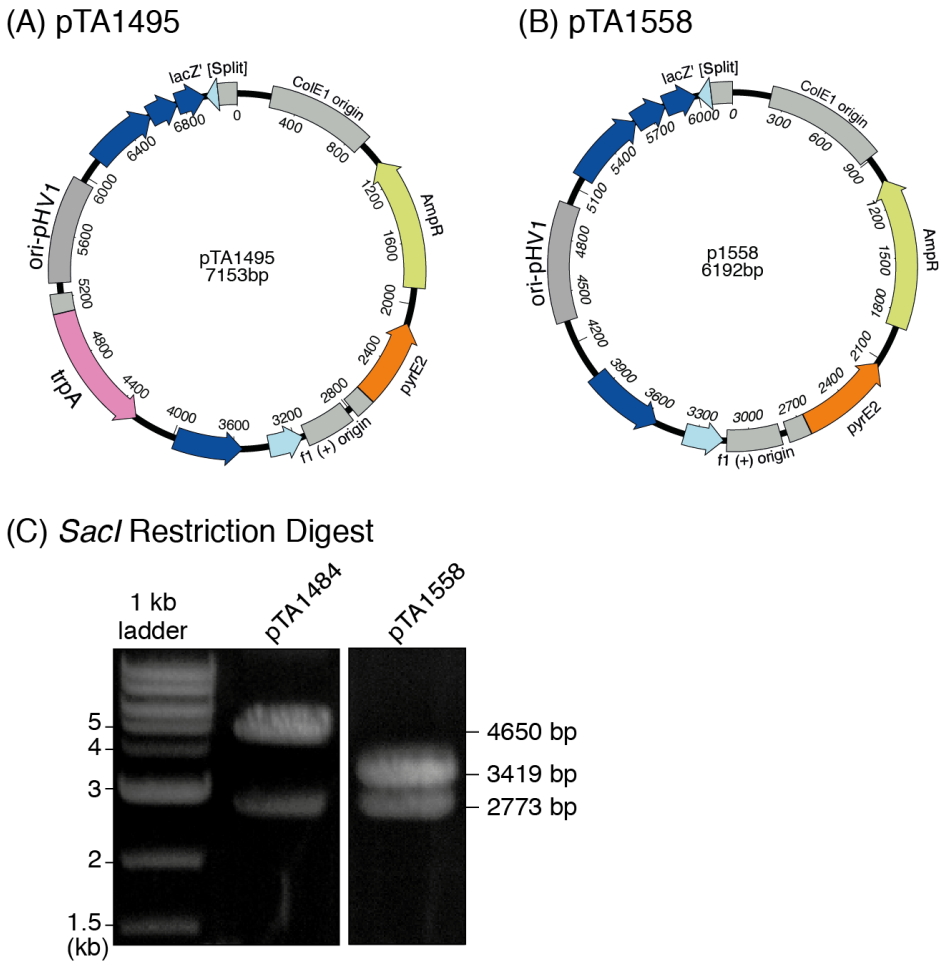


Figure 4.6 ***orc10* deletion constructs**. **(A)** pTA1495 *orc10* deletion construct with *trpA* marker. Constructed by Thorsten Allers in 2014 (unpublished data). **(B)** pTA1558 *orc10* deletion construct. **(C)** *SacI* restriction digest of pTA1484 producing band sizes of 4650 bp and 2773 bp and, pTA1558 producing band sizes of 3419 bp and 2773 bp, as predicted.

pTA1496 (*ori-pHV1 orc10::trpA+* deletion construct) and pTA1559 (*ori-pHV1 orc10* deletion construct)

The plasmid pTA1496 (*ori-pHV1 orc6::trpA+* deletion construct) was constructed by Thorsten Allers in 2014 (unpublished data). The genomic clone pTA1484 was digested using *BsmI* and *StuI*. The product was blunt-ended with Klenow and self-ligated to create pTA1559 *ori-pHv1 orc10* deletion construct (Figure 4.7).

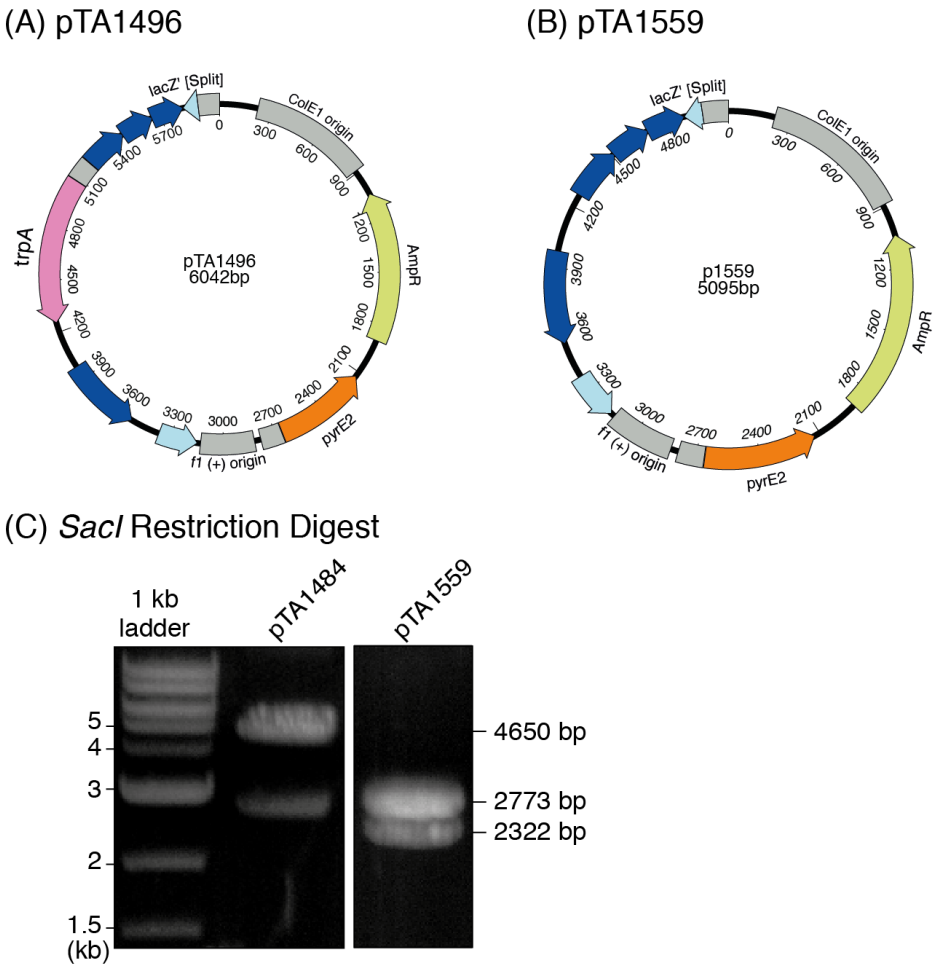


Figure 4.7 ***ori-pHV1* deletion constructs**. **(A)** pTA1494 *ori-pHV1* deletion construct with trpA marker. Constructed by Thorsten Allers in 2014 (unpublished data). **(B)** pTA1557 *ori-pHV1* deletion construct. **(C)** *SacI* restriction digest of pTA1484 producing band sizes of 4650 bp and 2773 bp and, pTA1559 producing band sizes of 2773 bp and 2322 bp, as predicted.

Deletion of *ori-pHV1* and *orc10* with a *trpA* marker

The strains H1945, H1941 and H1948 were generated by transforming H53 with the plasmids pTA1494 (*Δori-pHV1::trpA+*), pTA1495 (*Δorc10::trpA+*) and, pTA1496 (*Δori-pHV1 Δorc10::trpA+*), respectively. These plasmids were also used to transform H1804 to generate the strains H1939, H1949 and H1943, respectively. These strains were constructed by Thorsten Allers in 2014, however they were verified by colony hybridisation (not shown) and Southern blot as part of this study (Figure 4.8). They were verified using a pTA1484 *BssHII* fragment probe. It was found that all strain combinations could be generated as can be seen in Table 4.1

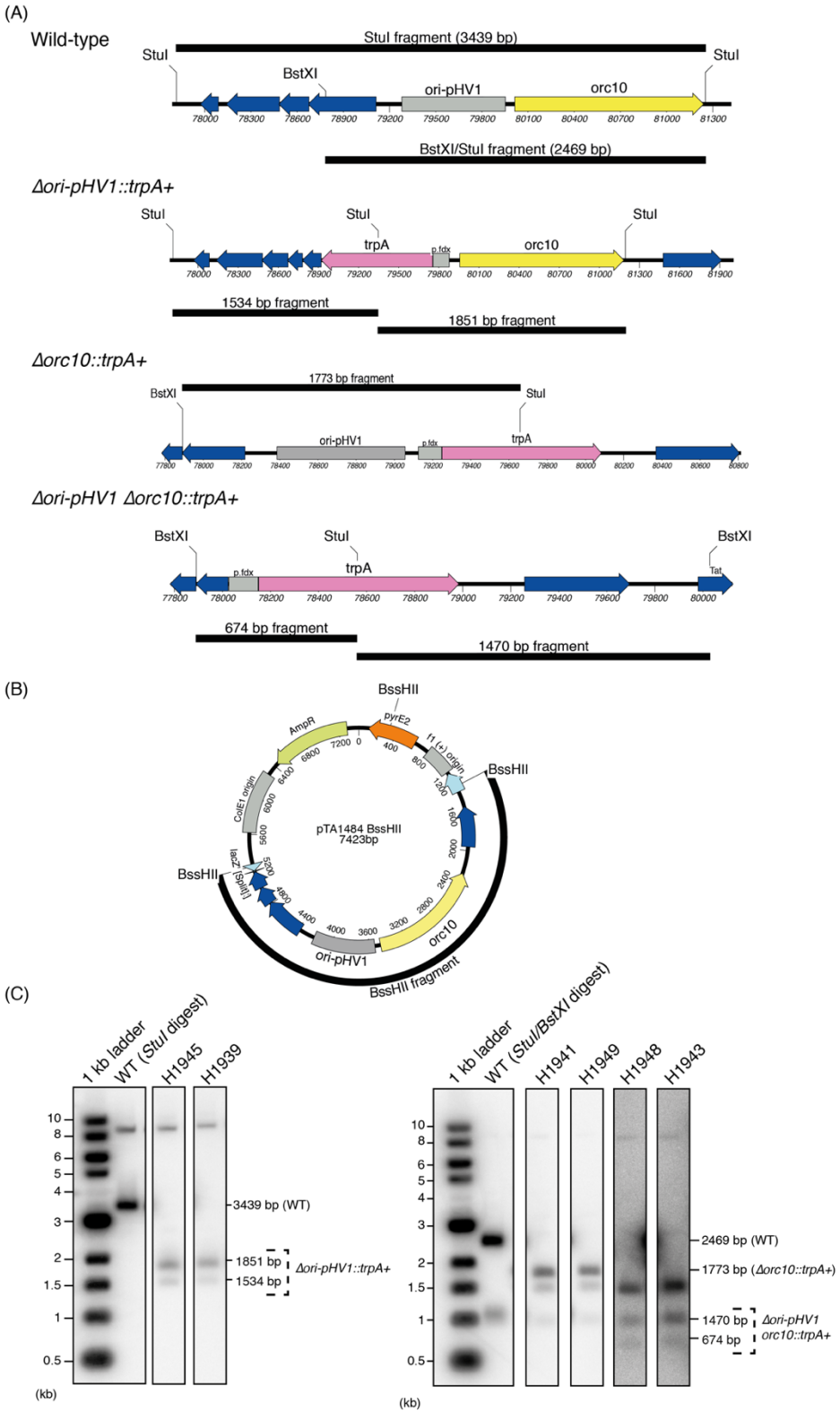


Figure 4.8 **Deletion of *ori-pHV1::trpA+***. **(A)** Expected Southern blot fragment sizes for wild-type (3439 bp band), *Δori-pHV1::trpA+* (1851 bp and 1534 bp bands), *Δorc10::trpA+* (1773 bp band) and *Δori-pHV1 Δorc10::trpA+* (1470 bp and 674 bp bands) strains when digested with either *StuI* or *StuI*/*BstXI*, as appropriate. **(B)** *BssHII* fragment of pTA1484 used for Southern blot probe which will hybridise to the *ori-pHV1* region including upstream and downstream regions and *orc10*. **(C)** Southern blot showing correct fragment sizes indicating deletions for H1939, H1941, H1943, H1945, H1948, H1949.

Table 4.1 ***ori-pHV1* and *orc10* deletion strains with *trpA* marker.**

Strain	Genotype	Background
H1945	*Δori-pHV1::trpA+*	H53 (wild-type)
H1939	*Δori-pHV1::trpA+*	H1804 (*ΔoriC1,2,3,pHV4*)
H1941	*Δorc10::trpA+*	H53 (wild-type)
H1949	*Δorc10::trpA+*	H1804 (*ΔoriC1,2,3,pHV4*)
H1948	*Δori-pHV1 Δorc10::trpA+*	H53 (wild-type)
H1943	*Δori-pHV1 Δorc10::trpA+*	H1804 (*ΔoriC1,2,3,pHV4*)

These results appear to disprove the hypothesis that recombination-dependent recombination is size dependent and has a threshold of ~500 kb. pHV1 is the smallest mega-plasmid at 85 kb and yet its origin and *orc10* gene can be deleted. By contrast, pHV3 is 438 kb and neither the origin nor the associated *orc6* gene can be deleted from this episomal mega-plasmid. pHV1 is thought to have been acquired by lateral gene transfer, since it has synonymous codon usage that is different to the rest of the *Haloferax volcanii* genome (Hartman *et al*. 2010). Because pHV1 is an entirely foreign element, there may be other factors at play. Therefore, the replication of pHV1 was investigated further.

4.3.2 Loss of pHV1

Removal of *trpA* marker from the deletion of *ori-pHV1* and *orc10*

The deletion of *ori-pHV1* and *orc10* had been achieved using the additional selection of the *trpA* marker. This selectable marker had to be removed from the deletion so that it could be reused in the construction of subsequent strains. To accomplish this, the strains detailed in Table 4.1 were transformed with the appropriate unmarked (no *trpA*) deletion plasmid; pTA1557 (*Δori-pHV1*), pTA1558 (*Δorc10*) or pTA1559 (*Δori-pHV1 Δorc10*). The pop-outs of the transformants were patched on Hv-Ca plates with and without tryptophan, to determine which colonies had successfully lost the *trpA* marker. Two colonies displaying a *trp-* phenotype were picked from each strain background and analysed by PFGE and Southern blot. Resulting strains are detailed in Table 4.2.

The strains H2175 and H2481 (*ΔpryE2, ΔtrpA, ΔoriC1, ΔoriC2, ΔoriC3, Δori-pHV4, Δori-pHV1*) were generated by transforming H1939 with the plasmid

pTA1557 (*Δori-pHV1*). H2175 was confirmed by *trp-* phenotype, Southern blot and PFGE and H2481 was confirmed by *trp-* phenotype and by colony hybridisation. The Southern blot of H2175 is shown as an example in Figure 4.9, which was hybridised using a radiolabelled probe specific for the *ori-pHV1* region on pHV1 (*BssHII* fragment of pTA1484, Figure 4.9C).

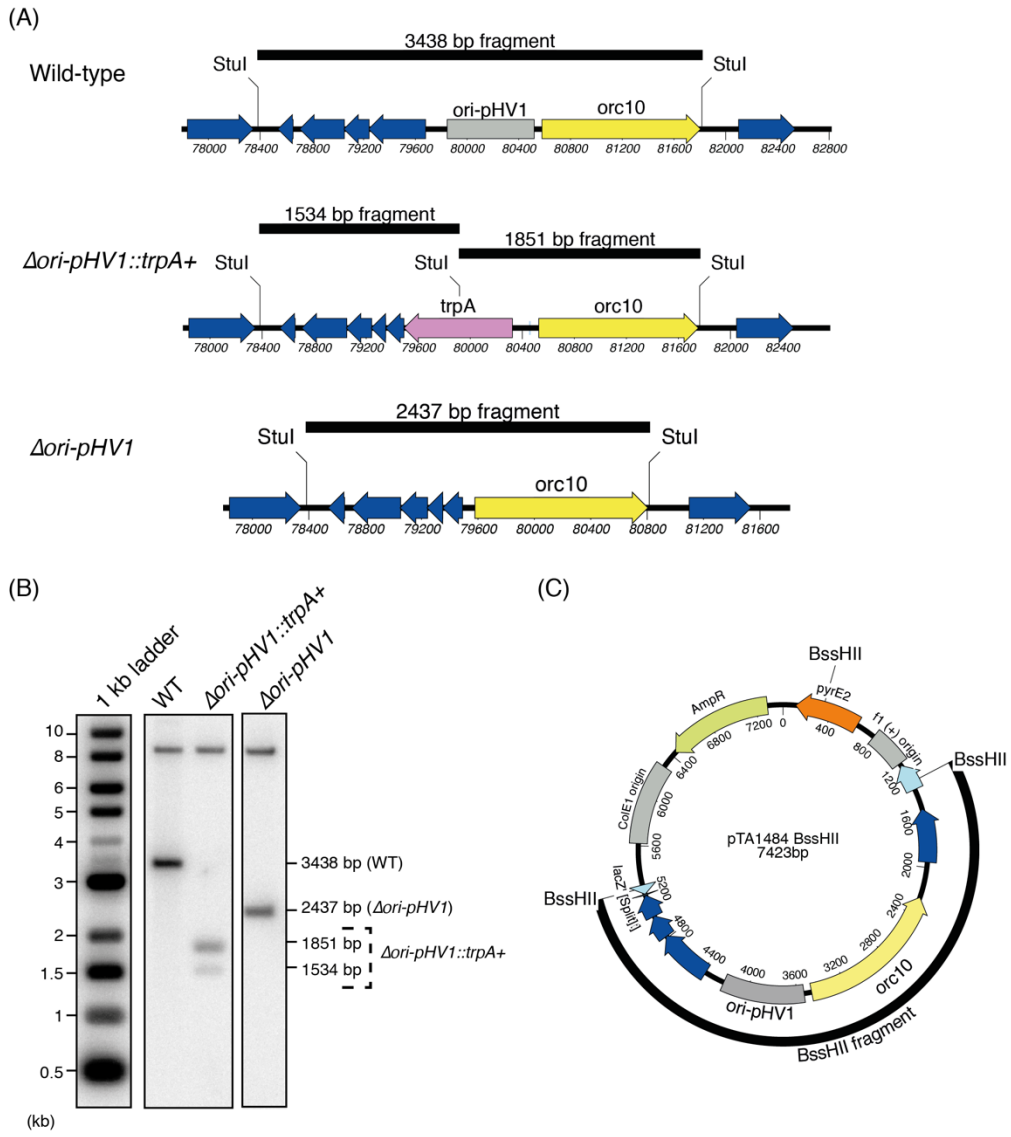


Figure 4.9 **Deletion of *ori-pHV1*. (A)** Expected Southern blot band sizes from *StuI* digest for wild-type (H53, 3438 bp band), *Δori-pHV1::trpA+* (H1939, 1851 bp and 1534 bp band) and *ΔoripHV1* (H2175, 2437 bp band). **(B)** Southern blot showing wild-type, *Δori-pHV1::trpA+* and *ΔoripHV1*, showing expected band sizes. The *ΔoripHV1* strain is H2475. **(C)** *BssHII* fragment of pTA1484 used for Southern blot probe which will hybridise to the *ori-pHV1* region including upstream and downstream regions and *orc10*.

Table 4.2 ***ori-pHV1* and *orc10* deletion strains with *trpA* marker removed.**

Strain	Genotype	Background
H2475	*Δori-pHV1*	H53 (wild-type)
H2481/ H2175	*Δori-pHV1*	H1804 (*ΔoriC1,2,3,pHV4*)
H2477	*Δorc10*	H53 (wild-type)
H2483	*Δorc10*	H1804 (*ΔoriC1,2,3,pHV4*)
H2479	*Δori-pHV1 Δorc10*	H53 (wild-type)
H2485	*Δori-pHV1 Δorc10*	H1804 (*ΔoriC1,2,3,pHV4*)

Deletion of mega-plasmid pHV1

Construction of the strain H2183 (*ΔpryE2, ΔtrpA, ΔoriC1, ΔoriC2, ΔoriC3, Δori-pHV4, ΔpHV1*) was achieved by transforming the strain H1985 respectively, with the deletion construct pTA1558 (*Δorc10*). H2184 (*ΔpryE2, ΔtrpA, ΔpHV1*) was constructed by transforming H1986 with the plasmid pTA1559 (*Δori-pHV1 Δorc10*). After the pop-out stage of strain construction there are a combination of strains produced. Some maintained the mega-plasmid pHV1 with the respective *ori-pHV1* and/or *orc10* deletion and some colonies lost the pHV1 mega-plasmid entirely. This was confirmed by Southern blot (not shown) and PFG with a probe amplified from genomic DNA using the primers pHV1probeF and pHV1probeR (1132 bp) (Figure 4.10).

In Figure 4.10, for H2183 (*ΔpHV1*) the two bands of 85 and 510 kb corresponding to pHV1 are missing. This is the same for H2184 (not shown). There is no hybridisation of the *ori-pHV1* specific probe, however this was not for technical reasons, as hybridisation was seen for the strains H53 (WT), H1939 (*Δori-pHV1::trpA+*) and H2175 (*Δori-pHV1*). This also shows that despite the deletion of *ori-pHV1* the mega-plasmid pHV1 is maintained in *Δori-pHV1::trpA+* and *Δori-pHV1* strains. Therefore, the probe design and hybridisation were performed successfully and the only other conclusion that could be drawn is that the strains H2183 and H2184 had lost the entire pHV1 mega-plasmid.

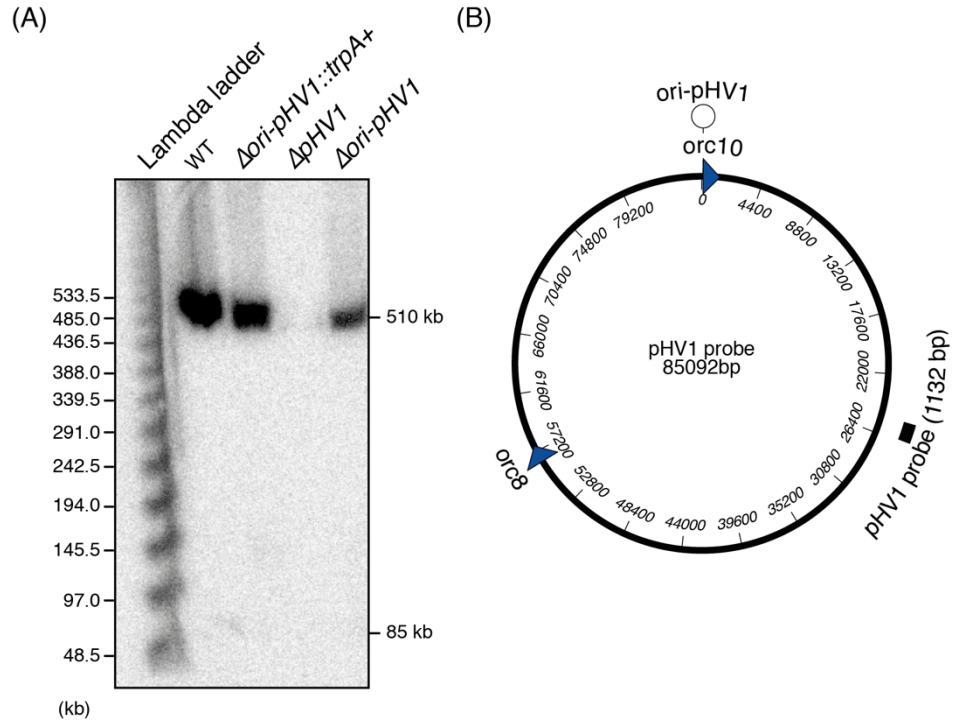


Figure 4.10 **ΔpHV1 strain checked by PFG. (A)** Strain that is missing pHV1 (H2183) does not hybridise to probe in a Southern blot of a PFG of whole genomic DNA, whereas strains that are either *Δori-pHV1::trpA+* (H1939) or *Δori-pHV1* (H2175) do hybridise to the probe and therefore still contain the pHV1 mega-plasmid. Methods described in *Chapter 2: Materials and Methods, Section 2.2.6 Genotype Screening*. **(B)** pHV1 specific probe generated by PCR using the primers pHV1probeF and pHV1probeR (1132 bp).

Table 4.3 **Loss of pHV1 from colonies.**

Strain	Number of *trp-* colonies	% of *trp-* colonies	Number of *trp-* colonies that also lost pHV1	% of *trp-* colonies that also lost pHV1
H1945	13/120	10.8%	7/13	53.8%
H1939	10/120	8.3%	5/10	50%
H1941	24/120	20%	9/24	37.5%
H1949	7/120	5.8%	3/7	42.9%
H1948	33/40	82.5%	6/33	18.2%
H1943	3/120	2.5%	3/3	100%

As Table 4.3 shows, the loss of pHV1 was found to be a common occurrence, however the percentage loss of the pHV1 mega-plasmid varied greatly between strains. Colonies were first patched ±tryptophan to test for the loss of the *trpA* marker gene from the *ori-pHV1* and/or *orc10* deletions. Those colonies that gave a *trp-* phenotype (i.e. were not able to grow on the plates lacking tryptophan),

were re-patched and probed via colony hybridisation using the pHV1 probe shown in Figure 4.10B, which hybridises to the pHV1 mega-plasmid rather than specifically to the *ori-pHV1*/ *orc10* region. Colonies that hybridise with the probe were deemed to have retained the pHV1 mega-plasmid, whereas those that had not hybridised were deemed to have lost pHV1.

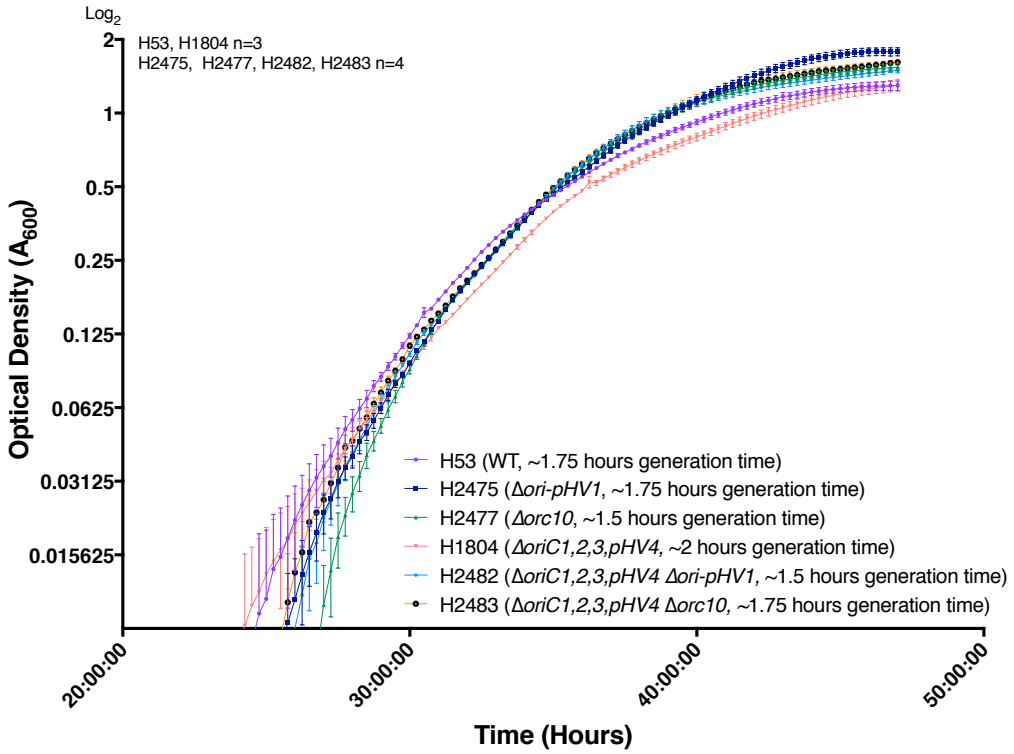


Figure 4.11 **Exponential growth rate of *ori-pHV1* and *orc10* deleted strains in wild-type and origin-deleted backgrounds compared to wild-type strains.** H53 (wild-type), H1804 (*ΔoriC1,2,3,pHV4*), H2475 (*ori-pHV1*), H2477 (*Δorc10*), H2481 (*ΔoriC1,2,3,pHV4, Δori-pHV1*), and H2483 (*ΔoriC1,2,3,pHV4, Δorc10*), were grown in liquid culture for 48 hours in Hv-YPC. Growth was measured by A_{600}. Generation time is indicated next to each strain. All strains and repeats were incubated on the same 48 well plate and measured simultaneously using an Epoch2 Microplate Spectrophotometer (BioTek). Error bars are calculated from standard error. For H53 and H1804 n=3, for H2475, H2477, H2482, H2483 n=4.

To determine if deleting *ori-pHV1* or *orc10* affected the growth rate of strains, H53 (wild-type), H1804 (*ΔoriC1,2,3,pHV4*), H2476 (*ori-pHV1*), H2477 (*Δorc10*), H2481 (*ΔoriC1,2,3,pHV4, Δori-pHV1*), and H2483 (*ΔoriC1,2,3,pHV4, Δorc10*), were grown in Hv-YPC. These cultures were then loaded onto the microplate spectrophotometer, a growth curve plotted and relative generation times calculated (Figure 4.11). This shows that deleting *ori-*

pHV1 or *orc10* in wild-type or origin-deleted backgrounds does not affect growth rate, meaning neither *ori-pHV1* or *orc10* is essential for normally cellular growth, and they must be performing another role within the cell.

Growth without pHV1

Since it is possible to cure pHV1 from *Haloferax volcanii*, it is clear that none of the genes found on pHV1 is essential to the cell. This leads to the question why *H. volcanii* maintains pHV1 if it is not essential? It could be that pHV1 confers a growth advantage to the host cell when it is present. To test this, cell growth during exponential phase was measured for strains lacking pHV1. A growth curve was plotted and relative generation times were calculated during exponential phase for each strain (Figure 4.12).

There were no noticeable growth differences in rich media (Hv-YPC) regardless of whether pHV1 was present in the strains or not. Therefore, this mega-plasmid is neither essential for the cell nor is required for efficient growth under these particular laboratory conditions. It is possible that pHV1 does not supply any additional growth benefits to the cell. The origin was found to be 'dominant' in an ARS assay (Norais *et al*. 2007), and therefore pHV1 appears to be acting as a parasite to the cell, ensuring that it is replicated with the chromosomal DNA. Despite this burden on the cell, pHV1 might confer possible phage and/or predator resistance, thereby protecting the cell from invading DNA or being destroyed by other organisms in its natural environment. These would not be seen as a growth advantage in laboratory conditions and would have to be tested specifically.

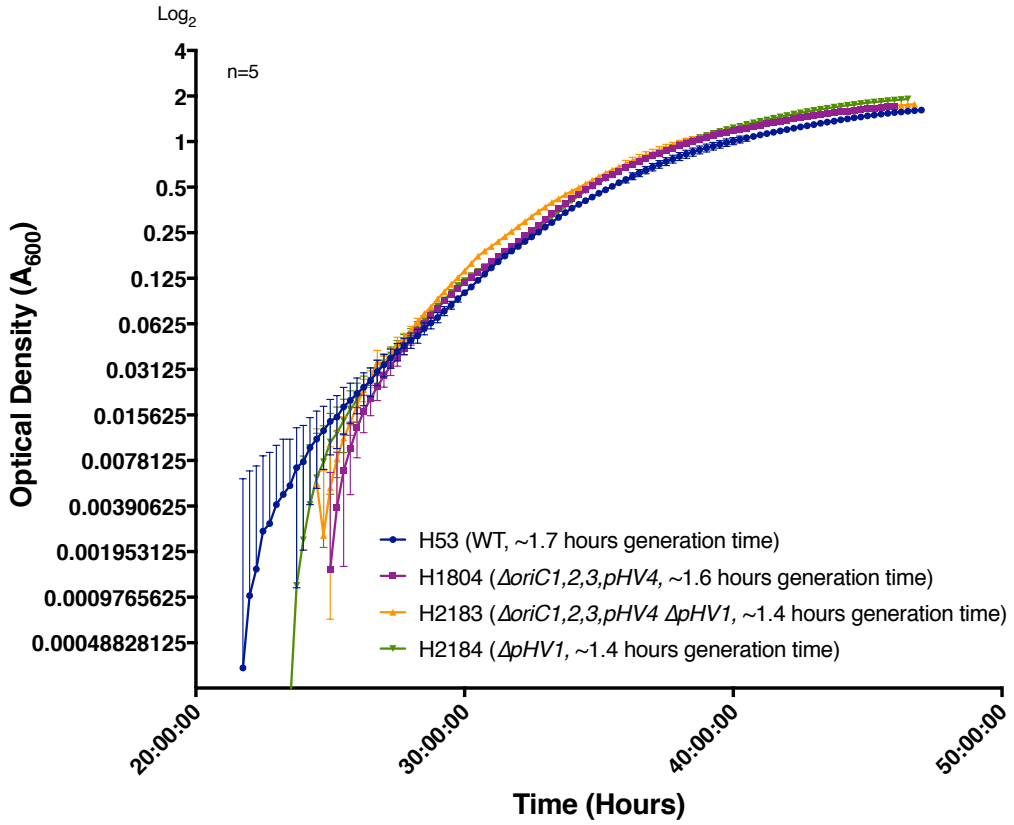


Figure 4.12 **Exponential growth rate of pHV1 deleted strains in wild-type and origin-deleted backgrounds compared to wild-type strains.** H53 (wild-type), H1804 (*ΔoriC1,2,3,pHV4*), H2183 (*ΔpHV1*) and H2184 (*ΔoriC1,2,3,pHV4 ΔpHV1*) were grown in liquid culture for 48 hours in Hv-YPC. Growth was measured by A_{600}. Generation time is indicated next to each strain. All strains and repeats were incubated on the same 48 well plate and measured simultaneously using an Epoch2 Microplate Spectrophotometer (BioTek). Error bars are calculated from standard error. For all strains n=5.

Stability of pHV1

It is assumed that pHV1 is stably maintained in cells when they are grown in the laboratory. However, it may be the case that pHV1 is naturally lost by cells over time; this is theoretically possible, since pHV1 is non-essential and confers neither a growth defect nor advantage. However, there may be mechanisms that ensure pHV1 is stably maintained, and the *ori-pHV1* origin, or the associated *orc10* gene, may confer stability to the mega-plasmid. If the origin or *orc* gene are deleted, this may make the plasmid unstable and therefore more easily lost.

To investigate this hypothesis, the *pyrE2* marker gene (for uracil biosynthesis) was integrated into a putatively-neutral region of pHV1. Criteria for a region to

be neutral included; it must not be near the origin or *orc* genes on pHV1, it must not interfere with genes of known function, and it must not be located in an IS-element, which are short DNA sequences that can act as transposons and are repeated throughout the *Haloferax volcanii* genome making them non-unique. The plasmid pTA1914 was generated according to these criteria and the region included the genes HVO_C0029, HVO_C0030, HVO_C0031, HVO_C0032, HVO_C0033 and HVO_C0034. The *pyrE2* marker was chosen as cultures that had lost pHV1 could easily be scored by plating onto media containing 5-FOA. The transformed strains were designated H3640 (wild-type background) and H3644 (*Δori-pHV1*).

pTA1878 pHV1 Region (HVO_C0029 – HVO_C0034)

H26 wild-type *H. volcanii* DNA was digested with *HindIII* and *SphI* to make a 2.9 kb fragment of a region of pHV1 covering HVO_C0029, HVO_C0030, HVO_C0031, HVO_C0032, HVO_C0033 and HVO_C0034, all annotated as hypothetical proteins (Figure 4.13). The digested DNA was run on a gel and fragments between 2.5-3.5 kb were extracted. The fragments were cloned into pTA1667 (a pBluescript II SK+ derivative) at *HindIII* and *NspI* sites. Transformants were patched and transferred onto a membrane via colony lift. This was then probed using a radiolabelled 961bp PCR products amplified from genomic DNA with the primers pHV1mkrprobeF and pHV1mkrprobeR. Patches that hybridised were restreaked and verified by sequencing and restriction digest.

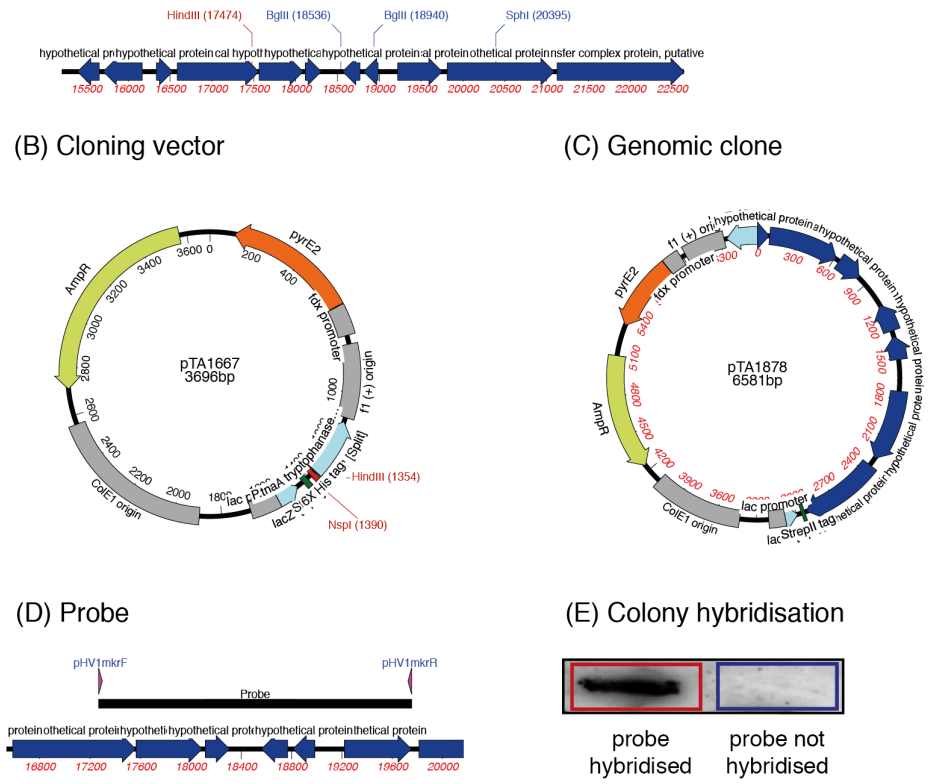


Figure 4.13 **pTA1878. (A)** The location of pHV1 HVO_C0031 region on the *H. volcanii* genome, genomic DNA was digested with *HindIII* and *SphI* sites. **(B)** pTA1667 plasmid where the *HindIII* to *SphI* fragment was inserted into *HindIII* and *NspI* sites. **(C)** Genomic clone pTA1878 constructed by inserting a *HindIII* to *SphI* fragment containing pHV1 HVO_C0031 region into pTA1667. **(D)** 961 bp probe was constructed by PCR using primers pHV1mkrF and pHV1mkrR, was radiolabelled and used to probe colonies. **(E)** Colony hybridisation showing a patch which the probe hybridised to (containing pHV1 HVO_C0031 region genomic clone) and a patch to which the probe did not hybridise to (does not contain pHV1 HVO_C0031 region genomic clone).

pTA1914 (pHV1 HVO_C0031 region::*pyrE2+* replacement construct)

The *pyrE2* marker was excised from pTA1677 using *BamHI* and ligated into the genomic clone pTA1878 at the *BglII* sites to create pTA1914 (Figure 4.14).

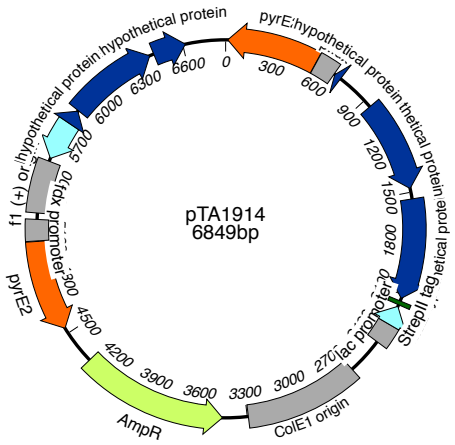


Figure 4.14 **pTA1914**. pTA1878 with additional *pyrE2* marker.

pyrE2 Marker on pHV1

Strains H53 (wild-type) and H2475 (*Δori-pHV1*) were transformed using a linear *EcoRV-ScaI* fragment of the plasmid pTA1914. As this was a one-step transformation pop-in and pop-out were performed within the transformation protocol and cells were plated on selective media appropriate for selection of pop-out colonies. Colonies were selected and verified by *ura+* phenotype (data not shown). This process generated strains H3640 (H53 background) and H3644 (H2475 background).

Loss Assay

Strains were grown over consecutive overnights in Hv-YPC broth (rich media) without selection, thus allowing the loss (or maintenance) of pHV1. Each day, the cultures were scored for loss of pHV1 by plating on selective media containing 5-FOA (and onto YPC for a viable count). A graph recording the percentage of 5FOA resistant colonies and therefore the percentage of colonies that are presumed to have lost of pHV1 (Figure 4.15) shows that this mega-plasmid appears to be much less stable when the *ori-pHV1* origin has been deleted.

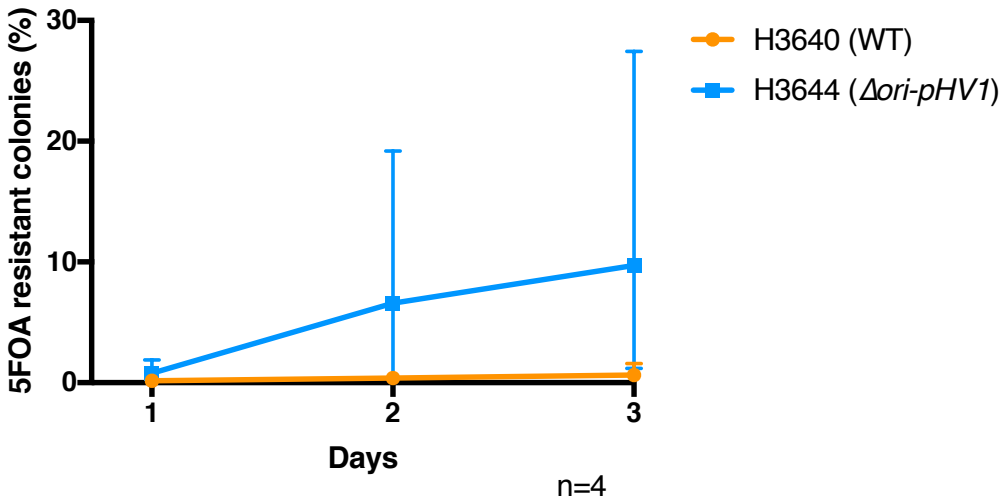


Figure 4.15 **Percentage of 5FOA resistant colonies.** Strains H3640 (wild-type) and H3644 (*Δori-pHV1*) were grown in 5 ml overnight cultures in Hv-YPC, diluted and plated on selective and non-selective media. Colonies were counted and percentage calculated using numbers of colonies that grew on media containing 5-FOA divided by viable counts multiplied by 100. 5FOA resistance suggests loss of pHV1, therefore H3644 (*Δori-pHV1*) seems more likely to lose pHV1.

A potential problem with using the *pyrE2* marker to score pHV1 loss is that it is susceptible to point mutations (Mackwan *et al*. 2007), leading to cells becoming 5-FOA resistant and therefore being able to grow on 5-FOA selective media without having lost pHV1. To assess the potential for false positive results, a number of 5-FOA-resistant colonies were patched and subjected to colony hybridisation with a pHV1-specific radiolabelled probe, generated by PCR using the primers pHV1probeF and pHV1probeR. The fraction of colonies that hybridised with the probe was calculated (Table 4.4), these colonies still contain pHV1 and are false positive results.

Table 4.4 **Percentage of colonies that contain pHV1.**

Strain	Genotype	Percentage that contain pHV1
H3640	Wild-type	99.6% (239/240)
H3644	*Δori-pHV1*	86.7% (208/240)

As Table 4.4 shows, most of the colonies in both strains hybridise with the pHV1-specific probe. This means that most of the results from the loss assay are in fact false positives. It is possible the recombination-dependent replication

used by pHV1 in the *Δori-pHV1* strain (H3644) is more mutagenic than origin-dependent replication, and this would account for the increase in false positives with point mutations in *pyrE2*. This is supported by Table 4.5 where it can be seen that the percentage mutation rate of *pyrE2* can be calculated for each strain, with the wild-type strain (H3640) giving a mutation rate of 0.49%, and the *Δori-pHV1* strain (H3644) giving a mutation rate of 7.28%. This shows that when *ori-pHV1* is deleted from pHV1, the mutation rate of the genes on pHV1 increases, indicating that origin-dependent replication may be less mutagenic that origin-independent replication. However, this trend is not seen on the main chromosome when all origins are deleted (Thorsten Allers, personal communication).

Table 4.5 **Percentage mutation rate of *pyrE2* in wild-type and *Δori-pHV1* strains.**

Strain	Viable Counts*	5FOA resistant colonies*	Number of colonies that have lost pHV1	% loss of pHV1	% mutation rate of *pyrE2*
H3640 (WT)	$1.76x10^8$	$8.64x10^5$	$3.46x10^3$	0.4%	0.49%
H3644 (*Δori-pHV1*)	$1.04x10^8$	$8.7x10^6$	$1.16x10^6$	13.3%	7.28%

*Day 3 average

In the *Δori-pHV1* strain (H3644) 13.3% of the colonies tested had lost pHV1, compared to 0.4% in the wild-type background (H3640). Therefore, it is likely that pHV1 is lost more easily from *Δori-pHV1* strains, and therefore *ori-pHV1* is important for the maintenance and stability of pHV1 in the cell. However, this needs to be investigated further, perhaps using a different marker gene in the pHV1 loss assay.

It may also be the case that deleting *ori-pHV1* causes in imbalance of Orc proteins in the cell, especially if Orc10 is able to bind to origins other that *ori-pHV1*. At this point it is not known where Orc10 binds or with what affinity. The mega-plasmid pHV1 also contains the gene for Orc8. It could be that Orc8 plays a role in binding to *ori-pHV1,* or to other origins, however it could also be the

case that there is a cryptic origin found upstream of the *orc8* gene, but this has yet to be investigated in detail. If there is cryptic origin found on pHV1 this could also have implications on the replication and stability/ maintenance of pHV1.

4.3.3 Pulsed-Field Gel Analysis of pHV1

Intact (undigested) genome of *Haloferax volcanii* can be visualised using pulsed-field gel electrophoresis (PFGE). The main chromosome and mega-plasmids can be distinguished from each other. The PFG can then be Southern blotted and probed with a radiolabelled probe to identify each component of the genome.

Optimisation of PFGE using Gamma Radiation

To improve the quality of the PFGs and the visualisation of mega-plasmids using radiolabelled probes, the quantity of intact genomic DNA loaded had to be increased. It was recognised that circular DNA does not run into PFGs and will be retained in the wells (Game & Mortimer 1984). In order to visualise intact circular DNA in *Haloferax volcanii*, the molecules should be linearised during DNA preparation. Gamma radiation (Caesium-137) was used as a controllable means of introducing double-strand breaks into the DNA (Beverley 1989).

To determine the optimum exposure time to ^{137}Cs that would break the DNA molecules once (so that they would enter the PFG more easily) and not degrade the DNA, a variety of exposure times were chosen according to the dose rate of 4.5 Gy per minute: 1 minute, 2 minutes, 5 minutes, 10 minutes and 20 minutes. These doses were based on calculations by Beverly (1989). From the results shown in Figure 4.16 it was concluded that the optimum time for the main chromosome of *Haloferax volcanii* was between 5 minutes and 10 minutes, and for future experiments a time of 8 minutes was used (35 Gy). This time minimised smearing on the gel due to DNA degradation, whilst allowing more DNA to enter the gel for visualisation.

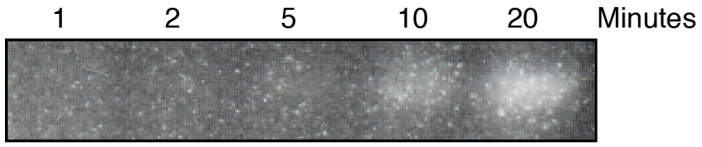


Figure 4.16 **Analysis of different exposure times of *Haloferax volcanii* genomic DNA to γ-radiation**. Whole genomic DNA was exposed to Caesium-137 for 1, 2, 3, 10 and 20 minutes and run on a Pulsed-field gel according to methods described in *Chapter 2: Materials and Methods, Section 2.2.6 Genotype Screening*. Figure shows chromosome band of *H. volcanii* genome, which becomes brighter as time increases due to more double strand breaks allowing more DNA to enter the gel.

PFG Analysis of pHV1

The mega-plasmid pHV1 is expected to migrate at 85 kb on a PFG but Norais *et al*. (2007) had observed two bands of ~500-700 kb and ~85 kb, which both hybridise with an *ori-pHV1* probe. To identify pHV1, a new probe was generated by PCR using the primers pHV1probeF and pHV1probeR, which amplifies a region not located near the origin or *orc* genes (Figure 4.17A). As can be seen in Figure 4.17B, the new pHV1-specific probe also hybridises to a band that migrates at around 510 kb, as well as 85 kb (the bands in Figure 4.17B are clearer than those seen by Norais *et al*. 2007 and can be sized more accurately).

One possible explanation for the 510 kb band is that there may have been a genome rearrangement involving pHV1 (Table 4.6). However, in this case pHV1 cannot have integrated into pHV4 as suggested by Norais *et al*. (2007), as pHV4 is 636 kb and therefore much larger than 510 kb. Moreover, in the strains used in this study, pHV4 had previously integrated into the main chromosome. Therefore, if pHV1 had also integrated into the chromosome or pHV4, this would give a band at around 3.59 Mb, which is not seen. Another explanation is the integration of pHV1 into the mega-plasmid pHV3, which would result in hybrid mega-plasmid of 523 kb in size (85 kb + 438 kb, respectively). Both pHV1 and pHV3 contain insertion sequence (IS) elements that could recombine with each other, similar to the integration of pHV4 onto the main chromosome. However, pHV3 has very few IS elements in comparison to its size (Hartman *et al*. 2010).

Table 4.6 **Possible sizes of integrated pHV1.**

DNA Molecules	Sizes of molecules	Total Size
pHV1 + Chromosome (+ integrated pHV4)	85 kb + 3.5 MB	3.59 MB
pHV1 + Chromosome	85 kb + 2.85 MB	2.94 MB
pHV1 + pHV4	85 kb + 636 kb	721 kb
pHV1 + pHV3	85 kb + 438 kb	523 kb

To investigate the possibility of pHV1 recombining with pHV3, the Southern blot of the PFG (Figure 4.17B) was left to decay until no signal could be detected and then re-probed with a radiolabelled probe specific for the *adh2* gene (*BamHI-NdeI* pTA1205 fragment), which is located on pHV3. The band seen at ~510 kb did not hybridise with the *adh2* probe and therefore cannot be the result of pHV1 integration into pHV3 (Figure 4.17C). However, a 438 kb band corresponding to pHV3 was seen, as expected.

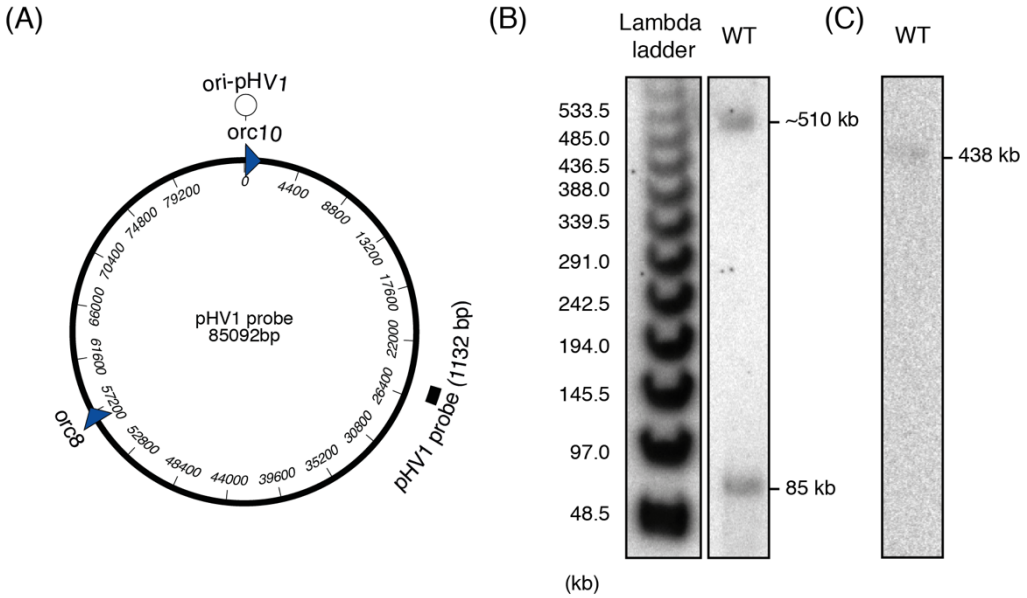


Figure 4.17 **pHV1 runs as two species on a PFG and is not integrated with other DNA molecules. (A)** Location of pHV1 specific probe used to probe Southern blot, generated by PCR using the primers pHV1probeF and pHV1probeR (1132 bp). **(B)** Whole genomic DNA was run on a pulsed-field gel and then subjected to Southern blot according to methods described in *Chapter 2: Materials and Methods, Section 2.2.6 Genotype Screening*. The Southern blot was probed with a pHV1 specific probe, and two bands were seen; one at 85 kb and one at around 510 kb. **(C)** The Southern blot was left to decay until no signal could be detected and re-probed with a pHV3 specific probe targeted at the *adh2* gene. This was found to hybridise to a band at 438 kb, which corresponds to the pHV3 mega-plasmid, not the band seen at ~510kb.

From this it was concluded that pHV1 has not integrated into any other mega-plasmid, or in fact the main chromosome, to produce a band with the size of 510 kb. Therefore, further investigation was required.

Concatemerisation of pHV1

Another explanation is that the 510 kb band is composed exclusively of pHV1 sequences. This would be possible if pHV1 exists as a concatemer in *H. volcanii*, in other words several copies of pHV1 could have joined together to form a larger DNA molecule. In fact, 510 kb is exactly 6 x 85 kb, suggesting that pHV1 forms a 6 x concatemer. What is not known from this is whether the 6 x concatemer of pHV1 is configured as one large circular DNA molecule or whether 6 circular, individual pHV1 molecules are interlinked. To test this, the DNA prepared for PFGE was digested with *AflII*, which cuts only once in pHV1. If the pHV1 mega-plasmid exists as a 6 x concatemer, then *AflII* should cut 6 times and there should only one band seen at 85 kb – at the same time, the band at 510 kb should disappear. As can be seen in Figure 4.18, this is in fact the case.

Figure 4.18 confirms that pHV1 exists as two species in *Haloferax volcanii*; one 85 kb in size and one a 6 x concatemer of 510 kb in size, and explains why pHV1 migrates as two species on a PFG. Furthermore, because pHV1 exists as a 6 x concatemer of 510 kb it fits with the model that there is a size threshold required for recombination-dependent replication, and therefore allowing *ori-pHV1* and *orc10* to be deleted, however it is not known whether *ori-pHV1* and *orc10* have been deleted from the concatemeric form of pHV1 or the single monomer form, or indeed if pHV1 uses recombination-dependent replication in the absence of its origin. However, as shown in Figure 4.10, even when *ori-pHV1* has been deleted from pHV1, pHV1 is still present in the cell in both monomeric and concatemeric forms.

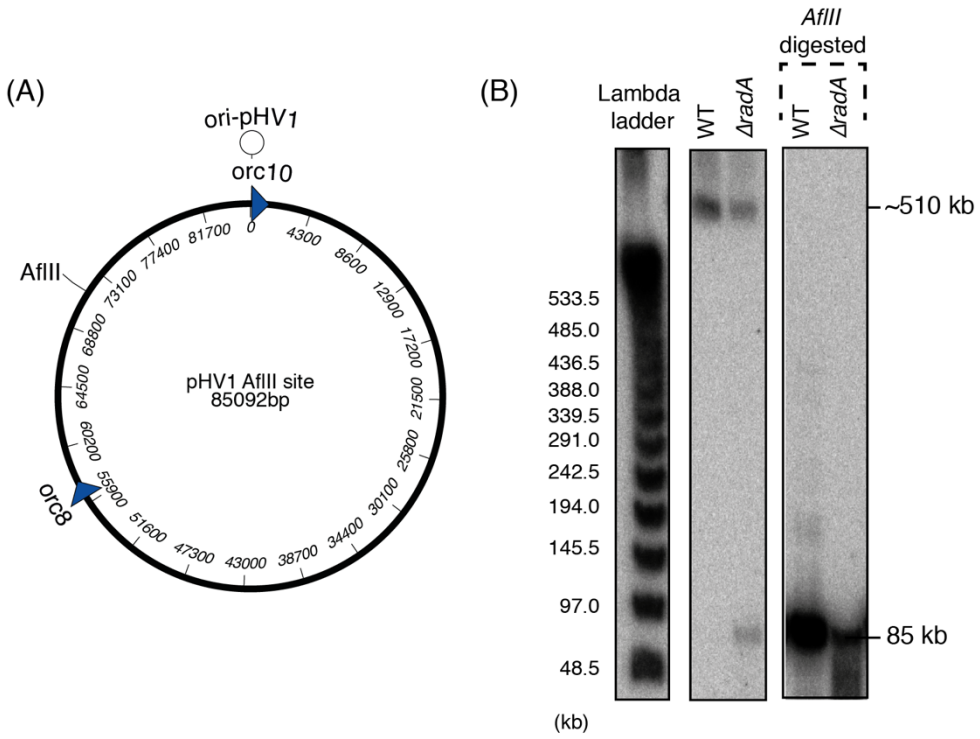


Figure 4.18 **pHV1 exists as a concatemer. (A)** The restriction enzyme *AflII* cuts once in pHV1. **(B)** Whole genomic DNA, and genomic DNA digests with *AflII* were run on a pulsed-field gel and then subjected to Southern blot according to methods described in *Chapter 2: Materials and Methods, Section 2.2.6 Genotype Screening*. The Southern blot was probed with a pHV1 specific probe, generated by PCR using the primers pHV1probeF and pHV1probeR. When digested, the pHV1 genome linearised into a single band at 85 kb, and the band at ~510 kb was no longer seen. Ladder is smeared due to unforeseen running conditions (>50°C for up to 12 hours).

In Figure 4.18, the 85 kb band cannot be seen in the wild-type strain (H53) undigested genomic DNA, whereas this band was seen clearly in the *ΔradA* strain H112. This may mean that the monomeric (85 kb) form of pHV1 is a transient species – since H112 is deficient in homologous recombination, any monomers present when *radA* was deleted would be incapable of any further recombination, thereby 'freezing' pHV1 in its two states as a monomer and a 6 x concatemer. If pHV1 is acting as a parasitic DNA molecule, for example it may be descended from a phage, deletion of *radA* would mean that it could no longer use homologous recombination to switch between monomeric and concatemeric forms. However, why does this needs to occur in the first place? If pHV1 was originally a phage, then its DNA may well be replicated as concatemers, which are resolved into monomers that are disseminated into the environment. The formation of concatemers may also have a repelling effect on

super-infection by another phage. As pHV1 is now a stable part of *H. volcanii* genome there may be no need to form monomers, however it is assumed that pHV1 retains the genes required for this is occur as monomers can be seen. It may be the case that pHV1 does not even use recombination-dependent replication as an origin-independent mode of DNA replication, and it may only require homologous recombination for the formation of monomers.

Other reasons why the 85 kb band is missing may be due to technical reasons, one of which was unforeseen running conditions of the gel. The cooling unit malfunctioned and the gel ran at >50°C for up to 12 hours. This caused the degradation and smearing of the ladder (Figure 4.18B) and may have degraded the DNA in the WT strain, causing the band to not be able to be seen. However, this does not necessarily explain why there is still DNA in the other lanes at 85 kb. Another possibility is that although care was taken to ensure optical density of the strains were the same, there may be less DNA present in the WT cells compared to the *ΔradA* strain, therefore causing the 85 kb band to be difficult to visualise due to insufficient DNA. This could also be due to the preparation of the plugs or loading of the gel and a lower quantity of DNA may have been loaded for the WT strain.

Chapter 5: Exploring Machinery used for Initiation of DNA Replication Including MCM and GINS

5.1 Background

The DNA replication machinery is conserved across all domains of life, but is more similar in eukaryotes and archaea than either are to bacteria. Because DNA replication in eukaryotes is rather complex, archaea such as *Haloferax volcanii* can be used as a model because their replication machinery is simpler.

In eukaryotes, the replicative helicase is a heterohexameric complex made up of six different proteins MCM2-7 and five accessory factors. It has a ring structure with two tiers made up of the N-terminal and C-terminal domains of the 6 MCM proteins, called the N-tier and the C-tier, respectively. The inactive MCM heterohexamer is loaded onto dsDNA by Cdt1 and Cdc6 during G1 phase to form the pre-replicative complex, this is known as origin licensing. It is then activated in S phase by cell-cycle dependent kinases and other accessory factors when it also interacts with Cdc45 and GINS to form the CMG complex, which unwinds DNA ahead of the replication fork. GINS and Cdc45 act as scaffolding proteins that stabilise the MCM2-7 helicase. The current accepted model is that the CMG complex translocates along DNA in a 3'-5' direction, although it has been proposed that the CMG complex is loaded head to head and moves in a 5'-3' orientation, with the C-tier motor pushing the N-tier along. This means that the two CMG complexes at a replication fork must pass each other and correctly remodel onto the correct DNA strand before the replication fork can continue (Sun *et al*. 2016; Yuan *et al*. 2016; Georgescu *et al*. 2017).

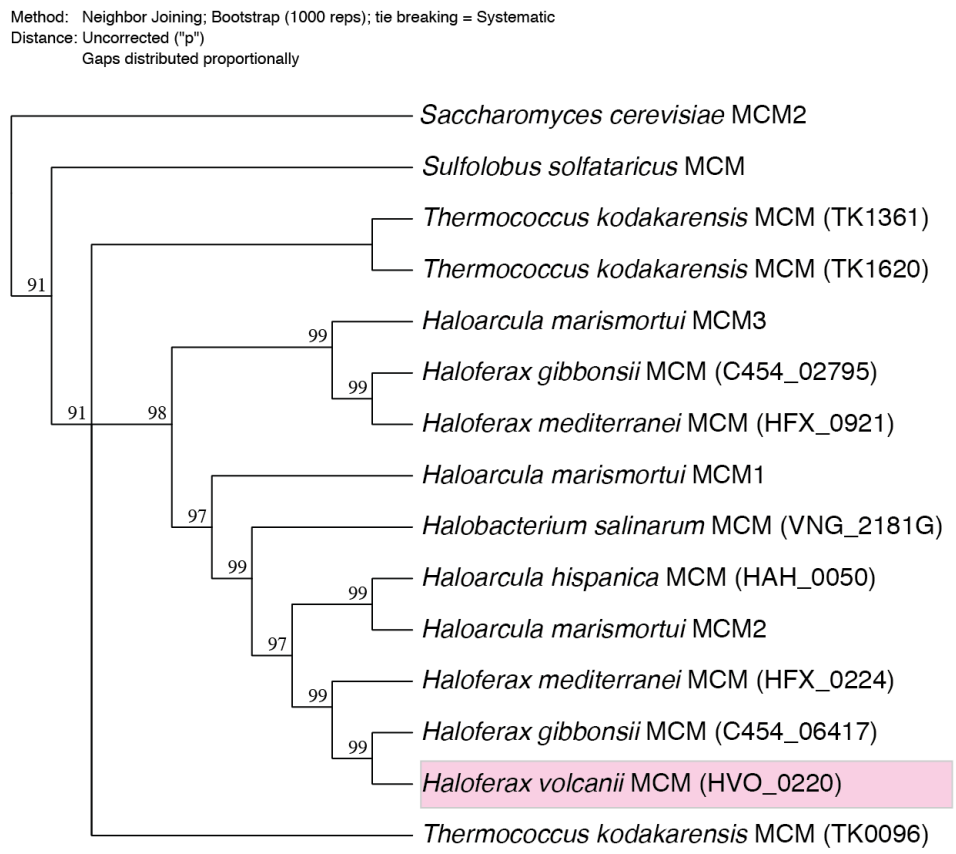


Figure 5.1 **Alignment of archaeal MCM proteins.** Tree is rooted to *Saccharomyces cerevisiae* MCM2. Halophile MCM proteins are closely related while *Thermococcus* and *Sulfolobus* are more distant. *Haloferax volcanii* MCM is highlighted in pink. Protein sequences were obtained from UniProt and alignment was created using MacVector software.

The replication machinery in archaea is similar to the eukaryotic machinery, with both MCM and GINS proteins present. MCM is generally a homohexamer, rather than a heterohexamer as seen in eukaryotes, although some organisms such as *Haloarcula marismortui* have more than one MCM present, but the role of these has yet to be explored (Figure 5.1). Alignments of *H. volcanii* MCM can be seen in Appendix *VII MCM*. MCM has an N-terminal domain and a C-terminal domain (Figure 5.2). The C-terminal domain is thought to interact with Orc proteins (Samson & Bell 2016) and is conserved across archaea and eukaryotes. It includes the AAA+ catalytic domain, which allows MCM to translocate by binding and hydrolysing ATP, and the winged-helix domain (Brewster *et al*. 2008).

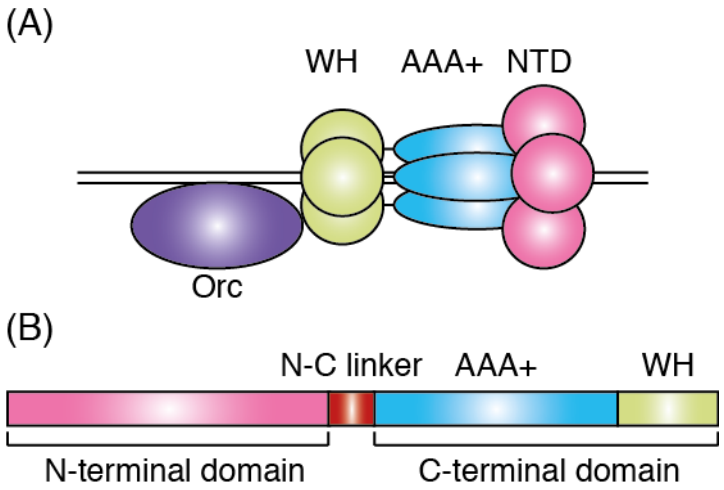


Figure 5.2 **Archaeal MCM diagram. (A)** A model showing the C-terminal domain on the MCM homohexamer interacts with Orc initiator proteins (purple). Adapted from Samson & Bell (2016). **(B)** Archaeal MCM has a classic MCM structure with an N-terminal domain (pink) and a C-terminal domain which includes a AAA+ domain (blue) and winged-helix domain (green). The two domains are bridged with an N-C linker region (red). Adapted from Brewster *et al*. 2008. Not to scale.

GINS in archaea is thought to be a four subunit protein similar to eukaryotes, however where eukaryotes have four different subunits, the archaeal GINS are split into two subunits; GINS23 and GINS51. *Sulfolobus, Pyrococcus* and *Thermococcus kodakarensis* have both the GINS23 and GINS51 subunits, but only one is found in *Haloferax volcanii*. Although in all cases GINS is still thought to act as a tetramer. *H. volcanii* GINS aligns with the GINS51 subunit of *Pyrococcus* and *Sulfolobus,* and has the GINS A domain located at the N-terminus like the archaeal GINS51/ eukaryotic Sld5 and Psf1 (Figure 5.3) (Appendix *VIII GINS*).

As in eukaryotes, MCM is loaded onto the leading strand of DNA synthesis and translocates 3'-5'. The homologue of Cdc45 is not certain but a GINS-associated nuclease (GAN) has been identified in *Thermococcus kodakarensis* that has homology to the bacterial RecJ nuclease. Nearly all archaea contain a RecJ homologue, which is the prime candidate to interact with MCM and GINS during DNA replication. In *Sulfolobus* it has been shown that GINS interacts with the N-terminus of MCM, and a Cdc45 homologue that resembles RecJ is present that interacts with GINS (Makarova *et al*. 2012; Xu *et al*. 2016; Nagata *et al*. 2017; Miller & Enemark 2015).

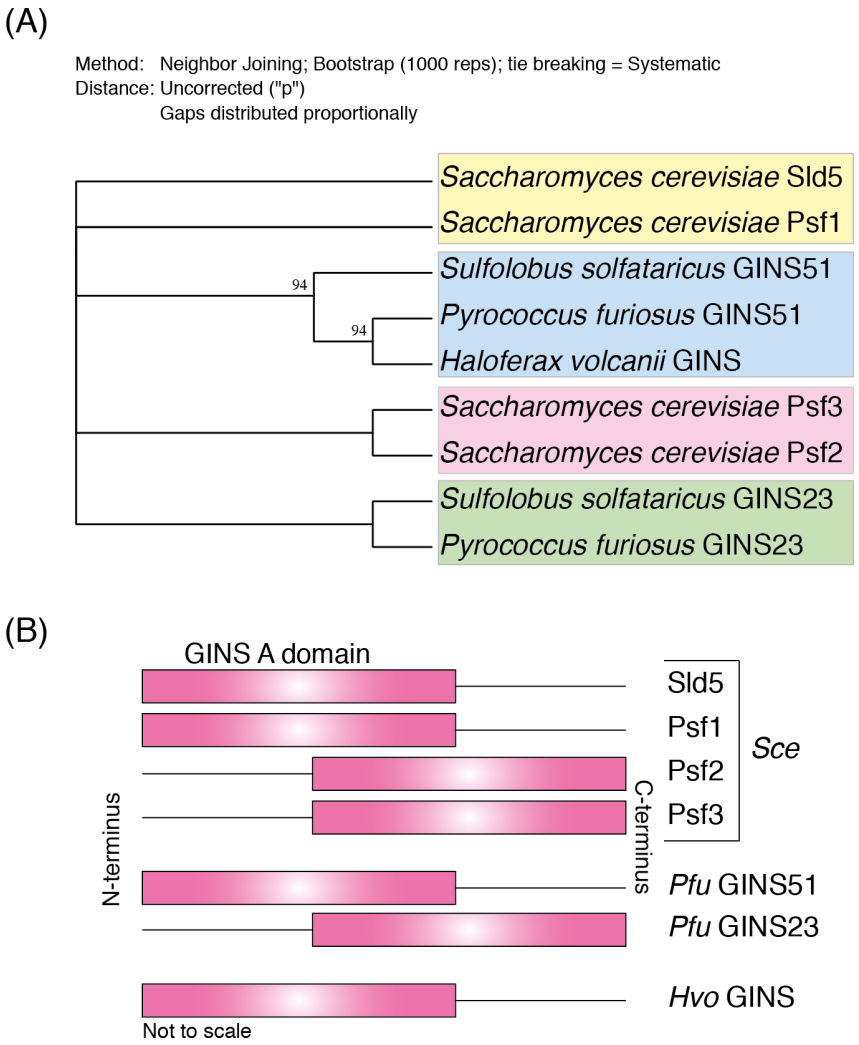


Figure 5.3 ***Haloferax volcanii* GINS is similar to archaeal GINS51 and eukaryotic Sld5 and Psf1. (A)** Alignment of eukaryotic and archaeal GINS proteins. Tree is rooted to *Saccharomyces cerevisiae* Sld5. *Haloferax volcanii* GINS is closely related to archaeal GINS51 (blue). Protein sequences were obtained from UniProt and alignment was created using MacVector software. **(B)**The GINS A domain (pink) is located at the N-terminus in *Haloferax volcanii* (*Hvo*) GINS (HVO_2698) similarly to *Pyrococcus furiosus* (*Pfu*) GINS51 (PF_RS04940) and *Saccharomyces cerevisiae* (*Sce*) Sld5 and Psf1, whereas *Pfu* GINS23 (PF_RS02475) and *Sce* Psf2 and Psf3 have the GINS A domain located at the C-terminus. Adapted from MacNeil 2010. Not to scale.

While it is clear that MCM and GINS are vital parts of the CMG (Cdc45-MCM-GINS) replicative helicase complex in eukaryotes and archaea, direct or indirect interactions between these two proteins in *H. volcanii* have yet to be explored.

5.2 Aims

Analysis of the replication machinery of *Haloferax volcanii*, specifically the replicative helicase complex, may give insights into the mechanisms used for origin-dependent replication and recombination-dependent replication. For example, it is possible that the faster growth seen in strains without origins might require different enzymes to carry out aspects of DNA replication. In yeast, break-induced replication relies upon the Pif1 helicase and not the normal replicative helicase, MCM (Saini *et al*. 2013). The aims of this chapter are to:

- Construct an inducible allele of the *mcm* gene and use this to replace the native *mcm* in strains with and without origins, to determine whether the amount of MCM in the cell affects growth rates and whether this differs in originless strains. The latter will indicate whether MCM is an essential part of the replication machinery for recombination-dependent replication.
- Tag MCM and GINS with epitope tags and express in *H. volcanii* from chromosomal alleles under the control of their native promoters. The affinity-purified MCM and GINS will be examined for interactions by co-purification. This may differ between wild-type and origin-less strains depending on whether the replication machinery used for recombination-dependent replication is the same as origin-dependent replication.

5.3 Results

5.3.1 Tryptophan Inducible MCM

p.tnaA-mcm

The strains H1794 (*ΔpyrE2, ΔhdrB, p.tnaA-mcm+::hdrB+*) and H1802 (*ΔpyrE2, ΔhdrB, ΔoriC1,C2,C3,pHV4, p.tnaA-mcm+::hdrB+*) were both constructed by Thorsten Allers in 2013, prior to this study (unpublished data). The plasmid pTA1411 was used to replace the native *mcm* gene with a construct where *mcm* is under the control of the tryptophan inducible promoter *p.tnaA*.

pTA1411 (*p.tnaA-mcm::hdrB+* gene replacement construct)

The plasmid pTA1411 (*p.tnaA-mcm::hdrB+* gene replacement construct) was constructed by Thorsten Allers in 2013 prior to this study (unpublished data) (Figure 5.4).

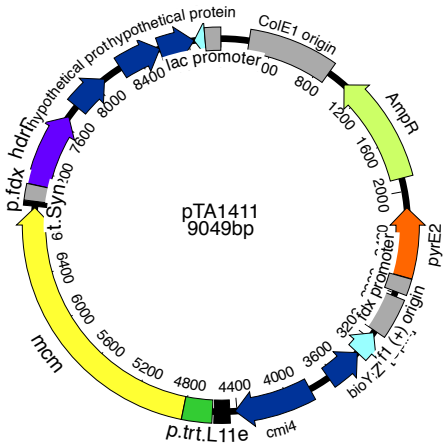


Figure 5.4 **pTA1411**. Gene replacement construct to put *mcm* (yellow) under the control of tryptophan inducible promoter *p.tnaA* (dark green) with *hdrB* marker (purple). Constructed by Thorsten Allers in 2013 (unpublished data).

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 inducer for this promoter is high, making this promoter ideal for control of other genes (Large *et al*. 2007; Allers *et al*. 2010). As part of this study they are checked by Southern blot (Figure 5.5). These strains were generated to examine the effect of reducing the levels of MCM in the cell, and to compare differences (if any) between wild-type and origin deleted strains in their requirement for MCM.

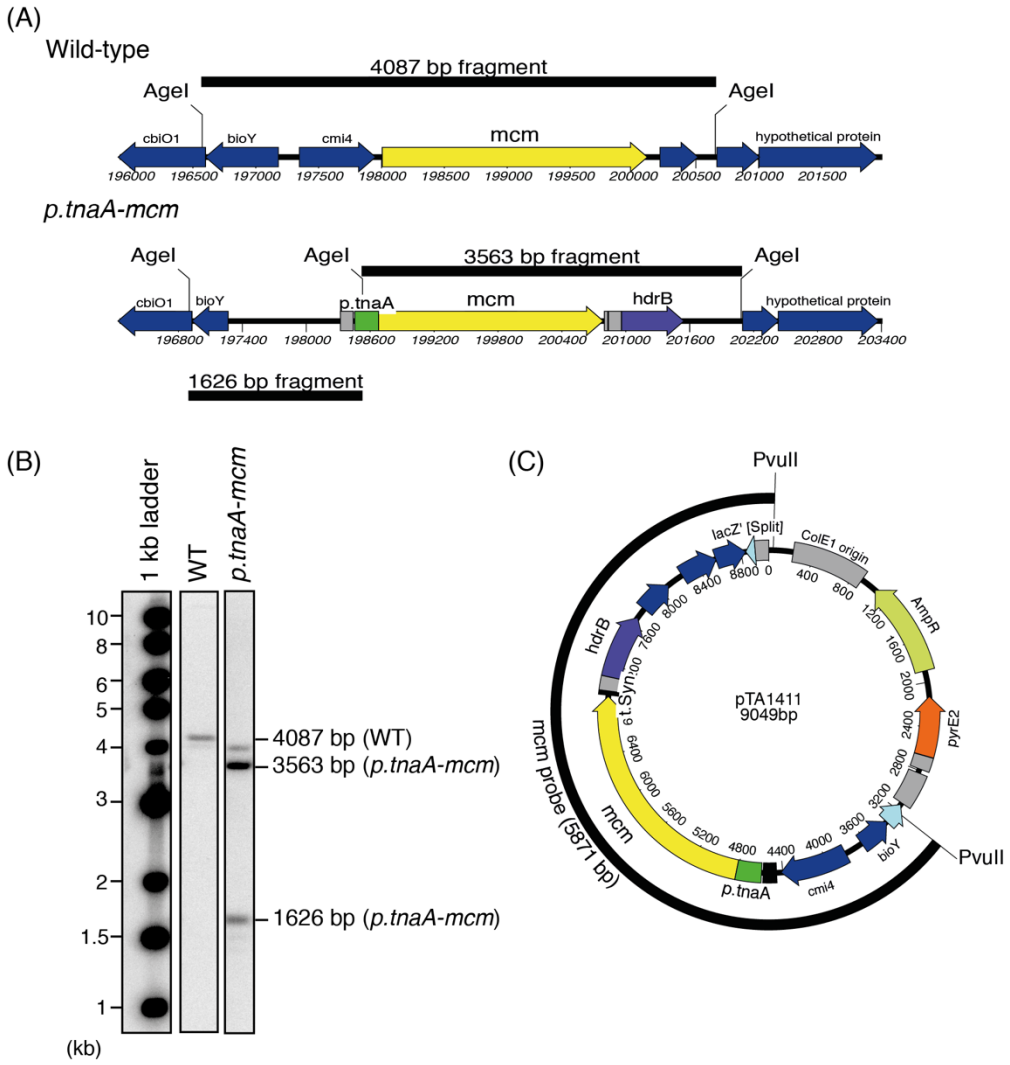


Figure 5.5 ***p.tnaA-mcm* strains checked by Southern blot. (A)** Predicted fragment sizes of wild-type (4087 bp) and *p.tnaA-mcm* strains (3563 bp and 1626 bp) with an *AgeI* digest of genomic DNA. **(B)** Example of Southern blot showing *p.tnaA-mcm* strain H1794 compared to wild-type, giving the expected band sizes (indicted on figure). **(C)** Probe used in Southern blot detection was *PvuII* fragment of pTA1411 (5871 bp).

H1794 and H1802 were plated onto tryptophan gradient plates (for details see *Chapter 2: Materials and Methods, Section 2.2.7 Phenotyping of Haloferax volcanii*) which are designed to form a gradient of tryptophan across the plate from 0 mM to a desired concentration on the other side. The range of maximum tryptophan concentrations used were; 0.25 mM, 0.5 mM, and 1 mM. The required strains are painted across the gradient and allowed to grow. If the gene under the *p.tnaA* tryptophan-inducible promoter is essential, strains will only grow on the side of the plate where tryptophan is present, since the *p.tnaA* promoter will be induced and therefore the associated gene transcribed and

translated. Conversely if the gene is non-essential, strains will also be able to grow where there is 0 mM tryptophan. It should be noted that both H1794 and H1802 are both are *trp+* (i.e. they are not *ΔtrpA*).

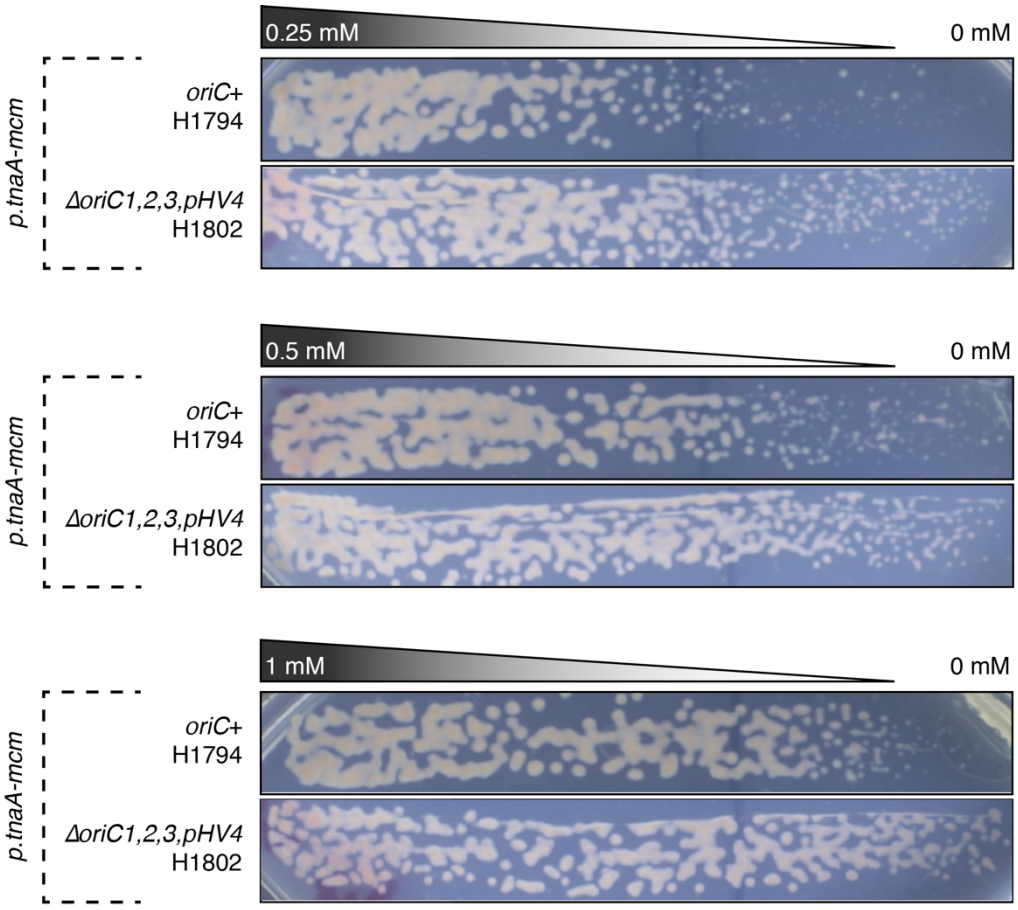


Figure 5.6 ***p.tnaA-mcm* strains on tryptophan gradient plates.** H1794 (*oriC+*) and H1802 (*ΔoriC1,2,3,pHV4*) were painted onto tryptophan gradient plates with tryptophan concentrations of 0.25 mM, 0.5 mM and 1 mM. Cells are viable across the whole gradient in all cases.

There was no observable difference in growth between the wild-type (H1794) and origin-deleted strain (H1802) at 0.5 mM and 1 mM tryptophan (Figure 5.6). At 0.25 mM there is a slight difference in growth between the two strains, however it is not definitive and may be due to slight differences in cell densities when plating. Moreover, cell growth was observed at 0 mM tryptophan where no MCM should be transcribed. Since MCM is regarded as essential (at least for origin-dependent DNA replication), this is unexpected. It is possible that the *p.tnaA* promoter is leaky and cells require very little MCM to be able to undergo successful DNA replication. To test this, a lower activity promoter was used

instead, which would lead to less induction of the *mcm* gene by tryptophan. By reducing the levels of MCM in the cell, it was hoped that a notable difference between wild-type and origin-deleted strains could be seen, so it can be determined whether MCM plays a role in recombination-dependent replication.

mcm Under the Control of the Promoter *p.tnaAM3*

The strains H1904 and H1911, wild-type and origin-deleted, respectively, were generated to examine the effect of reduced levels of MCM in the cell. The alternative tryptophan-inducible promoter *p.tnaAM3* (kindly donated by Stuart MacNeill, University of St. Andrews) was used so that lower levels of expression could be achieved, compared to the native tryptophan-inducible promoter *p.tnaA*. The promoter *p.tnaAM3* differs from the *p.tnaA* promoter by a single point mutation in the putative TATA box (Figure 5.7A), resulting in around 50% less activity upon induction; it also has a comparable reduction in basal activity (for details see *Chapter 3: Plasmid and Strain Construction, Section 3.1.2.3 Construction of Plasmids for Chapter 6: DNA Replication Machinery*).

pTA1460 (*p.tnaAM3-mcm::hdrB+* gene replacement construct)

The plasmid pNPM-tnaAM3-HfxMCM (courtesy of Stuart MacNeill, University of St. Andrews) contains the promoter *tnaAM3* which has a point mutation (T->G) in the TATA box making it 50% less active. pNPM-tnaAM3-HfxMCM was digested with *BclI* (blunt-ended) and *NdeI* and ligated into pTA1369 (constructed by Thorsten Allers in 2012, unpublished data) at the *ClaI* (blunt-ended) and *NdeI* sites to create pTA1451. The plasmid pTA1451 was digested with *XhoI* and *NdeI* and the fragment containing *p.tnaAM3* was ligated into pTA1411 at the *XhoI* and *NdeI* sites to create pTA1460 *p.tnaAM3-mcm::hdrB+* gene replacement construct (Figure 5.7).

(A)

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

(B)

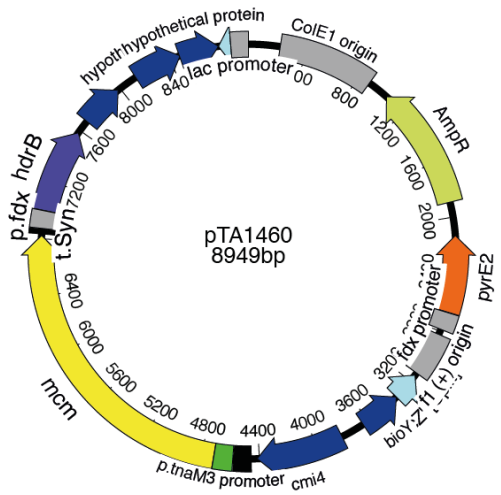


Figure 5.7 **pTA1460**. **(A)**TATA box of *tnaA* and *tnaAM3* promoters showing point mutation from T -> G, respectively. **(B)** pTA1460 gene replacement construct with *mcm* under the control of the *tnaAM3* promoter, with an *hdrB* marker.

The plasmid pTA1460 was used to transform the strain H1530, derived from H99, and, the strain H1608, derived from H53 to produce the strains H1904 (*ΔpyrE2, ΔhdrB, p.tnaAM3-mcm+::hdrB+*) and H1911 (*ΔpyrE2, ΔhdrB, ΔoriC1,C2,C3,pHV4, p.tnaAM3-mcm+::hdrB+*), respectively. This gene replacement puts *mcm* under the control of the low activity tryptophan inducible promoter *p.tnaAM3*. The strains were constructed by Thorsten Allers in 2014, but as part of this study they were checked by Southern blot (Figure 5.8).

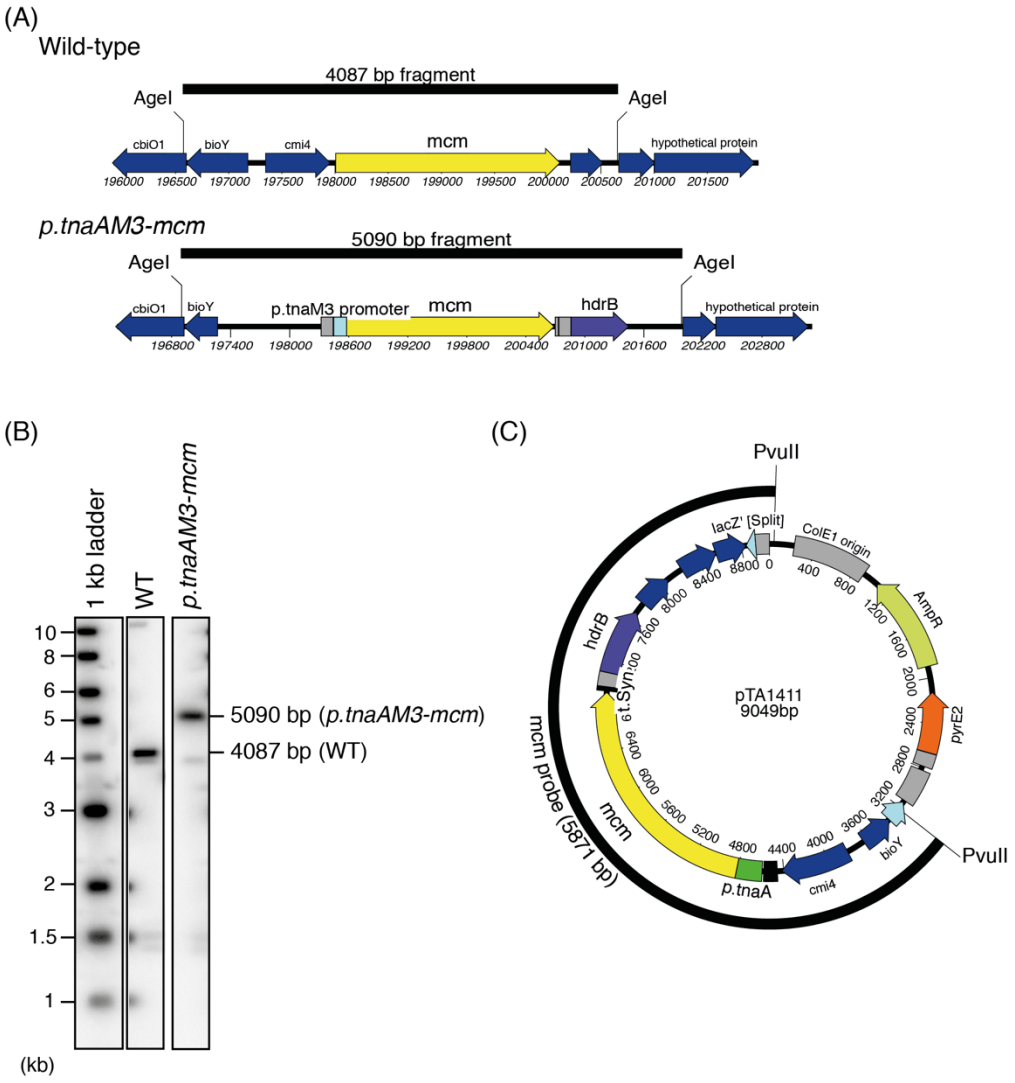


Figure 5.8 ***p.tnaAM3-mcm*** **strains checked by Southern blot. (A)** Predicted fragment sizes of wild-type (4087 bp) and *p.tnaAM3-mcm* strains (5090 bp) with an *AgeI* digest of genomic DNA. **(B)** Example of Southern blot showing *p.tnaAM3-mcm* strain H1904, compared to wild-type, giving the expected band sizes (indicted on figure). **(C)** Probe used in Southern blot detection was *PvuII* fragment of pTA1411 (5871 bp).

To visually compare growth differences between the strains H1904 and H1911, growth in liquid culture had to be achieved prior to plating on the tryptophan gradient plates. However, no growth was seen in 5 ml overnight liquid cultures and therefore the conditions were optimised. H1904 and H19011 were grown in Hv-Ca broth with different concentrations of tryptophan; 0.125 mM, 0.25 mM, 0.5 mM, 1 mM, and 2.5 mM, and the A_{650} was measured each day. As shown in Table 5.1, H1904 grew best in 0.25 mM tryptophan (which is the normal tryptophan concentration used in *H. volcanii* media) and H1911 grew best in 1 mM tryptophan (4 x normal concentration).

Table 5.1 **Growth in liquid of H1904 and H1911 with different tryptophan concentrations.**

Tryptophan Concentration (mM)	Average A_{650} after 24 hours growth	
	H1904	H1911
0.125	0.32	0.41
0.25	0.92	0.30
0.5	1.01	0.75
1	1.01	1.06
2.5	1.12	1.21

H1904 and H1911 were grown at their respective optimal concentrations of tryptophan and plated onto tryptophan gradient plates (Figure 5.9). A lower tryptophan concentration was used in the plates (0.125 mM), due to conclusions from previous experiments (using *p.tnaA-mcm*) that cells require very little induction of the *mcm* gene for growth. However, growth was very slow on the plates (9 days compared to 4 days) and it is likely that the tryptophan had diffused across the plate by the time that the colonies had grown, invalidating any results. However, there was an indication that the origin deletion strain H1911 requires more tryptophan and therefore more MCM for growth. This fits with the suggestion that faster growth in the absence of origins is due to more (recombination-dependent) replication, and this will require more DNA replication machinery; this also implies that MCM plays a key role in recombination-dependent replication.

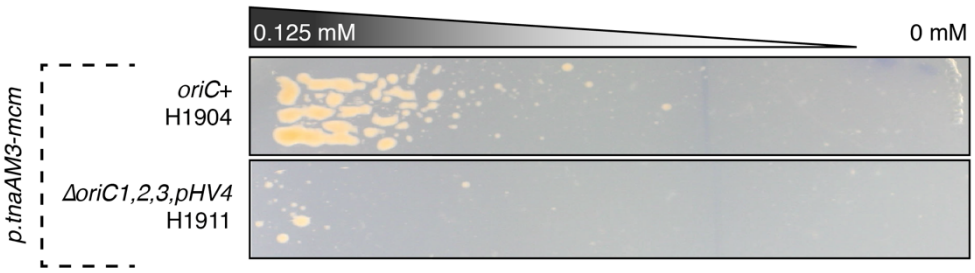


Figure 5.9 ***p.tnaAM3-mcm* strains on tryptophan gradient plates.** H1904 (*oriC+*) and H1911 (*ΔoriC1,2,3,pHV4*) were painted onto tryptophan gradient plates with a tryptophan concentration of 0.125 mM.

In this instance, tryptophan gradient plates proved to be an unsatisfactory method to measure differences in growth that are due to limiting MCM in wild-type vs origin-deleted strains. Therefore, growth in in liquid culture was measured instead. Strains were grown over two consecutive overnights (with 0.5 mM tryptophan) to give a vigorously-growing culture, and cell growth during exponential phase was then measured over a time course of 72 hours using a microplate spectrophotometer. A growth curve was plotted and relative generation times calculated during exponential phase (Figure 5.10).

Only a negligible difference was found in the generation times of H1904 and H1911, even when no tryptophan was added to the growth media. It was postulated that due to growing the strains in a relatively high concentration of 0.5 mM tryptophan prior to loading onto the microplate spectrophotometer, some tryptophan might have been carried over into the fresh media, thereby allowing cell growth. To avoid this, strains were grown in Hv-Ca with 0.5 mM tryptophan for one overnight and then diluted into fresh Hv-Ca with no tryptophan for another overnight. This culture was then loaded onto the microplate spectrophotometer, a growth curve plotted and relative generation times calculated (Figure 5.11). This method was chosen over an alternative method of pre-growing the cells in Hv-Ca with 0.5 mM tryptophan and then pelleting the cells, removing the supernatant and resuspending in fresh Hv-Ca broth with 0 mM tryptophan, as *Haloferax* cells are very sensitive to centrifugation, which puts them in a state of stress, therefore possibly affecting growth rates and not giving a true representation of growth regarding tryptophan concentrations.

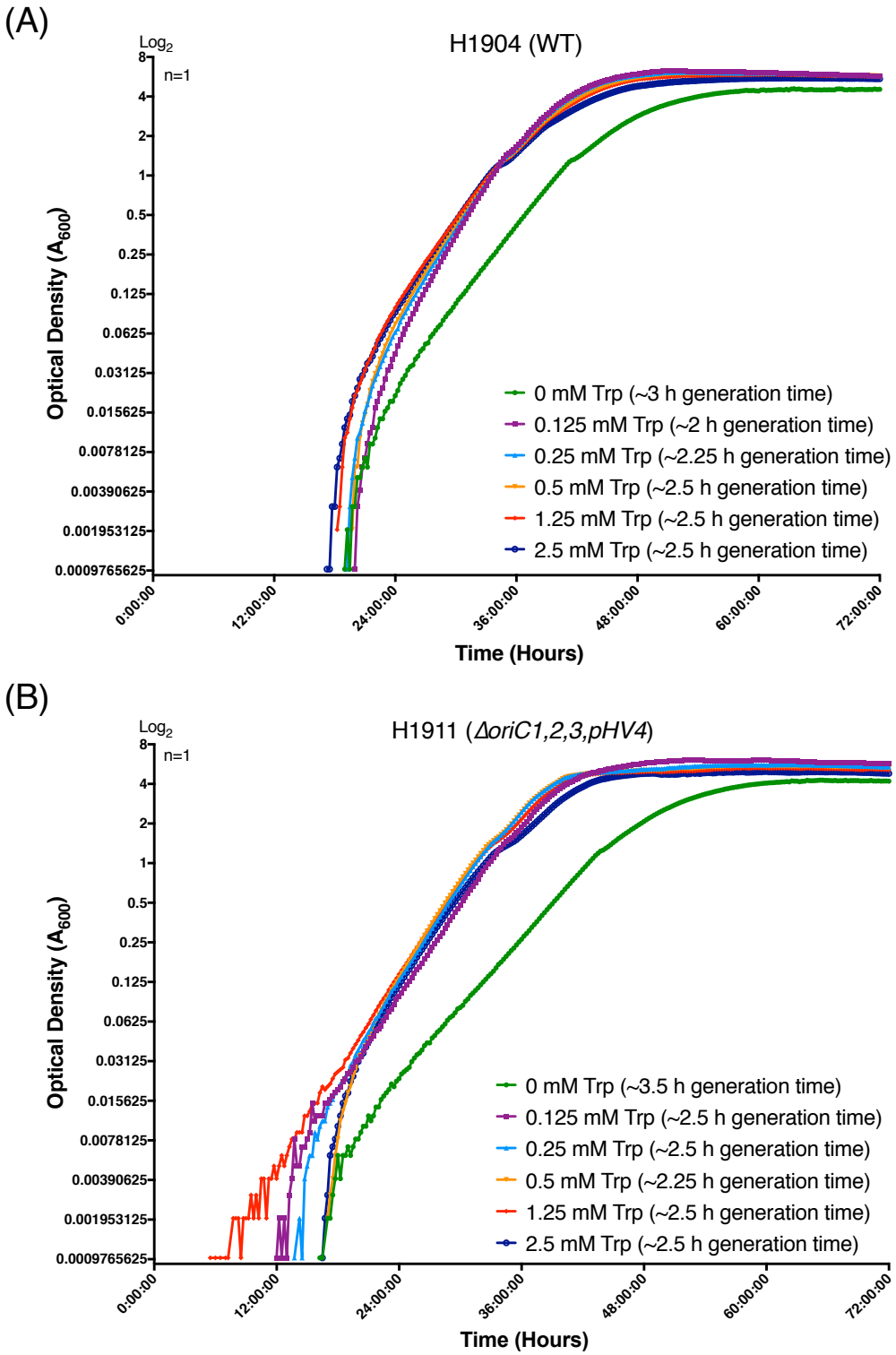


Figure 5.10 **Exponential growth rate of p.*tnaAM3-mcm* strains in different tryptophan concentrations. (A)** H1904 (wild-type) and **(B)** H1911 (*ΔoriC1,2,3,pHV4*) were grown in liquid culture for two consecutive overnights with 0.5 mM tryptophan and then diluted and grown for a further 72 hours with a range of tryptophan concentrations (0 mM, 0.125mM, 0.25 mM, 0.5 mM, 1.25 mM, 2.5 mM). Growth was measured by A_{600}. Generation time is indicated next to each tryptophan concentration. All strains (n=1) were incubated on the same 48 well plate and measured simultaneously using an Epoch2 Microplate Spectrophotometer (BioTek).

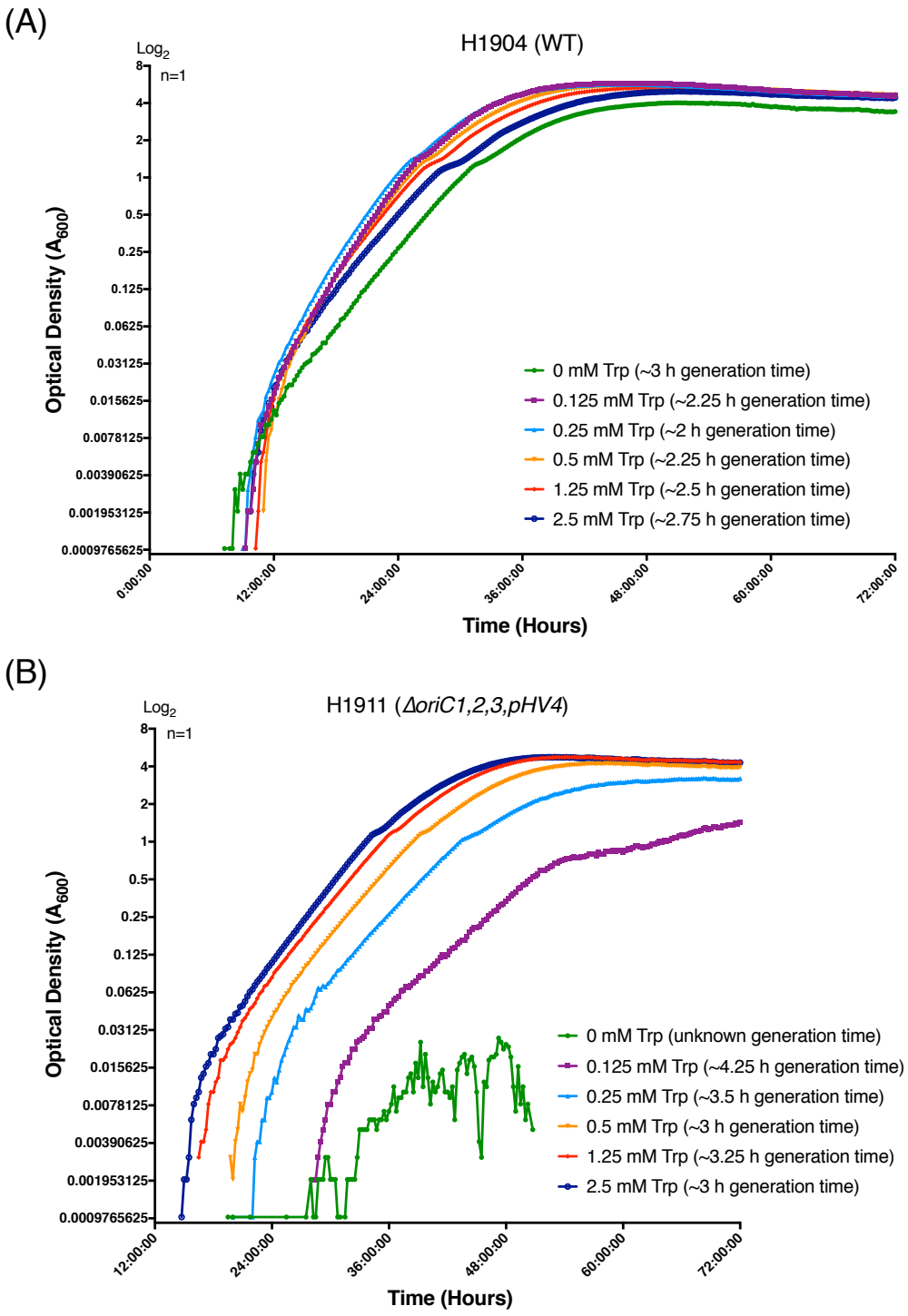


Figure 5.11 **Exponential growth rate of p.*tnaAM3-mcm* strains in different tryptophan concentrations pre-grown without tryptophan. (A)** H1904 (wild-type) and **(B)** H1911 (*ΔoriC1,2,3,pHV4*) were grown in liquid culture for two consecutive overnights; one overnight with 0.5 mM tryptophan and one overnight with 0 mM tryptophan. The strains were then diluted and grown for a further 72 hours with a range of tryptophan concentrations (0 mM, 0.125mM, 0.25 mM, 0.5 mM, 1.25 mM, 2.5 mM). Growth was measured by A_{600}. Generation time is indicated next to each tryptophan concentration, H1911 0mM tryptophan generation time was not able to be calculated. All strains (n=1) were incubated on the same 48 well plate and measured simultaneously using an Epoch2 Microplate Spectrophotometer (BioTek).

As depicted in Figure 5.11, these results show that in origin-deleted strain H1911, the generation time when grown at 0 mM tryptophan could not be calculated due to cell death, while the wild-type H1904 had a generation time of ~3 hours at 0 mM tryptophan. This indicates that when MCM is depleted in a strain where the chromosomal origins are deleted, there is a profound growth defect and even cell death due to lack of MCM. By contrast, in a wild-type strain where all the origins are present, depletion of MCM does not affect the growth rate (as much). Again, this supports the suggestion that faster growth is due to more (recombination-dependent) replication, which will require more MCM, and that MCM is required for origin-independent replication in *Haloferax volcanii*. If time had allowed, actual expression levels of MCM at each tryptophan concentration would have been checked by Western blot. This would show directly how the levels of tryptophan affect the levels of MCM present in the cell.

It is clear from the preliminary work undertaken in this study that MCM is required for origin-independent replication (most likely recombination-dependent replication), as its depletion severely affects cell growth in origin-deleted strain compared to a wild-type strain. This data also shows that despite this, only low levels of MCM are required in both the origin-deleted strain and the wild-type strain, however, the former requires more MCM. This is presumably due to initiation of recombination-dependent replication being able to take place all over the genome, compared to defined points (origins), and therefore there are higher numbers of replication forks, requiring more MCM to be present in the cell.

Deletion of tryptophanase

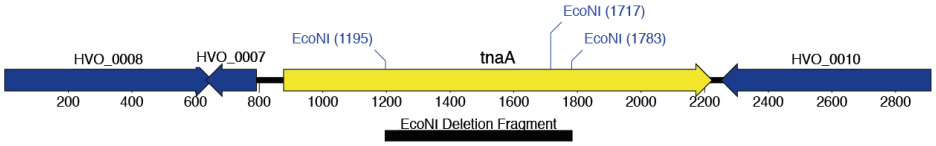
The presence of tryptophanase may be affecting growth rates of the wild-type and origin deleted strains, H1904 and H1911, respectively. Tryptophanase is the enzyme responsible for the breakdown of tryptophan within the cell, its transcription is tightly regulated by the promoter *p.tnaA* described above. Transcription from *p.tnaA* is induced by tryptophan, leading to the expression of tryptophanase and to the breakdown of tryptophan. Because tryptophan is used

to induce expression of the *mcm* gene, this will also induce expression of tryptophanase, which may in turn lead to reduced expression of MCM. This however, is more likely to be important for the over-expression of proteins, rather than the depletion of proteins in the cell. This is because for depletion, low levels of tryptophan are required and therefore tryptophanase will probably have less effect on levels of tryptophan than over-expression (which involves high concentrations of tryptophan).

pTA1730 (*tnaAΔEcoN* partial deletion construct)

Constructed by Thorsten Allers in 2015 (unpublished data) (Figure 5.12). This plasmid deletes part of the tryptophanase gene, *tnaA*, which renders *tnaA* inactive.

(A)



(B)

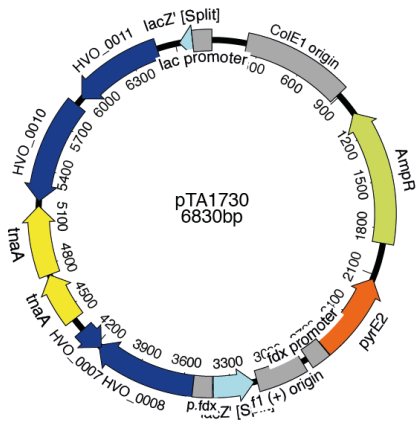


Figure 5.12 **pTA1730**. **(A)** *tnaA* region showing *EcoNI* sites and deleted fragment. **(B)** pTA1730 *tnaA* deletion construct with 5' and 3' ends of *tnaA* still present (constructed by Thorsten Allers in 2015, unpublished data).

To prevent the breakdown of tryptophan inducer, part of the tryptophanase gene *tnaA* was deleted which inactivates it causing the cell to become unable to breakdown tryptophan. H1904 and H1911 were transformed with the deletion construct pTA1730 to generate strains H2928 (*ΔpyrE2, ΔhdrB, p.tnaAM3-*

mcm+::hdrB+, tnaAΔEcoN) and H2931 (*ΔpyrE2, ΔhdrB, ΔoriC1,C2,C3,pHV4, p.tnaAM3-mcm+::hdrB+, tnaAΔEcoN*), respectively.

Growth Assay

The strains H2928 and H2931 were grown in the microplate spectrophotometer alongside H1904 and H1911 at a variety of different tryptophan concentrations. All cultures were pre-grown in Hv-Ca broth with 0.5 mM tryptophan. Growth curves were plotted and relative generation times calculated (Figure 5.13).

The growth curves show that H2928 and H2931 have a faster generation time than H1904 and H1911. This indicates that the absence of a functional tryptophanase gene improves growth. However, there was no difference in growth rate between H2928 and H2931, indicating that the presence or absence of chromosomal origins had no effect on growth (in this strain background). This was also the case for H1904 and H1911, and could be due to the pre-growth conditions discussed previously. For strains with *mcm* under the control of the *p.tnaAM3* promoter, it is clear that deleting the tryptophanase gene has no effect and in fact makes it difficult to see subtle growth differences. Nonetheless, deleting the tryptophanase gene may have a use in protein overexpression for other proteins, due to the potential for improved growth regardless of tryptophan concentration.

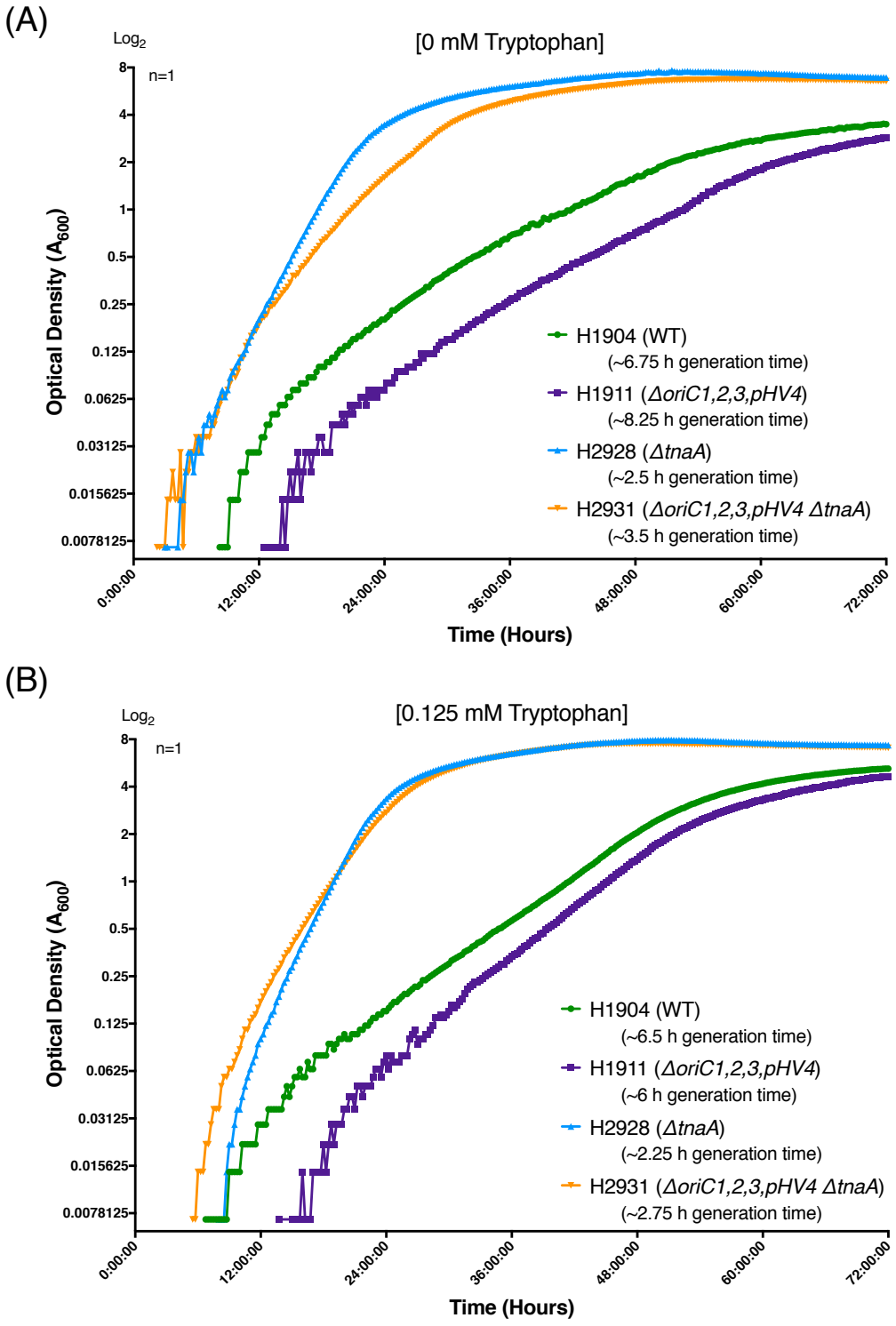


Figure 5.13 **Exponential growth rate of p.*tnaAM3-mcm* strains ±*tnaA* in different tryptophan concentrations.** H1904 (*oriC*+), H1911 (*ΔoriC1,2,3,pHV4*), H2928 (*oriC*+, *ΔtnaA*), and H2931 (*ΔoriC1,2,3,pHV4*, *ΔtnaA*) were grown in liquid culture for two consecutive overnights with 0.5 mM tryptophan. The strains were then diluted and grown for 72 hours with **(A)** 0 mM tryptophan and **(B)** 0.125 mM tryptophan. Growth was measured by A_{600}. Generation time is indicated next to each strain. All strains (n=1) were incubated on the same 48 well plate and measured simultaneously using an Epoch2 Microplate Spectrophotometer (BioTek).

5.3.2 Epitope-Tagged Replication Enzymes

Choice of Replication Enzymes

There is a wide range of proteins within the DNA replication fork that could be investigated further in *H. volcanii*, and recombination-dependent replication may use different proteins to replicate DNA compared to proteins found in the origin-dependent replication fork. In both processes, the double stranded DNA must unwound into single strands for replication to occur and it is likely to be carried out by the same replicative helicase. Alternatively, it is possible that recombination-dependent replication may use an atypical mode of DNA replication such as break-induced replication in yeast, which uses bubble migration of a D-loop structure catalysed by Pif1 helicase (Saini *et al*. 2013; Saini *et al*. 2017). Pif1may also have a role in recombination (Wilson *et al*. 2013), suggesting that an archaeal homologue of Pif1 may act in recombination-dependent replication; however, there are no clear Pif1 homologues in archaea. MCM and GINS were chosen as they are both important parts of the replicative helicase that is involved in normal origin-dependent replication in archaea. Furthermore, the results presented above using a tryptophan-inducible MCM indicate that this replicative helicase is also used for recombination-dependent replication. It is not yet clear which archaeal protein fulfils the role of Cdc45, the third component of the CMG complex in eukaryotes alongside MCM and GINS. However, there is some indication that archaeal RecJ (otherwise known as GINS-associated-nuclease, GAN) may have a role in the replicative helicase (Li *et al*. 2010; Oyama *et al*. 2016; Pellegrini 2017; Ogino *et al*. 2017). This must be explored further but was not chosen for investigation, since there are four RecJ proteins in *Haloferax volcanii* and it is unclear which one (if any) is acting as the equivalent of Cdc45.

Choice of Epitope Tag

There are many epitope tags available for affinity purification of recombinant (Table 5.2). *H. volcanii* proteins must be maintained in molar salt conditions to retain activity and correct folding, therefore many of these affinity tags are unsuitable. Smaller tags are also preferred due to fact they are less likely to

interfere with protein function and therefore physiological protein:protein interactions are more easily observed. For these reasons, the most appropriate tags for studying interactions of halophilic proteins are His_6 and StrepII.

Table 5.2 **Common epitope tags used in protein purification**

Tag	Size (aa)	Matrix
Hexahistidine (His_6)	6	Metal affinity
StrepII	8 (WSHPQFEK)	StrepTactin sepharose
FLAG	8 (DYKDDDDK)	Anti-FLAG M2 mAb agarose
Maltose binding protein (MBP)	396	Amylose
Glutathione *S*- transferase (GST)	218	GSH-sepharose
Calmodulin-binding peptide (CBP)	28	Calmodulin affinity
Chitin-binding domain (CBD)	51	Chitin
HA	9	Anti-HA epitope mAb

His_6 tags have previously been used to purify proteins from *H. volcanii* but this species contains many histidine-rich proteins that contaminate metal affinity chromatography purifications (Stroud *et al*. 2012). This can be mostly circumvented by the use of a specific protein-expression strain of *H. volcanii* in which most of these contaminant proteins have been removed or modified. For example, the histidine rich *pitA* gene has been replaced with an orthologue *Nph-pitA* from *Natronomonas pharaonis* that lacks a histidine-rich linker, and the *cdc34d* gene has been replaced with a modified version *cdc48d-ct* that is missing its histidine rich C-terminus. These modifications both avoid contamination by co-purification with His_6 tagged proteins (Allers *et al*. 2010). An octa-histidine tag could be used which would increase binding of the tag to the metal affinity column used in purification (Bornhorst & Falke 2000), however this would also mean that increased concentrations of imidazole must be used for elution of the tagged protein. Increased levels of imidazole may destabilise the protein and must be removed for further downstream applications, therefore a His_6 tag was chosen.

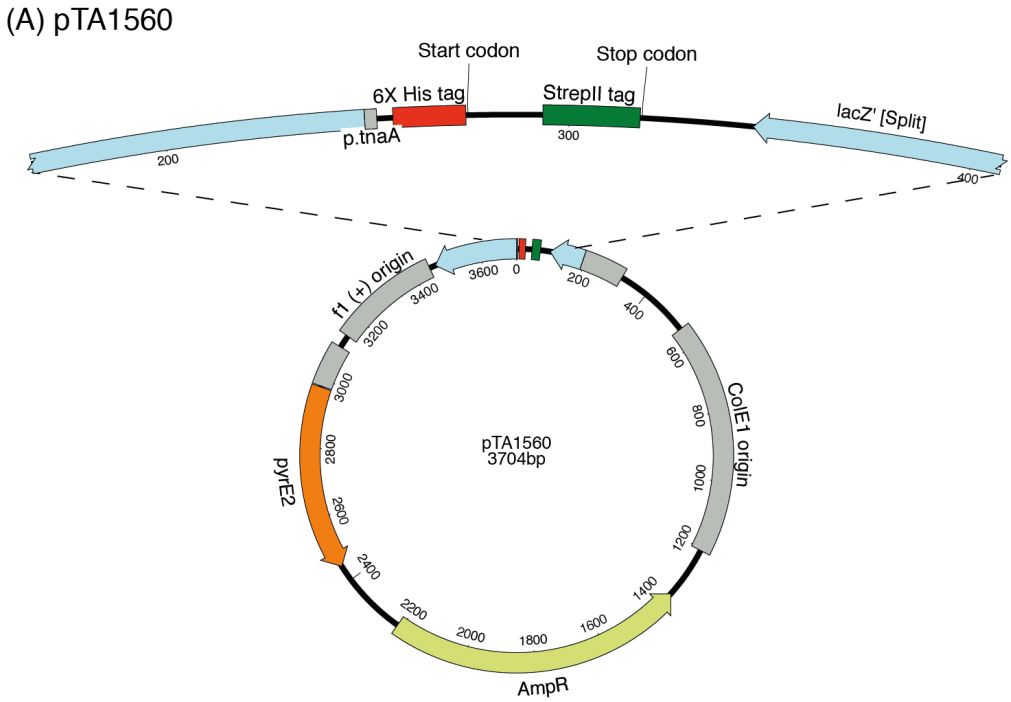
StrepII tags are stable up to 5 M NaCl and are therefore suitable for use in *H. volcanii*, they have previously been used to successfully purify proteins natively expressed in *H. volcanii* (Humbard *et al*. 2009). Proteins can be both His_6 tagged

and StrepII tagged, which helps to remove non-specific binding by histidine-rich proteins that may co-purify with a His_6 tag. Fractions purified with a metal affinity chromatography column can be subsequently applied to a StrepTactin column, which removes histidine rich contaminants as they cannot bind to the StrepTactin matrix.

An N-terminal His_6 tag and a C-terminal StrepII tag were used to tag MCM and GINS, in pTA1663 and pTA1791, respectively. The His_6 and StrepII tags were used together to allow a variety of purification options, for example purifying using the His_6 tag, the StrepII tag, or both in sequential purifications.

pTA1560 (N-terminal His_6 Tag, C-terminal StrepII tag)

The cloning vector pTA131 was digested with *PciI* and blunt-end filled with Klenow and religated to remove the *PciI* site to create pTA1553. The plasmid pTA1392 (constructed by Rebecca Gamble-Milner in 2014, Gamble-Milner 2016) was digested with *BamHI* and *FspI* and ligated into pTA1553 at the *BamHI* and *EcoRI* sites to create pTA1560 with 5'His_6 tag, 3'StrepII tag for the expression of double tagged gene(s) under the control of their native promoter on the chromosome (Figure 5.14).



(B) *BspHI/ NdeI* restriction digest

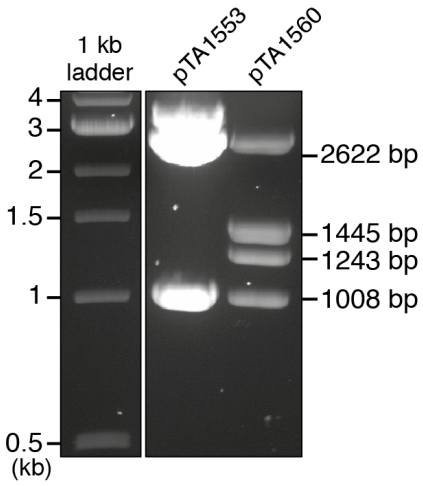


Figure 5.14 **pTA1560**. **(A)** 5'His_6 tag and 3'StrepII tag with start and stop codons for double tagged protein expression. **(B)***BspHI/ NdeI* restriction digest of pTA1560 giving bands of 1445 bp, 1243 bp and 1008 bp and pTA1553 giving bands of 2622 bp and 1008 bp, as expected.

Development of Tagged Strains

Tagged MCM

pTA1404 *mcm* (HVO_0220)

The *mcm* genomic clone, pTA1404, was constructed prior to this study in 2013 by Thorsten Allers (unpublished data) (Figure 5.15).

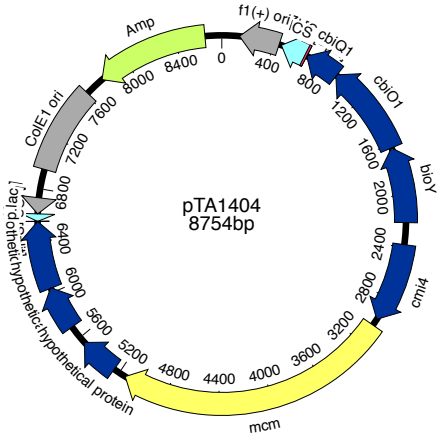


Figure 5.15 **pTA1404.** Genomic clone of *mcm*. Constructed by Thorsten Allers in 2013 (unpublished data).

pTA1663 (*His_6 tag-mcm-StrepII tag* gene replacement construct)

The genomic clone pTA1404 was used as a template to PCR amplify the *mcm* gene using the primers mcmtagGNcoF and mcmtagGNheR introducing *NcoI* and *NheI* sites respectively. The product was ligated into *PciI* and *NheI* sites in pTA1560 to create the plasmid pTA1627. The upstream region of *mcm* was PCR amplified from pTA1404 using the primers mcmtagUSKpNF and mcmtagUSNdeR introducing *KpnI* and *NdeI* sites respectively. The PCR product was ligated into pTA1627 at the *KpnI* and *NdeI* sites to create pTA1655. The downstream region of *mcm* was PCR amplified using the primers mcmtagDSBamF and mcmtagXbaR, introducing *BamHI* and *XbaI* sites respectively, from the genomic clone pTA1404. The product was ligated into the plasmid pTA1655 at the *BamHI* and *XbaI* sites to create the *His_6 tag-mcm-StrepII tag* gene replacement construct, pTA1663 (Figure 5.16).

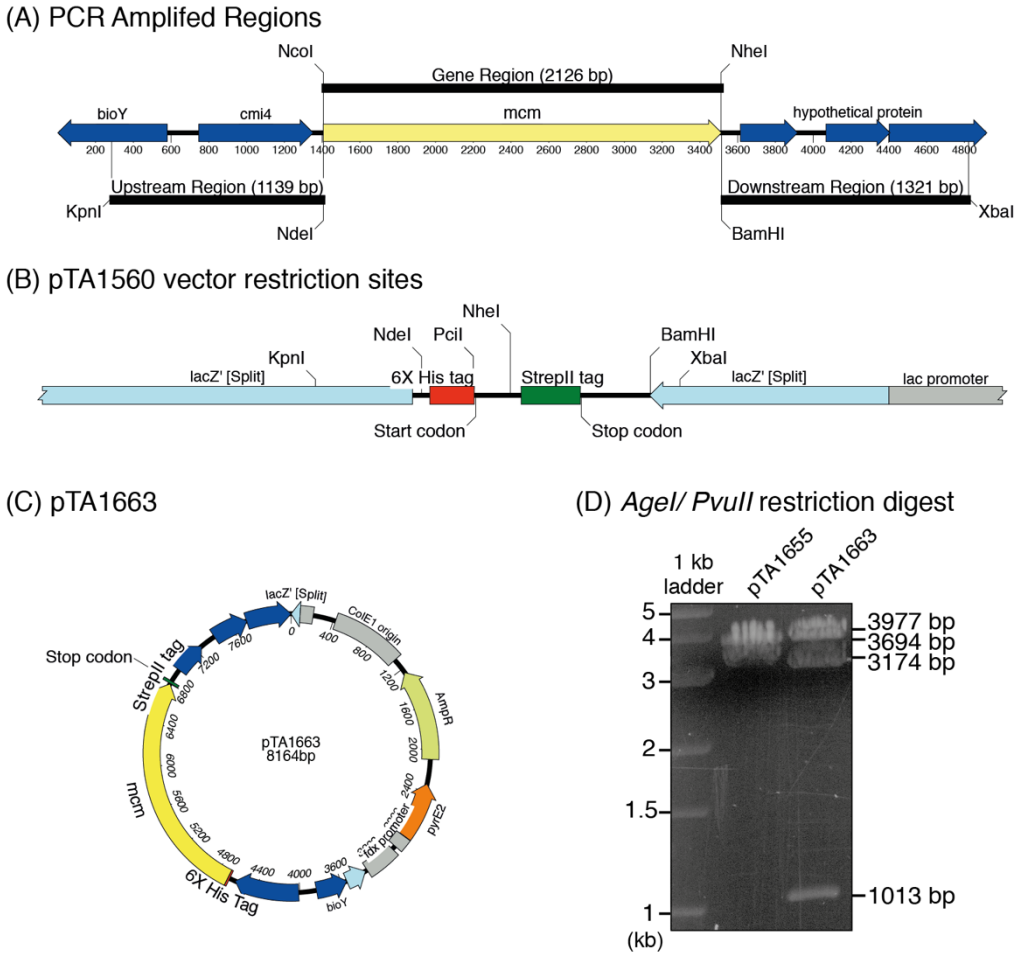


Figure 5.16 **pTA1663**. **(A)** *mcm* region showing upstream (1139 bp), downstream (1321 bp) and gene (2126 bp)fragments amplified by PCR with novel restriction sites. **(B)** pTA1560 vector showing restriction sites where fragments will be ligated into. **(C)** pTA1663 with 5'His_6 tagged, 3' StrepII tagged *mcm* gene replacement construct. **(D)** *AgeI/ PvuII* restriction digest of pTA1663 (3977 bp, 3174 bp and 1013 bp bands) and pTA1655 vector (3694 bp and 3174 bp bands).

pTA1663 was used to transform H2047, introducing the *His_6 tag-mcm-StrepII tag* construct into the genome to create the pop-in strain H2594. However, when the H2594 underwent pop-out to generate the strain H2660, it was found that only the 5'His_6 tag was present. In fact, it was found to be impossible to generate a strain with the 3'StrepII tagged *mcm* (Figure 5.17). This is not entirely surprising as there is evidence that the C-terminal domain of MCM in archaea interacts with the initiator protein Orc (Samson & Bell 2016), therefore any tag located at the C-terminus of MCM is likely to interfere with this interaction and disrupt DNA replication through steric hindrance. In this case the 3'StrepII tag

is located on the winged-helix domain of MCM (Figure 5.2). This region is disordered which may also cause this domain to not be viable for a protein tag.

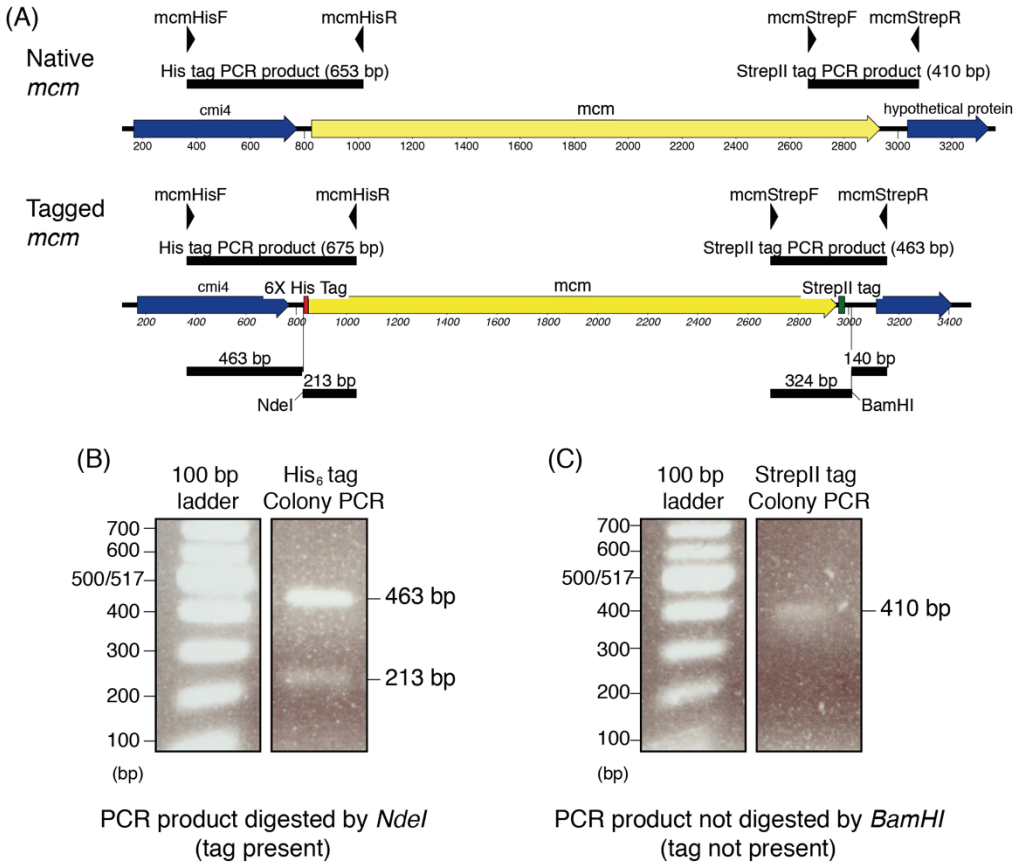


Figure 5.17 **Colony PCR of tagged MCM strain H2660. (A)** Location of PCR products produced with primers mcmHisF and mcmHisR (His tag PCR product) and, mcmStrepF and mcmStrepR (StrepII tag PCR product). In a strain with native *mcm* His tag PCR product is 653 bp and is unable to be digested by *NdeI*. Where *mcm* has been tagged the His tag PCR product is 675 bp and can be digested with *NdeI* to give two bands of 463 bp and 213 bp. In a strain with native *mcm* StrepII tag PCR product is 410 bp and is unable to be digested by *BamHI*. Where *mcm* has been tagged the StrepII tag PCR product is 463 bp and can be digested with *BamHI* to give two bands of 324 bp and 140 bp. **(B)** Colony PCR product can be digested with *NdeI* with bands at 463 bp and 213 bp, showing presence of N-terminal His_6 tag. **(C)** Colony PCR product cannot be digested with *BamHI* leaving a single wild-type band at 410 bp indicating an absence of C-terminal StrepII tag.

Tagged GINS

pTA1716 *ginS* (HVO_2698)

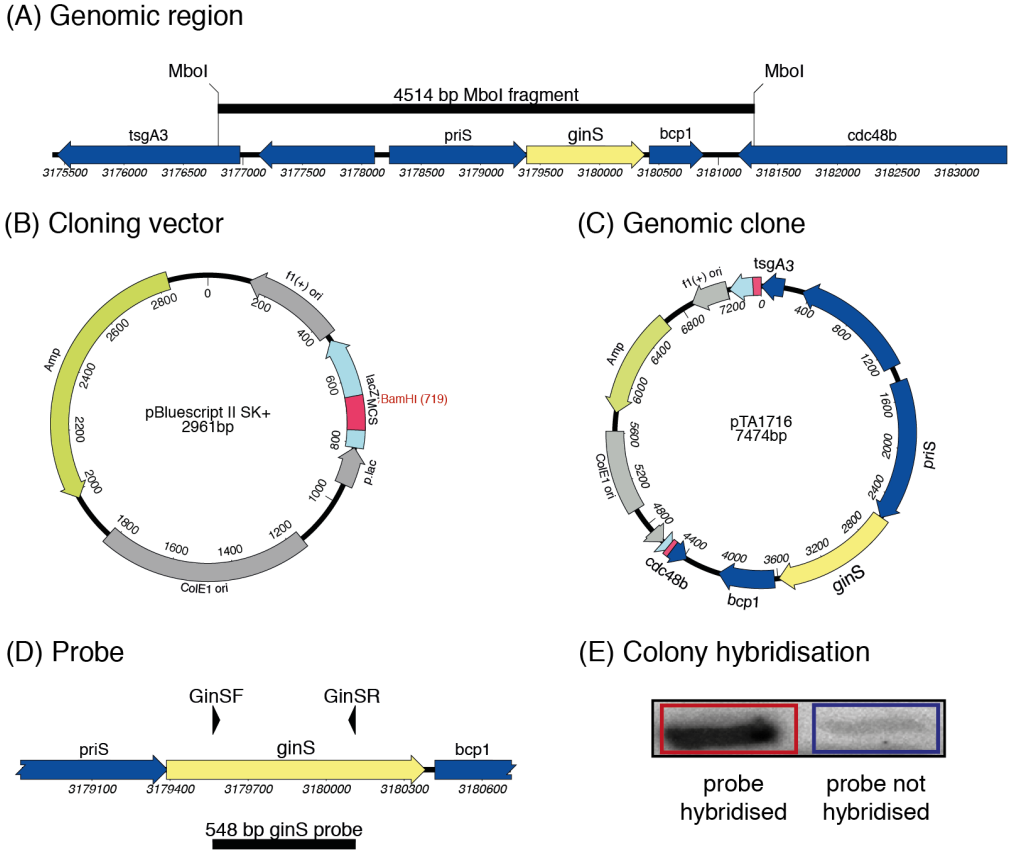


Figure 5.18 **pTA1716. (A)** The location of *ginS* on the *H. volcanii* genome, genomic DNA was digested with *MboI*. **(B)** pBluescript II SK+ plasmid where the *MboI* fragment was inserted into *BamHI* site. **(C)** Genomic clone pTA1716 constructed by inserting a *MboI* fragment containing *ginS* into pBluescript II SK+. **(D)** 548 bp *ginS* probe was constructed by PCR using primers GinSF and GinSR, was radiolabelled and used to probe colonies. **(E)** Colony hybridisation showing a patch which the probe hybridised to (containing *ginS* genomic clone) and a patch to which the probe did not hybridise to (does not contain *ginS* genomic clone).

Wild-type *H. volcanii* (H26) DNA was digested with *MboI* to produce a fragment of 4.5 kb containing *ginS* (HVO_2698), *priS* (HVO_2697), and upstream and downstream flanking regions (Figure 5.18). The fragments between 4-5 kb were gel extracted and cloned into pBluescript II SK+ at the *BamHI* site in the MCS. The transformants were patched onto LB+Amp agar and colony lifted onto a membrane. This membrane was then probed with a radiolabelled 548 bp PCR product using the primers GinSF and GinSR amplified

from genomic DNA. Patches that hybridised were restreaked and verified by sequencing.

pTA1791 (*His$_6$ tag-ginS-StrepII tag* gene replacement construct)

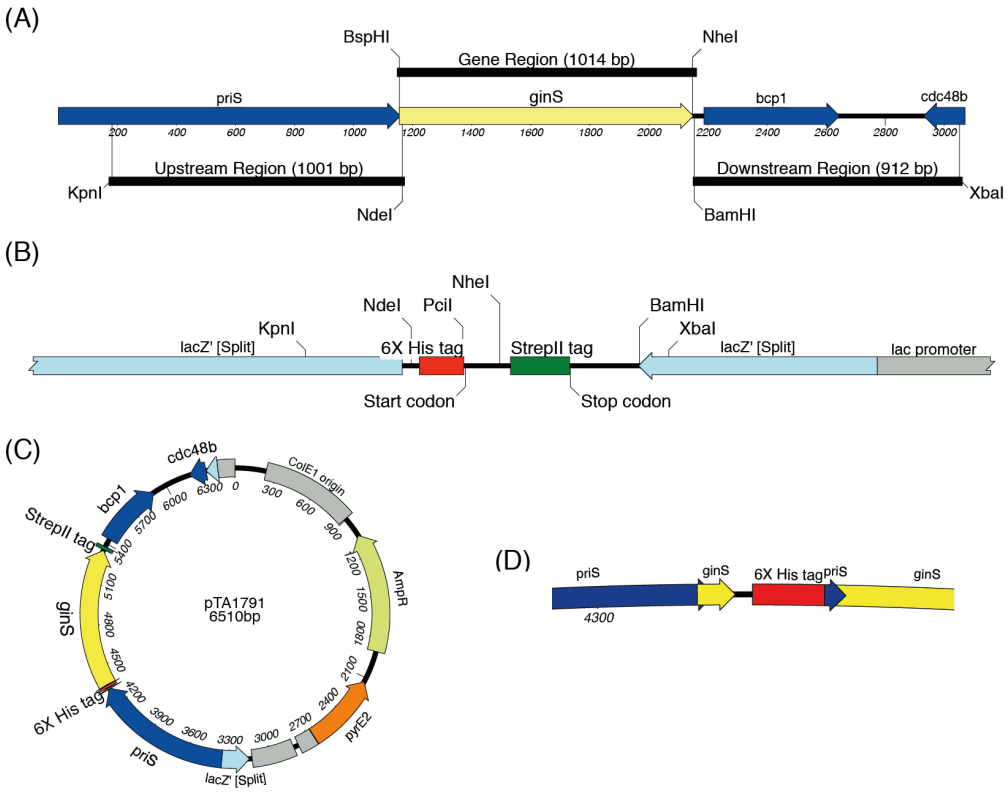


Figure 5.19 **pTA1791**. **(A)** *ginS* region showing upstream, downstream and gene fragments amplified by PCR with novel restriction sites. **(B)** pTA1560 vector showing restriction sites where fragments will be ligated into. **(C)** pTA1791 with 5'His$_6$ tagged, 3' StrepII tagged *ginS* gene replacement construct. **(D)** His$_6$ tag with overlap between *priS* and *ginS* leaving both genes functional.

The genomic clone pTA1716 was used as a template to PCR amplify the upstream region of *ginS* gene using the primers gintagUSKpnF and gintagUSNdeR introducing *KpnI* and *NdeI* sites, respectively. The product was ligated into pTA1560 at the *KpnI* and *NdeI* sites to create pTA1776. The primers gintagGBspF and gintagGNheR, introducing *BspI* and *NheI* sites respectively were used to PCR amplify the *ginS* gene from pTA1716. The PCR product was ligated into pTA1776 at the *PciI* and *NheI* sites to create pTA1790. The downstream region of *ginS* was PCR amplified from pTA1716 using the primers gintagDSBamF and gintagDSXbaR introducing novel *BamHI* and *XbaI* sites,

respectively. The product was ligated into pTA1790 at the *BamHI* and *XbaI* sites to create the plasmid pTA1791 *His_6 tag-ginS-StrepII tag* gene replacement construct (Figure 5.19). Care was taken when designing the gene replacement construct, to not interfere with the overlapping *priS* gene. The *priS* gene is often found in close proximity to *ginS* in archaea and encodes the PriS catalytic subunit of the archaeal primase PriSL. GINS B-domains and PriS C-terminal domains also share structural similarities, suggesting they are evolutionarily related to each other (Swiatek & MacNeill 2010).

H2047 was transformed with pTA1791, introducing the *His_6 tag-ginS-StrepII tag* construct into the genome (Figure 5.20). This successfully generated the protein expression strain H2964 with a 5'His_6 and 3'StrepII tagged *ginS*. Strains with single 5'His_6 (H2963) or 3'StrepII tags (H2962) were also generated. It was possible to tag *ginS* in wild-type and origin deleted backgrounds to create the strains H3222 and H3225, respectively. This allowed the comparison of interactions between GINS and other proteins in origin-dependent replication and recombination-dependent replication.

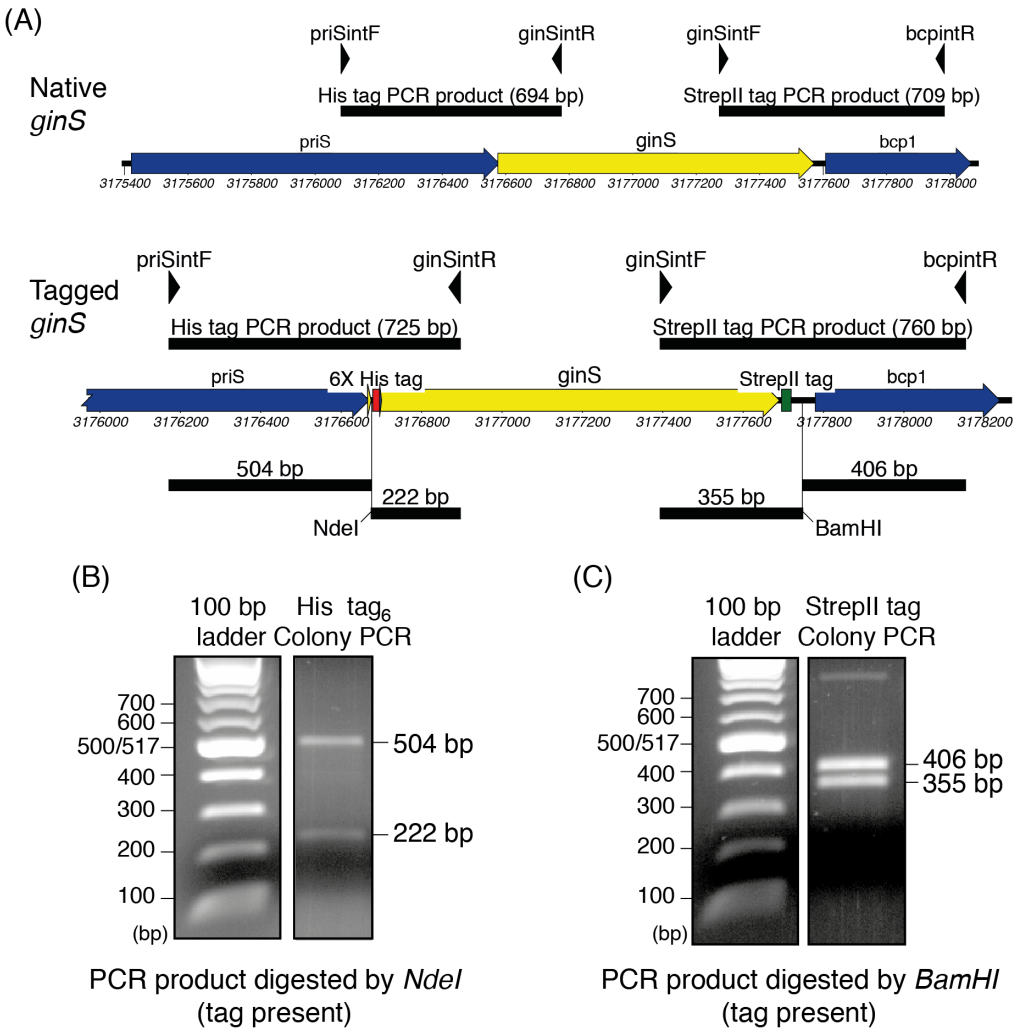


Figure 5.20 **Colony PCR of tagged *ginS* strain H2964. (A)** Location of PCR products produced with primers priSintF and ginSintR (His tag PCR product) and, ginSintF and bcpintR (StrepII tag PCR product). In a strain with native *ginS* His tag PCR product is 694 bp and is unable to be digested by *NdeI*. Where *ginS* has been tagged the His tag PCR product is 725 bp and can be digested with *NdeI* to give two bands of 504 bp and 222 bp. In a strain with native *ginS* StrepII tag PCR product is 709 bp and is unable to be digested by *BamHI*. Where *ginS* has been tagged the StrepII tag PCR product is 760 bp and can be digested with *BamHI* to give two bands of 406 bp and 355 bp. **(B)** Colony PCR product can be digested with *NdeI* with bands at 504 bp and 222 bp, showing presence of N-terminal His_6 tag. **(C)** Colony PCR product can be digested with *BamHI* giving two bands at 406 bp and 355 bp indicating the presence of a C-terminal StrepII tag.

Double Tagged Strain

The C-terminal StrepII-tagged *ginS* strain H2962 (*ginS+-StrepII tag*) was transformed with pTA1663 to introduce the tagged *mcm* construct, with the aim of creating a strain that was both *His_6 tag-mcm+* and *ginS+-StrepII tag*. The plasmid pTA1663 was used despite having a double tagged *mcm* construct (*His_6*

tag-mcm+-StrepII tag), since it had been shown that it is impossible to produce a viable C-terminal tagged *mcm* strain. The strain H3628 with tags on both *mcm* and *ginS* (*His_6 tag-mcm+ ginS+-StrepII tag*) was generated. The benefit of having two individually tagged proteins in the same strain is that after purification, Western blot can be used to look for interactions by probing separately for the His_6 and StrepII tagged proteins. As MCM and GINS are known to interact in other organisms, they can be used to test the method before testing other proteins such as the interaction of MCM or GINS with RecJ/Gan. In this study, only purification using the His_6-tag was carried out, although it would also be possible to carry out the purification and detection of proteins using StrepII tag purification.

Protein Expression and Purification

A 75-100 kDa protein was purified from the cell lysate of strain H3628 (*His_6 tag-mcm+ ginS+-StrepII tag*) using metal affinity chromatography. Cell lysate from the parental strain H2047 was also subjected to metal affinity chromatography as a control for possible contaminants. Due to time constraints, no other tagged proteins were purified. Gravity columns packed with Ni^{2+} charged Sepharose beads were used in small scale purifications, and protein was eluted in 100-500 mM imidazole (Figure 5.21).

The *Haloferax volcanii* MCM protein is ~78 kDa and GINS protein is ~34 kDa. There are 3 bands between 75-100 kDa in the E100 elution lane (100 mM imidazole) that might be MCM. There does not seem to be a band that corresponds in size to GINS, but it is possible that GINS is present in the 40-50 kDa band seen. Most of the bands are likely to be histidine-rich contaminants that have previously been described by Stroud *et al*. (2012). To verify purification of *His_6 tag-mcm+* and determine if *ginS+-StrepII tag* had co-purified, it was necessary to do a Western blot. Analysis of the bands using mass spectrometry would be more accurate than a Western blot, however this was not possible due to time constraints. It was decided to look for interacting partners by first using the His_6-tag rather than the StrepII tag as it is harder to see possible interacting partners with the StrepII tag protein purification method (Thorsten

Allers, personal communication). In the future, it would be possible to undertake a co-precipitation study using a His_6-tagged GINS as well.

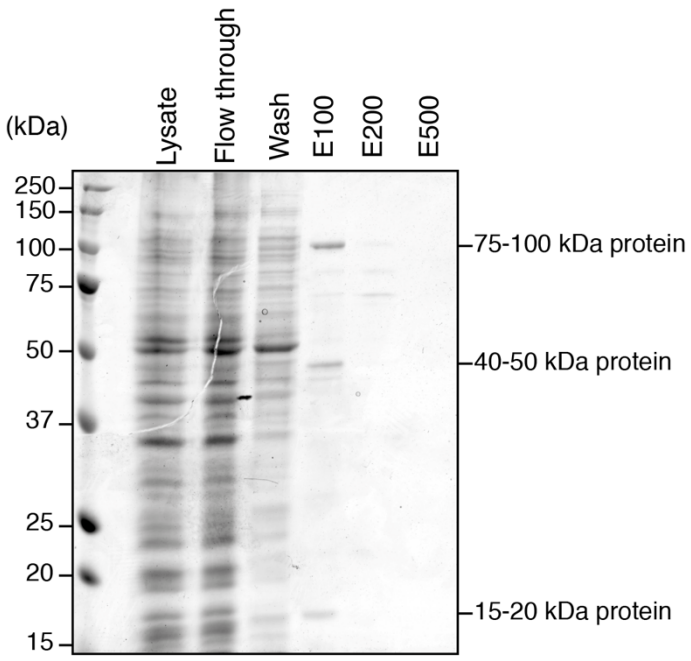


Figure 5.21 **SDS-PAGE gel of 75-100 kDa protein using metal affinity chromatography.** A 70-100 kDa protein (presumed to be MCM) was expressed and purified from H3628 (*His_6 tag-mcm+ ginS+-StrepII tag*) by metal affinity chromatography using a step elution with buffer containing imidazole. E refers to imidazole concentration in mM. MCM is approximately 78 kDa. Bands of 40-50 kDa and 15-20 kDa proteins were also seen indicating possible interacting partners.

Western Blotting

Western blotting was used to probe for purified *His_6 tag-mcm+* and separately for *ginS+-StrepII tag*, to investigate co-purification (and therefore interactions) of the two proteins (For details see *Chapter 2: Materials and Methods, Section 2.2.8 Protein Expression and Purification*). If *ginS+-StrepII tag* can be visualised on a Western blot using purified *His_6 tag-mcm+* samples, it is likely that these proteins interact *in vivo* in *Haloferax volcanii*.

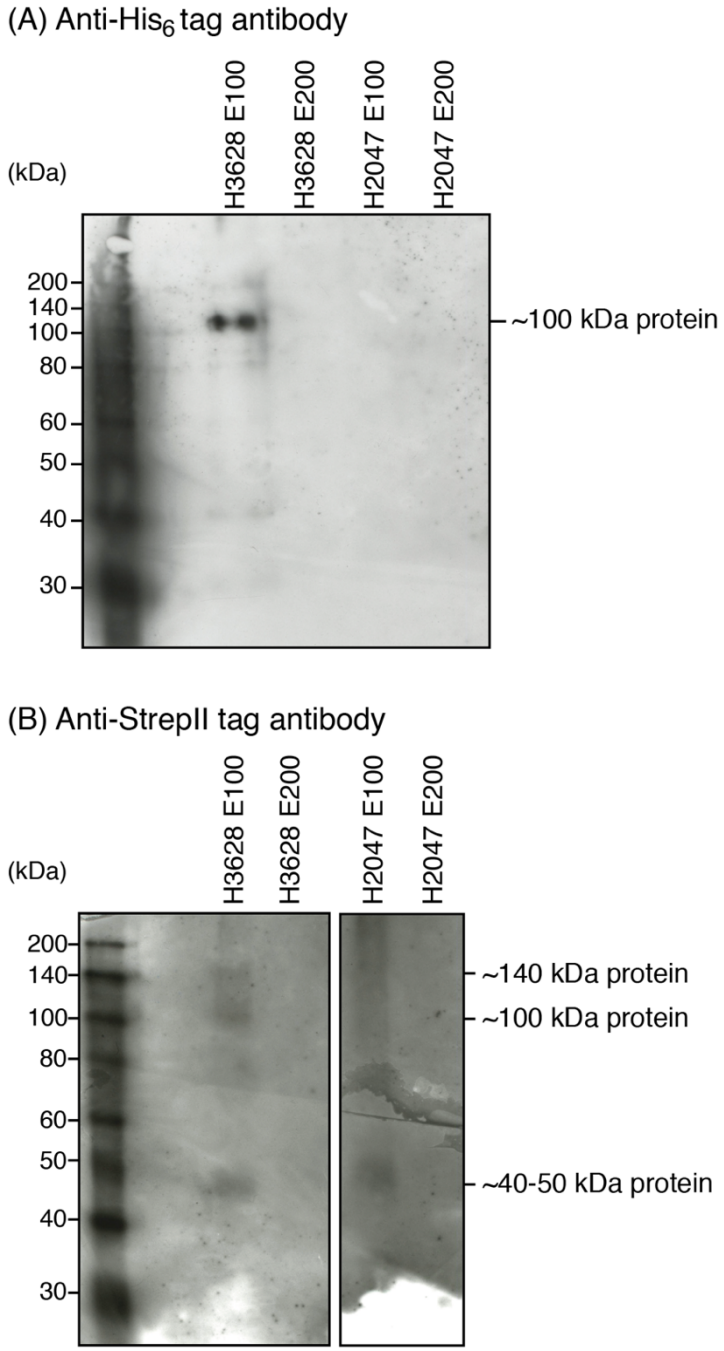


Figure 5.22 **Western blot of His_6 tag metal affinity chromatography purification. (A)** A His_6 tagged protein (~100 kDa) was detected by use of specific anti-His_6 antibody in the strain H3628 (*His_6 tag-mcm+ ginS+-StrepII tag*), but not in the wild-type strain H2047. This is presumed to be His_6-tagged MCM. **(B)** A ~40-50 kDa protein was detected in both H3628 and H2047 and presumed to be a contaminant. Two other proteins at ~100 kDa and ~140 kDa were also detected but thought to be ladder contamination, therefore StrepII tagged GINS was not detected with a specific anti-StrepII tag antibody.

As can be seen in Figure 5.22, a band was seen in the experimental samples from H3628 which is not present in the control strain H2047, when an SDS-PAGE gel was probed for the His_6 tag using 6x-His Tag monoclonal primary antibody

HIS.H8 (Thermo Fisher) and IgG peroxidase secondary antibody (Sigma) for detection with horseradish peroxidase (HRP). This band is likely to be MCM but migrates at ~100 kDa instead of the expected size of ~78 kDa; a similar sized band was seen on the SDS-PAGE gel (Figure 5.21). Bands can also be seen when probing for the StrepII tag using the primary antibody Anti-Strep-tag II antibody (Abcam) and IgG peroxidase secondary antibody (Sigma) for detection with HRP. The most prominent band (~40-50 kDa) is also seen in the control strain H2047, meaning that it is unlikely to be tagged GINS and most likely a contaminant. There are two other faint bands which can be seen in Figure 5.22B at ~100 kDa and ~140 kDa, which are not seen in the wild-type control strain, however as these bands correspond with the sizes of bands in the ladder these are likely to be caused by bleeding of the ladder into the lane. It is possible that GINS co-purified with MCM, however due to the Western blot protocol used it is hard to determine this.

Due to the use of the Biotinylated Protein Ladder Detection Pack (Cell Signalling Technologies) for the Western blotting, milk was not able to be used as a blocking agent as it would cross react with the antibodies used. BSA was used instead, however there are still high background levels seen which makes it difficult to distinguish bands. This could be due to inefficient washing or poor transfer of the proteins on to the membrane from the SDS-PAGE gel. It is possible that there was also a problem with the detection of the antibodies, therefore the protocol must be refined and the experiment repeated before definitive conclusions can be drawn. For example, it was suggested that the timing of the washes should be shorter to improve signal.

After the conclusion of this study is was found that a more sensitive secondary antibody (HRP-Conjugated 6*His, His-tag Monoclonal Antibody, Proteintech) should be used to improve the detection of halophilic proteins purified from *Haloferax volcanii*, which could improve the results gained from the Western blot method (Julie Maupin-Furlow, personal communication). It is also possible that the use of an alternative tag such as an octa-his tag rather than a hexa-his tag could provide stronger binding to the column, which may increase protein yield for easier visualisation.

Chapter 6: Discussion

6.1 Genome architecture

Size Threshold for Recombination-Dependent Replication

It has been suggested that *H. volcanii* uses recombination-dependent replication (RDR) to replicate its main chromosome in the absence of origins (Hawkins *et al*. 2013). In origin-dependent replication, each origin produces two bi-directional replication forks, whereas RDR produces one replication fork at a time (Figure 6.1).

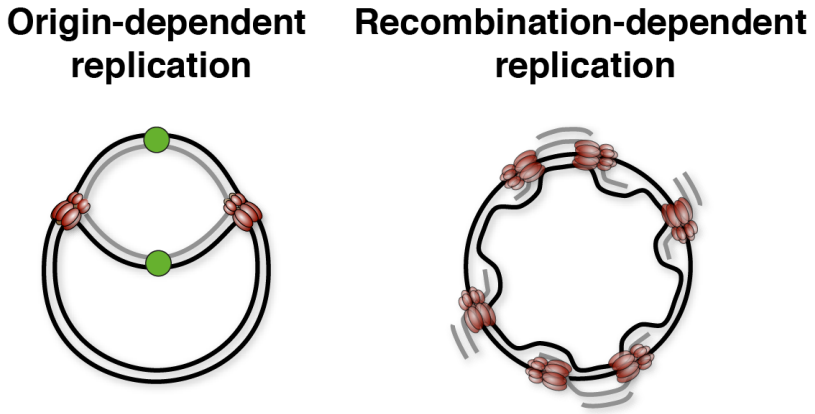


Figure 6.1 **Replication forks in origin-dependent replication and recombination-dependent replication.** In origin-dependent replication there are only two forks produced from a single origin firing event, whereas in recombination-dependent replication DNA replication initiation can be anywhere on the genome, meaning there are many replication forks.

Origin-dependent replication initiation is restricted to defined sites on the chromosome (origins), and these are programmed to fire on average once per cell cycle (Table 6.1). In *H. volcanii* with integrated pHV4, if all origins on the chromosome were to fire simultaneously there would be 8 replication forks. Conversely, RDR can theoretically initiate from anywhere on the genome, meaning that there could be an unlimited number of replication forks at any one time. But unlike origin firing, recombination is a stochastic process that is affected by the size of the DNA molecule. The longer the DNA molecule, the more the potential for homologous recombination, which will result in more DNA replication initiation.

Table 6.1 **Origin- vs recombination-dependent replication.**

	Origin	Recombination (D-loop)
Location	Site-specific	Random
Frequency	Once per cell cycle	Depends on DNA size

Having more replication forks would increase the amount of DNA replication and consequently have an effect on the growth rate of the cell. But if the size of the DNA molecule is small, the probability of recombination would be reduced proportionately and this may result in fewer than one replication initiation per molecule per cell cycle. In this case, the DNA molecule would be lost, most probably resulting in cell death, because it is below the threshold required for efficient RDR (Figure 6.2).

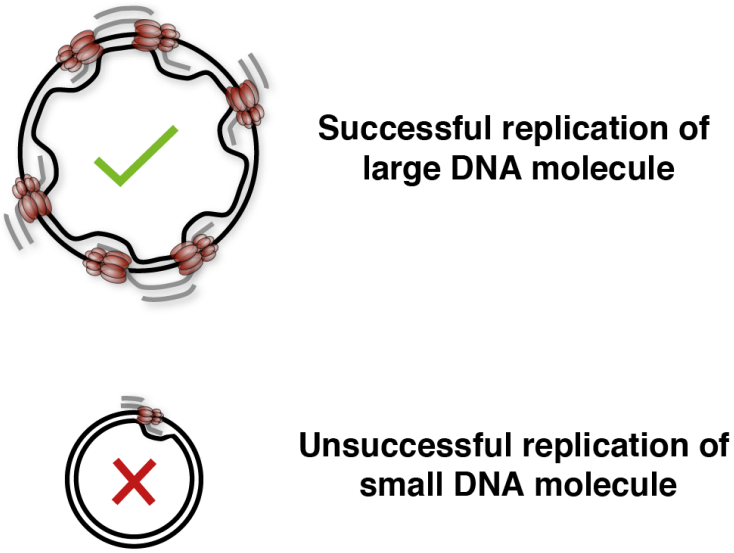


Figure 6.2 **Model for a size threshold for recombination-dependent replication.** Large DNA molecules can be efficiently replicated using recombination-dependent replication, whereas small DNA molecules cannot.

However, RDR would be rate-limited by availability factors such as the DNA replication machinery, nucleotides and ATP, which would limit the impact on generation time. Ploidy of the DNA molecule would also have an effect on the efficiency of RDR. If there is low ploidy then RDR may be difficult to coordinate, especially if homologous DNA molecules are spatially separated. *H. volcanii* is known to have high levels of ploidy with up to 20 copies of the genome, with double the number of pHV1 molecules compared to the rest of the genome (Breuert *et al*. 2006). This means that there are many copies of

homologous DNA molecules at any one time and they are likely to be spatially close to one another. Ensuring this could be a reason for pHV1 forming concatemers. It is not known whether at cell division the copies of the *H. volcanii* genome are fully separated or are interlinked, or how it is ensured that both daughter cells receive a copy of the full genome.

The data gathered from the origin deletions of integrated and episomal pHV3 and pHV4 suggests that there is a size limit for efficient recombination-dependent replication (RDR) that is between 438 kb (pHV3) and 636 kb (pHV4).

There could be other reasons why *ori-pHV3* is essential on an episomal pHV3 and *ori-pHV4* is not essential on episomal pHV4. It is currently unknown how RDR is initiated. Initiation at R-loops (as well as D-loops) is highly likely, however it has not yet been determined if this is the case. Levels of transcription on the mega-plasmids may be important, as increased transcription would lead to an increased possibility of R-loop formation, and therefore more RDR. There is a low level of transcription on pHV3 (Trieselmann & Charlebois 1992), which could mean that there may not be enough R-loops to efficiently initiate replication of pHV3 using RDR. To test this, the level of transcription on pHV3 would have to be artificially increased and then the deletion of the origin attempted again.

The origin and associated *orc* cannot be deleted from episomal pHV3, but can be deleted together from a strain where pHV3 has been integrated onto the main chromosome. This indicates that episomal pHV3 is below the size threshold required for RDR and cannot replicate without its origin or associated *orc* gene. However, this could be due to the site of the pHV3 integration in the main chromosome. The mega-plasmid pHV3 was integrated close to the chromosomal origin *oriC2* and a highly transcribed rRNA locus. Therefore, *ori-pHV3* becomes non-essential either because the integrated pHV3 is replicated by *oriC2*, or because increased levels of transcription (at rRNA) lead to higher local levels of RDR initiation by R -loops. If the site of the pHV3 integration were changed so it was further from an origin and a highly transcribed region, then *ori-pHV3* might still be essential. Currently there are no known essential genes on pHV3

apart from the origin and associated *orc* (when episomal). A possible gene to note on pHV3 is a predicted *mcm2* gene (HVO_B0008) found close to *ori-pHV3*. The translated protein sequence shares limited homology with Mcm (gene HVO_0220) but the gene is only 375 bp in length and there are no conserved protein domains (Appendix *IV pHV3*), indicating it is probably non-functional and therefore non-essential. With the lack of essential genes on pHV3, it is not known why it is important to the cell that pHV3 is maintained. In fact, in a previously-generated laboratory strain WFD11, where pHV2 was cured using ethidium bromide treatment, pHV3 is often lost spontaneously and the cells are slow growing. This shows that there must be some genes on pHV3 that are required for efficient growth (Wendoloski *et al*. 2001).

pHV4 is the largest mega-plasmid (636 kb) in the *H. volcanii* genome and is able to be replicated successfully without an origin, whether episomal or integrated onto the chromosome, and therefore must be large enough for RDR. The majority of investigated genes on pHV4 can be deleted, including biosynthesis pathway genes, transcription factors and CRISPR genes (Appendix *V pHV4*). The one known exception to this is a region encoding a sRNA, but its function is still unknown. There are many hypothetical and uncharacterised genes on pHV4, therefore it is possible that there are some phage-like replication proteins encoded, especially in regions designated as non-native to *H. volcanii*. However, no alternative replication genes have been characterised as yet.

On the other hand, the origin and associated *orc* gene on pHV1 can be deleted indicating that it is able to be replicated using RDR. However, pHV1 is only 85 kb in size and therefore well below the suggested size threshold for efficient RDR. It was found that pHV1 exists as a 6x concatemer and therefore increases its size to above the suggested size threshold for RDR. Both the 6x concatemer and the monomer versions of pHV1 can be seen on a pulsed-field gel, therefore it is unknown which form is most common and it may well be subject to a constant cycle of recombination. A recombination-deficient strain could provide a 'snapshot' of the levels of concatemers vs monomers, and in a *ΔradA* strain a higher proportion of monomer species was seen on a PFG compared to wild-type, suggesting that in wild-type strains pHV1 is more commonly a concatemer.

However, this is unknown at this stage and may have been due to slight differences in sample preparation. Concatemeric plasmids are seen in *E. coli* and *Bacillus subtilis*, which are arranged in linear 'head to tail' configurations and are the result of propagation via rolling circle replication (RCR) (Viret *et al*. 1991). It is not known whether pHV1 concatemers are present in the cell as linear or as circular molecules, although the monomer is known to be circular (Hartman *et al*. 2010). It could be the case that instead of RDR using homologous recombination intermediates for initiation, pHV1 is using RCR to replicate and this would explain the high numbers of concatemers in the cell. Wang *et al*. (2018) suggested that viral or plasmid genomes can become integrated into the host chromosome (in archaea), and their Rep proteins are repurposed as transposases. This could suggest why no known *rep* gene is present on pHV1, but high numbers of transposases are. If pHV1 uses RCR to replicate, it poses the question why pHV1 has a conventional archaeal origin and associated *orc* gene. Furthermore, how does it ensure that only 6 x concatemers are produced and not other numbers? It is unknown whether the concatemers are arranged in one molecule or 6 interlinked molecules. If pHV1 is a concatemer of 6 interlinked molecules, then it would suggest that it does not use RCR for DNA replication. There are currently no known RCR related genes, however there are many uncharacterised and hypothetical genes present on pHV1, so it is possible that it may encode the apparatus for RCR. It is also not known whether there are any phage-like genes present on pHV1 and if phage-like replication mechanisms are used, then this provides another explanation for how pHV1 could replicate to generate a 6 x concatemer.

H. volcanii has a very efficient homologous recombination mechanism and would in theory resolve concatemeric structures produced by RCR quickly into monomers, rather than maintaining them as concatemers. It has been noted in transfection of eukaryotic cells, if a plasmid is transfected as a concatemer, the level of transcription is increased compared to a comparable number of monomeric plasmids (Leahy *et al*. 1997). It is not known if this is the case for pHV1, but it may confer a benefit for pHV1 to replicate as a concatemer when its origin is present and has no need for RDR. To help determine whether R-

loops (generated by transcription) can initiate RDR, then the levels of transcription of pHV1 should be investigated.

Deletion of *ori-pHV1* and the Stability of the Mega-plasmid pHV1

Prior to this study, the *ori-pHV1* and *orc10* found on the mega-plasmid pHV1 had been deleted in a wild-type background using a deletion construct with a tryptophan-selectable marker (Norais *et al*. 2007). This study has found that they can also be deleted in an origin-deleted background using a tryptophan-selectable marker, both individually and together. It was also found that once the origin or *orc* had been deleted, the tryptophan marker could be removed allowing it to be used again in possible downstream deletions. These single deletions of *ori-pHV1* and *orc10* lead to no differences in generation time when the strains were grown in rich media, suggesting that pHV1 does not require its origin or associated *orc* for replication. This could suggest that pHV1 is perfectly capable of replicating without a traditional origin/Orc system, and retains an origin and associated Orc as a 'back-up' system for when its normal mode of replication is not functional or active. This could be the case if pHV1 is acting as a phage-like element within *Haloferax volcanii* and replicating using a phage-like origin-independent system rather than a host-encoded origin-dependent system. The mega-plasmid pHV1 also has a second *orc* gene, *orc8*. Although this gene was not investigated as part of this study, it could be the case that there is a cryptic origin associated with this *orc* gene. If this were the case, then this cryptic origin could take over replication of pHV1 when its main origin and *orc, ori-pHV1* and *orc10*, are deleted. It could also be the case that *orc8* plays a role at *ori-pHV1*, or *orc10* plays a role at the possible cryptic origin at *orc8*.

Many plasmids that naturally occur in bacteria contain genes that can be beneficial for the host, for example anti-microbial resistance genes. There are also so-called 'cryptic' plasmids that seem to confer no benefit in normal growth situations and persist without selective pressure or burden on the host. However, it is possible that they may confer benefits in certain growth conditions (Hall *et al*. 2017; Werbowy *et al*. 2017). This could be the case for pHV1 and although it was found that strains missing the pHV1 mega-plasmid showed no difference

in generation time in rich media, leading to the conclusion that pHV1 and all its genes are non-essential to *H. volcanii*. The growth of the strains missing pHV1 was not assessed in minimal media, so it could be possible that pHV1 confers a growth benefit in certain conditions. If pHV1 originated from a phage, it could be the case that the presence of pHV1 in the *H. volcanii* confers protection against invasion from other phage or plasmids.

H. volcanii is not the only archaeon to contain a mega-plasmid like pHV1, for example the thermophile *Thermus thermophilus* also contains a mega-plasmid that carries genes involved in sugar metabolism, carotenoid synthesis and DNA repair (Brüggemann & Chen 2006). The *T. thermophilus* mega-plasmid pTT27 can be lost from the cell, however this confers a growth defect compared to a wild-type strain, showing that the genes present on the mega-plasmid are vital for healthy growth (Ohtani *et al*. 2016). The bacterium *Vibrio cholerae* naturally contains two chromosomes, Chr1 and Chr2, which is unusual for bacteria. Chr2 is thought to have originated from a plasmid due to its plasmid-like origin, however it has since evolved into a chromosome with essential genes (and many with unknown function). There is also evidence to suggest that the genes on Chr1 and Chr2 are expressed under different growth conditions (Xie *et al*. 2017). Similarly, pHV1 may be providing benefits to the cell in certain growth conditions, however due to the large number of hypothetical genes a detailed analysis cannot be undertaken at this stage.

In *H. volcanii* other than the two *orc* genes, most of the other genes on pHV1 are uncharacterised, however there is a gene known to be involved in glycine synthesis, a predicted homologue to a virus protein, several predicted oxidoreductases and a predicted ParA1 protein. The predicted ParA1 has domains that have homology to plasmid RepA partitioning proteins as well as the protein BcsQ, which is a cellulose biosynthesis protein involved in chromosome partitioning (Le Quéré & Ghigo 2009) (Appendix *III pHV1*). In bacteria, the ParAB system is important for the coordination of chromosome segregation with cell division and DNA replication (Donczew *et al*. 2016). This indicates that the *parA1* gene on pHV1 may be important for the stability and

maintenance of the pHV1 mega-plasmid. It would be interesting to delete this gene from pHV1 and repeat the stability assay.

In bacteria is has been suggested that beneficial genes are likely to be 'captured' by the chromosome, which leads to the loss of the plasmid that may be energetically costly for the cell to maintain (Hall *et al*. 2017). *H. volcanii* is known to have a very efficient homologous recombination pathway and can use it for DNA replication – recombination-dependent replication (Hawkins *et al*. 2013). Therefore, it is likely to behave in a similar manner to bacteria and 'capture' any beneficial genes from pHV1, leaving pHV1 redundant. Norais *et al*. (2007) found that *ori-pHV1* is the dominant origin isolated in an ARS assay. This leads to the suggestion that *ori-pHV1* may be acting as a selfish element within *H. volcanii*, ensuring the replication of the pHV1 mega-plasmid. pHV1 contains many IS-elements and transposases, is non-native compared to the *H. volcanii* chromosome (by synonymous codon usage) and has a relative copy number of 2 compared to the rest of the genome (Hartman *et al*. 2010). The mega-plasmid found in *T. thermophilus* also has a high proportion of mobile genetic elements (Brüggemann & Chen 2006). It has been suggested that Chr2 in *Vibrio cholerae* is acting as a selfish genetic element and can naturally fuse with the main *V. cholerae* chromosome Chr1, resulting in a strain with one chromosome (Xie *et al*. 2017). Therefore, it is likely that pHV1 is a foreign element that acts selfishly to maintain itself within *H. volcanii*.

This leads to the hypothesis that the presence of *ori-pHV1* and/or *orc10* on pHV1 helps to maintain the stability of this mega-plasmid in *H. volcanii* cells. To test this a *pyrE2* marker was placed in a putatively-neutral region of pHV1 in both wild-type and *ori-pHV1* deleted strains. These strains were then grown and selected for colonies that had lost pHV1. There were many false positive results, probably due to point mutations in the *pyrE2* gene (Makwan *et al*. 2007). Despite this, there are some indication that the strain with deletion of *ori-pHV1* lost the pHV1 mega-plasmid more easily than the wild-type strain. It was also found that when *ori-pHV1* was deleted, the mutation rate of the *pyrE2* marker gene was increased compared to the wild-type strain. This phenomenon is not observed on the main chromosome when all of the chromosomal origins are deleted

(Thorsten Allers, personal communication), indicating that *ori-pHV1* is not only required to maintain stability of the pHV1 mega-plasmid in the cell, but to also control mutation rates. If pHV1 is using a phage-like mechanism of RDR that is more mutagenic than host-encoded origin-dependent replication, this could explain the presence of an origin on pHV1. However, these results are far from conclusive and it may require the use of a marker gene that is less susceptible to point mutations to draw firm conclusions regarding the role of *ori-pHV1* in the maintenance and stability of pHV1.

6.2 DNA Replication Machinery

Expression of MCM

MCM was expressed under control of the tryptophan inducible promoters *p.tnaA* and *p.tnaAM3*, from alleles stably integrated on the chromosome. It was found that when *mcm* was controlled by the *p.tnaA* promoter, cells were able to grow on gradient plates where there should have been no tryptophan present. The promoter *p.tnaA* is known to be tightly regulated by tryptophan levels (Large *et al*. 2007), however, in this case background levels of transcription were enough to fulfil the requirement for MCM. This suggests that *Haloferax volcanii* cells require little transcription of the *mcm* gene for successful cell growth. In the eukaryote *Xenopus*, MCM is found to be relatively abundant at >40 complexes per origin, which is similar to *Drosophila* and yeast. However, in human primary T cells there are estimated to be between 2-5 MCM helicases per origin, showing that high levels of MCM are not always required. In *Xenopus*, MCM can be depleted to 90% less than original levels and replication is still initiated efficiently *in vitro*. This has led to the suggestion that most of the MCM in *Xenopus* is redundant and only one MCM helicase is required per origin for DNA replication initiation. This was also suggested for *Drosophila* (Crevel *et al*. 2007; Orr *et al*. 2010). It is possible that in *Haloferax volcanii*, the requirement for MCM in DNA replication is very low. This prompted the use of the lower activity promoter, *p.tnaAM3,* for regulating expression of *mcm* so that an effect of the reduction of MCM could be seen.

Using the *p.tnaAM3* promoter to regulate *mcm* expression allowed differences to be noted between wild-type and origin-deleted cells. However, they were noted in liquid growth and not by using tryptophan gradient plates. The gradient plates were unsuccessful in showing any meaningful differences between wild-type and origin-deleted strains, although there was some indication that there was a higher requirement for tryptophan in the strain where all chromosomal origins had been deleted, compared to wild-type. This implies that there may be a greater demand for MCM in origin-deleted cells but it does not show that recombination-dependent replication itself requires more MCM than origin-dependent replication – the origin-deleted cells grow faster than the wild-type and therefore carry out more DNA replication, this alone accounts for the greater demand for MCM. The use of tryptophan gradient plates for these strains were not ideal due to long growth times (9 days rather than 4 days) and the possibility that the tryptophan had diffused across the plate removing any gradient.

When wild-type and origin-deleted *p.tnaAM3::mcm* strains were grown in liquid cultures in an Epoch2 Microplate Spectrophotometer (BioTek), there were observable differences between wild-type and origin-deleted strains. The results showed that even when there was no tryptophan present in the media some growth was still possible. This could either be due to some background activity of the promoter allowing very small amounts of MCM to be produced, or as cells die there will be a small release of tryptophan into the surrounding media that may be taken up by living cells. Despite this it was noted that in a wild-type strain, changing the levels of tryptophan (0 mM - 2.5 mM),and therefore MCM expression had no effect on the growth rate (~2.5 hours generation time). However, in the strain where all chromosomal origins were deleted (including the origin found on the integrated mega-plasmid pHV4), when tryptophan levels were depleted the growth rate of the strain decreased dramatically (~6.25 hours generation time at 0 mM tryptophan). This indicates that MCM plays an important role in cell growth in cells where origins are deleted, and that such cells require more MCM than the wild-type.

In origin-deleted cells, there are high levels of recombination due to the switch from origin-dependent replication to recombination-dependent replication.

Haloferax volcanii can be seen as an analogy for cancer cells due to its lack of cell cycle and high ploidy (Delmas *et al*. 2009). In human cancer cells, it has been found that expression of MCM is abnormal with a >20-fold excess of MCM, causing increased levels of MCM loading and genomic instability. Depletion of MCM-7 in human primary T cells leads to increased DNA damage, which is thought to be due to lack of MCM binding to DNA and not caused by specific MCM proteins, chromosome rearrangements or changes in ploidy. This in turn induces DNA repair mechanisms including non-homologous end joining (NHEJ) and homologous recombination (HR) that were vital to allow the cell cycle to continue (Orr *et al*. 2010). Recombination-dependent replication in *Haloferax volcanii* is thought to be initiated by DNA breaks that are converted into D-loops. Therefore, it is not surprising that MCM depletion causes a reduction in growth rate in strains that use recombination-dependent replication. Bailis *et al*. (2008) suggest that a loss of MCM function causes replication fork breakdown and therefore MCM proteins are important for maintaining replication fork stability, both during normal DNA replication and after replication fork stalling (where they act in replication restart). In the same study, it was shown that in fission yeast, the MCM complex interacts directly with the Rad51 homologue Rhp51. This interaction was also observed with human MCM and Rad51 proteins. Li *et al*. (2010) noted that in the archaeon *Thermococcus kodakarensis*, MCM was seen to interact with known DNA repair proteins indicating this is not isolated to the eukaryotes. This suggests that the reduction of growth rate seen upon MCM depletion in an origin-deleted *H. volcanii* strain compared to wild-type could be due to two reasons. Firstly, recombination-dependent replication leads to a faster growth rate and requires more resources from the cell, due to being able to initiate DNA replication from anywhere on the chromosome and therefore increase the number of concurrent replication forks. This is compared to origin-dependent replication, where DNA replication initiation is limited to defined regions of DNA (origins), therefore has a limited number of concurrent replication forks, and will require fewer resources. This greater demand for resources in recombination-dependent replication will include MCM, so as the levels of MCM are reduced the growth rate will be slowed as there is not enough MCM to satisfy the needs of recombination-dependent replication. Alternatively, the reduction in growth rate may be caused

by the lack of MCM leading to replication fork instability. As the cells undergo high levels of recombination, which itself is linked to genome instability, it may be that depleting MCM causes so much instability during DNA replication that it leads to a growth defect. Further investigation is required to determine the cause of the growth defect in origin-deleted cells that are depleted for MCM. The question also remains whether MCM is used in recombination-dependent replication as the replicative helicase. It seems likely that this is the case due to its close links to homologous recombination and replication fork integrity in normal replication, and when cells are stressed (Ibarra *et al*. 2008), but this must also be investigated further.

Interaction of MCM and GINS

Haloferax volcanii MCM protein was tagged with an N-terminal His_6 tag, however it could not be tagged on the C-terminus with a StrepII tag. Being unable to tag MCM on C-terminus is not unexpected as the current model for archaeal MCM loading is that, at origins, MCM is loaded onto DNA strands as a head to head (N-terminal to N-terminal) double hexamers. As shown in Figure 6.3, it is predicted that the C-terminal winged helix domains of MCM interact with the winged helix domains of Orc1 (Samson & Bell 2016). When the winged helix domain of MCM is deleted, recruitment of MCM by Orc1 is abrogated, which makes it likely that the C-terminus of archaeal MCM is important for protein:protein interactions and therefore recruitment to origins (Samson *et al*. 2016). MCM proteins in *Thermococcus kodakarensis* have also been successfully tagged on the N-terminus using a His_6 tag, as well as MCM in *Pyrococcus furiosus* (Li *et al*. 2010; Yoshimochi *et al*. 2008).

GINS was successfully tagged both on the N-terminus (His_6 tag) and on the C-terminus (StrepII tag), in all combinations. This is consistent with Li *et al*. (2010), who were able to tag GINS in *T. kodakarensis* with a His_6 tag.

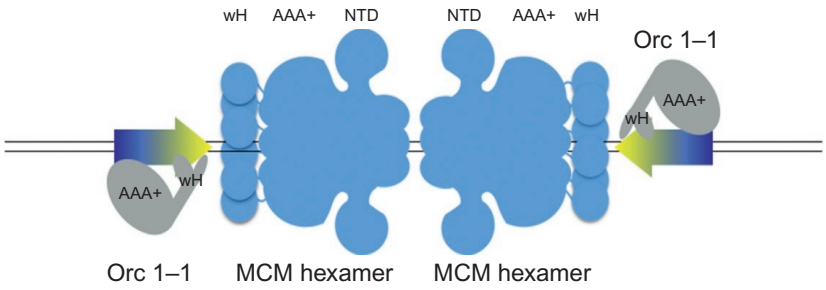


Figure 6.3 **Archaeal MCM interacts with Orc.** A model showing the C-terminal domain on the MCM homohexamer interacts with Orc initiator proteins. Adapted from Samson & Bell (2016).

The purification of His_6 tagged MCM using metal affinity chromatography was deemed successful, however further optimisation of the Western blot protocol is required. It is likely that GINS does co-purify with MCM, however it was not possible to establish this during this study. This may be due to the fact that MCM was not overexpressed (it was expressed in at native levels), therefore there may not enough protein to visualise GINS efficiently. Another reason may be that the interaction between MCM and GINS is transient and therefore only small amounts of protein interact within cells, which Western blotting may not be sensitive enough to detect. An alternative method of detecting interactions may be to purify both proteins individually and look for interaction *in vitro*, however this may still be inconclusive if other partners are involved such as Orc1 or RecJ.

In previous studies in archaea, interaction between MCM and GINS proteins were seen in a variety of species including *Pyrococcus furiosus* and *Sulfolobus solfataricus* (Yoshimochi *et al*. 2008; Marinsek *et al*. 2006). However, it was seen that GINS interacts with many other replication fork-associated proteins such as Orc1, primase, PolD, PCNA, and homologs of the eukaryotic protein Cdc45 and bacterial RecJ exonuclease (Yoshimochi *et al*. 2008; Marinsek *et al*. 2006; Li *et al*. 2010; Xu *et al*. 2016; Oyama *et al*. 2016). Therefore, it is highly likely that there is at least one other component present in the replicative helicase complex of *Haloferax volcanii* alongside MCM. It has been established in eukaryotes that MCM acts as part of the CMG (Cdc45-MCM-GINS) complex (Onesti & MacNeill 2013), therefore it may be fruitful to investigate a possible Cdc45 homolog in *H. volcanii* that could bridge the interaction between MCM and GINS, rather than attempt to observe direct interactions between MCM and

GINS. In bacteria, RecJ is an exonuclease and it is suggested that Cdc45 is the eukaryotic orthologue of RecJ, indicating that RecJ homologues could be acting as the third component of the CMG complex in archaea (Sanchez-Pulido & Ponting 2011; Makarova *et al*. 2012). RecJ2 in *Thermoplasma acidophilum* has been shown to form a complex *in vivo* with GINS and MCM, and maintains its exonuclease activity (Ogino *et al*. 2017). In *Thermococcus kodakarensis*, an exonuclease has been shown to associate with GINS and therefore has been designated GAN (GINS-Associated-Nuclease). This also indicates strong links between DNA replication and DNA repair (Oyama *et al*. 2016). It has been shown that a CMG-like complex can be formed in *T. kodakarensis* that includes GAN, GINS and MCM, and that the nuclease activity of GAN is not important in cells (Nagata *et al*. 2017). *Haloferax volcanii* has 4 copies of a RecJ homologue, two with a nuclease function and two lacking nuclease activity. Investigating one or more of these RecJ proteins may provide insights into MCM and GINS interactions at the replication fork. Experiments such as second site suppressor studies and/ or mutation studies of replication proteins will also enable further investigation into the role of MCM and GINS in *Haloferax volcanii*.

Chapter 7: Conclusions and Future Work

7.1 Genome architecture

7.1.1 Conclusions

It was discovered that the origins found on the episomal mega-plasmid pHV4 could be deleted, however the origin on episomal pHV3 could not. This suggests that DNA must be above a size threshold before efficient recombination can occur; pHV4 (636 kb) is large enough to use recombination-dependent replication (RDR), while pHV3 (438 kb) is not. The origin from pHV1 could be deleted, which seemingly argues against the hypothesis of a size requirement for RDR as it is well below the threshold set by pHV3 at only 85 kb. However, it was found that pHV1 can exist as a 6 x concatemer in *H. volcanii* as well as being present as a monomer. As a 6 x concatemer, the pHV1 DNA molecule is 510 kb in size, which puts it above the proposed threshold for efficient RDR. This poses the question of why pHV1 has an origin at all? It was found that the origin on pHV1 may be required for stability and maintenance of the mega-plasmid, and that pHV1 may be acting as a selfish element within *H. volcanii*, since it is unknown what benefits (if any) pHV1 confers to the cell.

7.2.2 Future Work

Several areas of this study require further investigation.

Firstly, the stability of the pHV3 integration must be assured by the deletion of one or both of the *leuB* alleles produced from the integration. This would ensure the integration is permanent and would avoid recombination between the two alleles making pHV3 episomal again. The integration of pHV3 should also be attempted at a different locus to make sure that the location of the integration does not affect DNA replication or transcription due to interrupted gene regions.

The suggestion of a size threshold for recombination-dependent recombination of between 438 kb and 636 kb needs to be explored further. This could be

achieved in two ways; making pHV1 smaller and assessing whether it still maintains a concatemer size of 510 kb, and making pHV3 larger so as to increase its size above the threshold for RDR. Possible initiation by R-loops also requires further investigation. The rRNA operon is a highly transcribed region, which could be inserted onto pHV3 to increase the levels of transcription, therefore increasing the numbers of R-loops and the possibility of RDR initiation. The origin and/or associated *orc* could then be attempted to be deleted. If R-loops are important for initiation of RDR and the rRNA operon increases the number of R-loops to a sufficient level, these deletions should be possible, which would show the pHV3 and RDR as a whole has a higher requirement for R-loops than first thought.

Deletions of *ori-pHV3* and/or *orc6* from integrated pHV3 must be carried out in all strain backgrounds with deletion constructs carrying the tryptophan selectable marker. This must be followed by growth assays of both the integrated pHV3 strains and the strains with *Δori-pHV3* and/or *Δorc6* to assess whether these genomic rearrangements and deletions affect their growth.

The stability assay of pHV1 should be repeated with a different marker that will hopefully give fewer false positive results and give a clearer indication about the stability of pHV1, with and without its origin. The *trpA* or *leuB* markers are good candidates for this. The *parA1* gene found on pHV1 should be deleted and the stability assay repeated to see if it confers any stability/ maintenance to pHV1, due to its connection with chromosome partitioning.

Growing a *ΔpHV1* strain in variety of growth conditions will help determine if pHV1 really is non-essential or confers a growth advantage in certain conditions. The *ΔpHV1* strain should also be tested for vulnerability to phage infection compared to a wild-type strain. This would indicate whether or not the presence of pHV1 in the cell confers phage resistance. There should also be further investigation into the genes present on pHV1 to see if there are any clues as to how it forms concatemers and carries out both origin-dependent and origin-independent DNA replication.

These experiments will help further the understanding into the relationships of origins, their associated *orc* genes and recombination-dependent replication in *Haloferax volcanii*.

7.2 DNA Replication Machinery

7.2.1 Conclusions

In this study, it was found that MCM plays a role in recombination-dependent replication in *Haloferax volcanii*. This was shown by the depletion of MCM using a tryptophan inducible promoter, which caused a growth defect in an origin-deleted strain compared to wild-type. This hints at an important role for MCM during recombination-dependent replication, although it must be investigated further as to whether recombination-dependent replication requires more MCM than origin-dependent replication, or whether higher levels of recombination within the cell plus the depletion of MCM leads to higher levels of genome instability, thus causing a growth defect.

MCM is known to interact with different proteins associated with DNA replication and DNA repair in a variety of organisms. Unfortunately, during the time frame of this study it was not possible to be determine whether MCM interacts with GINS in *Haloferax volcanii*. However, it was shown that a protein thought to be MCM can be successfully tagged and purified using an N-terminal His_6 tag, which will be useful in future work in determining the role of MCM in recombination-dependent replication. It was also shown that GINS may be tagged both on the C-terminus and the N-terminus, which will be valuable for future work looking at DNA replication mechanisms in *Haloferax volcanii*.

7.2.2 Future Work

To further investigate the role of MCM in recombination-dependent replication, it would be beneficial to look at two areas; the effect of MCM on growth rates and the interaction of MCM with replication machinery.

The effect of the levels of MCM on growth rates requires the determination of how the low activity tryptophan inducible promoter, *p.tnaAM3*, actually affects levels of MCM within the cell. This could be achieved by combining *p.tnaAM3::mcm* strains with N-terminal His_6 tagged MCM and using Western blot to look at levels of MCM when the cells are grown in different concentrations of tryptophan. In an origin-deleted strain it may also be worth investigating the level of ploidy when MCM has been depleted, as the levels of MCM may have a direct effect on DNA replication versus cell division; do strains grow slower and maintain ploidy, or continue cell division and reduce ploidy possibly to the extent where they can no longer efficiently used recombination-dependent replication? As MCM has been shown in this study to be required for origin-independent replication, a ChipSeq study could be carried out to see where MCM is binding in an origin-deleted strain versus a wild-type strain. This would show where initiation takes place in an origin-deleted strain and whether or not there are 'hotspots' of initiation.

Strains where *mcm* is regulated by the *p.tnaA* promoter would also be useful in look at what happens when MCM is overexpressed rather than depleted. Using the Microplate Spectrophotometer is useful as an indication of growth rates, but other methods such as growth competition assays are more accurate and can determine smaller changes in cell growth (Delmas *et al*. 2009). These methods would be useful for studies investigating looking at overexpression or depletion of MCM.

To further investigate the role of the replication machinery in recombination-dependent replication, it may be worthwhile to use the *p.tnaAM3* promoter to regulate expression of GINS. This would determine whether the reduction in growth in origin-deleted strains where MCM is depleted is due to the lack of resources required, in which case the same would be true of depleting GINS, or whether MCM alone is needed to genome and replication fork stability, in which case levels of GINS should not affect growth in the same manner.

Exploring the interaction of MCM with other proteins could provide insights into how this protein acts in recombination-dependent replication and DNA

repair pathways such as non-homologous end joining and homologous recombination. The method used for Western blotting of *Haloferax volcanii* proteins must be optimised, as this method did not produce reliable results in this study. It would then be possible to repeat the co-precipitation study with N-terminal His_6 tagged MCM and C-terminal StrepII tagged GINS to determine their interaction. Other proteins could be tagged that are involved in DNA replication (such as the RecJ proteins, or DNA polymerases), and DNA repair (such as RadA). DNA replication proteins could be tagged in wild-type and origin-deleted strain backgrounds and compared to see the if there are any differences in abundance, binding or interactions when the cell is using origin-dependent replication or recombination-dependent replication. When combined with improved techniques for Western blotting, Northern blotting to monitor transcription, and biophysical methods for examining protein:protein and protein:DNA interactions (e.g. microscale thermophoresis using a Nanotemper), these combinations of tagged MCM and GINS would provide further insights into how these and other proteins act during origin-dependent and recombination-dependent replication.

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.

Allers, T. 2010. Overexpression and purification of halophilic proteins in Haloferax volcanii. Bioengineered bugs, 1, 288-290.

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., 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.

Alonso, J. C., Cardenas, P. P., Sanchez, H., Hejna, J., Suzuki, Y. & Takeyasu, K. 2013. Early steps of double-strand break repair in Bacillus subtilis. DNA Repair (Amst), 12, 162-76.

Ammar, R., Torti, D., Tsui, K., Gebbia, M., Durbic, T., Bader, G. D., Giaever, G. & Nislow, C. 2012. Chromatin is an ancient innovation conserved between Archaea and Eukarya. Elife, 1, e00078.

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.

Bailey, L. J. & Doherty, A. J. 2017. Mitochondrial DNA replication: a PrimPol perspective. Biochem Soc Trans, 45, 513-529.

Bailey, R., Priego Moreno, S. & Gambus, A. 2015. Termination of DNA replication forks: "Breaking up is hard to do". Nucleus, 6, 187-96.

Bailis, J. M., Luche, D. D., Hunter, T. & Forsburg, S. L. 2008. Minichromosome maintenance proteins interact with checkpoint and recombination proteins to promote s-phase genome stability. Mol Cell Biol, 28, 1724-38.

Barry, E. R. & Bell, S. D. 2006. DNA replication in the archaea. Microbiol Mol Biol Rev, 70, 876-87.

Bartlett, E. J., Brissett, N. C., Plocinski, P., Carlberg, T. & Doherty, A. J. 2016. Molecular basis for DNA strand displacement by NHEJ repair polymerases. Nucleic Acids Res, 44, 2173-86.

Beattie, T. R. & Reyes-Lamothe, R. 2015. A Replisome's journey through the bacterial chromosome. Front Microbiol, 6, 562.

Bell, J. C. & Kowalczykowski, S. C. 2016. RecA: Regulation and Mechanism of a Molecular Search Engine. Trends Biochem Sci, 41, 491-507.

Bell, S. D. & Jackson, S. P. 2001. Mechanism and regulation of transcription in archaea. Curr Opin Microbiol, 4, 208-13.

Benelli, D. & Londei, P. 2011. Translation initiation in Archaea: conserved and domain-specific features. Biochem Soc Trans, 39, 89-93.

Bentchikou, E., Servant, P., Coste, G. & Sommer, S. 2010. A major role of the RecFOR pathway in DNA double-strand-break repair through ESDSA in Deinococcus radiodurans. PLoS Genet, 6, e1000774.

Bernstein, K. A., Reid, R. J., Sunjevaric, I., Demuth, K., Burgess, R. C. & Rothstein, R. 2011. The Shu complex, which contains Rad51 paralogues, promotes DNA repair through inhibition of the Srs2 anti-recombinase. Mol Biol Cell, 22, 1599-607.

Betermier, M., Bertrand, P. & Lopez, B. S. 2014. Is non-homologous end-joining really an inherently error-prone process? PLoS Genet, 10, e1004086.

Beverley, S. M. 1989. Estimation of circular DNA size using gamma-irradiation and pulsed-field gel electrophoresis. Anal Biochem, 177, 110-4.

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.

Bornhorst, J. A. & Falke, J. J. 2000. Purification of Proteins Using Polyhistidine Affinity Tags. Methods Enzymol, 326, 245-254.

Bowater, R. & Doherty, A. J. 2006. Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. PLoS Genet, 2, e8.

Breuert, S., Allers, T., Spohn, G. & Soppa, J. 2006. Regulated polyploidy in halophilic archaea. PLoS ONE, 1, e92.

Brewster, A. S., Wang, G., Yu, X., Greeleaf, W. B., Carazo, J. M., Tjajadi, M., Klein, M. G. & Chen, X. S. 2008. Crystal structure of a near-full-length archaeal MCM: Functional insights for an AAA+ hexameric helicase. PNAS, 105 (51), 20191–20196.

Bruand, C. & Ehrlich, S. D. 2000. UvrD-dependent replication of rolling-circle plasmids in Escherichia coli. Mol Microbiol, 35, 204-10.

Bruggemann, H. & Chen, C. 2006. Comparative genomics of Thermus thermophilus: Plasticity of the megaplasmid and its contribution to a thermophilic lifestyle. J Biotechnol, 124, 654-61.

Bryson, V. & Szybalski, W. 1952. Microbial Selection. Science, 116, 45-51.

Burgers, P. M. J. & Kunkel, T. A. 2017. Eukaryotic DNA Replication Fork. Annu Rev Biochem, 86, 417-438.

Byrd, A. K. & Raney, K. D. 2017. Structure and function of Pif1 helicase. Biochem Soc Trans.

Byrne, R. T., Schuller, J. M., Unverdorben, P., Forster, F. & Hopfner, K. P. 2014. Molecular architecture of the HerA-NurA DNA double-strand break resection complex. FEBS Lett, 588, 4637-44.

Cejka, P. 2015. DNA End Resection: Nucleases Team Up with the Right Partners to Initiate Homologous Recombination. J Biol Chem, 290, 22931-8.

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 Borellia. Annu Rev Microbiol, 64, 185–202.

Chandler, M., de la Cruz, F., Dyda, F., Hickman, A. B., Moncalian, G. & Ton-Hoang, B. 2013. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. Nat Rev Microbiol, 11, 525-38.

Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. 2017. Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat Rev Mol Cell Biol, 18, 495-506.

Charlebois, R. L., Lam, W. L., Cline, S. W. & Doolittle, W. F. 1987. Characterization of pHV2 from Halobacterium volcanii and its use in demonstrating transformation of an archaebacterium. Proc Natl Acad Sci U S A, 84, 8530-4.

Chen, X. J. & Clark-Walker, G. D. 2017. Unveiling the mystery of mitochondrial DNA replication in yeasts. Mitochondrion.

Chu, S. F., Shu, H. Y., Lin, L. C., Chen, M. Y., Tsay, S. S. & Lin, G. H. 2006. Characterization of a rolling-circle replication plasmid from Thermus aquaticus NTU103. Plasmid, 56, 46-52.

Cline, S. W., Schalkwyk, L. C. & Doolittle, W. F. 1989. Transformation of the archaebacterium Halobacterium volcanii with genomic DNA. J Bacteriol, 171, 4987-91.

Cooke, M. S., Evans, M. D., Dizdaroglu, M. & Lunec, J. 2003. Oxidative DNA damage: mechanisms, mutation, and disease. FASEB J, 17, 1195-214.

Costa, A., Renault, L., Swuec, P., Petojevic, T., Pesavento, J. J., Ilves, I., MacLellan-Gibson, K., Fleck, R. A., Botchan, M. R. & Berger, J. M. 2014. DNA binding polarity, dimerization, and ATPase ring remodeling in the CMG helicase of the eukaryotic replisome. Elife, 3, e03273.

Coster, G., Frigola, J., Beuron, F., Morris, E. P. & Diffley, J. F. 2014. Origin licensing requires ATP binding and hydrolysis by the MCM replicative helicase. Mol Cell, 55, 666-77.

Crevel, G., Hashimoto, R., Vass, S., Sherkow, J., Yamaguchi, M., Heck, M. M. & Cotterill, S. 2007. Differential requirements for MCM proteins in DNA replication in Drosophila S2 cells. PLoS One, 2, e833.

Crnigoj, M., Podlesek, Z., Zorko, M., Jerala, R., Anderluh, G. & Ulrih, N. P. 2013. Interactions of archaeal chromatin proteins Alba1 and Alba2 with nucleic acids. PLoS One, 8, e58237.

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.

d'Adda di Fagagna, F., Weller, G. R., Doherty, A. J. & Jackson, S. P. 2003. The Gam protein of bacteriophage Mu is an orthologue of eukaryotic Ku. EMBO Rep, 4, 47-52.

De Bont, R. & van Larebeke, N. 2004. Endogenous DNA damage in humans: a review of quantitative data. Mutagenesis, 19, 169-85.

De Falco, M., Catalano, F., Rossi, M., Ciaramella, M. & De Felice, M. 2015. NurA Is Endowed with Endo- and Exonuclease Activities that Are Modulated by HerA: New Insight into Their Role in DNA-End Processing. PLoS One, 10, e0142345.

del Pilar Garcillan-Barcia, M., Bernales, I., Mendiola, M. V. & de la Cruz, F. 2001. Single-stranded DNA intermediates in IS91 rolling-circle transposition. Mol Microbiol, 39, 494-501.

del Solar, G., Giraldo, R., Ruiz-Echevarria, M. J., Espinosa, M. & Diaz-Orejas, R. 1998. Replication and control of circular bacterial plasmids. Microbiol Mol Biol Rev, 62, 434-64.

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.

Dewar, J. M., Budzowska, M. & Walter, J. C. 2015. The mechanism of DNA replication termination in vertebrates. Nature, 525, 345-50.

Dewar, J. M. & Walter, J. C. 2017. Mechanisms of DNA replication termination. Nat Rev Mol Cell Biol, 18, 507-516.

Dillingham, M. S. & Kowalczykowski, S. C. 2008. RecBCD enzyme and the repair of double-stranded DNA breaks. Microbiol Mol Biol Rev, 72, 642-71, Table of Contents.

Donczew, M., Mackiewicz, P., Wróbel, A., Flärdh, K., Zakrzewska-Czerwinska, J. & Jakimowicz, D. 2016. ParA and ParB coordinate chromosome segregation with cell elongation and division during Streptomyces sporulation. Open Biol, 6, 150263.

Donczew, R., Zakrzewska-Czerwinska, J. & Zawilak-Pawlik, A. 2014. Beyond DnaA: the role of DNA topology and DNA methylation in bacterial replication initiation. J Mol Biol, 426, 2269-82.

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.

Elkins, J. G., Podar, M., Graham, D. E., Makarova, K. S., Wolf, Y., Randau, L., Hedlund, B. P., Brochier-Armanet, C., Kunin, V., Anderson, I., Lapidus, A., Goltsman, E., Barry, K., Koonin, E. V., Hugenholtz, P., Kyrpides, N., Wanner, G., Richardson, P., Keller, M. & Stetter, K. O. 2008. A korarchaeal genome reveals insights into the evolution of the Archaea. Proc Natl Acad Sci U S A, 105, 8102-7.

Embley, T. M. & Williams, T. A. 2015. Evolution: Steps on the road to eukaryotes. Nature, 521, 169-70.

Erauso, G., Marsin, S., Benbouzid-Rollet, N., Baucher, M. F., Barbeyron, T., Zivanovic, Y., Prieur, D. & Forterre, P. 1996. Sequence of plasmid pGT5 from the archaeon Pyrococcus abyssi: evidence for rolling-circle replication in a hyperthermophile. J Bacteriol, 178, 3232-7.

Ettenauer, J. D., Jurado, V., Pinar, G., Miller, A. Z., Santner, M., Saiz-Jimenez, C. & Sterflinger, K. 2014. Halophilic microorganisms are responsible for the rosy discolouration of saline environments in three historical buildings with mural paintings. PLoS One, 9, e103844.

Fachinetti, D., Bermejo, R., Cocito, A., Minardi, S., Katou, Y., Kanoh, Y., Shirahige, K., Azvolinsky, A., Zakian, V. A. & Foiani, M. 2010. Replication termination at eukaryotic chromosomes is mediated by Top2 and occurs at genomic loci containing pausing elements. Mol Cell, 39, 595-605.

Fernandez-Lopez, C., Bravo, A., Ruiz-Cruz, S., Solano-Collado, V., Garsin, D. A., Lorenzo-Diaz, F. & Espinosa, M. 2014. Mobilizable Rolling-Circle Replicating Plasmids from Gram-Positive Bacteria: A Low-Cost Conjugative Transfer. Microbiol Spectr, 2, 8.

Fijalkowska, I. J., Schaaper, R. M. & Jonczyk, P. 2012. DNA replication fidelity in Escherichia coli: a multi-DNA polymerase affair. FEMS Microbiol Rev, 36, 1105-21.

Flach, J., Bakker, S. T., Mohrin, M., Conroy, P. C., Pietras, E. M., Reynaud, D., Alvarez, S., Diolaiti, M. E., Ugarte, F., Forsberg, E. C., Le Beau, M. M., Stohr, B. A., Mendez, J., Morrison, C. G. & Passegue, E. 2014. Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells. Nature, 512, 198-202.

French, S. L., Santangelo, T. J., Beyer, A. L. & Reeve, J. N. 2007. Transcription and translation are coupled in Archaea. Mol Biol Evol, 24, 893-5.

Friedberg, E. C., McDaniel, L. D. & Schultz, R. A. 2004. The role of endogenous and exogenous DNA damage and mutagenesis. Curr Opin Genet Dev, 14, 5-10.

Game, J. C., Sitney, K. C., Cook, V. E. & Mortimer, R. K. 1989. Use of a ring chromosome and pulsed-field gels to study interhomolog recombination, double-strand DNA breaks and sister-chromatid exchange in yeast. Genetics, 123, 695-713.

Georgescu, R., Yuan, Z., Bai, L., de Luna Almeida Santos, R., Sun, J., Zhang, D., Yurieva, O., Li, H. & O'Donnell, M. E. 2017. Structure of eukaryotic CMG helicase at a replication fork and implications to replisome architecture and origin initiation. Proc Natl Acad Sci U S A, 114, E697-E706.

Gobbini, E., Cassani, C., Villa, M., Bonetti, D. & Longhese, M. P. 2016. Functions and regulation of the MRX complex at DNA double-strand breaks. Microb Cell, 3, 329-337.

Gowrishankar, J. 2015. End of the beginning: elongation and termination features of alternative modes of chromosomal replication initiation in bacteria. PLoS Genet, 11, e1004909.

Gray, M. W. 2012. Mitochondrial Evolution. Cold Spring Harb Perspect Biol, 4, a011403.

Gregor, D. & Pfeifer, F. 2001. Use of a halobacterial bgaH reporter gene to analyse the regulation of gene expression in halophilic archaea. Microbiology, 147, 1745-54.

Guglielmetti, S., Mora, D. & Parini, C. 2007. Small rolling circle plasmids in Bacillus subtilis and related species: organization, distribution, and their possible role in host physiology. Plasmid, 57, 245-64.

Haldenby, S., White, M. F. & Allers, T. 2009. RecA family proteins in archaea: RadA and its cousins. Biochem Soc Trans, 37, 102-7.

Hall, J. P. J., Brockhurst, M. A., Dytham, C. & Harrison, E. 2017. The evolution of plasmid stability: Are infectious transmission and compensatory evolution competing evolutionary trajectories? Plasmid, 91, 90-95.

Hartman, A. L., Norais, C., Badger, J. H., Delmas, S., Haldenby, S., Madupu, R., Robinson, J., Khouri, H., Ren, Q., Lowe, T. M., Maupin-Furlow, J., Pohlschroder, M., Daniels, C., Pfeiffer, F., Allers, T. & Eisen, J. A. 2010. The complete genome sequence of Haloferax volcanii DS2, a model archaeon. PLoS ONE, 5, e9605.

Hawkins, M. 2009. DNA replication origins in Haloferax volcanii. University of Nottingham.

Hawkins, M., Malla, S., Blythe, M. J., Nieduszynski, C. A. & Allers, T. 2013. Accelerated growth in the absence of DNA replication origins. Nature, 503, 544-7.

Hedlund, B. P., Murugapiran, S. K., Alba, T. W., Levy, A., Dodsworth, J. A., Goertz, G. B., Ivanova, N. & Woyke, T. 2015. Uncultivated thermophiles: current status and spotlight on 'Aigarchaeota'. Curr Opin Microbiol, 25, 136-45.

Hefferin, M. L. & Tomkinson, A. E. 2005. Mechanism of DNA double-strand break repair by non-homologous end joining. DNA Repair (Amst), 4, 639-48.

Helleday, T., Eshtad, S. & Nik-Zainal, S. 2014. Mechanisms underlying mutational signatures in human cancers. Nat Rev Genet, 15, 585-98.

Heyer, W. D., Ehmsen, K. T. & Liu, J. 2010. Regulation of homologous recombination in eukaryotes. Annu Rev Genet, 44, 113-39.

Hoggard, T., Shor, E., Muller, C. A., Nieduszynski, C. A. & Fox, C. A. 2013. A Link between ORC-origin binding mechanisms and origin activation time revealed in budding yeast. PLoS Genet, 9, e1003798.

Holmes, M. L. & Dyall-Smith, M. L. 2000. Sequence and expression of a halobacterial beta-galactosidase gene. Mol Microbiol, 36, 114-122.

Holmes, M. L., Nuttall, S. D. & Dyall-Smith, M. L. 1991. Construction and use of halobacterial shuttle vectors and further studies on Haloferax DNA gyrase. J Bacteriol, 173, 3807-13.

Holt, I. J. & Jacobs, H. T. 2014. Unique features of DNA replication in mitochondria: a functional and evolutionary perspective. Bioessays, 36, 1024-31.

Holt, I. J. & Reyes, A. 2012. Human mitochondrial DNA replication. Cold Spring Harb Perspect Biol, 4.

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.

Hu, J., Lieb, J. D., Sancar, A. & Adar, S. 2016. Cisplatin DNA damage and repair maps of the human genome at single-nucleotide resolution. Proc Natl Acad Sci U S A, 113, 11507-11512.

Huang, Q., Liu, L., Liu, J., Ni, J., She, Q. & Shen, Y. 2015. Efficient 5'-3' DNA end resection by HerA and NurA is essential for cell viability in the crenarchaeon Sulfolobus islandicus. BMC Mol Biol, 16, 2.

Huber, H., Hohn, M. J., Stetter, K. O. & Rachel, R. 2003. The phylum Nanoarchaeota: present knowledge and future perspectives of a unique form of life. Res Microbiol, 154, 165-71.

Humbard, M. A., Zhou, G. & Maupin-Furlow, J. A. 2009. The N-terminal penultimate residue of 20S proteasome alpha1 influences its N(alpha) acetylation and protein levels as well as growth rate and stress responses of Haloferax volcanii. J Bacteriol, 191, 3794-803.

Ibarra, A., Schwob, E. & Mendez, J. 2008. Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. Proc Natl Acad Sci U S A, 105, 8956-61.

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.

Jaakkola, S. T., Ravantti, J. J., Oksanen, H. M. & Bamford, D. H. 2016. Buried Alive: Microbes from Ancient Halite. Trends Microbiol, 24, 148-60.

Jacob, F., Brenner, S. & Cuzin, F. 1963. On the regulation of DNA replication in bacteria. Cold Spring Harb Symp Quant Biol, 28, 329-348.

Jasin, M. & Rothstein, R. 2013. Repair of strand breaks by homologous recombination. Cold Spring Harb Perspect Biol, 5, a012740.

Johansson, E. & Dixon, N. 2013. Replicative DNA polymerases. Cold Spring Harb Perspect Biol, 5.

Jun, S. H., Reichlen, M. J., Tajiri, M. & Murakami, K. S. 2011. Archaeal RNA polymerase and transcription regulation. Crit Rev Biochem Mol Biol, 46, 27-40.

Kapitonov, V. V. & Jurka, J. 2001. Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci U S A, 98, 8714-9.

Kapitonov, V. V. & Jurka, J. 2007. Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet, 23, 521-9.

Kasho, K. & Katayama, T. 2013. DnaA binding locus datA promotes DnaA-ATP hydrolysis to enable cell cycle-coordinated replication initiation. Proc Natl Acad Sci U S A, 110, 936-41.

Kelman, Z. & White, M. F. 2005. Archaeal DNA replication and repair. Curr Opin Microbiol, 8, 669-76.

Khan, S. A. 1997. Rolling-circle replication of bacterial plasmids. Microbiol Mol Biol Rev, 61, 442-55.

Khan, S. A. 2005. Plasmid rolling-circle replication: highlights of two decades of research. Plasmid, 53, 126-36.

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.

Kolinjivadi, A. M., Sannino, V., de Antoni, A., Techer, H., Baldi, G. & Costanzo, V. 2017. Moonlighting at replication forks - a new life for homologous recombination proteins BRCA1, BRCA2 and RAD51. FEBS Lett, 591, 1083-1100.

Komori, K., Miyata, T., DiRuggiero, J., Holley-Shanks, R., Hayashi, I., Cann, I. K., Mayanagi, K., Shinagawa, H. & Ishino, Y. 2000. Both RadA and RadB are involved in homologous recombination in Pyrococcus furiosus. J Biol Chem, 275, 33782-90.

Konopa, G., Baranska, S., Wegrzyn, A. & Wegrzyn, G. 2000. Bacteriophage and host mutants causing the rolling-circle lambda DNA replication early after infection. FEBS Lett, 472, 217-20.

Koonin, E. V. 2015. Archaeal ancestors of eukaryotes: not so elusive any more. BMC Biol, 13, 84.

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.

Krejci, L., Altmannova, V., Spirek, M. & Zhao, X. 2012. Homologous recombination and its regulation. Nucleic Acids Res, 40, 5795-818.

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.

Lam, W. L. & Doolittle, W. F. 1989. Shuttle vectors for the archaebacterium Halobacterium volcanii. Proc Natl Acad Sci U S A, 86, 5478-82.

Large, A., Stamme, C., Lange, C., Duan, Z., Allers, T., Soppa, J. & Lund, P. A. 2007. Characterization of a tightly controlled promoter of the halophilic archaeon Haloferax volcanii and its use in the analysis of the essential cct1 gene. Mol Microbiol, 66, 1092-106.

Laurens, N., Driessen, R. P., Heller, I., Vorselen, D., Noom, M. C., Hol, F. J., White, M. F., Dame, R. T. & Wuite, G. J. 2012. Alba shapes the archaeal genome using a delicate balance of bridging and stiffening the DNA. Nat Commun, 3, 1328.

Leahy, P., Carmichael, G. G. & Rossomando, E. F. 1997. Transcription from plasmid expression vectors is increased up to 14-fold when plasmids are transfected as concatemers. Nucleic Acids Res, 25, 449-50.

LeBowitz, J. H. & McMacken, R. 1986. The Escherichia coli dnaB replication protein is a DNA helicase. J Biol Chem, 261, 4738-48.

Lee, D. Y. & Clayton, D. A. 1998. Initiation of mitochondrial DNA replication by transcription and R-loop processing. J Biol Chem, 273, 30614-21.

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.

Leman, A. R. & Noguchi, E. 2013. The replication fork: understanding the eukaryotic replication machinery and the challenges to genome duplication. Genes (Basel), 4, 1-32.

Leonard, A. C. & Grimwade, J. E. 2009. Initiating chromosome replication in E. coli: it makes sense to recycle. Genes Dev, 23, 1145-50.

Le Quéré, B. & Ghigo, J.-M. 2009. BcsQ is an essential component of the Escherichia coli cellulose biosynthesis apparatus that localizes at the bacterial cell pole. Mol Micro, 72, 3, 724-740.

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.

Lewis, S. C., Joers, P., Willcox, S., Griffith, J. D., Jacobs, H. T. & Hyman, B. C. 2015. A rolling circle replication mechanism produces multimeric lariats of mitochondrial DNA in Caenorhabditis elegans. PLoS Genet, 11, e1004985.

Li, X. & Heyer, W. D. 2008. Homologous recombination in DNA repair and DNA damage tolerance. Cell Res, 18, 99-113.

Li, Z., Santangelo, T. J., Cubonova, L., Reeve, J. N. & Kelman, Z. 2010. Affinity purification of an archaeal DNA replication protein network. MBio, 1.

Lieber, M. R., Ma, Y., Pannicke, U. & Schwarz, K. 2003. Mechanism and regulation of human non-homologous DNA end-joining. Nat Rev Mol Cell Biol, 4, 712-20.

Lindås, A. C. & Bernander, R. 2013. The cell cycle of archaea. Nat Rev Microbiol, 11, 627-38.

Liu, H., Wu, Z., Li, M., Zhang, F., Zheng, H., Han, J., Liu, J., Zhou, J., Wang, S. & Xiang, H. 2011a. Complete genome sequence of Haloarcula hispanica, a Model Haloarchaeon for studying genetics, metabolism, and virus-host interaction. J Bacteriol, 193, 6086-7.

Liu, J., Ehmsen, K. T., Heyer, W. D. & Morrical, S. W. 2011b. Presynaptic filament dynamics in homologous recombination and DNA repair. Crit Rev Biochem Mol Biol, 46, 240-70.

Liu, J. & Morrical, S. W. 2010. Assembly and dynamics of the bacteriophage T4 homologous recombination machinery. Virol J, 7, 357.

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.

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.

Macheret, M. & Halazonetis, T. D. 2018. Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. Nature, 555, 112-116.

Mackiewicz, P., Zakrzewska-Czerwinska, J., Zawilak, A., Dudek, M. R. & Cebrat, S. 2004. Where does bacterial replication start? Rules for predicting the oriC region. Nucleic Acids Res, 32, 3781-91.

Mackwan, R. R., Carver, G. T., Drake, J. W. & Grogan, D. W. 2007. An unusual pattern of spontaneous mutations recovered in the halophilic archaeon Haloferax volcanii. Genetics, 176, 697-702.

MacNeil, S. A. 2010. Structure and function of the GINS complex, a key component of the eukaryotic replisome. Biochem J, 425, 489-500.

Madern, D., Ebel, C. & Zaccai, G. 2000. Halophilic adaptation of enzymes. Extremophiles, 4, 91-8.

Makarova, K. S. & Koonin, E. V. 2003. Comparative genomics of Archaea: how much have we learned in six years, and what's next? Genome Biol, 4, 115.

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.

Malkova, A. & Ira, G. 2013. Break-induced replication: functions and molecular mechanism. Curr Opin Genet Dev, 23, 271-9.

Marini, V. & Krejci, L. 2010. Srs2: the "Odd-Job Man" in DNA repair. DNA Repair (Amst), 9, 268-75.

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. & Allers, T. 2016. Archaea and the meaning of life. Microbiology Today, 43, 74-77.

Mattiroli, F., Bhattacharyya, S., Dyer, P. N., White, A. E., Sandman, K., Burkhart, B. W., Byrne, K. R., Lee, T., Ahn, N. G., Santangelo, T. J., Reeve, J. N. & Luger, K. 2017. Structure of histone-based chromatin in Archaea. Science, 357, 609-612.

McHenry, C. S. 2011. Bacterial replicases and related polymerases. Curr Opin Chem Biol, 15, 587-94.

Meijer, W. J., Wisman, G. B., Terpstra, P., Thorsted, P. B., Thomas, C. M., Holsappel, S., Venema, G. & Bron, S. 1998. Rolling-circle plasmids from Bacillus subtilis: complete nucleotide sequences and analyses of genes of pTA1015, pTA1040, pTA1050 and pTA1060, and comparisons with related plasmids from gram-positive bacteria. FEMS Microbiol Rev, 21, 337-68.

Messer, W. 2002. The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. FEMS Microbiol Rev, 26, 355-74.

Michel, B. & Bernander, R. 2014. Chromosome replication origins: do we really need them? Bioessays, 36, 585-90.

Miller, J. M. & Enemark, E. J. 2015. Archaeal MCM Proteins as an Analog for the Eukaryotic Mcm2-7 Helicase to Reveal Essential Features of Structure and Function. Archaea, 2015, 305497.

Miura, T., Yamana, Y., Usui, T., Ogawa, H. I., Yamamoto, M. T. & Kusano, K. 2012. Homologous recombination via synthesis-dependent strand annealing in yeast requires the Irc20 and Srs2 DNA helicases. Genetics, 191, 65-78.

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.

Morimatsu, K., Wu, Y. & Kowalczykowski, S. C. 2012. RecFOR proteins target RecA protein to a DNA gap with either DNA or RNA at the 5' terminus: implication for repair of stalled replication forks. J Biol Chem, 287, 35621-30.

Mullakhanbhai, M. F. & Larsen, H. 1975. Halobacterium volcanii spec. nov., a Dead Sea halobacterium with a moderate salt requirement. Arch Microbiol, 104, 207-14.

Mulugu, S., Potnis, A., Shamsuzzaman, Taylor, J., Alexander, K. & Bastia, D. 2001. Mechanism of termination of DNA replication of Escherichia coli involves helicase-contrahelicase interaction. Proc Natl Acad Sci U S A, 98, 9569-74.

Munson-McGee, J. H., Field, E. K., Bateson, M., Rooney, C., Stepanauskas, R. & Young, M. J. 2015. Nanoarchaeota, Their Sulfolobales Host, and Nanoarchaeota Virus Distribution across Yellowstone National Park Hot Springs. Appl Environ Microbiol, 81, 7860-8.

Nagata, M., Ishino, S., Yamagami, T., Ogino, H., Simons, J.-R., Kanai, T., Atomi, H. & Ishino, Y. 2017. 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.

Naor, A., Lapierre, P., Mevarech, M., Papke, R. T. & Gophna, U. 2012. Low Species Barriers in Halophilic Archaea and the Formation of Recombinant Hybrids. Curr Biol, 22, 1444-1448.

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.

Northall, S. J., Buckley, R., Jones, N., Penedo, J. C., Soultanas, P. & Bolt, E. L. 2017. DNA binding and unwinding by Hel308 helicase requires dual functions of a winged helix domain. DNA Repair (Amst), 57, 125-132.

Novick, R. P. 1998. Contrasting lifestyles of rolling-circle phages and plasmids. Trends Biochem Sci, 23, 434-8.

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.

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.

Ogino, H., Ishino, S., Kohda, D. & Ishino, Y. 2017. The RecJ2 protein in the thermophilic archaeon Thermoplasma acidophilum is a 3'-5' exonuclease that associates with a DNA replication complex. J Biol Chem, 292, 7921-7931.

Ohtani, N., Tomita, M. & Itaya, M. 2015. Curing the Megaplasmid pTT27 from Thermus thermophilus HB27 and Maintaining Exogenous Plasmids in the Plasmid-Free Strain. Appl Environ Microbiol, 82, 1537-48.

Olsen, G. J. & Woese, C. R. 1997. Archaeal genomics: an overview. Cell, 89, 991-4.

Onesti, S. & MacNeill, S. A. 2013. Structure and evolutionary origins of the CMG complex. Chromosoma, 122, 47-53.

Oren, A. 2008. Microbial life at high salt concentrations: phylogenetic and metabolic diversity. Saline Systems, 4, 2.

Orr, S. J., Gaymes, T., Ladon, D., Chronis, C., Czepulkowski, B., Wang, R., Mufti, G. J., Marcotte, E. M. & Thomas, N. S. 2010. Reducing MCM levels in human primary T cells during the G(0)-->G(1) transition causes genomic instability during the first cell cycle. Oncogene, 29, 3803-14.

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., 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.

Ozaki, S. & Katayama, T. 2012. Highly organized DnaA-oriC complexes recruit the single-stranded DNA for replication initiation. Nucleic Acids Res, 40, 1648-65.

Peeters, E. & Charlier, D. 2010. The Lrp family of transcription regulators in archaea. Archaea, 2010, 750457.

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.

Pester, M., Schleper, C. & Wagner, M. 2011. The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. Curr Opin Microbiol, 14, 300-6.

Pitcher, R. S., Brissett, N. C. & Doherty, A. J. 2007. Nonhomologous end-joining in bacteria: a microbial perspective. Annu Rev Microbiol, 61, 259-82.

Poirier, M. C. 2004. Chemical-induced DNA damage and human cancer risk. Nat Rev Cancer, 4, 630-7.

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.

Rizvi, I., Choudhury, N. R. & Tuteja, N. 2015. Insights into the functional characteristics of geminivirus rolling-circle replication initiator protein and its interaction with host factors affecting viral DNA replication. Arch Virol, 160, 375-87.

Rosenshine, I., Tchelet, R. & Mevarech, M. 1989. The mechanism of DNA transfer in the mating system of an archaebacterium. Science, 245, 1387-9.

Rotman, E., Khan, S. R., Kouzminova, E. & Kuzminov, A. 2014. Replication fork inhibition in seqA mutants of Escherichia coli triggers replication fork breakage. Mol Microbiol, 93, 50-64.

Ruiz-Maso, J. A., Macho, N. C., Bordanaba-Ruiseco, L., Espinosa, M., Coll, M. & Del Solar, G. 2015. Plasmid Rolling-Circle Replication. Microbiol Spectr, 3, PLAS-0035-2014.

Rzechorzek, N. J., Blackwood, J. K., Bray, S. M., Maman, J. D., Pellegrini, L. & Robinson, N. P. 2014. Structure of the hexameric HerA ATPase reveals a mechanism of translocation-coupled DNA-end processing in archaea. Nat Commun, 5, 5506.

Saini, N., Ramakrishnan, S., Elango, R., Ayyar, S., Zhang, Y., Deem, A., Ira, G., Haber, J. E., Lobachev, K. S. & Malkova, A. 2013. Migrating bubble during break-induced replication drives conservative DNA synthesis. Nature, 502, 389-92.

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. 2011. Cell cycles and cell division in the archaea. Curr Opin Microbiol, 14, 350-6.

Samson, R. Y. & Bell, S. D. 2016. Archaeal DNA Replication Origins and Recruitment of the MCM Replicative Helicase. Enzymes, 39, 169-90.

Sanchez-Pulido, L. & Ponting, C. P. 2011. Cdc45: the missing RecJ ortholog in eukaryotes? Bioinformatics, 27, 1885-8.

Sandman, K. & Reeve, J. N. 2005. Archaeal chromatin proteins: different structures but common function? Curr Opin Microbiol, 8, 656-61.

Sanger, F., Nicklen, S. & Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A, 74, 5463-7.

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.

Sinha, R. P. & Hader, D. P. 2002. UV-induced DNA damage and repair: a review. Photochem Photobiol Sci, 1, 225-36.

Smith, G. R. 2012. How RecBCD enzyme and Chi promote DNA break repair and recombination: a molecular biologist's view. Microbiol Mol Biol Rev, 76, 217-28.

Soler, N., Gaudin, M., Marguet, E. & Forterre, P. 2011. Plasmids, viruses and virus-like membrane vesicles from Thermococcales. Biochem Soc Trans, 39, 36-44.

Soler, N., Justome, A., Quevillon-Cheruel, S., Lorieux, F., Le Cam, E., Marguet, E. & Forterre, P. 2007. The rolling-circle plasmid pTN1 from the hyperthermophilic archaeon Thermococcus nautilus. Mol Microbiol, 66, 357-70.

Spang, A., Saw, J. H., Jorgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., van Eijk, R., Schleper, C., Guy, L. & Ettema, T. J. G. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature, 521, 173-179.

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.

Sun, J., Yuan, Z., Georgescu, R., Li, H. & O'Donnell, M. 2016. The eukaryotic CMG helicase pumpjack and integration into the replisome. Nucleus, 7, 146-54.

Sung, P. & Klein, H. 2006. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. Nat Rev Mol Cell Biol, 7, 739-50.

Swiatek, A. & MacNeill, S. A. 2010. The archaeo-eukaryotic GINS proteins and the archaeal primase catalytic subunit share a common domain. Biology Direct, 5, 17.

Taylor, K. & Wegrzyn, G. 1995. Replication of coliphage lambda DNA. FEMS Microbiol Rev, 17, 109-19.

Timpson, L. M., Liliensiek, A. K., Alsafadi, D., Cassidy, J., Sharkey, M. A., Liddell, S., Allers, T. & Paradisi, F. 2012. A comparison of two novel alcohol dehydrogenase enzymes (ADH1 and ADH2) from the extreme halophile Haloferax volcanii. Appl Microbiol Biotechnol.

Trieselmann, B. A. & Charlebois, R. L. 1992. Transcriptionally Active Regions in the Genome of the Archaebacterium Haloferax volcanii. J Bact, 174, 30-34.

Viret, J. F., Bravo, A. & Alonso, J. C. 1991. Recombination-dependent concatemeric plasmid replication. Microbiol Rev, 55, 675-83.

Waldminghaus, T., Weigel, C. & Skarstad, K. 2012. Replication fork movement and methylation govern SeqA binding to the Escherichia coli chromosome. Nucleic Acids Res, 40, 5465-76.

Wang, Y., Chen, B., Cao, M., Sima, L., Prangishvili, D., Chen, X. & Krupovic, M. 2018. Rolling-circle replication initiation protein of haloarchaeal sphaerolipovirus SNJ1 is homologous to bacterial transposases of the IS91 family insertion sequences. J Gen Virol, 10.1099/jgv.0.001009.

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.

Wegrzyn, K., Fuentes-Perez, M. E., Bury, K., Rajewska, M., Moreno-Herrero, F. & Konieczny, I. 2014. Sequence-specific interactions of Rep proteins with ssDNA in the AT-rich region of the plasmid replication origin. Nucleic Acids Res, 42, 7807-18.

Wendoloski, D., Ferrer, C. & Dyall-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.

Werbowy, O., Werbowy, S. & Kaczorowski, T. 2017. Plasmid stability analysis based on a new theoretical model employing stochastic simulations. PLoS One, 12, e0183512.

White, M. F. & Bell, S. D. 2002. Holding it together: chromatin in the Archaea. Trends Genet, 18, 621-6.

Williams, T. A., Foster, P. G., Cox, C. J. & Embley, T. M. 2013. An archaeal origin of eukaryotes supports only two primary domains of life. Nature, 504, 231-6.

Wilson, M. A., Kwon, Y., Xu, Y., Chung, W. H., Chi, P., Niu, H., Mayle, R., Chen, X., Malkova, A., Sung, P. & Ira, G. 2013. Pif1 helicase and Poldelta promote recombination-coupled DNA synthesis via bubble migration. Nature, 502, 393-6.

Wingfield, P. T. 2015. Overview of the purification of recombinant proteins. Curr Protoc Protein Sci, 80, 6 1 1-35.

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.

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.

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.

Wu, Z., Liu, J., Yang, H. & Xiang, H. 2014. DNA replication origins in archaea. Front Microbiol, 5, 179.

Xie, G., Johnson, S. L., Davenport, K. W., Rajavel, M., Waldminghaus, T., Detter, J. C., Chain, P. S. & Sozhamannan, S. 2017. Exception to the Rule: Genomic Characterization of Naturally Occurring Unusual Vibrio cholerae Strains with a Single Chromosome. Int J Genomics, 2017, 8724304.

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.

Yang, H., Wu, Z., Liu, J., Liu, X., Wang, L., Cai, S. & Xiang, H. 2015. Activation of a dormant replication origin is essential for Haloferax mediterranei lacking the primary origins. Nat Commun, 6, 8321.

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, 1601-1609.

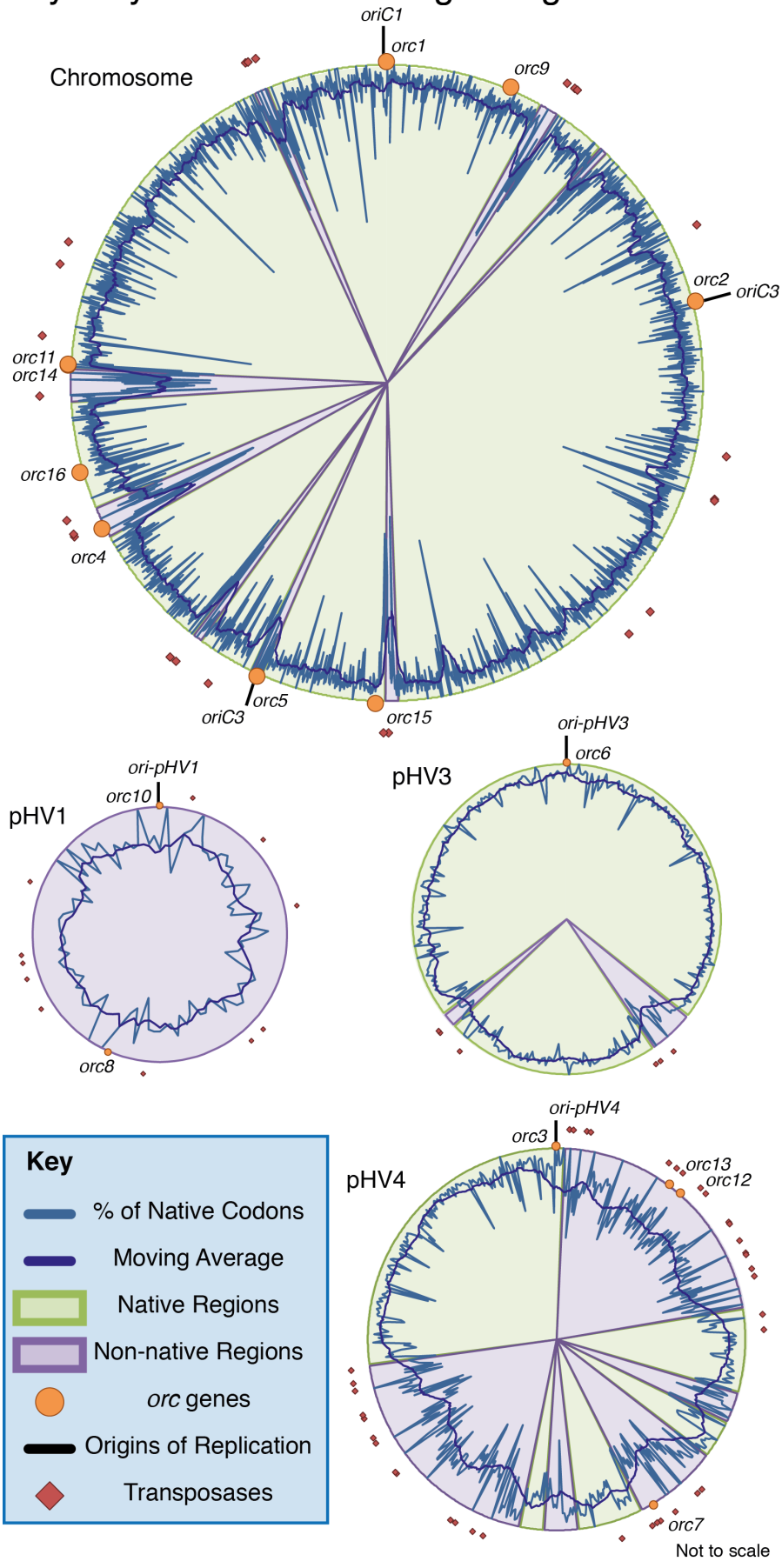
Yuan, Z., Bai, L., Sun, J., Georgescu, R., Liu, J., O'Donnell, M. E. & Li, H. 2016. Structure of the eukaryotic replicative CMG helicase suggests a pumpjack motion for translocation. Nat Struct Mol Biol, 23, 217-24.

Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K. W., Anantharaman, K., Starnawski, P., Kjeldsen, K. U., Stott. M. B., Nunoura, T., Banfield, J. F., Schramm, A., Baker, B. J., Spang, A. & Ettema, T. J. G. 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature, 541, 353-358.

Zhao, X., Li, G. & Liang, S. 2013. Several affinity tags commonly used in chromatographic purification. J Anal Methods Chem, 2013, 581093.

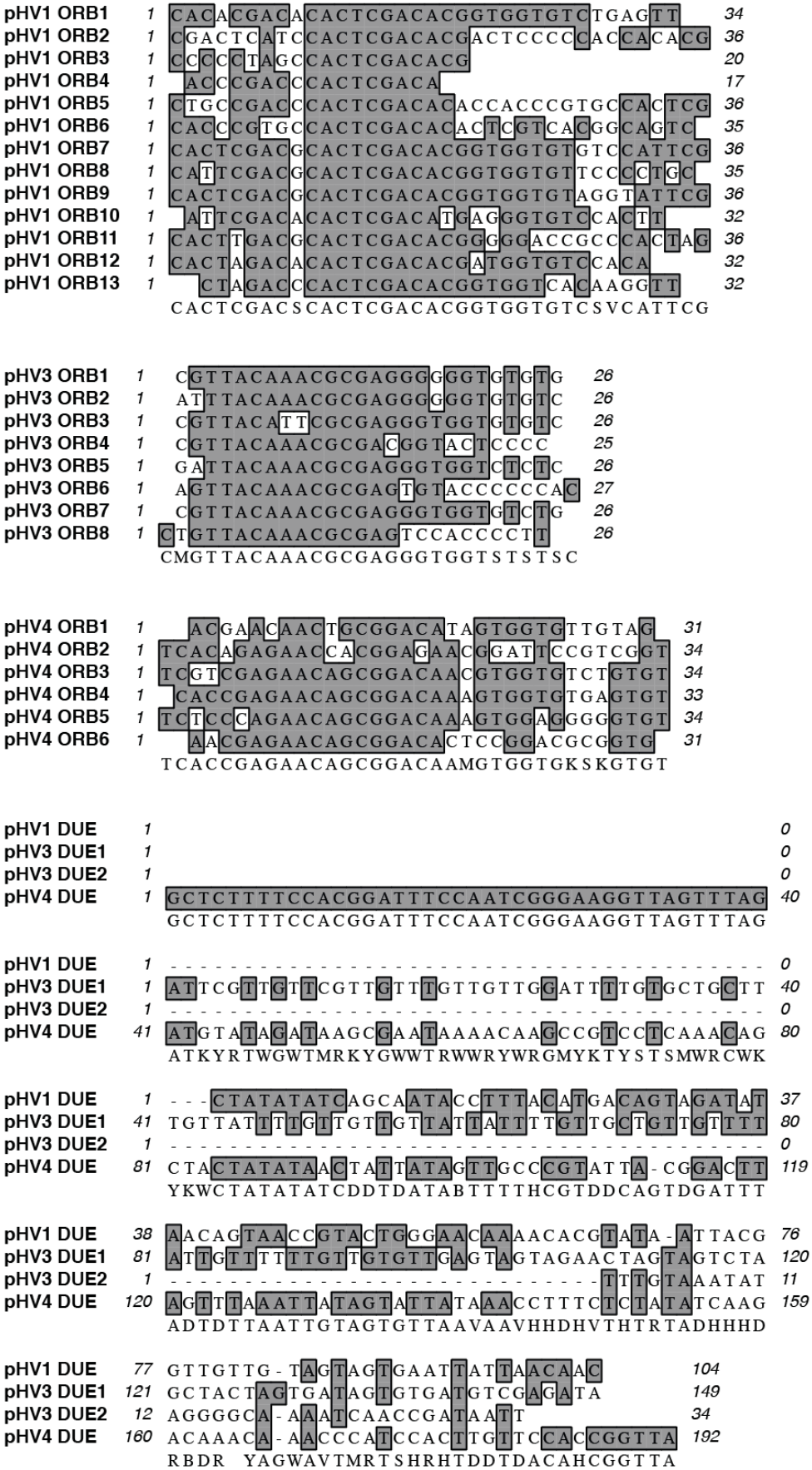
Zhou, M., Xiang, H., Sun, C., Li, Y., Liu, J. & Tan, H. 2004. Complete sequence and molecular characterization of pNB101, a rolling-circle replicating plasmid from the haloalkaliphilic archaeon Natronobacterium sp. strain AS7091. Extremophiles, 8, 91-8.

I Synonymous Codon Usage Diagrams



Synonymous codon usage of the *Haloferax volcanii* genome. Percentage of native codons by synonymous codon usage has been plotted for each gene on each replicon (light blue line) and a moving average has been taken for each replicon (dark blue line). For the main chromosome this is 20 genes, for pHV1 5 genes, pHV3 10 genes and for pHV4 20 genes. Regions that contain 10 or more genes with >85% native codons have been designated native regions (green shading) and regions that contain 10 or more genes with <85% native codons were designated non-native regions (purple shading). Origins of replication (black line), *orc* genes (orange circle) and genes annotated as transposases (red diamond) have also been indicated. Adapted from Hartman *et al*. (2010).

II DUE and ORB Alignments

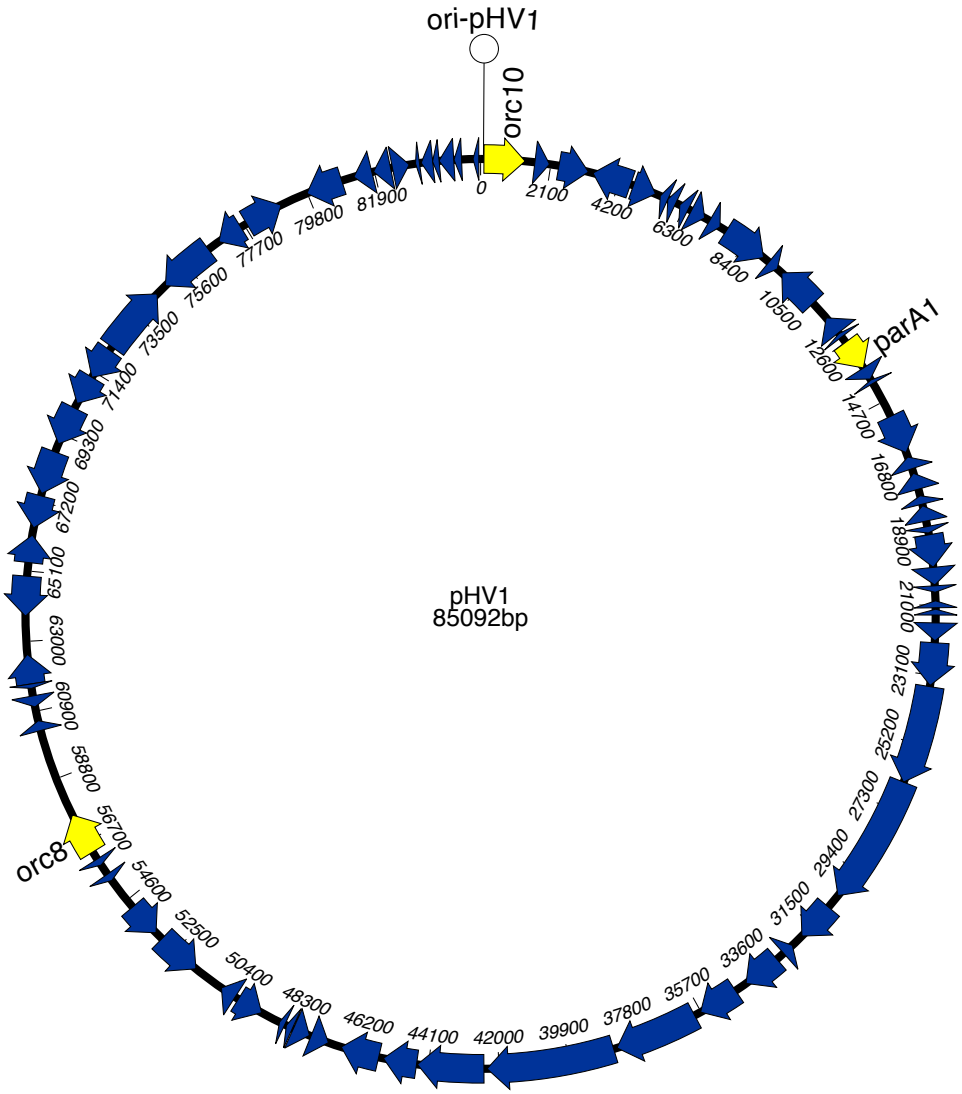


III pHV1

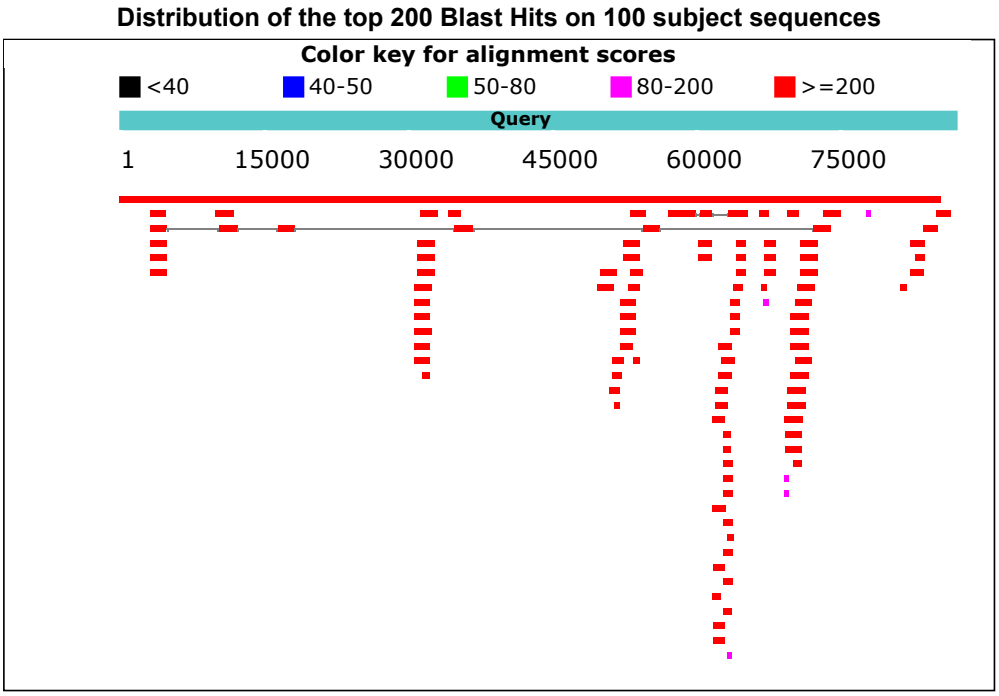
Gene number	Gene name	Gene type	Strand	arCOG annotation	Deleted
HVO_C0001	*orc10*	CDS	1	Cdc6-related protein, AAA superfamily ATPase	Yes
HVO_C0001ori	*ori-pHV1*	ORI	1		Yes
HVO_C0002	-	CDS	1		
HVO_C0003	-	CDS	1	Uncharacterized conserved protein	
HVO_C0004	-	CDS	-1	Transposase	
HVO_C0005	-	CDS	1		
HVO_C0005_A					
HVO_C0006	-	CDS	-1	Predicted RNA binding protein (dsRBD-like fold), HicA family	
HVO_C0007	-	CDS	-1	Predicted nuclease of the RNAse H fold, HicB family	
HVO_C0008	-	CDS	-1		
HVO_C0009	-	CDS	1	Uncharacterized conserved protein	
HVO_C0010	-	CDS	1		
HVO_C0011	-	CDS	1		
HVO_C0012	-	CDS	1	RHH-fold DNA-binding ptotein	
HVO_C0013	-	CDS	-1	Transposase	
HVO_C0014	-	CDS	1		
HVO_C0015	-	CDS	-1	Uncharacterized conserved protein	
HVO_C0016	-	CDS	-1		
HVO_C0017	*parA1*	CDS	1	ATPase involved in chromosome partitioning, ParA family	
HVO_C0018	-	CDS	1	Uncharacterized conserved protein	
HVO_C0019	-	CDS	-1		
HVO_C0020	-	CDS	1		
HVO_C0020_A	-	CDS	1		
HVO_C0021	-	CDS	1	Transposase	
HVO_C0022	-	CDS	-1		
HVO_C0023	-	CDS	-1		
HVO_C0024	-	CDS	-1		
HVO_C0025	-	CDS	-1		
HVO_C0026	-	CDS	-1		
HVO_C0027	-	CDS	1	Predicted metal-dependent hydrolase	
HVO_C0028	-	CDS	1		
HVO_C0029	-	CDS	1	PIN domain containing protein	
HVO_C0030	-	CDS	1	Calcineurin-like phosphoesterase superfamily protein	
HVO_C0031	-	CDS	-1		
HVO_C0032	-	CDS	-1		
HVO_C0033	-	CDS	1	Uncharacterized conserved protein	
HVO_C0034	-	CDS	1	Uncharacterized conserved protein	
HVO_C0035	-	CDS	1	Uncharacterized conserved protein	
HVO_C0036	-	CDS	1	TraG/TraD/VirD4 family enzyme, ATPase	
HVO_C0037	-	CDS	1	Transposase	
HVO_C0038	-	CDS	-1		
HVO_C0039	-	CDS	1	Transposase	
HVO_C0040	-	CDS	1	Site-specific DNA methylase	
HVO_C0041	-	CDS	1	Uncharacterized conserved protein	
HVO_C0042	-	CDS	1	DNA or RNA helicase of superfamily II	
HVO_C0043	-	CDS	1	Uncharacterized conserved protein	
HVO_C0044	-	CDS	1	Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthase or related enzyme	
HVO_C0045	-	CDS	1	Small-conductance mechanosensitive channel	
HVO_C0046	-	CDS	-1	Predicted membrane-associated trancriptional regulator	
HVO_C0047	-	CDS	-1		
HVO_C0048	-	CDS	1		
HVO_C0049	-	CDS	1		
HVO_C0050	-	CDS	-1	Uncharacterized conserved protein	
HVO_C0051	-	CDS	1	Fe2+/Zn2+ uptake regulation protein, fur/PerR	
HVO_C0052	-	CDS	-1		
HVO_C0053	-	CDS	-1	Transposase	
HVO_C0054	-	CDS	-1		
HVO_C0054_A	-	CDS	1		
HVO_C0055	-	CDS	-1		
HVO_C0056	-	CDS	-1	Uncharacterized conserved protein	
HVO_C0057	*orc8*	CDS	1	Cdc6-related protein, AAA superfamily ATPase	
HVO_C0057s	*hts_467*	sRNA	1		Yes
HVO_C0057s2	*hts_468*	sRNA	1		Yes
HVO_C0057A					
HVO_C0058	-	CDS	-1		
HVO_C0059	-	CDS	-1		
HVO_C0060	-	CDS	1	Antitoxin (DNA-binding domain) of toxin-antitoxin stability system	
HVO_C0061	-	CDS	1		

HVO_C0062	-	CDS	-1	Uncharacterized conserved protein	
HVO_C0063	-	CDS	-1		
HVO_C0064	-	CDS	1	Transposase	
HVO_C0065	-	CDS	1		
HVO_C0066	*hts_469*	CDS	1		Yes
HVO_C0067	*soxA3*	CDS	-1	Glycine/D-amino acid oxidase (deaminating)	
HVO_C0068	-	CDS	1	Transposase, IS5 family	
HVO_C0069	*gfo1*	CDS	-1	Predicted dehydrogenase	
HVO_C0070	-	CDS	-1	Transposase	
HVO_C0071	-	CDS	-1	Transposase	
HVO_C0072	*dppC16*	CDS	-1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	
HVO_C0073	*dppB16*	CDS	-1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	
HVO_C0074	*dppDF16*	CDS	1	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	
HVO_C0075	*dppA16*	CDS	-1	ABC-type transport system, periplasmic component	
HVO_C0076	*arcR22*	CDS	-1	Transcriptional regulator (IclR family)	
HVO_C0077	*soxA2*	CDS	1	Glycine/D-amino acid oxidase (deaminating)	
HVO_C0078	-	CDS	-1		
HVO_C0079	-	CDS	-1		
HVO_C0080	-	CDS	-1	Transposase	
HVO_C0080_A	-	CDS	-1		
HVO_C0081	-	CDS	-1	Predicted membrane-associated trancriptional regulator	
HVO_C0082	-	CDS	-1		
HVO_C0082_A					
HVO_C0083	-	CDS	1		
HVO_C0083_A					
HVO_C0084	-	CDS	-1		
HVO_C0085	-	CDS	-1		
HVO_C0086	-	CDS	-1		
HVO_C0087	-	CDS	-1		
HVO_C0088	-	CDS	-1		
HVO_C0089	-	CDS	-1		

Note – All genes on pHV1 are non-essential since pHV1 can be lost in its entirety.



BLASTn search of pHV1 nucleotide sequence

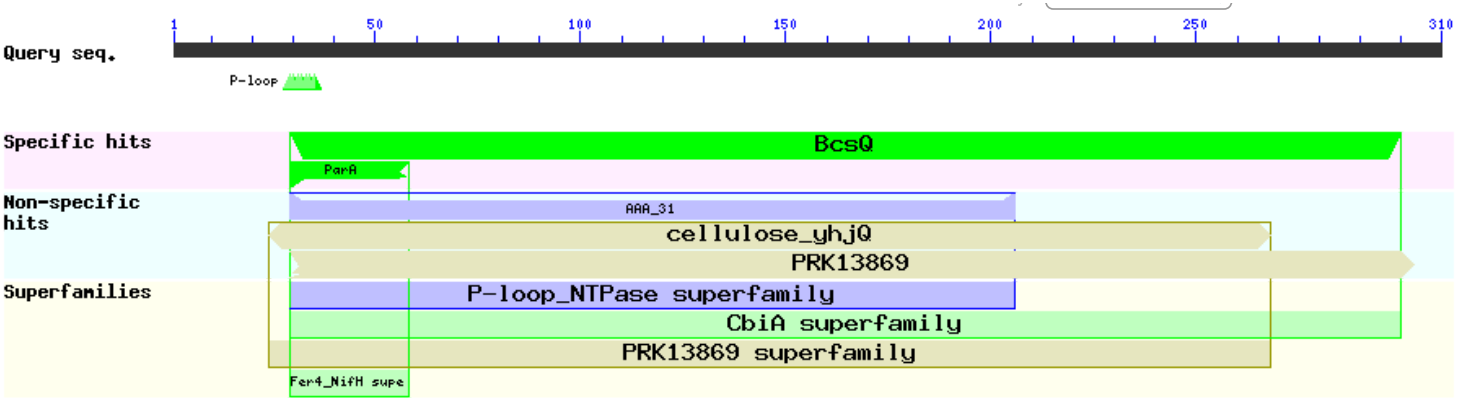


Description	Max score	Total score	Query cover	E value	Ident	Accession
Haloferax volcanii DS2 plasmid pHV1, complete sequence	1.571e+05	2.160e+05	100%	0.0	100%	CP001957.1
Haloferax gibbonsii strain ARA6 plasmid pHG4, complete sequence	3650	6755	5%	0.0	92%	CP011951.1
Haloferax volcanii DS2, complete genome	2604	2.061e+05	14%	0.0	99%	CP001956.1
Haloferax volcanii DS2 plasmid pHV4, complete sequence	2593	2.130e+05	15%	0.0	99%	CP001955.1
Haloferax volcanii DS2 plasmid pHV3, complete sequence	2457	25631	10%	0.0	98%	CP001953.1
H. volcanii insertion sequence-like repetitive element ISH51	2130	18872	8%	0.0	94%	X04389.1
Halophilic archaeon DL31 plasmid phalar01, complete sequence	2078	1.795e+05	9%	0.0	93%	CP002989.1
Salinigranum rubrum strain GX10 plasmid unnamed3, complete sequence	2036	14488	12%	0.0	93%	CP026312.1
Halobacterium salinarum R1 plasmid PHS2 complete genome	1925	25999	10%	0.0	91%	AM774417.1
Halobacterium salinarum R1 plasmid PHS1 complete genome	1925	25894	12%	0.0	91%	AM774416.1
Halobacterium sp. NRC-1, complete genome	1871	35714	10%	0.0	91%	AE004437.1
Halobacterium sp. NRC-1 plasmid pNRC200 complete genome	1871	78252	10%	0.0	91%	AE004438.1
Halobacterium sp. NRC-1 plasmid pNRC100, complete sequence	1871	62070	12%	0.0	91%	AF016485.1
Halobacterium salinarum R1 plasmid PHS3 complete genome	1869	37435	10%	0.0	91%	AM774418.1
Halobacterium halobium insertion element ISH27-1	1840	9037	8%	0.0	90%	X54432.1
Haloferax mediterranei ATCC 33500 plasmid pHM500, complete sequence	1805	2885	1%	0.0	98%	CP001871.1
Haloferax mediterranei ATCC 33500, complete genome	1801	6718	2%	0.0	98%	CP001868.2
Halorubrum lacusprofundi ATCC 49239 chromosome 2, complete sequence	1690	1.250e+05	10%	0.0	89%	CP001366.1
Halobacterium sp. DL1 plasmid, complete sequence	1504	49219	11%	0.0	86%	CP007061.1
Halophilic archaeon DL31, complete genome	1498	15685	11%	0.0	86%	CP002988.1
Haloarcula hispanica N601 plasmid pHH126, complete sequence	1493	15396	10%	0.0	86%	CP006886.1
Haloarcula hispanica ATCC 33960 chromosome II, complete sequence	1493	22402	10%	0.0	86%	CP002922.1
Halorhabdus tiamatea SARL4B plasmid pHTIA complete sequence	1483	13854	11%	0.0	91%	HF571521.1
Natrialbaceae archaeon JW/NM-HA 15, complete genome	1476	39805	10%	0.0	87%	CP019893.1
Halorubrum lacusprofundi ATCC 49239 chromosome 1, complete sequence	1439	22815	9%	0.0	85%	CP001365.1
Halorubrum lacusprofundi ATCC 49239 plasmid pHLAC01, complete sequence	1437	1.025e+05	9%	0.0	85%	CP001367.1
Halobacterium sp. DL1, complete genome	1435	36193	8%	0.0	85%	CP007060.1
Halostagnicola larsenii XH-48, complete genome	1435	7963	7%	0.0	87%	CP007055.1
Halophilic archaeon True-ADL chromosome, complete genome	1434	83826	10%	0.0	85%	CP024845.1

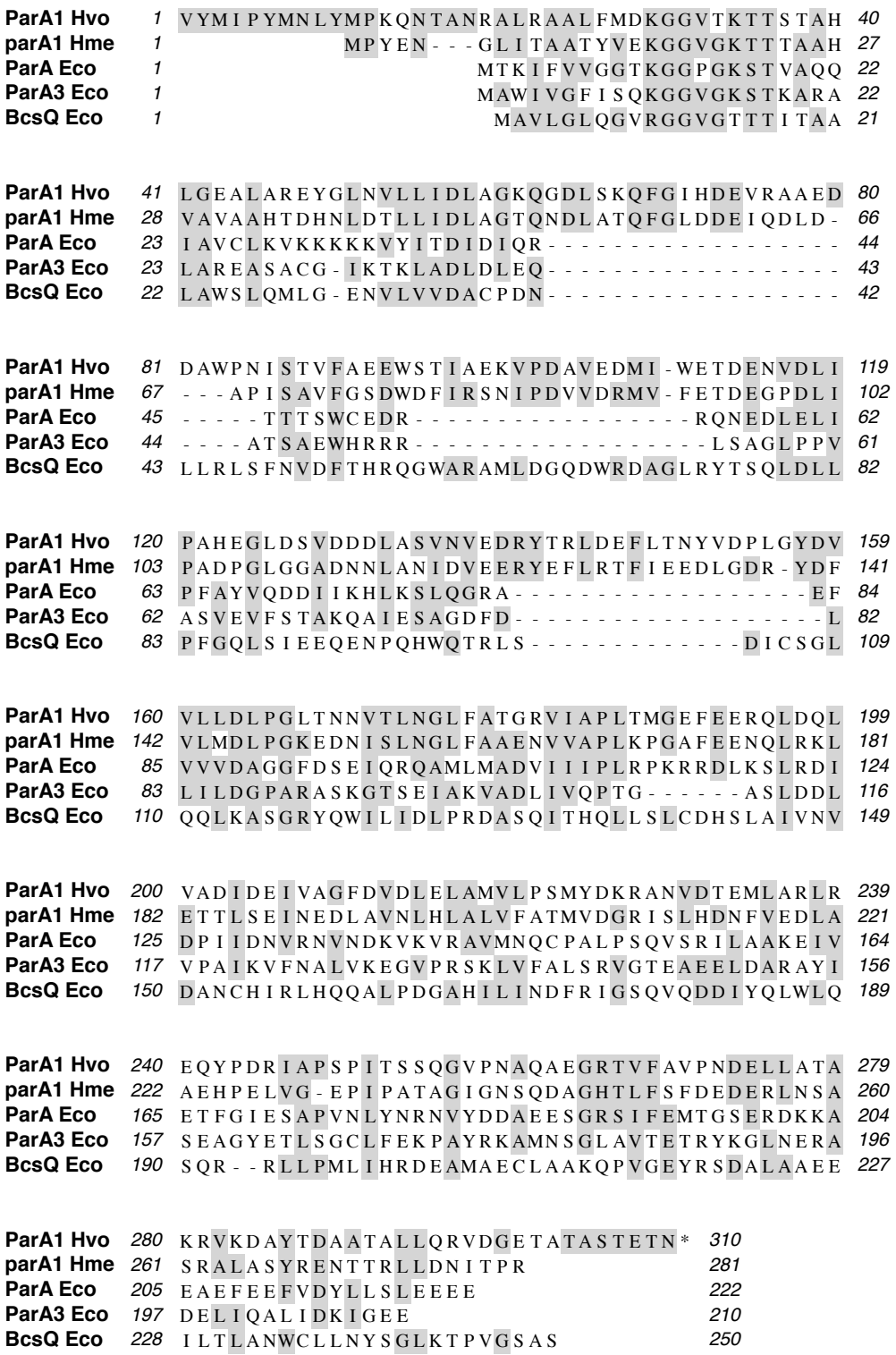
Description	Max score	Total score	Query cover	E value	Ident	Accession
Halorubrum lacusprofundi strain HLS1 plasmid pR1SE2, complete sequence	1434	30099	9%	0.0	85%	KX906370.1
Halophilic archaeon DL31 plasmid phalar02, complete sequence	1434	7237	8%	0.0	85%	CP002990.1
Halorubrum lacusprofundi strain R1S1 plasmid pR1SE, complete sequence	1432	7245	8%	0.0	85%	KX687704.1
Halomicrobium mukohataei DSM 12286, complete genome	1415	13866	10%	0.0	86%	CP001688.1
Natronococcus occultus SP4, complete genome	1339	2517	3%	0.0	88%	CP003929.1
Haloarcula hispanica N601 chromosome 2, complete sequence	1323	7060	7%	0.0	85%	CP006885.2
Natrinema pellirubrum DSM 15624, complete genome	1266	47767	13%	0.0	83%	CP003372.1
Halobacterium salinarum R1 complete genome	1245	20063	9%	0.0	84%	AM774415.1
Haloarcula marismortui ATCC 43049 chromosome I, complete sequence	1243	4583	4%	0.0	86%	AY596297.1
Haloarcula hispanica N601 chromosome 1, complete sequence	1234	6540	8%	0.0	84%	CP006884.1
Haloarcula hispanica ATCC 33960 chromosome I, complete sequence	1234	6540	8%	0.0	84%	CP002921.1
Halobacterium halobium insertion element ISH27-2	1218	6294	7%	0.0	84%	X54433.1
Halobacterium halobium insertion element ISH27-3	1201	6406	7%	0.0	83%	X54434.1
Salinigranum rubrum strain GX10 plasmid unnamed2, complete sequence	1199	2399	1%	0.0	90%	CP026311.1
Natrinema pellirubrum DSM 15624 plasmid pNATPE01, complete sequence	1199	14607	12%	0.0	84%	CP003373.1
Natronobacterium gregoryi SP2, complete genome	1182	31715	5%	0.0	90%	CP003377.1
Haloarcula marismortui ATCC 43049 plasmid pNG500, complete sequence	1166	9762	10%	0.0	85%	AY596294.1
Salinigranum rubrum strain GX10 plasmid unnamed1, complete sequence	1149	1612	2%	0.0	82%	CP026310.1
Halogeometricum borinquense DSM 11551 plasmid pHBOR02, complete sequence	1147	4484	0%	0.0	94%	CP001692.1
Halogeometricum borinquense DSM 11551 plasmid pHBOR04, complete sequence	1142	1142	0%	0.0	94%	CP001694.1
Haloferax mediterranei ATCC 33500 plasmid HMPLAS3, complete sequence	1125	4571	1%	0.0	94%	CP007552.1
Haloferax mediterranei ATCC 33500 plasmid pHM100, complete sequence	1125	3627	1%	0.0	94%	CP001869.1
Halobiforma lacisalsi AJ5 plasmid pHLAJ5I, complete sequence	1090	2170	1%	0.0	82%	CP019286.1
Haloferax gibbonsii strain ARA6 plasmid pHG1, complete sequence	1085	1261	1%	0.0	93%	CP011948.1
Halobiforma lacisalsi AJ5, complete genome	1079	11264	11%	0.0	81%	CP019285.1
Haloferax gibbonsii strain ARA6 plasmid pHG3, complete sequence	1068	1068	0%	0.0	93%	CP011950.1
Natrinema pellirubrum DSM 15624 plasmid pNATPE02, complete sequence	1053	1053	1%	0.0	81%	CP003374.1
Haloferax mediterranei ATCC 33500 genome	1038	3115	0%	0.0	92%	CP007551.1
Haloquadratum walsbyi DSM 16790 complete genome	968	2222	6%	0.0	85%	AM180088.1
Halorhabdus tiamatea SARL4B complete genome	957	3294	6%	0.0	85%	HF571520.1
Halogeometricum borinquense DSM 11551, complete genome	922	922	0%	0.0	94%	CP001690.1

Description	Max score	Total score	Query cover	E value	Ident	Accession
Natronomonas moolapensis 8.8.11 complete genome	920	3551	3%	0.0	83%	HF582854.1
Natronomonas pharaonis DSM 2160 plasmid PL131 complete genome	918	2816	3%	0.0	85%	CR936258.1
Haloarcula marismortui ATCC 43049 plasmid pNG400, complete sequence	911	1664	2%	0.0	83%	AY596293.1
Natrialba magadii ATCC 43099, complete genome	852	2730	3%	0.0	82%	CP001932.1
Halomicrobium mukohataei DSM 12286 plasmid pHmuk01, complete sequence	843	2455	3%	0.0	82%	CP001689.1
Halopenitus persicus plasmid pCBA1233 DNA, complete genome, strain: CBA1233	817	3906	3%	0.0	81%	AP017559.1
Haloarcula sp. CBA1115, complete genome	797	6926	3%	0.0	81%	CP010529.1
Haloterrigena turkmenica DSM 5511 plasmid pHTUR03, complete sequence	791	5790	6%	0.0	84%	CP001863.1
Natrialba magadii ATCC 43099 plasmid pNMAG02, complete sequence	741	1956	1%	0.0	83%	CP001934.1
Haloferax mediterranei ATCC 33500 plasmid HMPLAS2, complete sequence	671	2568	1%	0.0	90%	CP007553.1
Haloferax mediterranei ATCC 33500 plasmid pHM300, complete sequence	671	2568	1%	0.0	90%	CP001870.1

Note – only sequences with a E value of 0.0 are shown



Domain search of *Haloferax volcanii* ParA1 (gene number HVO_C0017). A conserved ParA domain was found at the N-terminus, with the majority of the protein having a conservation with BcsQ (cellulose biosynthesis protein involved in chromosome partitioning. Search was conducted via NCBI Conserved Domain (CD) search using translated protein sequence (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi). Diagram shows standard results.


ParA1 Hvo 1 VYMIPYMNLYMPKQNTANRALRAALFMDKGGVTKTTSTAH 40
parA1 Hme 1 MPYEN---GLITAATYVEKGGVGKTTTAAH 27
ParA Eco 1 MTKIFVVGGTKGGPGKSTVAQQ 22
ParA3 Eco 1 MAWIVGFISQKGGVGKSTKARA 22
BcsQ Eco 1 MAVLGLQGVRGGVGTTTITAA 21
ParA1 Hvo 41 LGEALAREYGLNVLLIDLAGKQGDLSKQFGIHDEVRAAED 80
parA1 Hme 28 VAVAAHTDHNLDTLLIDLAGTQNDLATQFGLDDEIQDLD- 66
ParA Eco 23 IAVCLKVKKKKKVYITDIDIQR---------- 44
ParA3 Eco 23 LAREASACG-IKTKLADLDLEQ---------- 43
BcsQ Eco 22 LAWSLQMLG-ENVLVVDACPDN---------- 42
ParA1 Hvo 81 DAWPNISTVFAEEWSTIAEKVPDAVEDMI-WETDENVDLI 119
parA1 Hme 67 ---APISAVFGSDWDFIRSNIPDVVDRMV-FETDEGPDLI 102
ParA Eco 45 -----TTTSWCEDR---------- 62
ParA3 Eco 44 ----ATSAEWHRRR---------- 61
BcsQ Eco 43 LLRLSFNVDFTHRQGWARAMLDGQDWRDAGLRYTSQLDLL 82
ParA1 Hvo 120 PAHEGLDSVDDDLASVNVEDRYTRLDEFLTNYVDPLGYDV 159
parA1 Hme 103 PADPGLGGADNNLANIDVEERYEFLRTFIEEDLGDR-YDF 141
ParA Eco 63 PFAYVQDDIIKHLKSLQGRA---------- 84
ParA3 Eco 62 ASVEVFSTAKQAIESAGDFD---------- 82
BcsQ Eco 83 PFGQLSIEEQENPQHWQTRLS---------- 109
ParA1 Hvo 160 VLLDLPGLTNNVTLNGLFATGRVIAPLTMGEFEERQLDQL 199
parA1 Hme 142 VLMDLPGKEDNISLNGLFAAENVVAPLKPGAFEENQLRKL 181
ParA Eco 85 VVVDAGGFDSEIQRQAMLMADVIIIPLRPKRRDLKSLRDI 124
ParA3 Eco 83 LILDGPARASKGTSEIAKVADLIVQPTG------ASLDDL 116
BcsQ Eco 110 QQLKASGRYQWILIDLPRDASQITHQLLSLCDHSLAIVNV 149
ParA1 Hvo 200 VADIDEIVAGFDVDLELAMVLPSMYDKRANVDTEMLARLR 239
parA1 Hme 182 ETTLSEINEDLAVNLHLALVFATMVDGRISLHDNFVEDLA 221
ParA Eco 125 DPIIDNVRNVNDKVKVRAVMNQCPALPSQVSRILAAKEIV 164
ParA3 Eco 117 VPAIKVFNALVKEGVPRSKLVFALSRVGTEAEELDARAYI 156
BcsQ Eco 150 DANCHIRLHQQALPDGAHILINDFRIGSQVQDDIYQLWLQ 189
ParA1 Hvo 240 EQYPDRIAPSPITSSQGVPNAQAEGRTVFAVPNDELLATA 279
parA1 Hme 222 AEHPELVG-EPIPATAGIGNSQDAGHTLFSFDEDERLNSA 260
ParA Eco 165 ETFGIESAPVNLYNRNVYDDAEESGRSIFEMTGSERDKKA 204
ParA3 Eco 157 SEAGYETLSGCLFEKPAYRKAMNSGLAVTETRYKGLNERA 196
BcsQ Eco 190 SQR--RLLPMLIHRDEAMAECLAAKQPVGEYRSDALAAEE 227
ParA1 Hvo 280 KRVKDAYTDAATALLQRVDGETATASTETN* 310
parA1 Hme 261 SRALASYRENTTRLLDNITPR 281
ParA Eco 205 EAEFEEFVDYLLSLEEEE 222
ParA3 Eco 197 DELIQALIDKIGEE 210
BcsQ Eco 228 ILTLANWCLLNYSGLKTPVGSAS 250

IV pHV3

Gene number	Gene name	Gene type	Strand	arCOG annotation	Deleted
HVO_B0001	*orc6*	CDS	1	Cdc6-related protein, AAA superfamily ATPase	No*
HVO_B0001ori	*ori-pHV3*	ORI	1		No*
HVO_B0002	-	CDS	1	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase or related deacylase	
HVO_B0003	-	CDS	-1	Uncharacterized protein, possibly involved in utilization of glycolate and propanediol	
HVO_B0004	*glyA2*	CDS	-1	Glycine/serine hydroxymethyltransferase	
HVO_B0005	*ilvA2*	CDS	1	Threonine dehydratase	
HVO_B0006	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0007	-	CDS	1	Predicted transcriptional regulator, C-terminal HTH-like domain	
HVO_B0008	-	CDS	-1	Predicted ATPase involved in replication control, Cdc46/Mcm family	
HVO_B0009	-	CDS	-1	Phenylpropionate dioxygenase or related ring-hydroxylating dioxygenase, large terminal subunit	
HVO_B0010	-	CDS	1		
HVO_B0011	*soxA1*	CDS	1	Glycine/D-amino acid oxidase (deaminating)	
HVO_B0012	*betI2*	CDS	1	Choline-glycine betaine transporter	
HVO_B0013	-	CDS	1	Uncharacterized conserved protein	
HVO_B0014	*gcvT1*	CDS	1	Glycine cleavage system T protein (aminomethyltransferase)	
HVO_B0015	-	CDS	-1	PIN domain containing protein	
HVO_B0016	-	CDS	-1	Antitoxin MazE superfamily protein	
HVO_B0017	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0018	*soj*	CDS	-1	ATPase involved in chromosome partitioning, ParA family	
HVO_B0019	-	CDS	1	Na+/H+ antiporter NhaD or related arsenite permease	
HVO_B0020	-	CDS	1	Signal transduction regulator, containsPAS and HTH domains	
HVO_B0021	*dppA10*	CDS	1	ABC-type transport system, periplasmic component	
HVO_B0022	*dppB10*	CDS	1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	
HVO_B0023	*dppC10*	CDS	1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	
HVO_B0024	*dppDF10*	CDS	1	ABC-type oligopeptide transport system, ATPase component	
HVO_B0025	-	CDS	-1	Threonine dehydrogenase or related Zn-dependent dehydrogenase	
HVO_B0026	-	CDS	1	Short-chain alcohol dehydrogenase	
HVO_B0027	-	CDS	1	Fumarylacetoacetate (FAA) hydrolase family protein	
HVO_B0028	-	CDS	1	Predicted dehydrogenase	
HVO_B0029	-	CDS	1	Predicted dehydrogenase	
HVO_B0030	-	CDS	-1	Gluconolactonase	
HVO_B0031	-	CDS	-1	Short-chain alcohol dehydrogenase	
HVO_B0032	-	CDS	1	Nucleoside-diphosphate-sugar epimerase	
HVO_B0033	-	CDS	-1		
HVO_B0034	*tsgA7*	CDS	1	ABC-type sugar transport system, periplasmic component	
HVO_B0035	*tsgB7*	CDS	1	ABC-type sugar transport system, permease component	
HVO_B0036	*tsgC7*	CDS	1	ABC-type sugar transport system, permease component	
HVO_B0037	*tsgE7*	CDS	1	ABC-type spermidine/putrescine transport system, ATPase component	Yes
HVO_B0038	*tsgD7*	CDS	1	ABC-type spermidine/putrescine transport system, ATPase component	
HVO_B0038A	-	CDS	-1	L-alanine-DL-glutamate epimerase or related enzyme of enolase superfamily	
HVO_B0039	-	CDS	1	Lactaldehyde dehydrogenase, Succinate semialdehyde dehydrogenase or other NAD-dependent aldehyde dehydrogenase	
HVO_B0040	*arcR14*	CDS	-1	Transcriptional regulator (IclR family)	Yes
HVO_B0041	*iucC*	CDS	-1	Siderophore synthetase component	
HVO_B0042	*hxyA*	CDS	-1	Lysine/ornithine N-monooxygenase	
HVO_B0043	*iucB*	CDS	-1	Acetyltransferase, RimL family	
HVO_B0044	*iucA*	CDS	-1	Siderophore synthetase component	
HVO_B0045	*bdb*	CDS	-1	L-tyrosine decarboxylase, PLP-dependent protein	
HVO_B0046	*dat*	CDS	-1	4-aminobutyrate aminotransferase or related aminotransferase	
HVO_B0047	-	CDS	-1	ABC-type Fe3+-hydroxamate transport system, periplasmic component	
HVO_B0048	*cbiE*	CDS	-1	Precorrin-6B methylase 1	
HVO_B0049	*cbiC*	CDS	-1	Precorrin isomerase	
HVO_B0050	*cobN*	CDS	-1	Cobalamin biosynthesis protein CobN or related Mg-chelatase	
HVO_B0051	*hmcA*	CDS	1	AAA family ATPase	
HVO_B0052	-	CDS	-1	WD40-like repeat	
HVO_B0053	-	CDS	-1	Uncharacterized conserved protein in cobalamin operon	
HVO_B0054	*cbiK*	CDS	-1	Sirohydrochlorin iron chelatase fused to [2Fe-2S] Ferredoxin	
HVO_B0055	-	CDS	-1	Uncharacterized conserved protein in cobalamin operon	
HVO_B0056	-	CDS	-1	Ferredoxin	
HVO_B0057	*cobJ*	CDS	-1	Precorrin-3B methylase	
HVO_B0058	*cobH*	CDS	-1	Precorrin-3B methylase	
HVO_B0059	-	CDS	-1	Cobalamin biosynthesis protein CbiG	
HVO_B0060	*cbiF*	CDS	-1	Precorrin-4 methylase	
HVO_B0061	*cbiL*	CDS	-1	Precorrin-2 methylase	
HVO_B0062	*cbiT*	CDS	-1	Precorrin-6B methylase 2	
HVO_B0063	-	CDS	1	Probable cobalt transporter subunit, CbtB	
HVO_B0064	-	CDS	1	Probable cobalt transporter subunit, CbtA	
HVO_B0065	-	CDS	1	Sirohydrochlorin iron chelatase fused to [2Fe-2S] Ferredoxin	
HVO_B0066	*trh10*	CDS	1	Transcriptional regulator, IclR family	
HVO_B0067	-	CDS	-1	Amino acid transporter	
HVO_B0068	-	CDS	-1		
HVO_B0069	*gabD*	CDS	-1	Lactaldehyde dehydrogenase, Succinate semialdehyde dehydrogenase or other NAD-dependent aldehyde dehydrogenase	
HVO_B0070	*gabT1*	CDS	1	4-aminobutyrate aminotransferase or related aminotransferase	
HVO_B0071	*adh2*	CDS	-1	Threonine dehydrogenase or related Zn-dependent dehydrogenase	Yes
HVO_B0072	-	CDS	-1	Predicted transcriptional regulator, C-terminal HTH-like domain	
HVO_B0073	-	CDS	-1	Transcriptional regulator (Lrp/AsnC family)	
HVO_B0074	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0075	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0076	*hyuA4*	CDS	-1	N-methylhydantoinase A/acetone carboxylase, beta subunit	
HVO_B0077	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0078	*dppF11*	CDS	-1	ABC-type oligopeptide transport system, ATPase component	
HVO_B0079	*dppD11*	CDS	-1	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	
HVO_B0080	*dppC11*	CDS	-1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	
HVO_B0081	*dppB11*	CDS	-1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	
HVO_B0082	*dppA11*	CDS	-1	ABC-type transport system, periplasmic component	
HVO_B0083	*gnl*	CDS	1	Gluconolactonase	
HVO_B0084	-	CDS	-1	Uncharacterized protein, possibly involved in utilization of glycolate and propanediol	
HVO_B0085	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0086	-	CDS	1	Arylsulfatase A or related enzyme	
HVO_B0087	-	CDS	1	L-alanine-DL-glutamate epimerase or related enzyme of enolase superfamily	

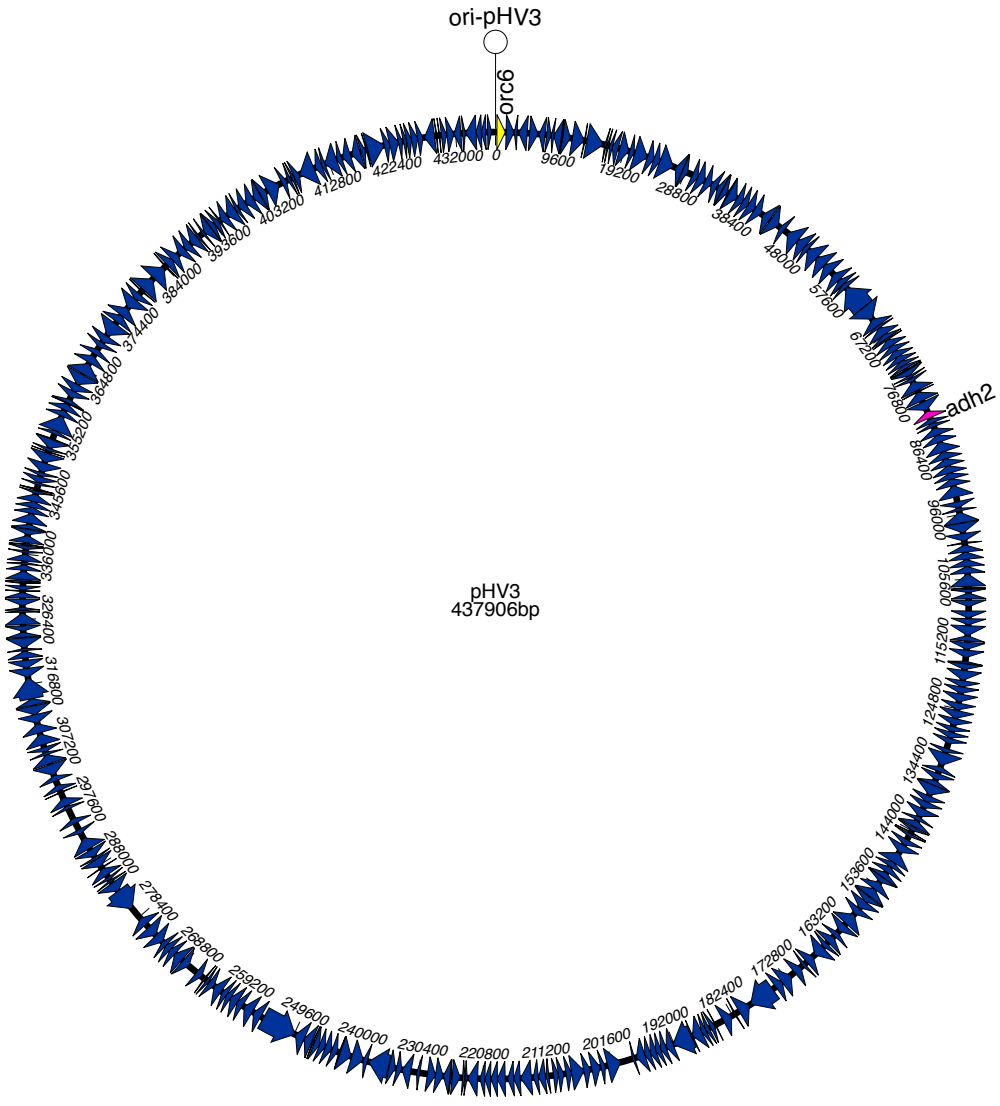
HVO_B0088	-	CDS		-1		
HVO_B0089	*dppF12*	CDS		-1	ABC-type oligopeptide transport system, ATPase component	
HVO_B0090	*dppD12*	CDS		-1	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	
HVO_B0091	*dppC12*	CDS		-1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	
HVO_B0092	*dppB12*	CDS		-1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	
HVO_B0093	*dppA12*	CDS		-1	ABC-type transport system, periplasmic component	
HVO_B0094	-	CDS		-1	Uncharacterized conserved protein	
HVO_B0095	*uspA32*	CDS		1	Nucleotide-binding protein, UspA family	
HVO_B0096	-	CDS		-1	Nucleoside-diphosphate-sugar epimerase	
HVO_B0097	-	CDS		1	Dihydrodipicolinate synthase/N-acetylneuraminate lyase	
HVO_B0098	-	CDS		1	Uncharacterized conserved protein	
HVO_B0099	*mdcM*	CDS		1	Predicted permease	
HVO_B0100	*aldY3*	CDS		1	Lactaldehyde dehydrogenase, Succinate semialdehyde dehydrogenase or other NAD-dependent aldehyde dehydrogenase	Yes
HVO_B0101	*arcR15*	CDS		-1	Transcriptional regulator (IclR family)	
HVO_B0102	-	CDS		1	Uncharacterized conserved protein	
HVO_B0103	-	CDS		1	Predicted dehydrogenase	
HVO_B0104	-	CDS		-1	Uncharacterized conserved protein	
HVO_B0105	-	CDS		1	Predicted metal-dependent hydrolase of the TIM-barrel fold	
HVO_B0106	*tsgA8*	CDS		1	ABC-type sugar transport system, periplasmic component	
HVO_B0107	*tsgB8*	CDS		1	ABC-type sugar transport system, permease component	
HVO_B0108	*tsgC8*	CDS		1	ABC-type sugar transport system, permease component	
HVO_B0109	*tsgD8*	CDS		1	ABC-type spermidine/putrescine transport system, ATPase component	
HVO_B0110	-	CDS		1	Threonine dehydrogenase or related Zn-dependent dehydrogenase	
HVO_B0111	-	CDS		1	L-alanine-DL-glutamate epimerase or related enzyme of enolase superfamily	
HVO_B0112	-	CDS		1		
HVO_B0113	-	CDS		1	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase or related flavin-dependent oxidoreductase	Yes
HVO_B0114	*arcR16*	CDS		1	Transcriptional regulator (IclR family)	
HVO_B0115	-	CDS		1	2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1,7-dioic acid hydratase (catechol pathway)	
HVO_B0116	-	CDS		1	Short-chain alcohol dehydrogenase	
HVO_B0116s	*sRNA288*	sRNA		1		Yes
HVO_B0117	-	CDS		1	Uncharacterized conserved protein	
HVO_B0118	*sph2*	CDS		1	DNA sulfur modification protein DndD, ATPase	
HVO_B0119	*arcR17*	CDS		-1	Transcriptional regulator (IclR family)	
HVO_B0120	-	CDS		1	Metal-dependent amidase/aminoacylase/carboxypeptidase	
HVO_B0121	-	CDS		-1	Fe2+ dependent formamide hydrolase, cofactor Fo biosynthesis enzyme	
HVO_B0122	*dppA13*	CDS		1	ABC-type transport system, periplasmic component	
HVO_B0123	*dppB13*	CDS		1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	
HVO_B0124	*dppC13*	CDS		1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	
HVO_B0125	*dppD13*	CDS		1	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	
HVO_B0126	-	CDS		-1	Predicted xylanase/chitin deacetylase	
HVO_B0127	-	CDS		-1	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase or related flavin-dependent oxidoreductase	
HVO_B0128	-	CDS		1	Short-chain alcohol dehydrogenase	
HVO_B0129	-	CDS		-1		
HVO_B0130	-	CDS		1	Predicted DNA-binding protein with PD1-like DNA-binding motif	
HVO_B0131	-	CDS		1	Predicted CRISPR related protein, COG1517 family	
HVO_B0132	-	CDS		1	WD40-like repeat	
HVO_B0133	-	CDS		-1	Uncharacterized conserved protein	
HVO_B0134	-	CDS		-1	ABC-type multidrug transport system, ATPase component	
HVO_B0135	-	CDS		-1	Uncharacterized conserved membrane protein	
HVO_B0136	*pcbD2*	CDS		-1	Alpha/beta superfamily hydrolase	
HVO_B0137	-	CDS		1	Predicted transcriptional regulator, C-terminal HTH-like domain	
HVO_B0138	-	CDS		-1	WD40-like repeat	
HVO_B0139	-	CDS		-1	WD40-like repeat	
HVO_B0140	-	CDS		1	WD40-like repeat	
HVO_B0141	-	CDS		1	WD40-like repeat	
HVO_B0142	-	CDS		1	HEAT repeats containing protein	
HVO_B0143	*htlD4*	CDS		-1	Signal transduction histidine kinase, contains REC and PAS domains	
HVO_B0144	-	CDS		1	ABC-type Fe3+-hydroxamate transport system, periplasmic component	
HVO_B0145	-	CDS		-1	Fe2+ dependent formamide hydrolase, cofactor Fo biosynthesis enzyme	
HVO_B0146	-	CDS		-1	Selenocysteine lyase/Cysteine desulfurase	
HVO_B0147	-	CDS		1		
HVO_B0148	-	CDS		1	Predicted transcriptional regulator, C-terminal HTH-like domain	
HVO_B0149	*orf3*	CDS		1	Phytoene dehydrogenase or related enzyme	
HVO_B0150	-	CDS		-1	ABC-type Fe3+-hydroxamate transport system, periplasmic component	
HVO_B0150A						
HVO_B0151	-	CDS		1	Transcriptional regulator, CopG/Arc/MetJ family (DNA-binding and a metal-binding domains)	
HVO_B0152	-	CDS		-1	Transposase	
HVO_B0152_A	-	CDS		-1		
HVO_B0153	*capB*	CDS		-1	UDP-N-acetylmuramyl tripeptide synthase and Folylpolyglutamate synthase	
HVO_B0153_A	-	CDS		-1	Uncharacterized conserved protein	
HVO_B0153B						
HVO_B0154	-	CDS		1	Signal transduction histidine kinase	
HVO_B0155	-	CDS		-1	Transposase	
HVO_B0156	-	CDS		1	Uncharacterized conserved protein	
HVO_B0157	-	CDS		-1	Uncharacterized conserved protein	
HVO_B0158	-	CDS		-1	Uncharacterized conserved protein	
HVO_B0159	*boa9*	CDS		-1	Predicted transcriptional regulator, C-terminal HTH-like domain	
HVO_B0160	-	CDS		1	Uncharacterized conserved protein	
HVO_B0161	*narB*	CDS		1	Rieske Fe-S protein	
HVO_B0162	*qcrB*	CDS		1	Cytochrome b subunit of the bc complex	
HVO_B0163	-	CDS		1	Uncharacterized conserved protein	
HVO_B0164	*narG*	CDS		1	Nitrate reductase alpha subunit	
HVO_B0165	*narH*	CDS		1	Nitrate reductase beta subunit	
HVO_B0166	-	CDS		1	Ethylbenzene dehydrogenase	
HVO_B0167	*narJ*	CDS		1	Uncharacterized component of anaerobic dehydrogenase	
HVO_B0168	-	CDS		1	Uncharacterized conserved protein	
HVO_B0169	*prp*	CDS		1	HEAT repeats containing protein	
HVO_B0170	*mrp2*	CDS		1	Mrp family protein, ATPase, contains iron-sulfur cluster	
HVO_B0171	-	CDS		1	Uncharacterized conserved protein	
HVO_B0172	*phoT3*	CDS		-1		
HVO_B0173	*sph4*	CDS		-1	DNA sulfur modification protein DndD, ATPase	

HVO_B0174	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0175	*kinA6*	CDS	-1	Signal transduction histidine kinase	
HVO_B0176	*cbiQ3*	CDS	-1	ABC-type cobalt transport system, permease component CbiQ or related transporter	
HVO_B0177	*cbiO3*	CDS	-1	ABC-type cobalt transport system, ATPase component	
HVO_B0178	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0179	-	CDS	-1	Sugar phosphate isomerase/epimerase	
HVO_B0180	-	CDS	-1	Phosphosulfolactate phosphohydrolase or related enzyme	
HVO_B0181	*fjo21*	CDS	1	Predicted membrane protein	
HVO_B0182	-	CDS	1	Deoxyhypusine synthase	
HVO_B0183	*racD*	CDS	1	Aspartate racemase	
HVO_B0184	*dppA14*	CDS	1	ABC-type transport system, periplasmic component	
HVO_B0185	*dppB14*	CDS	1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	
HVO_B0186	*dppC14*	CDS	1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	
HVO_B0187	-	CDS	1	ABC-type Fe3+ transport system, periplasmic component	
HVO_B0188	-	CDS	1	ABC-type sugar transport system, permease component	
HVO_B0189	-	CDS	1	ABC-type Fe3+ transport system, permease component	
HVO_B0190	-	CDS	1	ABC-type spermidine/putrescine transport system, ATPase component	
HVO_B0190A					
HVO_B0191	-	CDS	1	Uncharacterized conserved protein	
HVO_B0192	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0193	-	CDS	-1	Transcriptional regulator containing HTH domain, ArsR family	
HVO_B0194	-	CDS	-1	Outer membrane lipoprotein-sorting protein	
HVO_B0195	-	CDS	-1		
HVO_B0196	*kinA8*	CDS	1	Signal transduction histidine kinase, contains PAS domain	
HVO_B0197	-	CDS	-1	ABC-type Fe3+-siderophore transport system, permease component	
HVO_B0198	-	CDS	-1	ABC-type Fe3+-hydroxamate transport system, periplasmic component	
HVO_B0199	-	CDS	1	Cytosine/adenosine deaminase	
HVO_B0200	*aceB2*	CDS	1	Beta-methylmalyl-CoA lyase (Citrate lyase beta subunit family)	
HVO_B0201	*arcR18*	CDS	-1	Transcriptional regulator (IclR family)	
HVO_B0202	-	CDS	-1	Signal transduction histidine kinase with PAS domain	
HVO_B0203	-	CDS	1	FAD/FMN-containing dehydrogenase	
HVO_B0204	-	CDS	1	Membrane protein DegA family	
HVO_B0205	*lccA*	CDS	-1	Multicopper oxidase	Yes
HVO_B0206	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0207	-	CDS	1		
HVO_B0208	-	CDS	-1	Sugar phosphate isomerase/epimerase	
HVO_B0209	-	CDS	-1	4-hydroxybenzoate polyprenyltransferase or related prenyltransferase	
HVO_B0210	-	CDS	-1	Sugar phosphate isomerase/epimerase	
HVO_B0211	-	CDS	-1	Predicted metal-dependent hydrolase with the TIM-barrel fold	
HVO_B0212	*hybF*	CDS	1		
HVO_B0213	-	CDS	1	Myo-inositol-1-phosphate synthase	
HVO_B0214	*idi*	CDS	1	Phosphodiesterase of AP superfamily	
HVO_B0215	-	CDS	-1	HEAT repeats containing protein	
HVO_B0216	-	CDS	1	Uncharacterized conserved protein	
HVO_B0217	*livJ6*	CDS	1	ABC-type branched-chain amino acid transport system, periplasmic component	
HVO_B0218	*livH6*	CDS	1	Branched-chain amino acid ABC-type transport system, permease component	
HVO_B0219	*livM6*	CDS	1	ABC-type branched-chain amino acid transport system, permease component	
HVO_B0220	*livG6*	CDS	1	ABC-type branched-chain amino acid transport system, ATPase component	
HVO_B0221	*livF6*	CDS	1	ABC-type branched-chain amino acid transport system, ATPase component	
HVO_B0222	-	CDS	1	WD40-like repeat	
HVO_B0223	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0224	*yoaN*	CDS	-1	Protein containing cupin domain	
HVO_B0225	-	CDS	1		
HVO_B0226	-	CDS	-1	Predicted permease	
HVO_B0227	*tsgD9*	CDS	-1	ABC-type spermidine/putrescine transport system, ATPase component	
HVO_B0228	*tsgA9*	CDS	1	ABC-type sugar transport system, periplasmic component	
HVO_B0229	*tsgB9*	CDS	1	ABC-type sugar transport system, permease component	
HVO_B0230	*tsgC9*	CDS	1	ABC-type sugar transport system, permease component	
HVO_B0231	-	CDS	1	Uncharacterized conserved protein	
HVO_B0232	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0233	*xsa*	CDS	-1	Uncharacterized conserved protein	
HVO_B0234	-	CDS	1	Transposase	
HVO_B0234s	*hts_494*	sRNA	-1		Yes
HVO_B0234A					
HVO_B0234B					
HVO_B0235	*fdhG*	CDS	-1	Uncharacterized anaerobic dehydrogenase	
HVO_B0236	*nuoF*	CDS	-1	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	
HVO_B0237	-	CDS	1	Uncharacterized conserved protein	
HVO_B0238	-	CDS	-1	Putative translation initiation inhibitor, yjgF family	
HVO_B0239	*fdh*	CDS	1	Threonine dehydrogenase or related Zn-dependent dehydrogenase	
HVO_B0240	-	CDS	1	Uncharacterized conserved protein	
HVO_B0241	-	CDS	-1	Transposase	
HVO_B0242	-	CDS	-1	Transposase	
HVO_B0243	*cobW2*	CDS	-1	GTPase (G3E type)	
HVO_B0244	*aor1*	CDS	1	Aldehyde:ferredoxin oxidoreductase	
HVO_B0245	*mobA1*	CDS	-1	Molybdopterin-guanine dinucleotide biosynthesis protein A	
HVO_B0246	-	CDS	1		
HVO_B0247	-	CDS	1		
HVO_B0248	-	CDS	1	Short-chain alcohol dehydrogenase	
HVO_B0249	-	CDS	-1		
HVO_B0250	*cbg1*	CDS	1	Bifunctional b-D-xylosidase/a-L-arabinosidase or related glycosidase	
HVO_B0251	-	CDS	1	Permease of the major facilitator superfamily	
HVO_B0252	-	CDS	1		
HVO_B0253	*suhB3*	CDS	1	Archaeal fructose-1,6-bisphosphatase or related enzyme of inositol monophosphatase family	
HVO_B0255	*arcR19*	CDS	1	Transcriptional regulator (IclR family)	
HVO_B0256	*aor3*	CDS	-1	Aldehyde:ferredoxin oxidoreductase	
HVO_B0257	*gabT5*	CDS	1	4-aminobutyrate aminotransferase or related aminotransferase	
HVO_B0258	-	CDS	1	Putative translation initiation inhibitor, yjgF family	
HVO_B0259	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0260	*betL1*	CDS	1	Choline-glycine betaine transporter	
HVO_B0261	*gabT4*	CDS	1	4-aminobutyrate aminotransferase or related aminotransferase	

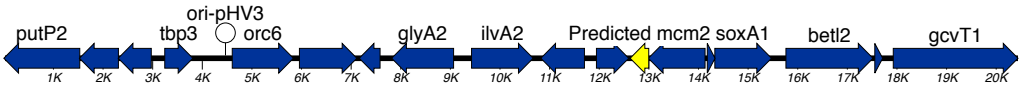
HVO_B0262	*ilvA3*	CDS	-1	Threonine dehydratase	
HVO_B0263	-	CDS	-1	Metal-dependent amidase/aminoacylase/carboxypeptidase	
HVO_B0264	*serA4*	CDS	1	Phosphoglycerate dehydrogenase or related dehydrogenase	
HVO_B0265	-	CDS	1	FAD/FMN-containing dehydrogenase	
HVO_B0266	*gdhB*	CDS	-1	Glutamate dehydrogenase/leucine dehydrogenase	
HVO_B0267	-	CDS	-1	Xaa-Pro aminopeptidase	
HVO_B0268	-	CDS	1	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase or related flavin-dependent oxidoreductase	
HVO_B0269	-	CDS	1	Uncharacterized conserved protein	
HVO_B0270	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0271	-	CDS	1		
HVO_B0271_A					
HVO_B0272	*boa10*	CDS	-1	Rec domain	
HVO_B0273	*kinA7*	CDS	1	Signal transduction histidine kinase, contains PAS domain	
HVO_B0274	*cdsA*	CDS	1	Uncharacterized conserved protein	
HVO_B0275	-	CDS	1	Uncharacterized conserved protein	
HVO_B0275s	*sRNA30*	sRNA	1		Yes
HVO_B0276	-	CDS	-1	Permease of the drug/metabolite transporter (DMT) superfamily	
HVO_B0277	-	CDS	1	Predicted transcriptional regulator	
HVO_B0278	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0279	-	CDS	-1	SAM-dependent methyltransferase	
HVO_B0280	-	CDS	-1		
HVO_B0281	-	CDS	1		
HVO_B0282	-	CDS	1	Predicted regulatory protein containing HTH domain	
HVO_B0283	-	CDS	1	Predicted transcriptional regulator, C-terminal HTH-like domain	
HVO_B0284	-	CDS	1	Translation initiation factor 1 (eIF-1/SUI1)	Yes
HVO_B0285	*tfb3*	CDS	1	Transcription initiation factor TFIIB, Brf1 subunit/Transcription initiation factor TFIIB	Yes
HVO_B0286	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0287	-	CDS	-1	Predicted RNA-binding protein, contains TRAM domain	
HVO_B0288	-	CDS	1		
HVO_B0289	*bioB*	CDS	1	Biotin synthase or related enzyme	
HVO_B0290	*suhB4*	CDS	-1	Archaeal fructose-1,6-bisphosphatase or related enzyme of inositol monophosphatase family	
HVO_B0291	-	CDS	1	Glycerophosphoryl diester phosphodiesterase	
HVO_B0292	*ugpB*	CDS	1	ABC-type trehalose transport system, periplasmic component	
HVO_B0293	*ugpA*	CDS	1	ABC-type sugar transport system, permease component	
HVO_B0294	*ugpE*	CDS	1	ABC-type sugar transport system, permease component	
HVO_B0295	*ugpC*	CDS	1	ABC-type spermidine/putrescine transport system, ATPase component	
HVO_B0296	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0297	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0298	-	CDS	1		
HVO_B0299	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0300	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0301	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0302	*hyuH*	CDS	-1	Dihydroorotase or related cyclic amidohydrolase	
HVO_B0303	*pbuX*	CDS	-1	Xanthine/uracil permease	
HVO_B0304	-	CDS	-1		
HVO_B0305	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0306	*amaB4*	CDS	-1	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase or related deacylase	
HVO_B0307	-	CDS	1		
HVO_B0308	-	CDS	1	Aerobic-type carbon monoxide dehydrogenase, small subunit CoxS/CutS homolog	
HVO_B0309	*xdhD2*	CDS	1	Aerobic-type carbon monoxide dehydrogenase, large subunit CoxL/CutL homolog	
HVO_B0310	*coxM, cutM*	CDS	1	Aerobic-type carbon monoxide dehydrogenase, middle subunit CoxM/CutM homolog	
HVO_B0311	-	CDS	1	Xanthine and CO dehydrogenase maturation factor, XdhC/CoxF family	
HVO_B0312	-	CDS	1		
HVO_B0313	*glmU*	CDS	1	GT-A family glycosyltransferase involved in molybdopterin guanine dinucleotide biosynthesis	
HVO_B0314	*tsgC13*	CDS	-1	Uncharacterized ABC-type transport system, permease component	
HVO_B0315	*tsgB13*	CDS	-1	ABC-type uncharacterized transport system, permease component	
HVO_B0316	*tsgD13*	CDS	-1	ABC-type uncharacterized transport system, ATPase component	
HVO_B0317	*ade*	CDS	1	Adenine deaminase	
HVO_B0318	*tsgA13*	CDS	1	Uncharacterized ABC-type transport system, periplasmic component/surface lipoprotein	
HVO_B0319	*arcR20*	CDS	1	Transcriptional regulator (IclR family)	
HVO_B0320	*arcR21*	CDS	-1	Transcriptional regulator (IclR family)	Yes
HVO_B0321	-	CDS	-1	Predicted dehydrogenase	
HVO_B0322	*pelZ*	CDS	-1	Extracellular protein containing Kelch and FN3 domains	
HVO_B0323	-	CDS	1	Predicted dehydrogenase	
HVO_B0324	-	CDS	1	Uncharacterized conserved protein	
HVO_B0325	-	CDS	-1	Lipoprotein involved in nitrous oxide reduction NosD- and NosL-like domains	
HVO_B0326	-	CDS	-1		
HVO_B0326s	*sRNA63*	sRNA	-1		Yes
HVO_B0327	-	CDS	-1		
HVO_B0328	*dppA15*	CDS	1	ABC-type transport system, periplasmic component	
HVO_B0329	*dppBC15*	CDS	1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	
HVO_B0330	-	CDS	-1		
HVO_B0331	*dppD15*	CDS	1	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	
HVO_B0332	*dppF15*	CDS	1	ABC-type oligopeptide transport system, ATPase component	
HVO_B0333	-	CDS	-1	Predicted dehydrogenase	
HVO_B0334	-	CDS	-1	Predicted dehydrogenase	
HVO_B0335	-	CDS	-1	Permease of the major facilitator superfamily	
HVO_B0336	-	CDS	-1	Altronate dehydratase	
HVO_B0337	*garD*	CDS	-1	SAF domain, altronate dehydratase related	
HVO_B0338	-	CDS	-1	Altronate dehydratase	
HVO_B0339	-	CDS	1	Uncharacterized conserved protein	
HVO_B0340	-	CDS	1	Uncharacterized conserved protein	
HVO_B0341	-	CDS	1	Uncharacterized conserved protein	
HVO_B0342	*mer2*	CDS	-1	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase or related flavin-dependent oxidoreductase	Yes
HVO_B0343	*lplD*	CDS	-1	Alpha-galactosidase/6-phospho-beta-glucosidase, family 4 of glycosyl hydrolase	
HVO_B0344	*uspA33*	CDS	-1	Nucleotide-binding protein, UspA family	
HVO_B0345	-	CDS	-1	L-alanine-DL-glutamate epimerase or related enzyme of enolase superfamily	
HVO_B0346	-	CDS	-1	L-alanine-DL-glutamate epimerase or related enzyme of enolase superfamily	
HVO_B0347	-	CDS	-1	Predicted dehydrogenase	
HVO_B0348	-	CDS	1	Dihydrodipicolinate synthase/N-acetylneuraminate lyase	
HVO_B0349	-	CDS	1	Uncharacterized conserved protein	

HVO_B0350	*palA*	CDS	1	Trehalose utilization protein	
HVO_B0351	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0352	-	CDS	1	Protein containing cupin domain	
HVO_B0353	-	CDS	1	Predicted amidohydrolase	
HVO_B0354	-	CDS	-1	Pentapeptide repeats containing protein	
HVO_B0355	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0356	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0357	*speE*	CDS	-1	Predicted spermidine synthase with an N-terminal membrane domain	Yes
HVO_B0358	-	CDS	-1	Uncharacterized conserved protein	
HVO_B0359	-	CDS	-1	Predicted permease	
HVO_B0360	-	CDS	-1	Glycosyltransferase	
HVO_B0361	-	CDS	1	Predicted transcriptional regulator, C-terminal HTH-like domain	
HVO_B0362	-	CDS	1	Uncharacterized conserved protein	
HVO_B0363	*dmsA*	CDS	1	Anaerobic dehydrogenase	
HVO_B0364	*hmoA*	CDS	1	Fe-S-cluster-containing dehydrogenase component	
HVO_B0365	*moz*	CDS	1	Polysulphide reductase	
HVO_B0366	-	CDS	1	Uncharacterized component of anaerobic dehydrogenase	
HVO_B0367	-	CDS	1	Sulfite oxidase or related enzyme	
HVO_B0368	*mobA2*	CDS	1	Molybdopterin-guanine dinucleotide biosynthesis protein A	
HVO_B0369	-	CDS	1	ABC-type molybdate transport system, periplasmic component	
HVO_B0370	-	CDS	1	ABC-type sulfate transport system, permease component	
HVO_B0370A					
HVO_B0371	-	CDS	-1	Lactaldehyde dehydrogenase, Succinate semialdehyde dehydrogenase or other NAD-dependent aldehyde dehydrogenase	
HVO_B0372	-	CDS	1		
HVO_B0373	-	CDS	-1		
HVO_B0374	-	CDS	1		
HVO_B0375	-	CDS	1	Uncharacterized conserved protein	
HVO_B0376	-	CDS	1	Predicted oxidoreductase (related to aryl-alcohol dehydrogenase)	
HVO_B0377	-	CDS	-1	Peptidoglycan interpeptide bridge formation enzyme	
HVO_B0378	-	CDS	1	Predicted flavin-nucleotide-binding protein	
HVO_B0379	*putP2*	CDS	-1	Na+/proline symporter	
HVO_B0380	-	CDS	-1	Transcriptional activator TenA	
HVO_B0381	-	CDS	-1	Transcriptional activator TenA	
HVO_B0382	*tbp3*	CDS	1	TATA-box binding protein (TBP), component of TFIID and TFIIIB	Yes

**ori-pHV3* and *orc6* can both be deleted when pHV3 has been integrated onto the main chromosome.



Predicted *mcm2* gene (HVO_B0008)



Predicted *mcm2* gene (HVO_B0008) found on pHV3. The predicted *mcm2* gene is located near to *ori-pHV3* and *orc6*, is 375 bp in length and points towards the origin.

Predicted Mcm2 HVO_B0008	*1*		*0*
Haloferax volcanii MCM (HVO_0220)	*1*	MAQAPQNRDLTERFIEFYRNYYREEIGTLAQQYPKEKRSL	*40*
Predicted Mcm2 HVO_B0008	*1*	----------	*0*
Haloferax volcanii MCM (HVO_0220)	*41*	HIDYDDLYRFDSELADDYITKPGQFQEYAEEALRLFDLPA	*80*
Predicted Mcm2 HVO_B0008	*1*	----------	*0*
Haloferax volcanii MCM (HVO_0220)	*81*	DVKLGQAHVRMRNLPETVDIRNLRVNDDHIGTLISVQGIV	*120*
Predicted Mcm2 HVO_B0008	*1*	----------	*0*
Haloferax volcanii MCM (HVO_0220)	*121*	RKATDVRPKITEAAFECQRCGTMSYIPQGDGGFQEPHECQ	*160*
Predicted Mcm2 HVO_B0008	*1*	----------	*0*
Haloferax volcanii MCM (HVO_0220)	*161*	GCERQGPFRIDFDQSNFVDSQKLRVQESPEGLRGGETPQS	*200*
Predicted Mcm2 HVO_B0008	*1*	----MPDDESAKLAEKPHAGVVTCPACDLHVSVTEPNDAV	*36*
Haloferax volcanii MCM (HVO_0220)	*201*	IDINLSDDVTGKVTAGDHVTVVGVLHIEQQTSGNEKTPVF	*240*
Predicted Mcm2 HVO_B0008	*37*	DLYRRHANVTGHDVEWERVAFD-------VDVESD-----	*64*
Haloferax volcanii MCM (HVO_0220)	*241*	DYYMEGISLTIEDEEFEDMEISDEDVAEIVELSNDPAIYE	*280*
Predicted Mcm2 HVO_B0008	*65*	----------GVKTALTELG----	*74*
Haloferax volcanii MCM (HVO_0220)	*281*	KMVESVAPAIYGYEQEKIAMILQLFSGVTKHLPDGSRIRG	*320*
Predicted Mcm2 HVO_B0008	*75*	----------	*74*
Haloferax volcanii MCM (HVO_0220)	*321*	DLHMLLIGDPGTGKSQMLSYIRHIAPRSVYTSGKGSSSAG	*360*
Predicted Mcm2 HVO_B0008	*75*	-------EDHPDG--VELGRLAAALADNGVAIGETLDAVR	*105*
Haloferax volcanii MCM (HVO_0220)	*361*	LTAAAVRDDFGDGQQWTLEAGALVLADKGIAAVDELDKMR	*400*
Predicted Mcm2 HVO_B0008	*106*	----------	*105*
Haloferax volcanii MCM (HVO_0220)	*401*	PEDRSAMHEGLEQQQISVSKAGINATLKSRCSLLGAANPK	*440*
Predicted Mcm2 HVO_B0008	*106*	----------	*105*
Haloferax volcanii MCM (HVO_0220)	*441*	YGRFDQYEPIGEQIDLEPALISRFDLIFTVTDDPDPDEDS	*480*
Predicted Mcm2 HVO_B0008	*106*	----------	*105*
Haloferax volcanii MCM (HVO_0220)	*481*	KLADHILKTNYAGELNTQRTNVANSEFTEQQVDAVTDEVA	*520*
Predicted Mcm2 HVO_B0008	*106*	----------	*105*
Haloferax volcanii MCM (HVO_0220)	*521*	PTIDADLLRKYIAYAKRTCYPTMTDEAKEVIRDFYVDFRA	*560*
Predicted Mcm2 HVO_B0008	*106*	----------	*105*
Haloferax volcanii MCM (HVO_0220)	*561*	RGADEDAPVPVTARKLEALVRLGEASARVRLSDKVTREDA	*600*
Predicted Mcm2 HVO_B0008	*106*	----------	*105*
Haloferax volcanii MCM (HVO_0220)	*601*	ERVTGIVESCLRDIGMDPETGEFDADIVETGRSKTQRDRI	*640*
Predicted Mcm2 HVO_B0008	*106*	----------	*105*
Haloferax volcanii MCM (HVO_0220)	*641*	KNLLELIRTMQEEYEEGAPHEEVLERANSELNMDEKTVND	*680*
Predicted Mcm2 HVO_B0008	*106*	---DLRMSGEIYEPQDDYVLAV	*124*
Haloferax volcanii MCM (HVO_0220)	*681*	QLDKLKMKGDIYEPRGDVYRAT	*702*



Domain search of *Haloferax volcanii* predicted Mcm2 (gene number HVO_B0008). There are no conserved domains found. Search was conducted via NCBI Conserved Domain (CD) search using translated protein sequence (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi). Diagram shows standard results.

V pHV4

Gene number	Gene name	Gene type	Strand	arCOG annotation	Deleted
HVO_A0001	*orc3*	CDS	1	Cdc6-related protein, AAA superfamily ATPase	Yes
HVO_A0001ori	*ori-pHV4*	ORI	1		Yes
HVO_A0002	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0003	*parA3*	CDS	-1	ATPase involved in chromosome partitioning, ParA family	
HVO_A0004	*msrA2*	CDS	-1		
HVO_A0005	-	CDS	1	Integrase	
HVO_A0006	-	CDS	1		
HVO_A0007	-	CDS	1	Transposase	
HVO_A0007_A					
HVO_A0008	-	CDS	1	Transposase	
HVO_A0009	-	CDS	1	SAM-dependent methyltransferase	
HVO_A0010	-	CDS	-1	Transcriptional regulator, contains wHTH domain	
HVO_A0011	-	CDS	-1	Predicted transcriptional regulator, contains HTH domain	
HVO_A0012	-	CDS	-1		
HVO_A0012s	*sRNA168*	sRNA	-1		Yes
HVO_A0013	-	CDS	1		
HVO_A0014	-	CDS	1	Transposase	
HVO_A0015	-	CDS	-1		
HVO_A0016	-	CDS	-1	Transposase	
HVO_A0017	-	CDS	1	Uncharacterized conserved protein	
HVO_A0018	-	CDS	1	Transposase	
HVO_A0019	-	CDS	1	Single-stranded DNA-binding replication protein A (RPA), large (70 kD) subunit or related ssDNA-binding protein	
HVO_A0019A					
HVO_A0020	-	CDS	1	Uncharacterized conserved protein	
HVO_A0021	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0022	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0023	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0024	-	CDS	1		
HVO_A0025	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0026	*tfb9*	CDS	1	Transcription initiation factor TFIIB, Brf1 subunit/Transcription initiation factor TFIIB	Yes
HVO_A0027	-	CDS	-1	HTH-domain containing transcriptional regulator	
HVO_A0028	-	CDS	1	Zn-finger protein	
HVO_A0029	-	CDS	1	Uncharacterized conserved protein	
HVO_A0030	-	CDS	1	Uncharacterized conserved protein	
HVO_A0031	-	CDS	1	Uncharacterized conserved protein	
HVO_A0032	-	CDS	1	Uncharacterized conserved protein	
HVO_A0033	-	CDS	1	Uncharacterized conserved protein	
HVO_A0034	-	CDS	1	Zn finger containing protein	
HVO_A0035	*ftsZ6*	CDS	-1	Cell division GTPase	Yes
HVO_A0035_A					
HVO_A0036	-	CDS	1	Transcriptional regulator containing HTH domain, ArsR family	
HVO_A0037	-	CDS	1	Uncharacterized conserved protein	
HVO_A0038	-	CDS	-1	Uncharacterized membrane protein	
HVO_A0039	-	CDS	1		
HVO_A0040	-	CDS	-1		
HVO_A0041	-	CDS	-1		
HVO_A0042	-	CDS	-1	Predicted transcriptional regulator related to sugar-specific regulator TrmB	
HVO_A0043	-	CDS	-1	Nucleotide-binding protein, UspA family	
HVO_A0044	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0045	-	CDS	-1	Zn-dependent protease with chaperone function	
HVO_A0046	-	CDS	-1	Membrane protein DegA family	
HVO_A0047	*uspA26*	CDS	1	Nucleotide-binding protein, UspA family	
HVO_A0048	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0049	-	CDS	1	Hemolysins or related protein containing CBS domains	
HVO_A0050	-	CDS	1	C4-type Zn-finger protein	
HVO_A0051	-	CDS	1	Uncharacterized conserved protein	
HVO_A0052	-	CDS	1	Uncharacterized conserved protein	
HVO_A0053	-	CDS	1	Uncharacterized conserved protein	
HVO_A0054	-	CDS	1	Uncharacterized conserved protein	
HVO_A0055	-	CDS	1	Uncharacterized conserved protein	
HVO_A0056	-	CDS	1	Uncharacterized conserved protein	
HVO_A0056_A					
HVO_A0057	-	CDS	1	Transposase, IS5 family	
HVO_A0058	-	CDS	1	Uncharacterized conserved protein	
HVO_A0059	-	CDS	1	Uncharacterized conserved protein	
HVO_A0060	*ltrC1*	CDS	1	Zincin superfamily protease	
HVO_A0061	-	CDS	1	SWIM zinc finger containing protein	
HVO_A0062	-	CDS	-1	Transposase	
HVO_A0063	-	CDS	1	Uncharacterized conserved protein	
HVO_A0064	*orc13*	CDS	-1	Cdc6-related protein, AAA superfamily ATPase	
HVO_A0065	*polB2*	CDS	-1	DNA polymerase elongation subunit (family B), inactivated	
HVO_A0066	-	CDS	-1	Transposase	
HVO_A0067	-	CDS	-1	Uncharacterized conserved protein associated with inactivated PolB-like polymerase	
HVO_A0068	-	CDS	-1	Inactivated RecA-like ATPase associated with anactivated PolB gene	
HVO_A0069	-	CDS	-1	Predicted DNA repair protein, contains HHH and endonuclease domain	
HVO_A0070	-	CDS	-1	Icc family phosphoesterase	
HVO_A0071	-	CDS	1	Uncharacterized conserved protein	
HVO_A0072	*orc12*	CDS	-1	Cdc6-related protein, AAA superfamily ATPase	
HVO_A0073	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0074	-	CDS	1	ATPase of the AAA+ class	Yes
HVO_A0075	-	CDS	-1		
HVO_A0076	-	CDS	1		
HVO_A0076_A	-	CDS	-1		
HVO_A0077	-	CDS	-1	Transposase	
HVO_A0078	-	CDS	-1	Superfamily II DNA/RNA helicase, SNF2 family	
HVO_A0079	-	CDS	-1	Type II restriction enzyme, methylase subunit	
HVO_A0080	-	CDS	1	Transposase	
HVO_A0081	-	CDS	-1	Transposase	
HVO_A0082	*arcR6*	CDS	-1	Transcriptional regulator (IclR family)	
HVO_A0083	-	CDS	-1	Ferredoxin subunit of nitrite reductase or ring-hydroxylating dioxygenase	
HVO_A0084	-	CDS	-1	Predicted metal-dependent hydrolase of the TIM-barrel fold	
HVO_A0085	-	CDS	1	Uncharacterized conserved protein	
HVO_A0086	*uspA27*	CDS	-1	Nucleotide-binding protein, UspA family	
HVO_A0087	*xlnF*	CDS	-1	2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1,7-dioic acid hydratase (catechol pathway)	
HVO_A0088	-	CDS	-1	FAD/FMN-containing dehydrogenase	
HVO_A0089	-	CDS	-1		
HVO_A0090	*phdI*	CDS	-1	Cupin domain containing protein	

HVO_A0091	-	CDS	1	Na+/H+ antiporter NhaD or related arsenite permease	
HVO_A0092	*acd6*	CDS	-1	Acyl-CoA dehydrogenase	
HVO_A0093	*arcR7*	CDS	1	Transcriptional regulator (IclR family)	
HVO_A0094	-	CDS	1	Uncharacterized conserved protein	
HVO_A0095	-	CDS	1	Predicted metal-dependent hydrolase of the TIM-barrel fold	
HVO_A0096	*acd7*	CDS	1	Acyl-CoA dehydrogenase	
HVO_A0097	*hbd2*	CDS	1	3-hydroxyacyl-CoA dehydrogenase, some fused to Enoyl-CoA hydratase	
HVO_A0097A					
HVO_A0098	*tfb12*	CDS	-1	Transcription initiation factor TFIIB, Brf1 subunit/Transcription initiation factor TFIIB	Yes
HVO_A0099	-	CDS	-1	Transposase	
HVO_A0100	-	CDS	-1	Transposase	
HVO_A0101	-	CDS	1	Uncharacterized conserved protein	
HVO_A0102	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0103	-	CDS	1		
HVO_A0104	*ltrC2*	CDS	-1	Zincin superfamily protease	
HVO_A0105	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0106	-	CDS	-1	C4-type Zn-finger protein	
HVO_A0107	-	CDS	-1		
HVO_A0108	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0109	-	CDS	1		
HVO_A0110	-	CDS	-1	Transposase	
HVO_A0111	-	CDS	-1		
HVO_A0112	-	CDS	-1		
HVO_A0113	-	CDS	-1		
HVO_A0113A					
HVO_A0114	-	CDS	-1	Predicted transcriptional regulator	
HVO_A0115	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0116	-	CDS	-1	Transposase	
HVO_A0117	-	CDS	1	Transposase, IS5 family	
HVO_A0118	-	CDS	1	Predicted transcriptional regulator, PadR family	
HVO_A0119	-	CDS	-1	Kynurenine formamidase	
HVO_A0120	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0121	-	CDS	1	Predicted transcriptional regulator, C-terminal HTH-like domain	
HVO_A0122	-	CDS	-1		
HVO_A0123	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0124	-	CDS	-1		
HVO_A0125	-	CDS	1	Transposase	
HVO_A0126	-	CDS	-1	XerD/XerC family integrase	
HVO_A0126s	*sRNA529*	sRNA	1		Yes
HVO_A0127	*trpS2*	CDS	1	Tryptophanyl-tRNA synthetase	
HVO_A0128	-	CDS	-1	Transposase	
HVO_A0129	-	CDS	1	Uncharacterized conserved protein	
HVO_A0130	-	CDS	1	Uncharacterized conserved protein	
HVO_A0131	-	CDS	-1		
HVO_A0131_A					
HVO_A0132	-	CDS	-1		
HVO_A0132_A	-	CDS	-1		
HVO_A0133	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0133s	*sRNA168*	sRNA	-1		
HVO_A0134	-	CDS	1		
HVO_A0135	-	CDS	1	Predicted transcriptional regulator	
HVO_A0136	-	CDS	1	PemK/MazF family toxin	
HVO_A0137	*tfb8*	CDS	-1	Transcription initiation factor TFIIB, Brf1 subunit/Transcription initiation factor TFIIB	
HVO_A0137s	*sRNA362(1)*	sRNA	1		No
HVO_A0138	-	CDS	1	Uncharacterized conserved protein	
HVO_A0139	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0140	-	CDS	-1	Predicted hydrolase of HD superfamily	
HVO_A0141	*ltrC3*	CDS	-1	Zincin superfamily protease	
HVO_A0142	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0143	-	CDS	-1	C4-type Zn-finger protein	
HVO_A0144	-	CDS	1	Transposase	
HVO_A0144A					
HVO_A0145	*tsgA4*	CDS	-1	ABC-type spermidine/putrescine transport system, ATPase component	
HVO_A0146	*tsgB4*	CDS	-1	ABC-type sugar transport system, permease component	
HVO_A0147	*tsgC4*	CDS	-1	ABC-type sugar transport system, permease component	
HVO_A0148	*tsgD4*	CDS	-1	ABC-type sugar transport system, periplasmic component	
HVO_A0149	*cga2*	CDS	1	Glycosyl hydrolase family 15	
HVO_A0150	*trmB1*	CDS	1	Sugar-specific transcriptional regulator TrmB	
HVO_A0151	-	CDS	1	Transposase	
HVO_A0152	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0153	*uspA28*	CDS	-1	Nucleotide-binding protein, UspA family	
HVO_A0154	*ssf*	CDS	-1	Sodium:solute symporter	
HVO_A0155	-	CDS	-1	Uncharacterized small membrane protein	
HVO_A0156	*acs7*	CDS	-1	Acyl-coenzyme A synthetase/AMP-(fatty) acid ligase	Yes
HVO_A0157	*boa4*	CDS	-1	Signal transduction regulator, containsPAS and HTH domains	Yes
HVO_A0158	*acs8*	CDS	-1	Acyl-coenzyme A synthetase/AMP-(fatty) acid ligase	Yes
HVO_A0159	*aspC4*	CDS	1	Aspartate/tyrosine/aromatic aminotransferase	
HVO_A0160	*htlD1*	CDS	-1	Signal transduction histidine kinase, contains REC and PAS domains	
HVO_A0161	-	CDS	1	Transcriptional regulator, TetR/AcrR family	
HVO_A0162	*matE1*	CDS	1	Na+-driven multidrug efflux pump	
HVO_A0163	*prr*	CDS	-1	Diaminopimelate epimerase	
HVO_A0164	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0165	*PF0371*	CDS	-1	Na+-dependent transporter of the SNF family	Yes
HVO_A0166	-	CDS	1		
HVO_A0167	-	CDS	1		
HVO_A0168	*boa5*	CDS	-1	Predicted transcriptional regulator, C-terminal HTH-like domain	
HVO_A0169	-	CDS	1	OxaA/SpoJ/YigC translocase/secretase, sec-independent itegration of nascent memrane proteins into membrane	
HVO_A0170	-	CDS	1	Uncharacterized conserved protein	
HVO_A0171	*clc*	CDS	-1	Chloride channel protein EriC	
HVO_A0172	-	CDS	-1		
HVO_A0173	-	CDS	1	Predicted metal-dependent protease of the PAD1/JAB1 superfamily	
HVO_A0174	*sec61a2, secY2*	CDS	1	Preprotein translocase subunit SecY	
HVO_A0175	*cat3*	CDS	1	Amino acid transporter	
HVO_A0176	-	CDS	-1	ABC-type multidrug transport system, permease component	
HVO_A0177	-	CDS	-1	ABC-type multidrug transport system, ATPase component	
HVO_A0178	*atpD2*	CDS	1	Archaeal/vacuolar-type H+-ATPase subunit D	
HVO_A0179	-	CDS	1	Uncharacterized conserved protein	
HVO_A0180	*sph1*	CDS	1	DNA sulfur modification protein DndD, ATPase	
HVO_A0181	-	CDS	-1	Uncharacterized conserved protein	

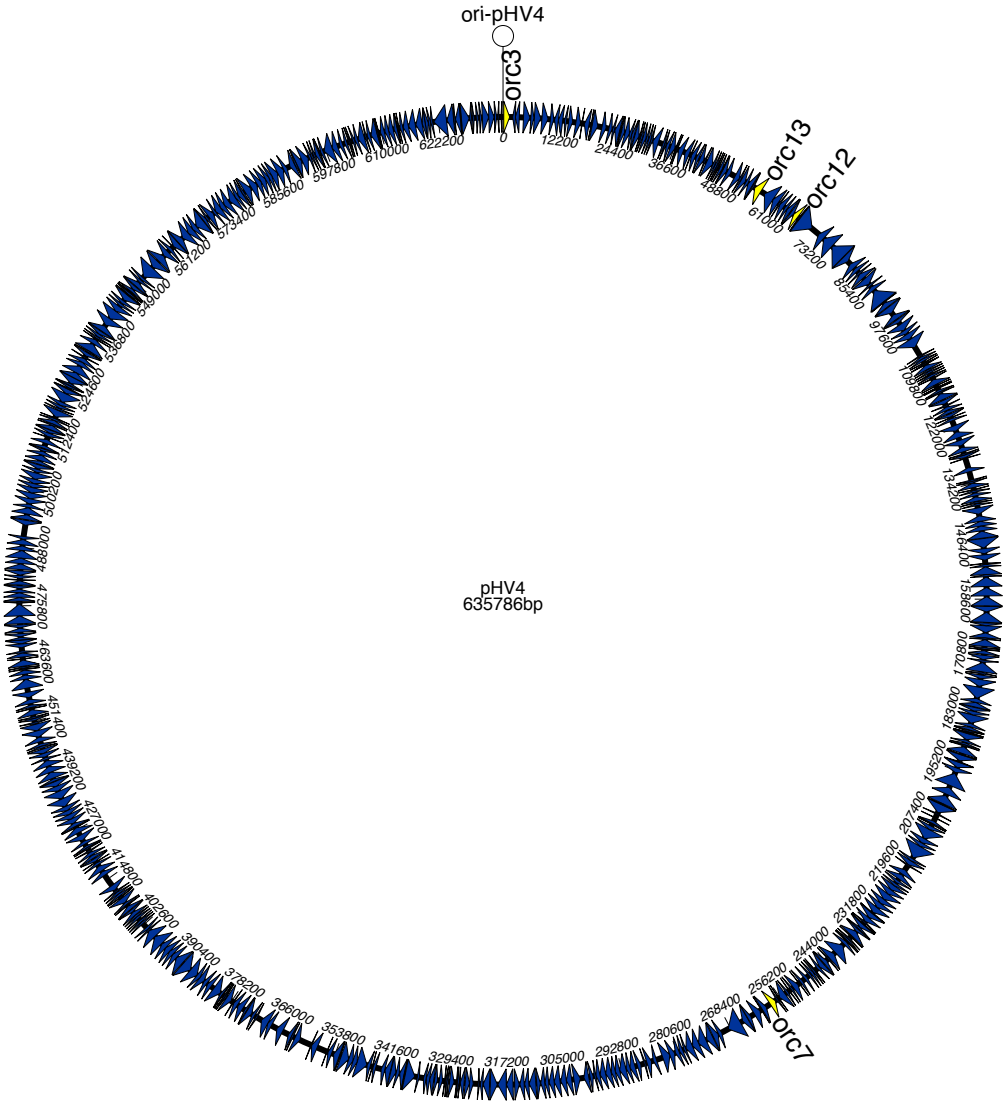
HVO_A0182	-	CDS	-1	Excinuclease-associated helix-hairpin-helix domain	
HVO_A0183	-	CDS	1	CopG-family DNA-binding protein	
HVO_A0184	-	CDS	-1		
HVO_A0185	-	CDS	1		
HVO_A0186	-	CDS	1	Predicted ATP-grasp enzyme	
HVO_A0187	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0188	-	CDS	-1	PIN domain containing protein	
HVO_A0189	-	CDS	-1	RHH family protein, predicted antitoxin of TA system	
HVO_A0190	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0191	-	CDS	1		
HVO_A0192	-	CDS	1	Uncharacterized conserved protein	
HVO_A0193	-	CDS	-1		
HVO_A0194	-	CDS	1	Glycosyltransferase	
HVO_A0195	-	CDS	1	Predicted membrane protein	
HVO_A0196	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0196 (3'-UTR)	*p225.2R*	sRNA	-1		Yes
HVO_A0197	-	CDS	-1	Polysaccharide biosynthesis protein, Mvin family	
HVO_A0198	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0199	-	CDS	-1		
HVO_A0200	-	CDS	1	Asparagine synthase (glutamine-hydrolyzing)	
HVO_A0200c	*CRISPR locus P1*	CRISPR	1		
HVO_A0201	-	CDS	-1		
HVO_A0202	-	CDS	-1		
HVO_A0203	-	CDS	-1		
HVO_A0204	-	CDS	-1		
HVO_A0204s	*sRNA479*	sRNA	1		Yes
HVO_A0205	*cas6*	CDS	1	CRISPR system related protein Cas6, RAMP superfamily	Yes
HVO_A0206	*cas8b*	CDS	1	CRISPR-associated protein, large subunit of CASCADE complex, MTH1090 family	Yes
HVO_A0207	*cas7*	CDS	1	CRISPR system related protein Cas7, RAMP superfamily	Yes
HVO_A0208	*cas5*	CDS	1	CRISPR system related protein Cas5, RAMP superfamily	Yes
HVO_A0209	*cas3*	CDS	1	CRISPR-associated helicase Cas3	Yes
HVO_A0210	*cas4*	CDS	1	CRISPR-associated protein Cas4, RecB family exonuclease	Yes
HVO_A0211	*cas1*	CDS	1	CRISPR-associated protein Cas1	Yes
HVO_A0212	*cas2*	CDS	1	CRISPR-associated protein Cas2	Yes
HVO_A0212c	*CRISPR locus P2*	CRISPR	1		
HVO_A0213	-	CDS	-1	Glycosyltransferase	Yes
HVO_A0214	-	CDS	-1	Lipid-A-disaccharide synthase related glycosyltransferase	
HVO_A0215	-	CDS	1	Predicted membrane-bound metal-dependent hydrolase	
HVO_A0216	-	CDS	1	Acetyltransferase (isoleucine patch superfamily)	
HVO_A0217	*PF0767*	CDS	1	Predicted pyridoxal phosphate-dependent enzyme apparently involved in regulation of cell wall biogenesis	
HVO_A0218	-	CDS	1	Predicted dehydrogenase	
HVO_A0219	*PAB0776*	CDS	1	UDP-N-acetyl-D-mannosaminuronate dehydrogenase	
HVO_A0220	-	CDS	1	UDP-N-acetylglucosamine 2-epimerase	
HVO_A0221	-	CDS	1	Predicted xylanase/chitin deacetylase	
HVO_A0222	-	CDS	1	Prenyltransferase family protein containing thioredoxin domain	
HVO_A0223	-	CDS	1		
HVO_A0224	-	CDS	1	Peptidoglycan interpeptide bridge formation enzyme	
HVO_A0225	-	CDS	1	Peptidoglycan interpeptide bridge formation enzyme	
HVO_A0226	-	CDS	-1	Transcriptional regulator containing HTH domain, ArsR family	
HVO_A0227	-	CDS	1		
HVO_A0228	*surE2*	CDS	1	Predicted acid phosphatase	
HVO_A0229	-	CDS	-1	CBS domain	
HVO_A0230	*msrA1*	CDS	1	Peptide methionine sulfoxide reductase	
HVO_A0231	-	CDS	-1	Rolling circle replication protein Rep	
HVO_A0232	-	CDS	-1		
HVO_A0233	-	CDS	-1		
HVO_A0234	-	CDS	-1	Predicted transcriptional regulator, contains HTH domain	
HVO_A0235	-	CDS	-1		
HVO_A0236	-	CDS	-1		
HVO_A0237	-	CDS	-1	Type I restriction-modification system methyltransferase subunit	
HVO_A0238	-	CDS	-1	Transposase	
HVO_A0239	-	CDS	1		
HVO_A0239_A	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0240	-	CDS	-1	Transposase, IS5 family	
HVO_A0241	-	CDS	1		
HVO_A0242	-	CDS	1	Uncharacterized conserved membrane protein	
HVO_A0242s	*sRNA450*	sRNA	-1		Yes
HVO_A0242A					
HVO_A0243	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0244	*doc*	CDS	-1	Prophage maintenance system killer protein	
HVO_A0245	-	CDS	-1	Transcriptional regulator containing HTH domain, ArsR family	
HVO_A0246	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0247	-	CDS	-1	Uncharacterized protein	
HVO_A0248	-	CDS	-1		
HVO_A0249	-	CDS	-1	Predicted transcriptional regulator related to sugar-specific regulator TrmB	
HVO_A0249s	*sRNA503*	sRNA	-1		Yes
HVO_A0249_A	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0249B					
HVO_A0250	-	CDS	-1	Transposase	
HVO_A0251	-	CDS	-1	Predicted transcriptional regulator	
HVO_A0252	-	CDS	1		
HVO_A0253	-	CDS	-1	Fe2+/Zn2+ uptake regulation protein, fur/PerR	
HVO_A0254	*PAB1516*	CDS	1	AAA+ superfamily ATPase fused to HTH and RecB nuclease domains	
HVO_A0254_A	-	CDS	1	Uncharacterized conserved protein	
HVO_A0255	-	CDS	1		
HVO_A0256	-	CDS	1	Uncharacterized conserved protein	
HVO_A0257	*orc7*	CDS	-1	Cdc6-related protein, AAA superfamily ATPase	
HVO_A0258	-	CDS	-1	Transposase	
HVO_A0259	-	CDS	-1		
HVO_A0260	-	CDS	-1	Transposase	
HVO_A0260_A	-	CDS	-1		
HVO_A0261	-	CDS	-1	Predicted transcriptional regulator	
HVO_A0262	*pilA5*	CDS	1	Transposase	Yes
HVO_A0263	*pilA6*	CDS	1	PKD repeats containing protein	Yes
HVO_A0264	-	CDS	-1		
HVO_A0265	-	CDS	-1		
HVO_A0266	*arcR8*	CDS	-1	Transcriptional regulator (IclR family)	
HVO_A0267	-	CDS	-1	L-alanine-DL-glutamate epimerase or related enzyme of enolase superfamily	
HVO_A0268	*Fuca*	CDS	1	Fuculose-1-phosphate aldolase	

HVO_A0269	*gpdA2*	CDS	1	Glycerol-3-phosphate dehydrogenase		Yes
HVO_A0270	*gpdB2*	CDS	1	Anaerobic glycerol-3-phosphate dehydrogenase		
HVO_A0271	*gpdC2*	CDS	1	Fe-S oxidoreductase		
HVO_A0272	-	CDS	1	Uncharacterized protein, possibly involved in utilization of glycolate and propanediol		
HVO_A0273	-	CDS	1	Trehalose utilization protein		
HVO_A0274	*suhB2*	CDS	1	Archaeal fructose-1,6-bisphosphatase or related enzyme of inositol monophosphatase family		
HVO_A0275	-	CDS	-1	Uncharacterized conserved protein		
HVO_A0276	-	CDS	-1	Sugar (pentulose and hexulose) kinase		
HVO_A0277	*rnhD*	CDS	1			
HVO_A0278	*sfsA*	CDS	1	DNA-binding protein, stimulates sugar fermentation		
HVO_A0279	-	CDS	-1	Transposase		
HVO_A0279_A	-	CDS	-1	DNA-directed RNA polymerase, subunit M/Transcription elongation factor TFIIS		
HVO_A0280	*arcR9*	CDS	-1	Transcriptional regulator (IclR family)		
HVO_A0281	*tsgD5*	CDS	1	ABC-type spermidine/putrescine transport system, ATPase component		
HVO_A0282	-	CDS	1	Fe2+ dependent formamide hydrolase, cofactor Fo biosynthesis enzyme		
HVO_A0283	*tsgA5*	CDS	1	ABC-type sugar transport system, periplasmic component		
HVO_A0284	*tsgB5*	CDS	1	ABC-type sugar transport system, permease component		
HVO_A0285	*tsgC5*	CDS	1	ABC-type sugar transport system, permease component		
HVO_A0286	-	CDS	1	Uncharacterized conserved protein		
HVO_A0287	-	CDS	1	L-alanine-DL-glutamate epimerase or related enzyme of enolase superfamily		
HVO_A0288	-	CDS	1	Short-chain alcohol dehydrogenase		
HVO_A0289	-	CDS	1	Putative translation initiation inhibitor, yjgF family		
HVO_A0290	*kdgA2*	CDS	-1	2-keto-3-deoxy-6-phosphogluconate aldolase		
HVO_A0290A						
HVO_A0291	-	CDS	1	Uncharacterized conserved protein		
HVO_A0292	-	CDS	-1	Amino acid transporter		
HVO_A0293	*potA1*	CDS	1	ABC-type spermidine/putrescine transport system, ATPase component		
HVO_A0294	*potA2*	CDS	1	ABC-type spermidine/putrescine transport system, ATPase component		
HVO_A0295	*amaB2*	CDS	1	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase or related deacylase		
HVO_A0295_A	-	CDS	1	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase or related flavin-dependent oxidoreductase		
HVO_A0296	-	CDS	1	Short-chain alcohol dehydrogenase		
HVO_A0297	*potC*	CDS	1	ABC-type spermidine/putrescine transport system, permease component II		
HVO_A0298	-	CDS	-1	Short-chain alcohol dehydrogenase		
HVO_A0299	*potD*	CDS	1			
HVO_A0300	*potB*	CDS	1	ABC-type spermidine/putrescine transport system, permease component I		
HVO_A0301	-	CDS	1	Predicted xylanase/chitin deacetylase		
HVO_A0302	*ansB*	CDS	-1	L-asparaginase/archaeal Glu-tRNAGln amidotransferase subunit D		
HVO_A0303	*dht*	CDS	1	Dihydroorotase or related cyclic amidohydrolase		
HVO_A0304	*groL*	CDS	1	Chaperonin GroEL (HSP60 family)		
HVO_A0305	*mmsA*	CDS	-1	Lactaldehyde dehydrogenase, Succinate semialdehyde dehydrogenase or other NAD-dependent aldehyde dehydrogenase		
HVO_A0306	*gabT6*	CDS	1	4-aminobutyrate aminotransferase or related aminotransferase		
HVO_A0307	*trh9*	CDS	1	Transcriptional regulator, IclR family		
HVO_A0308	-	CDS	-1			
HVO_A0309	-	CDS	1			
HVO_A0310	-	CDS	-1		1492	
HVO_A0311	*halC8*	CDS	-1	Predicted metal-binding protein		
HVO_A0312	-	CDS	1	Uncharacterized conserved protein		
HVO_A0313	-	CDS	-1	Predicted transcriptional regulator, xre family		
HVO_A0314	-	CDS	-1			
HVO_A0315	-	CDS	-1			
HVO_A0316	-	CDS	-1	Uncharacterized conserved protein		
HVO_A0317	-	CDS	1	Transcriptional regulator containing HTH domain, ArsR family		
HVO_A0318	-	CDS	1	Uncharacterized conserved protein		
HVO_A0319	-	CDS	-1			
HVO_A0320	-	CDS	-1	Uncharacterized conserved membrane protein		
HVO_A0321	-	CDS	-1	Uncharacterized conserved protein		
HVO_A0322	-	CDS	1	ABC-type transport system involved in multi-copper enzyme maturation, permease component		
HVO_A0323	-	CDS	1	ABC-type multidrug transport system, ATPase component		
HVO_A0324	-	CDS	1	Uncharacterized conserved protein		
HVO_A0325	-	CDS	-1	Extracellular protein containing Kelch and FN3 domains		
HVO_A0326	*bgaH*	CDS	-1	Uncharacterized protein conserved, contains glutamine amidotransferase (GATase1)-like domain		Yes
HVO_A0327	-	CDS	1			
HVO_A0328	*pfk1*	CDS	1	Sugar kinase, ribokinase family		
HVO_A0329	*kdgA3*	CDS	-1	2-keto-3-deoxy-6-phosphogluconate aldolase		
HVO_A0330	*gfo2*	CDS	-1	Predicted dehydrogenase		
HVO_A0331	-	CDS	1	L-alanine-DL-glutamate epimerase or related enzyme of enolase superfamily		
HVO_A0332	*arcR10*	CDS	1	Transcriptional regulator (IclR family)		
HVO_A0333	-	CDS	1	Zn-dependent metalloprotease, SprT family		
HVO_A0334	-	CDS	-1	Uncharacterized conserved protein		
HVO_A0335	-	CDS	1	N-terminal nucleophile hydrolase superfamily protein		
HVO_A0336	*dppDF7*	CDS	-1	ABC-type oligopeptide transport system, ATPase component		
HVO_A0337	*dppC7*	CDS	-1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component		
HVO_A0338	*dppB7*	CDS	-1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component		
HVO_A0339	*dppA7*	CDS	-1	ABC-type transport system, periplasmic component		
HVO_A0340	-	CDS	-1			
HVO_A0341	*amaB3*	CDS	1	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase or related deacylase		
HVO_A0342	*arcR11*	CDS	1	Transcriptional regulator (IclR family)		
HVO_A0343	-	CDS	1			
HVO_A0344	-	CDS	-1			
HVO_A0345	*cat4*	CDS	-1			
HVO_A0346	-	CDS	1	XerD/XerC family integrase		
HVO_A0347	-	CDS	1			
HVO_A0348	-	CDS	-1			
HVO_A0348A						
HVO_A0349	-	CDS	-1	Transposase		
HVO_A0350	-	CDS	1			
HVO_A0351	-	CDS	1			
HVO_A0352	-	CDS	-1			
HVO_A0353	-	CDS	1			
HVO_A0354	-	CDS	1			
HVO_A0355	-	CDS	1	Transposase		
HVO_A0356	-	CDS	1			
HVO_A0357	-	CDS	1	Eukaryotic-type DNA primase, catalytic (small) subunit		
HVO_A0358	-	CDS	1	Uncharacterized protein		
HVO_A0359	-	CDS	1			
HVO_A0359A						
HVO_A0360	-	CDS	-1	Uncharacterized conserved protein		
HVO_A0361	-	CDS	1			
HVO_A0362	-	CDS	-1	WD40-like repeat		

HVO_A0363	-	CDS	1	Uncharacterized conserved protein	
HVO_A0364	-	CDS	1	Transposase	
HVO_A0365	-	CDS	1	Uncharacterized conserved protein	
HVO_A0366	-	CDS	1		
HVO_A0367	-	CDS	-1	Transposase	
HVO_A0368	-	CDS	-1	Cytotoxic translational repressor of toxin-antitoxin stability system	
HVO_A0369	-	CDS	-1	Antitoxin (DNA-binding domain) of toxin-antitoxin stability system	
HVO_A0370	-	CDS	1	Predicted antitoxins containing the HTH domain	
HVO_A0371	-	CDS	1	Uncharacterized conserved protein, contains Zn ribbon C-terminal domain	
HVO_A0372	-	CDS	-1	Predicted exonuclease of the beta-lactamase fold involved in RNA processing	
HVO_A0373	-	CDS	-1		
HVO_A0374	-	CDS	1	Single-stranded DNA-binding replication protein A (RPA), large (70 kD) subunit or related ssDNA-binding protein	Yes
HVO_A0375	*tfb11*	CDS	1	Transcription initiation factor TFIIB, Brf1 subunit/Transcription initiation factor TFIIB	
HVO_A0376	-	CDS	-1	Xaa-Pro aminopeptidase	
HVO_A0377	-	CDS	-1	Aspartate racemase	
HVO_A0378	*hyuB2*	CDS	-1	N-methylhydantoinase B/acetone carboxylase, alpha subunit	
HVO_A0379	*hyuA2*	CDS	-1	N-methylhydantoinase A/acetone carboxylase, beta subunit	
HVO_A0380	*dppA8*	CDS	1	ABC-type transport system, periplasmic component	
HVO_A0381	*dppB8*	CDS	1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	
HVO_A0382	*dppC8*	CDS	1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	
HVO_A0383	*dppD8*	CDS	1	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	
HVO_A0384	*dppF8*	CDS	1	ABC-type oligopeptide transport system, ATPase component	
HVO_A0385	*hyuA3*	CDS	1	N-methylhydantoinase A/acetone carboxylase, beta subunit	
HVO_A0386	*hyuB3*	CDS	1	N-methylhydantoinase B/acetone carboxylase, alpha subunit	
HVO_A0387	-	CDS	1	Cupin domain containing protein	
HVO_A0388	-	CDS	-1	Transcriptional regulator (Lrp/AsnC family)	
HVO_A0389	-	CDS	1		
HVO_A0390	-	CDS	-1		
HVO_A0390_A					
HVO_A0391	-	CDS	1	Uncharacterized conserved protein	
HVO_A0392	-	CDS	1	Prophage maintenance system killer protein	
HVO_A0393	-	CDS	1	Uncharacterized conserved protein	
HVO_A0394	-	CDS	1	Transcriptional regulator containing HTH domain, ArsR family	
HVO_A0395	-	CDS	1	Uncharacterized conserved protein	
HVO_A0396	-	CDS	1	SWIM zinc finger containing protein	
HVO_A0397	-	CDS	-1	HTH-domain containing transcriptional regulator	
HVO_A0398	-	CDS	1	Zn-finger protein	
HVO_A0399	-	CDS	1	Uncharacterized conserved protein	
HVO_A0400	-	CDS	1	Uncharacterized conserved protein	
HVO_A0401	-	CDS	-1	Predicted transciptional regulator, contains Fic and DeoR-like domains	
HVO_A0402	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0403	-	CDS	1	Transposase	
HVO_A0404	-	CDS	1	Transposase	
HVO_A0405	-	CDS	1	Uncharacterized conserved protein	
HVO_A0406	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0407	-	CDS	-1		
HVO_A0408	-	CDS	1		
HVO_A0408A					
HVO_A0409	-	CDS	1	Single-stranded DNA-binding replication protein A (RPA), large (70 kD) subunit or related ssDNA-binding protein	
HVO_A0410	-	CDS	1	WD40-like repeat	
HVO_A0411	-	CDS	1	Uncharacterized conserved protein	
HVO_A0412	-	CDS	1		
HVO_A0413	-	CDS	-1	Predicted transcriptional regulator, C-terminal HTH-like domain	
HVO_A0414	-	CDS	1	Uncharacterized conserved protein	
HVO_A0415	-	CDS	1	Kynurenine formamidase	
HVO_A0416	-	CDS	-1	Predicted transcriptional regulator, PadR family	
HVO_A0417	-	CDS	1	Uncharacterized conserved protein	
HVO_A0418	-	CDS	1	Predicted transcriptional regulator	
HVO_A0418A					
HVO_A0419	-	CDS	1	Minimal nucleotidyltransferase fused to HTH domain	
HVO_A0420	-	CDS	-1	Transposase	
HVO_A0421	-	CDS	1		
HVO_A0422	-	CDS	-1		
HVO_A0423	-	CDS	1	Transposase	
HVO_A0424	*boa6*	CDS	1	Predicted transcriptional regulator, C-terminal HTH-like domain	
HVO_A0425	*dppF9*	CDS	-1	ABC-type oligopeptide transport system, ATPase component	
HVO_A0426	*dppD9*	CDS	-1	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	
HVO_A0426_A	*dppC9*	CDS	-1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	
HVO_A0427	*dppB9*	CDS	-1	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	
HVO_A0428	*dppA9*	CDS	-1	ABC-type transport system, periplasmic component	
HVO_A0429	*ureX*	CDS	-1	N-acyl-D-aspartate/D-glutamate deacylase	
HVO_A0429_A	-	CDS	-1	L-alanine-DL-glutamate epimerase or related enzyme of enolase superfamily	
HVO_A0430	-	CDS	-1	NAD dependent epimerase/dehydratase family	
HVO_A0430_A	-	CDS	1	Uncharacterized conserved protein	
HVO_A0431	-	CDS	1	Uncharacterized conserved protein	
HVO_A0432	-	CDS	1	Uncharacterized conserved protein	
HVO_A0433	-	CDS	1	Transposase	
HVO_A0434	-	CDS	1	SAM-dependent methyltransferase	
HVO_A0435	-	CDS	-1	Transcriptional regulator, contains wHTH domain	
HVO_A0436	-	CDS	-1	Predicted transcriptional regulator, contains HTH domain	
HVO_A0437	-	CDS	1	RHH-fold DNA-binding ptotein	
HVO_A0438	-	CDS	1	PIN domain containing protein	
HVO_A0439	-	CDS	1	Uncharacterized conserved protein	
HVO_A0440	-	CDS	-1	Nuclease of restriction endonuclease-like fold, implicated in nucleotide excision repair	
HVO_A0441	*rad25a*	CDS	-1	ssDNA-dependent ATPase, helicase superfamily II	Yes
HVO_A0442	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0443	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0444	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0444A					
HVO_A0445	-	CDS	1	Transposase	
HVO_A0446	-	CDS	1		
HVO_A0447	-	CDS	1	Membrane associated serine protease	
HVO_A0448	*tfb10*	CDS	1	Transcription initiation factor TFIIB, Brf1 subunit/Transcription initiation factor TFIIB	
HVO_A0449	-	CDS	-1	Transposase	
HVO_A0450	*uspA29*	CDS	-1		
HVO_A0451	*arsB4*	CDS	1	Di- and tricarboxylate transporter	
HVO_A0452	*uspA30*	CDS	1	Nucleotide-binding protein, UspA family	
HVO_A0453	-	CDS	1	Zn-finger protein	
HVO_A0454	-	CDS	1	Uncharacterized conserved protein	

HVO_A0455	-	CDS	1	Uncharacterized conserved protein	
HVO_A0456	-	CDS	1	Transposase	
HVO_A0457	-	CDS	1		
HVO_A0457_A	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0458	*sojE*	CDS	1	ATPase involved in chromosome partitioning, ParA family	
HVO_A0459	-	CDS	1	Uncharacterized conserved protein	
HVO_A0460	*phnE3*	CDS	-1	ABC-type phosphate/phosphonate transport system, permease component	
HVO_A0461	*phnE2*	CDS	-1	ABC-type phosphate/phosphonate transport system, permease component	
HVO_A0462	-	CDS	1		1492
HVO_A0463	*rnhC*	CDS	-1	Ribonuclease HI	
HVO_A0464	-	CDS	1	Uncharacterized conserved protein	
HVO_A0465	-	CDS	-1	Transcriptional regulator, TetR/AcrR family	
HVO_A0466	-	CDS	1	S-layer domain	
HVO_A0467	*PF0262*	CDS	1	Predicted exporter of the RND superfamily	
HVO_A0468	-	CDS	-1		
HVO_A0469	*acc1*	CDS	1	Pirin-related protein	
HVO_A0470	-	CDS	1	Lactoylglutathione lyase or related enzyme	
HVO_A0471	*estA*	CDS	1	Predicted esterase	
HVO_A0472	*trxB6*	CDS	-1	Thioredoxin reductase	
HVO_A0473	-	CDS	1	Permease of the drug/metabolite transporter (DMT) superfamily	
HVO_A0474	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0475	*sod1*	CDS	-1	Superoxide dismutase	
HVO_A0476	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0477	*pstS2*	CDS	1	ABC-type phosphate transport system, periplasmic component	
HVO_A0478	*pstC2*	CDS	1	ABC-type phosphate transport system, permease component	
HVO_A0479	*pstA2*	CDS	1	ABC-type phosphate transport system, permease component	
HVO_A0480	*pstB2*	CDS	1	ABC-type phosphate transport system, ATPase component	
HVO_A0481	*phoU4*	CDS	1	Phosphate uptake regulator	
HVO_A0482	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0483	-	CDS	1		
HVO_A0484	-	CDS	1		
HVO_A0485	-	CDS	-1	Serine-pyruvate aminotransferase/archaeal aspartate aminotransferase	
HVO_A0486	*hypE*	CDS	1	Hydrogenase maturation factor	
HVO_A0487	*cobB*	CDS	1	Cobyrinic acid a,c-diamide synthase	
HVO_A0488	*cobO*	CDS	1	ATP:corrinoid adenosyltransferase	
HVO_A0489	*nifV*	CDS	-1	Isopropylmalate/homocitrate/citramalate synthase	
HVO_A0490	*PA0882*	CDS	-1	Succinyl-CoA:mesaconate CoA-transferase or predicted acyl-CoA transferase/carnitine dehydratase or	
HVO_A0491	*tsgD6*	CDS	-1	ABC-type spermidine/putrescine transport system, ATPase component	
HVO_A0492	*tsgC6*	CDS	-1	ABC-type sugar transport system, permease component	
HVO_A0493	*tsgB6*	CDS	-1	ABC-type sugar transport system, permease component	
HVO_A0494	*tsgA6*	CDS	-1	ABC-type sugar transport system, periplasmic component	
HVO_A0495	*boa7*	CDS	-1	Signal transduction regulator, containsPAS and HTH domains	
HVO_A0496	*uspA31*	CDS	-1	Nucleotide-binding protein, UspA family	
HVO_A0497	-	CDS	1		
HVO_A0498	-	CDS	1	Uncharacterized conserved protein	
HVO_A0499	-	CDS	1		
HVO_A0500	*ppiB2*	CDS	-1	FKBP-type peptidyl-prolyl cis-trans isomerase 2	
HVO_A0501	-	CDS	1	SAM-dependent methyltransferase	
HVO_A0502	-	CDS	1		
HVO_A0503	-	CDS	1	Predicted metal-dependent hydrolase of the TIM-barrel fold	
HVO_A0504	-	CDS	1		
HVO_A0505	*fadA4*	CDS	-1	Enoyl-CoA hydratase/carnithine racemase	
HVO_A0506	*boa8*	CDS	-1	Signal transduction regulator, containsPAS and HTH domains	
HVO_A0507	*paaA*	CDS	1	Phenylacetic acid catabolic protein	
HVO_A0508	*pacF*	CDS	1	Uncharacterized conserved protein	
HVO_A0509	*paaI*	CDS	1	Phenylacetic acid catabolic protein	
HVO_A0510	-	CDS	1	Predicted metal-sulfur cluster biosynthetic enzyme	
HVO_A0511	-	CDS	1	Uncharacterized conserved protein	
HVO_A0512	*aldY5*	CDS	-1	Lactaldehyde dehydrogenase, Succinate semialdehyde dehydrogenase or other NAD-dependent aldehyde dehydrogenase	
HVO_A0513	-	CDS	-1	HGG motif-containing thioesterase, possibly involved in aromatic compounds catabolism	
HVO_A0514	*livF4*	CDS	-1	ABC-type branched-chain amino acid transport system, ATPase component	
HVO_A0515	*livG4*	CDS	-1	ABC-type branched-chain amino acid transport system, ATPase component	
HVO_A0516	*livHM4*	CDS	-1	ABC-type branched-chain amino acid transport system, permease component	
HVO_A0517	-	CDS	-1	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase or related flavin-dependent oxidoreductase	
HVO_A0518	-	CDS	1		
HVO_A0519	*maoC2*	CDS	-1	Mesaconyl-CoA hydratase (Acyl dehydratase family)	
HVO_A0520	-	CDS	1	HGG motif-containing thioesterase, possibly involved in aromatic compounds catabolism	
HVO_A0521	*paaK1*	CDS	-1	Coenzyme F390 synthetase	
HVO_A0522	*acaB4*	CDS	1	Acetyl-CoA acetyltransferase	
HVO_A0523	-	CDS	1	3-hydroxy-3-methylglutaryl CoA synthase	
HVO_A0524	*hbd4*	CDS	1	3-hydroxyacyl-CoA dehydrogenase	
HVO_A0525	*fadA5*	CDS	1	Enoyl-CoA hydratase/carnithine racemase	
HVO_A0526	-	CDS	-1		
HVO_A0527	*arcR12*	CDS	-1	Transcriptional regulator (IclR family)	
HVO_A0528	-	CDS	1	Uncharacterized conserved protein	
HVO_A0529	*ilvB2*	CDS	1	Acetolactate synthase large subunit or other thiamine pyrophosphate-requiring enzyme	
HVO_A0530	*galE6*	CDS	1	Nucleoside-diphosphate-sugar epimerase	Yes
HVO_A0531	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0532	-	CDS	-1	Uncharacterized conserved protein, contains double-stranded beta-helix domain	
HVO_A0533	-	CDS	-1	Methanopterin biosynthesis cyclic phosphodiesterase, HD-superfamily hydrolase	
HVO_A0534	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0535	-	CDS	1	Xaa-Pro aminopeptidase	
HVO_A0536	*hyuB1*	CDS	-1	N-methylhydantoinase B/acetone carboxylase, alpha subunit	
HVO_A0537	*hyuA1*	CDS	-1	N-methylhydantoinase A/acetone carboxylase, beta subunit	
HVO_A0538	*mntH*	CDS	-1	Mn2+ and Fe2+ transporter of the NRAMP family	
HVO_A0539	*orf5*	CDS	-1	Predicted antitoxin, copG family	
HVO_A0540	*htlD2*	CDS	-1	Signal transduction regulator, contains Rec and HTH domains	
HVO_A0541	-	CDS	-1	ABC-type Fe3+-hydroxamate transport system, periplasmic component	
HVO_A0542	-	CDS	-1		
HVO_A0543	-	CDS	-1	Membrane protein containing Zn-ribbon domain	
HVO_A0544	-	CDS	1	Uncharacterized conserved protein	
HVO_A0545	*tetR1*	CDS	1	Transcriptional regulator, TetR/AcrR family	
HVO_A0546	-	CDS	-1	Short-chain alcohol dehydrogenase	Yes
HVO_A0547	-	CDS	1	N-terminal nucleophile hydrolase superfamily protein	
HVO_A0548	*doxD1*	CDS	-1	Predicted membrane protein, DoxD family	
HVO_A0549	*nosL1*	CDS	-1	Uncharacterized conserved protein	
HVO_A0550	-	CDS	-1	Signal transduction histidine kinase, contains REC and PAS domains	
HVO_A0551	*acs9*	CDS	-1	Acyl-CoA synthetase (AMP-forming)/AMP-acid ligase II	
HVO_A0552	-	CDS	-1		

HVO_A0553	*cobQ*	CDS	1	Cobyric acid synthase	
HVO_A0554	-	CDS	-1	Permease of the major facilitator superfamily	
HVO_A0555	-	CDS	1	Uncharacterized conserved protein, DUF1121	
HVO_A0556	-	CDS	1	Uncharacterized conserved protein	
HVO_A0557	-	CDS	-1	ABC-type Fe3+-hydroxamate transport system, periplasmic component	
HVO_A0558	-	CDS	1	ABC-type Fe3+-hydroxamate transport system, periplasmic component	
HVO_A0559	*hutH*	CDS	-1	Histidine ammonia-lyase	Yes
HVO_A0560	*hutI*	CDS	-1	Imidazolonepropionase or related amidohydrolase	Yes
HVO_A0561	*hutG*	CDS	-1	Arginase family enzyme	Yes
HVO_A0562	*hutU*	CDS	-1	Urocanate hydratase	Yes
HVO_A0563	-	CDS	1	Predicted transcriptional regulator, C-terminal HTH-like domain	Yes
HVO_A0564	-	CDS	-1	Lactoylglutathione lyase or related enzyme	
HVO_A0565	-	CDS	1	Permease of the major facilitator superfamily	
HVO_A0566	-	CDS	-1	3-isopropylmalate dehydratase small subunit	
HVO_A0567	-	CDS	-1	Homoaconitate hydratase/3-isopropylmalate dehydratase large subunit family protein	
HVO_A0568	-	CDS	-1	Isopropylmalate/homocitrate/citramalate synthase	
HVO_A0569	*baiF1*	CDS	-1	Succinyl-CoA:mesaconate CoA-transferase or predicted acyl-CoA transferase/carnitine dehydratase or	
HVO_A0570	-	CDS	-1	Dihydroorotase or related cyclic amidohydrolase	
HVO_A0571	-	CDS	-1	Amidase related to nicotinamidase	
HVO_A0572	-	CDS	-1	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase or related flavin-dependent oxidoreductase	Yes
HVO_A0573	-	CDS	-1		
HVO_A0574	-	CDS	-1	Ketosteroid isomerase-related protein	
HVO_A0575	-	CDS	1	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit	
HVO_A0576	*livJ5*	CDS	1	ABC-type branched-chain amino acid transport system, periplasmic component	
HVO_A0577	*livH5*	CDS	1	Branched-chain amino acid ABC-type transport system, permease component	
HVO_A0578	*livM5*	CDS	1	ABC-type branched-chain amino acid transport system, permease component	
HVO_A0579	*livG5*	CDS	1	ABC-type branched-chain amino acid transport system, ATPase component	
HVO_A0580	*livF5*	CDS	1	ABC-type branched-chain amino acid transport system, ATPase component	
HVO_A0581	-	CDS	1	Uncharacterized protein involved in propionate catabolism	
HVO_A0582	*PA0881*	CDS	1	Uncharacterized protein involved in propionate catabolism	
HVO_A0583	*arcR13*	CDS	1	Transcriptional regulator (IclR family)	
HVO_A0584	*guaB2*	CDS	-1	IMP dehydrogenase/GMP reductase	
HVO_A0585	-	CDS	-1		
HVO_A0586	*graD3*	CDS	1	N-acetylglucosamine-1-phosphate uridyltransferase	Yes
HVO_A0587	*htlD3*	CDS	1	Sensory protein, contains PAS domain	
HVO_A0588	-	CDS	1	Predicted transcriptional regulator, contains C-terminal CBS domains	
HVO_A0589	-	CDS	-1	Predicted transcriptional regulator	
HVO_A0590	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0591	-	CDS	-1	Predicted sensor protein/domain	
HVO_A0592	-	CDS	1	Uncharacterized conserved protein	
HVO_A0593	-	CDS	-1	Transcriptional regulator containing HTH domain, ArsR family	
HVO_A0594	*glmS*	CDS	-1	Glucosamine 6-phosphate synthetase	
HVO_A0595	*graD4*	CDS	-1	N-acetylglucosamine-1-phosphate uridyltransferase	
HVO_A0596	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0597	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0598	-	CDS	1	Predicted transcriptional regulator, PadR family	
HVO_A0599	-	CDS	1		
HVO_A0600	-	CDS	1	Uncharacterized conserved protein	
HVO_A0601	*nosL2*	CDS	-1	Uncharacterized conserved protein	
HVO_A0602	*manB1*	CDS	-1	Phosphomannomutase	
HVO_A0603	-	CDS	1		
HVO_A0604	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0605	-	CDS	-1	Predicted dinucleotide-binding enzyme	
HVO_A0606	-	CDS	-1	Permease of the major facilitator superfamily	
HVO_A0607	*cheW2*	CDS	-1	Chemotaxis signal transduction protein	
HVO_A0607A					
HVO_A0608	-	CDS	1		
HVO_A0609	*znuB2*	CDS	-1	ABC-type Mn2+/Zn2+ transport system, permease component	
HVO_A0610	*znuC2*	CDS	-1	ABC-type Mn/Zn transport system, ATPase component	
HVO_A0611	*znuA2*	CDS	-1	ABC-type metal ion transport system, periplasmic component/surface adhesin	
HVO_A0612	-	CDS	-1		
HVO_A0613	-	CDS	-1		
HVO_A0614	-	CDS	1		
HVO_A0615	*cspD5*	CDS	-1	Cold shock protein, CspA family	
HVO_A0616	*tetR2*	CDS	-1	Transcriptional regulator, TetR/AcrR family	
HVO_A0617	-	CDS	-1	Permease of the drug/metabolite transporter (DMT) superfamily	
HVO_A0618	-	CDS	-1	CoA-dependent NAD(P)H Sulfur Oxidoreductase	
HVO_A0619	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0620	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0621	-	CDS	1	Uncharacterized conserved protein	
HVO_A0622	-	CDS	1	Uncharacterized conserved protein	
HVO_A0623	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0624	*zntA1*	CDS	-1	Cation transport ATPase	
HVO_A0625	-	CDS	-1	Uncharacterized conserved protein	
HVO_A0626	-	CDS	-1	Metal-binding domain	
HVO_A0627	-	CDS	-1	Uncharacterized protein probably involved in high-affinity Fe2+ transport	
HVO_A0628	-	CDS	1		
HVO_A0629	*htlD5*	CDS	1	Sensory protein, contains PAS domain	
HVO_A0630	-	CDS	-1	Lactoylglutathione lyase or related enzyme	
HVO_A0631	-	CDS	1	Uncharacterized conserved protein	
HVO_A0632	-	CDS	1	Predicted flagellin FlaG	Yes
HVO_A0633	-	CDS	1	Predicted flagellin FlaG	Yes
HVO_A0634	-	CDS	1	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase or related deacylase	
HVO_A0635	*Cysteine lyase*	CDS	1	Selenocysteine lyase/Cysteine desulfurase	Yes
HVO_A0636	-	CDS	1	Cys-tRNA(Pro)/Cys-tRNA(Cys) deacylase, ybaK family	
HVO_A0637	*tif1A2*	CDS	-1	Translation initiation factor 1 (IF-1)	Yes
HVO_A0638	-	CDS	-1		
HVO_A0638s	*hts_574*	sRNA	-1		Yes

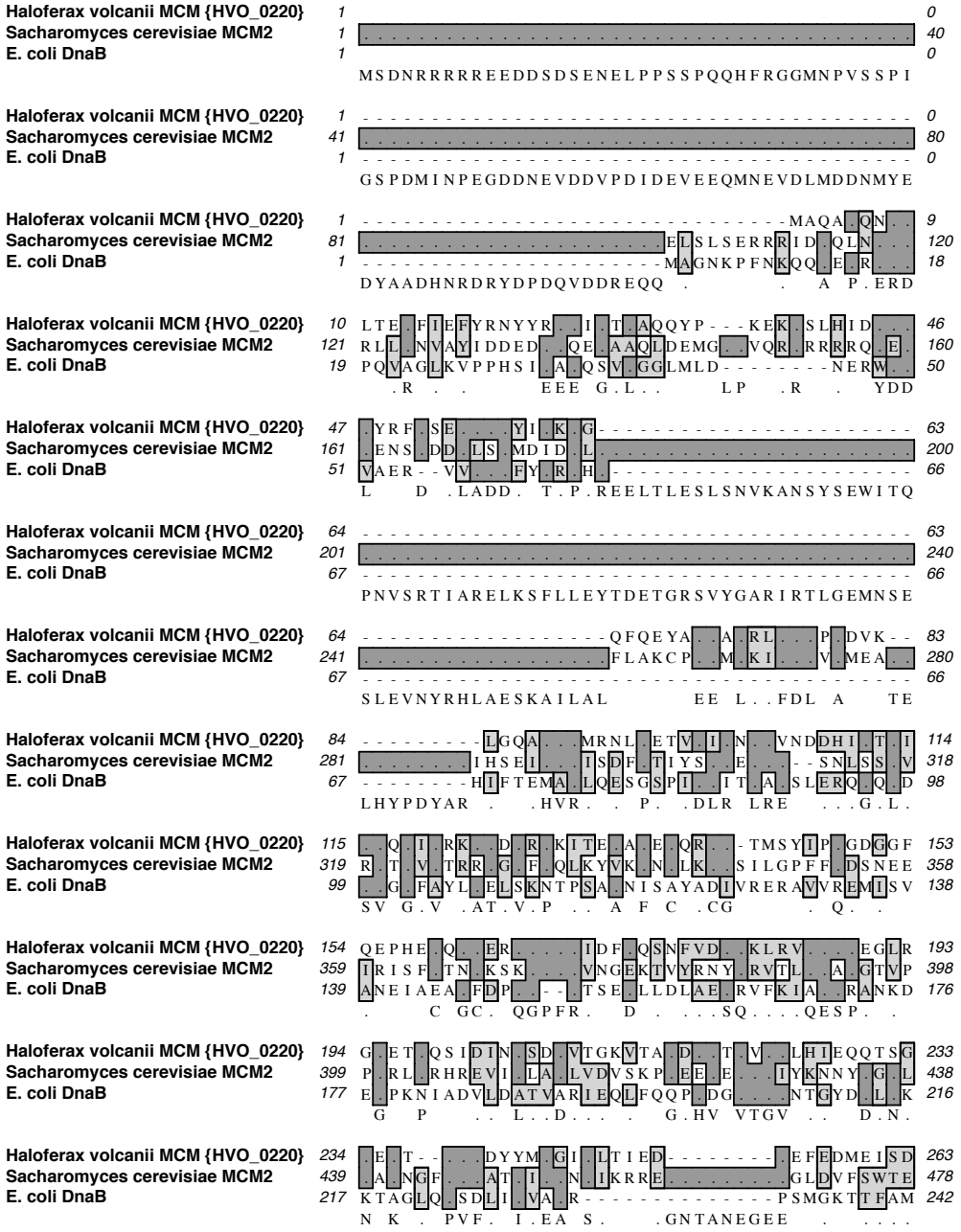

ori-pHV4
orc3
orc13
orc12
orc7
pHV4
635786bp
0
12200
24400
36600
48800
61000
73200
85400
97600
109800
122000
134200
146400
158600
170800
183000
195200
207400
219600
231800
244000
256200
268400
280600
292800
305000
317200
329400
341600
353800
366000
378200
390400
402600
414800
427000
439200
451400
463600
475800
488000
500200
512400
524600
536800
549000
561200
573400
585600
597800
610000
622200

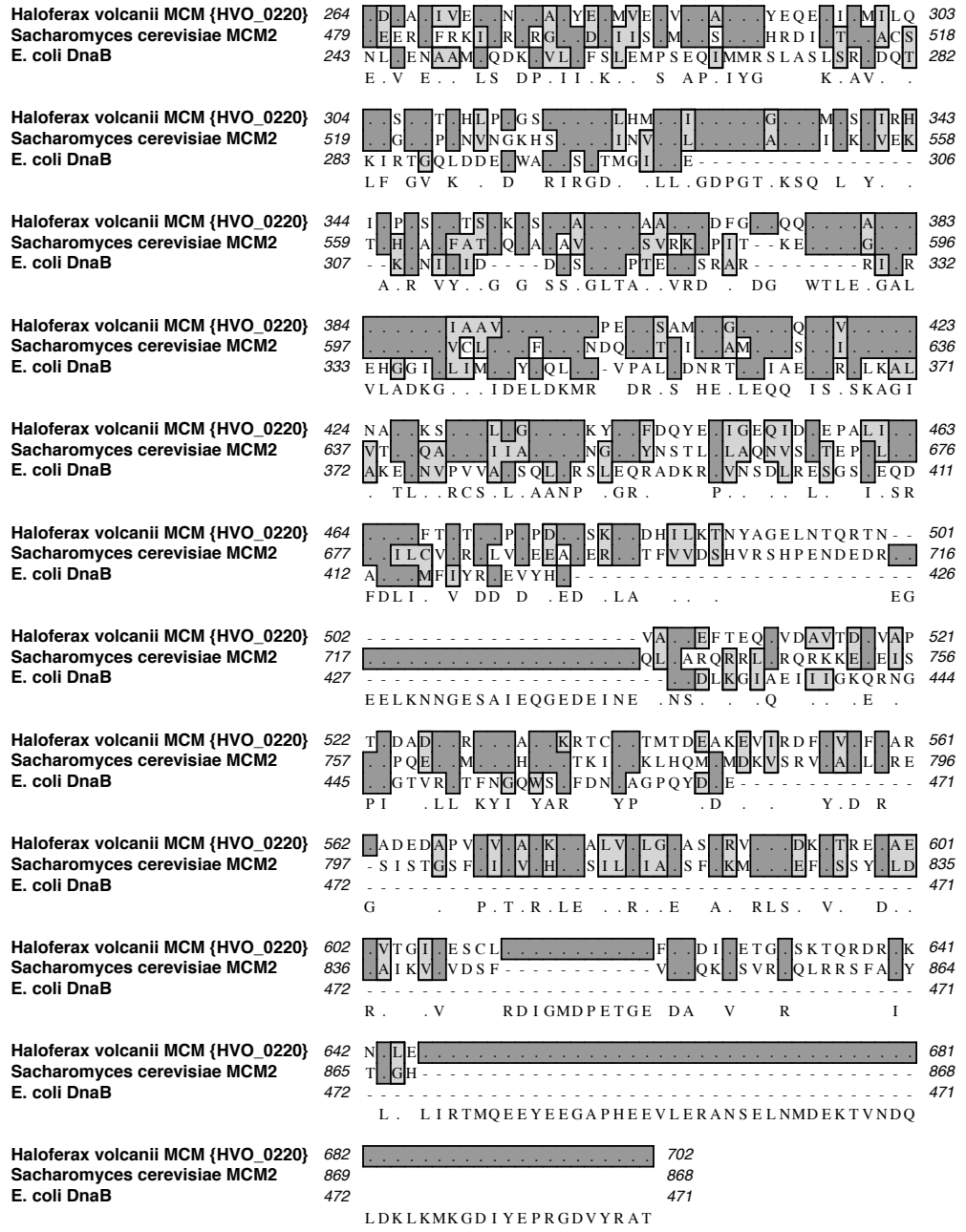
VI Sites of pHV3 Integration

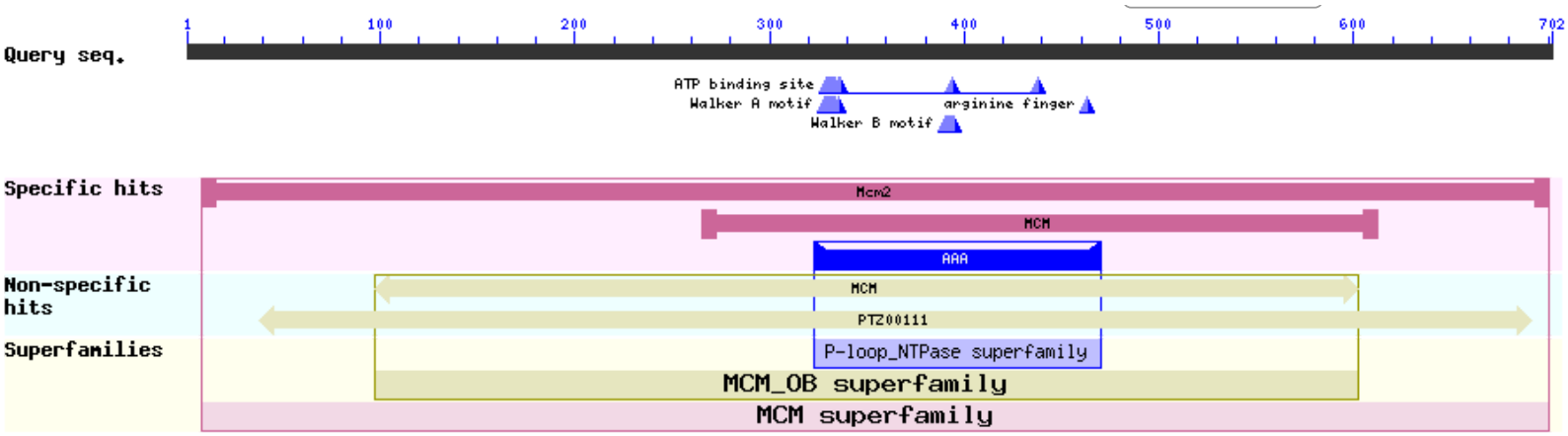
Sites of integration of pHV3 onto the main chromosome. (**A**) *adh2* region on pHV3. *adh2* was replaced with a *leuBΔ5'Dd* allele (pTA1621). (**B**) *leuB* region on the main chromosome. *leuB* deleted and replaced with *leuBΔ3'Bb* allele (pTA700). (**C**) Site of integration of pHV3 onto the chromosome with a functional *leuB* gene. (**D**) Non-functional *5'Δ3'ΔleuB* allele at the other site of pHV3 integration. Functional *leuB* is shown in green, non-functional *leuB* is shown in red, chromosome is shown in blue and pHV3 is shown in purple.

VII MCM

Alignments



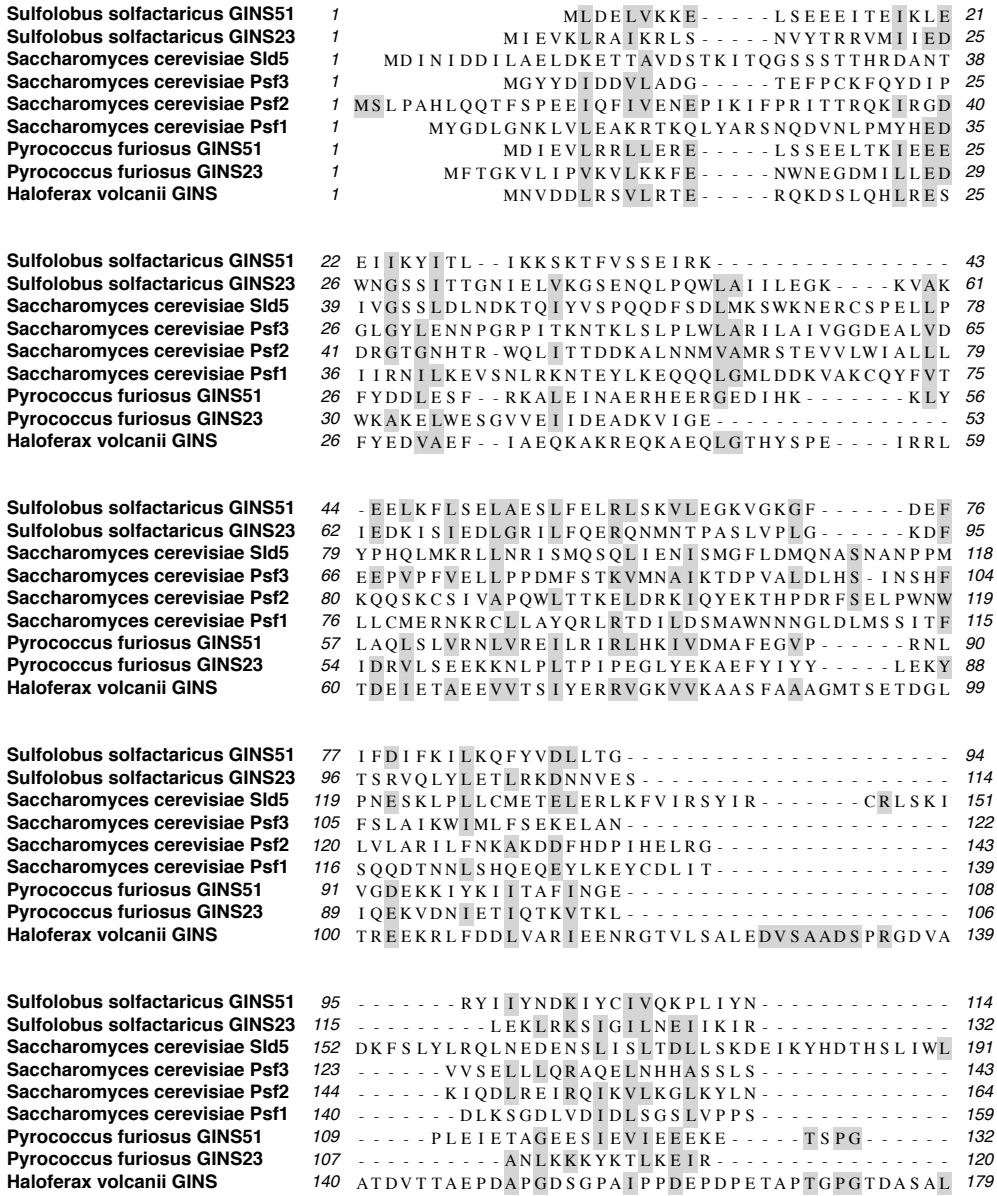

Haloferax volcanii MCM {HVO_0220}
Sacharomyces cerevisiae MCM2
E. coli DnaB



Domain search of *Haloferax volcanii* MCM (gene number HVO_0220). Domains show conserved MCM structure with a central AAA+ domain with Walker A and Walker B motifs for ATP binding and hydrolysis. Search was conducted via NCBI Conserved Domain (CD) search using translated protein sequence (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi). Diagram shows standard results.

VIII GINS

Alignments

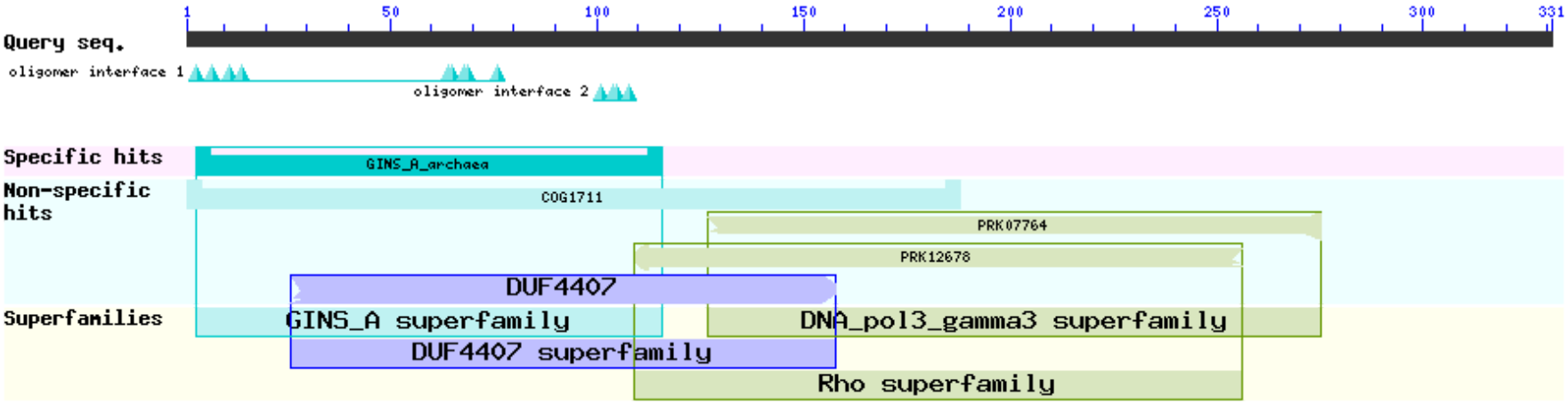


Sulfolobus solfactaricus GINS51	*115*	--	*114*
Sulfolobus solfactaricus GINS23	*133*	--	*132*
Saccharomyces cerevisiae Sld5	*192*	KLVNDSILKYMPEELQAINDTEG-----------------	*214*
Saccharomyces cerevisiae Psf3	*144*	--	*143*
Saccharomyces cerevisiae Psf2	*165*	--	*164*
Saccharomyces cerevisiae Psf1	*160*	--	*159*
Pyrococcus furiosus GINS51	*133*	--	*132*
Pyrococcus furiosus GINS23	*121*	--	*120*
Haloferax volcanii GINS	*180*	EAVTGDLAAGDVLADAMGGGTGGDASNAQSTPQSDADRPA	*219*

Sulfolobus solfactaricus GINS51	*115*	--	*114*
Sulfolobus solfactaricus GINS23	*133*	--	*132*
Saccharomyces cerevisiae Sld5	*215*	------------------------------------SVNM	*218*
Saccharomyces cerevisiae Psf3	*144*	--	*143*
Saccharomyces cerevisiae Psf2	*165*	--	*164*
Saccharomyces cerevisiae Psf1	*160*	--	*159*
Pyrococcus furiosus GINS51	*133*	--	*132*
Pyrococcus furiosus GINS23	*121*	--	*120*
Haloferax volcanii GINS	*220*	TEAAAVSGESGPSDSDAANSGSADSGPAIPPDGPDPETAA	*259*

Sulfolobus solfactaricus GINS51	*115*	---------------------------DHRVNEGDVLVL-	*126*
Sulfolobus solfactaricus GINS23	*133*	---------------------------LRKLIQLAFLNI-	*144*
Saccharomyces cerevisiae Sld5	*219*	IDEPDWNKFVFIHVNGPPDGKWNEDPLLQENEFGKPCYTV	*258*
Saccharomyces cerevisiae Psf3	*144*	--------------------------IDLNADSTGKNSA-	*156*
Saccharomyces cerevisiae Psf2	*165*	-------------------------ESHLQLDNLSLLEIN	*179*
Saccharomyces cerevisiae Psf1	*160*	---------------------------DVFIDVRVLKDAG	*172*
Pyrococcus furiosus GINS51	*133*	----------------------IIEAYLLRVDIPKILDE-	*149*
Pyrococcus furiosus GINS23	*121*	---------------------------FKKILEAVRLRP-	*132*
Haloferax volcanii GINS	*260*	PSAGDTDTVEPSADDPLSGLADERETVRITADVGTILGV-	*298*

Sulfolobus solfactaricus GINS51	*127*	PMREALPLIIASYLTP--------YKIDIEEQL*	*152*
Sulfolobus solfactaricus GINS23	*145*	DDQNLINGMTEEELLIYKTIKQLIKELYGDIIGNS*	*180*
Saccharomyces cerevisiae Sld5	*259*	TIPDLKEEVELTIGSIYVMRYEVIRDLLRDDKVALI	*294*
Saccharomyces cerevisiae Psf3	*157*	NTNIATSTFLLKLEEMEKEIYKKSHESYKDTKRWMFKK	*194*
Saccharomyces cerevisiae Psf2	*180*	ELRPFITEIMDKLREIHTASLTAGTENDEEEFNI	*213*
Saccharomyces cerevisiae Psf1	*173*	EIQTEYGVFNLIKDSQFFVRQSDVERLIQQGYLQKI	*208*
Pyrococcus furiosus GINS51	*150*	NLREYGPFKAGDLVVLPKSIGRVLIQRDAADKVLIQL*	*187*
Pyrococcus furiosus GINS23	*133*	NSMEILARLSPAEKRIYLEISKIRREWIGD*	*163*
Haloferax volcanii GINS	*299*	DEREY-DLATEDVVTLPAANADPLVERGAAERLD	*331*



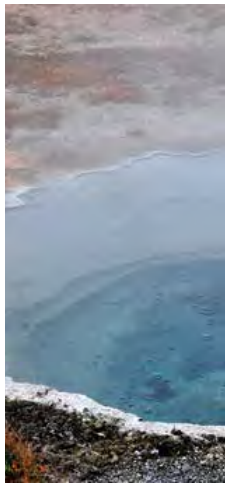
Domain search of *Haloferax volcanii* GINS (gene number HVO_2698). Domains show a conserved archaeal GINS A domain at the N-terminus. C-terminal domains show conservation with bacterial DNA polymerase III and the transcription termination protein Rho. Search was conducted via NCBI Conserved Domain (CD) search using translated protein sequence (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi). Diagram shows standard results.

Archaea and the meaning of life

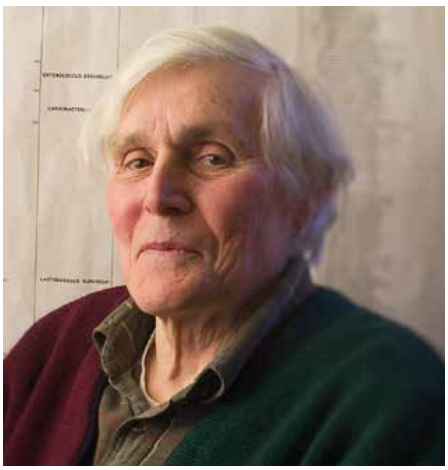
Hannah Marriott & Thorsten Allers



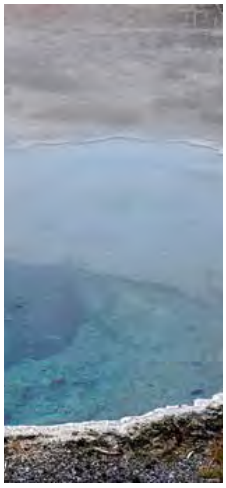




Humans are inordinately fond of dividing things into two – our approach to taxonomic classification is no exception. The earliest system published by Linnaeus divided living organisms into animals and plants, and by the 1960s this was superseded by a more fundamental split between prokaryotes and eukaryotes. So it should come as no surprise that the idea of a third domain of life – Archaea – met with fierce resistance when it was proposed in the 1970s.



Above **Carl Woese.** Don Hamerman, Institute for Genomic Biology, University of Illinois at Urbana-Champaign







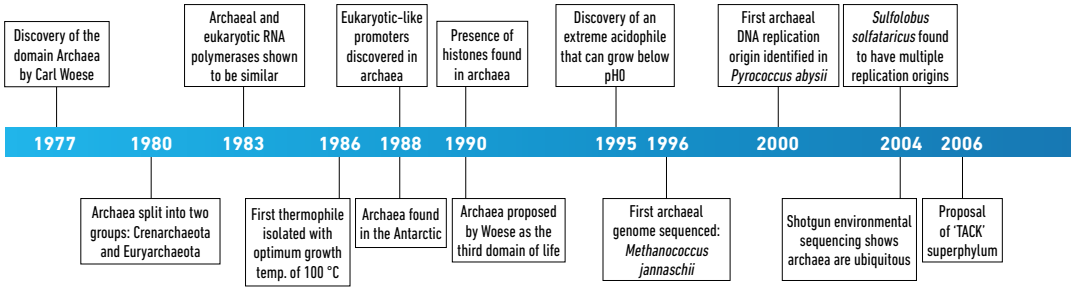
Geothermal areas and hot springs from Waiotapu geothermal area (Roturua, New Zealand) and Yellowstone National Park (USA). Far left **A mud pool, Waiotapu.** Second from left **Champagne Pool, Waiotapu.** Centre **A hot spring in Yellowstone National Park.** Above **Sulfur deposits, Waiotapu.** Right **Lady Knox Geyser, Waiotapu.** All Ian Haidl, except centre Sonja-Verena Albers

A brief history of Archaea: 1977 to present

Archaea are widespread on Earth yet relatively little is known about them, outside of the select groups of people that study these fascinating organisms. They are a mystery that is being slowly unravelled since their 'discovery' in 1977 by Carl Woese and his group, including George Fox, while working at the University of Illinois.

Methanogens (archaea that produce methane) and other groups of micro-organisms, including halobacteria (now called halophilic archaea) as well as thermophiles, had already been discovered, but they had been misclassified under the domain Bacteria. Woese found that these organisms did not just share a love for extreme environments, but also that they are phylogenetically related to each other. However, he was surprised to find that they are fundamentally distinct to bacteria.

Woese was using small-subunit rRNA (ribosomal RNA) to build a new phylogenic tree. Small-subunit rRNA is an essential component of all self-replicating organisms and shows remarkable sequence conservation. This made it a perfect choice for a molecular chronometer. At the time this molecular



Timeline of Archaea from their discovery to present day. Ian Haidl, Sonja Albers, Thorsten Allers

approach to phylogeny was novel – previous methods relied instead on visible characteristics such as cell shape or growth conditions. Woese found that the prokaryotes are not one coherent domain, but are made up of two distinct groups: Bacteria and Archaea. At the time they were named 'Eubacteria' and 'Archaebacteria', respectively, but these two groups of prokaryotes were found to be no more similar to each other than they were to eukaryotes. Woese proposed that the tree of life has three equal branches – Archaea, Bacteria and Eukarya – and that the term 'prokaryote' should be abandoned because it has no taxonomic meaning. Unsurprisingly, his ideas were not universally popular.

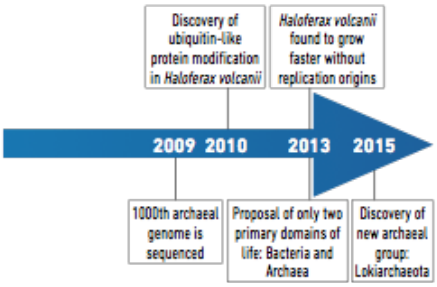
As with any new discovery there are sceptics, but biochemical data from Wolfram Zillig supported the 16S rRNA data collected by Woese. Over time, the new domain of Archaea was accepted by the scientific community. Interest in archaea increased further after whole genome sequencing took off in the 1990s, and researchers increasingly switched from bacteria or eukaryotes to working on these exotic micro-organisms. But contrary to popular belief, not all archaea are extremophiles. They have also been found in 'normal' environments such as soil and the ocean, and in environments where they cohabit with bacteria. For example, in the human gut archaea are responsible for producing methane! However, unlike bacteria, no pathogenic archaea have ever been found.

Almost 40 years have passed since their reclassification into a new domain but still many more species of archaea are being discovered. DNA sequencing has improved dramatically, meaning that archaea no longer have to be cultured to be characterised. This has led to the discovery of entirely new lineages. Based on phylogenetic data (16S rRNA and other genes) the domain Archaea was originally split into two groups; the Euryarchaeota and the Crenarchaeota. However, since 2006 three more lineages have been discovered: Thaumarchaeota, Aigarchaeota and Korarchaeota. These three new groups are often combined with the Crenarchaeota to form the 'TACK' superphylum (more on this later). Even more recently there have been reports of new lineages of 'nano' archaea, which are characterised by a small cell size with very few genes. The constant unearthing of new species and groups make the Archaea a very fluid domain, with the phylogenetic tree changing as each new discovery is made.

The origins of life?

Archaea may look like bacteria at a first glance and there are certainly many superficial similarities, but dig deeper and archaea have more in common with eukaryotes. In fact, it is now widely accepted that archaea are the ancestors of all eukaryotes.

Archaea, like bacteria, are single-celled organisms with a circular double-stranded DNA genome, and they have neither a nuclear membrane nor organelles. This means that they are similar to bacteria in terms of cell structure, although there are differences. Archaea do not have a bacterial-style cell wall and their plasma membrane is different to that found in both bacteria and eukaryotes. But on the inside of the cell, archaea show a striking family resemblance to eukaryotes. This is especially so for the enzymic machinery that processes genetic information – DNA packaging and replication, transcription into RNA, and translation into protein. All of these processes are essentially the same in archaea and eukaryotes, and are quite distinct from bacteria.



This prompts the question: if archaea are more closely related to eukaryotes than bacteria, then how do they fit in the tree of life?

The phylogenetic tree proposed by Woese was split into three equal domains, with Archaea and Eukarya sharing a common ancestor that had already diverged from Bacteria. However, biologists working on the origins of life have recently concluded that Eukarya and Archaea are not sister groups. Instead, eukaryotes are the direct descendants of archaea, and our long-lost ancestor belongs to the 'TACK' superphylum of Archaea. One of the most exciting new discoveries of the last year was the identification of a 'missing link' between Eukarya and Archaea. Called Lokiarchaeota, they were found near a hydrothermal vent at a site known as Loki's Castle in the Arctic Ocean.

Can we use our knowledge of archaea to trace the origins of complex life? Eukaryotic microfossils can be dated back to 1.8 billion years ago but biological methane has been found in rocks that are 3.4 billion years old. The only source of biological methane is methanogenic Euryarchaeota, so we know that archaea have been around since the very beginnings of life on Earth. As for life on other planets, it is tempting to speculate that archaea may have also colonised Mars – evidence is mounting that methane in the Martian atmosphere has a biological origin.

Woese's revolution

What do the recent discoveries mean for us humans? Given our desire for dichotomy, we should be relieved that the tree of life has been pruned back to just two primary branches – Bacteria and Archaea. And more than anything else, we should give credit to Carl Woese, whose taxonomic revolution has allowed us to trace our ancestors all the way back to their humble beginnings as archaeal cells.

Hannah Marriott & Thorsten Allers
School of Life Sciences, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK

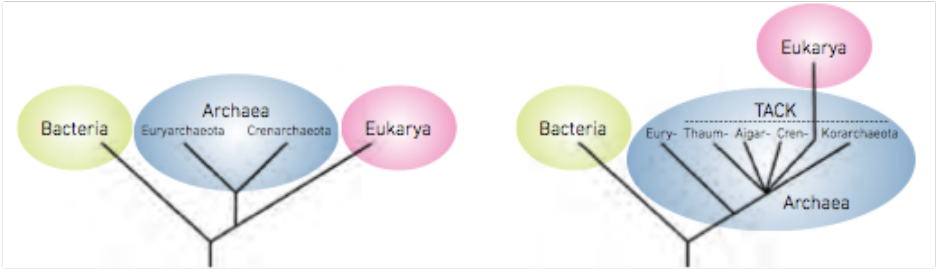
Further reading

Eme, L. & Doolittle, W. F. (2015). Archaea. *Curr Biol* **25**, R845–R875.

Spang, A. & others (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179.

Williams, T. A. & others (2013). An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* **504**, 231–236.

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 USA* **87**, 4576–4579.



Universal phylogenetic trees in rooted form, showing the three domains. Ian Haïdi, Sonja Albers, Thorsten Allers