Origin-independent replication in *Haloferax volcanii*: RadA localisation, R-loops, and genomic rearrangements

Katie Hodson, BSc (Hons)

Student ID: 14322881

Thesis submitted to the University of Nottingham for the degree of Master of Research

Initial submission: 30.09.2022

Submission with corrections: 24.01.2023

University of Nottingham

UK | CHINA | MALAYSIA

Contents

Abstract

DNA replication typically initiates at origins, DNA sequences at which replication initiator proteins bind and instigate formation of the DNA replication machinery. It was thought that origins of replication are essential across all domains of life. However, the halophilic archaeon *Haloferax volcanii* can survive in the absence of origins, and in fact, strains lacking origins grow more quickly than the wild-type strain. When all origins are deleted, the archaeal RecA family recombinase RadA becomes essential. It is for this reason that it is thought that in the absence of origins, *H. volcanii* employs recombination-dependent replication.

In this study, Chromatin Immunoprecipitation Sequencing (ChIP-Seq) was undertaken using FLAG-tagged RadA, in a strain in which all origins were present on the main chromosome. This revealed a relative lack of RadA binding upon mini chromosome pHV3, extensive binding across mini chromosome pHV1, and a greater proportion of peaks mapping to promotor-transcription start site regions than to transcription termination sites or within exons. These findings demonstrate a tentative link between transcription and recombination; however, aspects of the protocol require optimisation.

Prior to this study, *H. volcanii* strains were generated in which the level of transcription of the mini chromosome pHV3 was dramatically increased; it was then found that the origin of replication on pHV3 may be deleted. This work established that in these strains, pHV3 had integrated onto the main chromosome – a phenomenon not previously seen in attempts to delete the pHV3 origin. This suggests that significant changes in transcription may lead to genome rearrangements. Oxford Nanopore MinION sequencing showed that integration of pHV3 on the main chromosome most likely occurred via recombination between two homologous ISH18 elements (transposase genes), one on pHV3 and the other on the main chromosome. Similar recombination events leading to large scale rearrangements of the *H. volcanii* genome architecture have been noted previously.

Given that ChIP-Seq revealed a lack of RadA binding to wilt-type pHV3, it could be hypothesised that the significant increase in transcription of pHV3 had led to an increase in RadA binding, thereby facilitating recombination of pHV3 with the main chromosome.

Overall, this study shows the interplay between recombination, transcription, and origin-independent replication. This study also demonstrates the efficacy of Oxford Nanopore MinION sequencing for *H. volcanii* strain verification.

Acknowledgements

Firstly, I would like to thank Thorsten, Ambika, and Stephen. Thorsten, your teaching and guidance have been invaluable; I've learned so much. I'm extremely grateful for the opportunity and for your belief in me as a student. Thank you for always entertaining my scientific queries like "why are there no pathogenic archaeal species?" and obviously most importantly, for wearing such cool outfits. Thank you for everything. Ambika, thank you for never getting tired of my questions and for helping me to develop belief in my own abilities. I would never have gotten so much out of the project without you. Diya is going to be an expert in all things archaeal by the time she even goes to school. Stephen, thank you for your support! I also appreciate you taking the time to explain the cool things you have going on in your lab.

Thank you also to the other members of the Allers lab! In particular Laura, thank you for all of your help, both in the lab and with so much more. I will forever be jealous of your radiation skills. Thank you also to Cath, Callum, Andy and Sophia.

To everyone in D119, I've had a great year and I'm grateful to have met all of you! Ryan, you've been so kind and I'm thankful to have had you in the lab for the past year. Thank you also to Harriet, Mahari, Nikita, Barbora, and Hanan.

Thank you to my friends and family outside of the lab. Amy, I'm not sure that anything I could write would convey how grateful I am. Thank you for always being there. Thank you to Sian for being so supportive and understanding, and to Roo for being the most incredible sister ever. Dad, Chantelle, Anita and Brian, thank you so much for all of your support.

A final thanks to Hayley Ferax (even despite the genomic rearrangements).

Chapter 1- Introduction

1.1- Archaea

Until 1977, it was thought that the tree of life comprised two major branches: prokaryotes and eukaryotes. Work by Carl Woese and George Fox on small-subunit ribosomal RNA (16S rRNA) lead to the reclassification of prokaryotes into two distinct domains, namely, Bacteria and Archaea, and a tree of life instead consisting of 3 primary domains: Eukarya, Bacteria, and Archaea[1] (Figure 1.1). This was later-supported by work from Wolfram Zillig on RNA polymerases[2] and Otto Kandler on cell membranes[3]. Additionally, further work showed that despite Archaea having previously been misclassified as Bacteria, Archaea were genetically more similar to Eukarya[4]. Archaea and bacteria are both prokaryotic cells and they share morphological features, but the cellular machinery used to process information, such as the proteins involved in DNA replication, transcription, and translation, is more similar to simplified eukaryotic mechanisms[5,6].

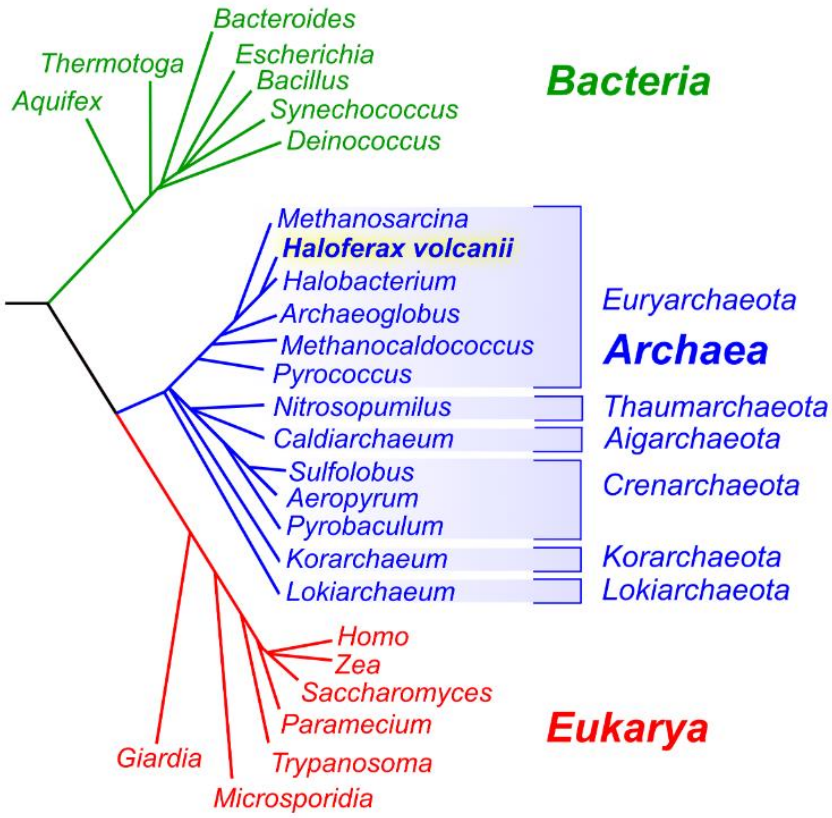


Figure 1.1: The proposed three-domain tree of life. 16S rRNA sequencing defined three domains: Bacteria, Archaea, and Eukarya.

Figure adapted from [7].

In 1991, additional work on 16S rRNA lead to the discovery of two distinct phyla within the archaeal domain: Euryarchaeota and Crenarchaeota[8]. However, this has now been revised as further phyla have been discovered. It is thought that a superphylum, known as TACK, exists and is recognised as a cluster of four main phyla (Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota)[9]. More recently, additional lineages falling within the Lokiarchaeota clade have been identified[10,11] and metagenomic profiling has identified further phyla such as Thorarchaeota, Heimdallarchaeota and Odinarchaeota[10,12]. The Thorarchaeota phylum was first described upon phylogenetic analysis of rRNA genes obtained from the sulphate-methane transition zone in estuary sediments[12]. Metagenomes obtained from hot springs alone contained contigs corresponding to species falling within the Odinarchaeota, whereas contigs corresponding to species falling within the Heimdallarchaeota were identified in marine sediments such as Loki's Castle and Aarhus Bay[10]. Together, these phyla form the Asgard superphylum (Figure 1.2) and within it, Lokiarchaeota represent the closest ancestor to eukaryotes which provides support for the two-domain tree of life. Proteins have been found in species falling within the Asgard superphylum that were previously thought to be specific to eukaryotes[10].

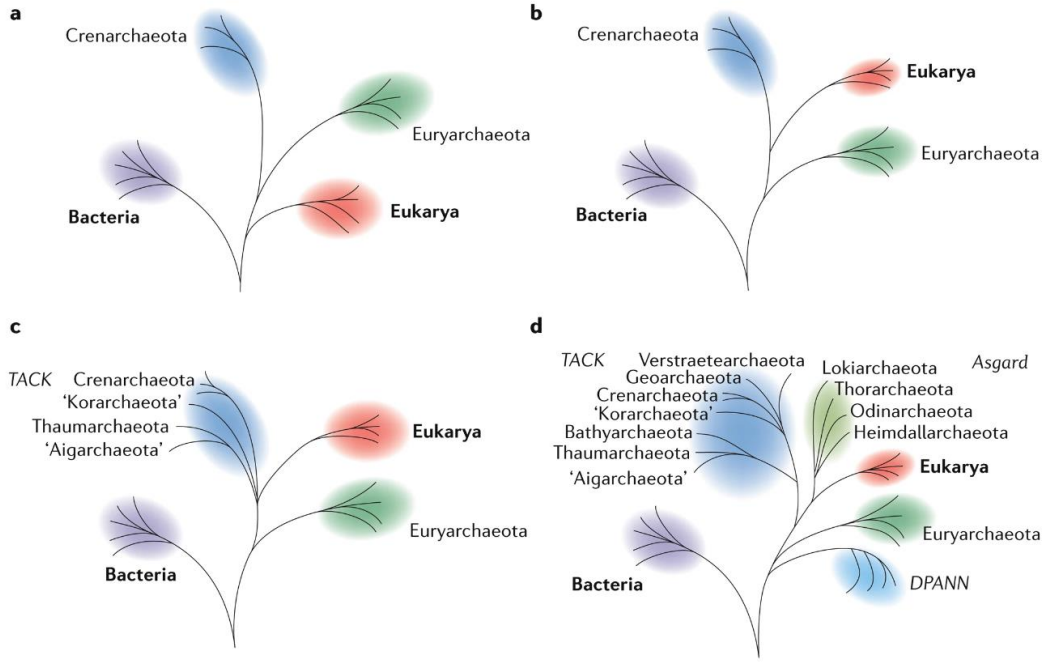


Figure 1.2: Changes in our understanding of the tree of life. Bacteria and Eukarya are denoted in purple and red, respectively, and archaeal lineages in blue and green. **a**) The three-domain tree in which Archaea and Eukarya share a unique common ancestor, not shared by Bacteria. **b**) A two-domain tree of life proposed when the Archaea was thought to only comprise Euryarchaeota and Crenarchaeota phyla. **c**) The remodelled tree of life proposed in the 2010s upon the finding that eukaryotes branch within or as a sister to the TACK superphylum. **d**) The tree of life proposed when Eukarya were strongly suggested to branch from within or as a sister to the ASGARD archaea.

Figure taken from [13].

The most current phylogenetic tree consists of eukaryotes descending directly from or as a sister to the Asgard superphylum, and is therefore a tree consisting of two domains, Bacteria and Archaea, contrasting the three-domain tree proposed by Woese and Fox [13] (Figure 1.2). This hypothesis is derived from comparisons between eukaryotic signature proteins including those involved in ubiquitin signalling, trafficking machinery, methods of RNA interference and cytoskeletal structures (Figure 1.3).

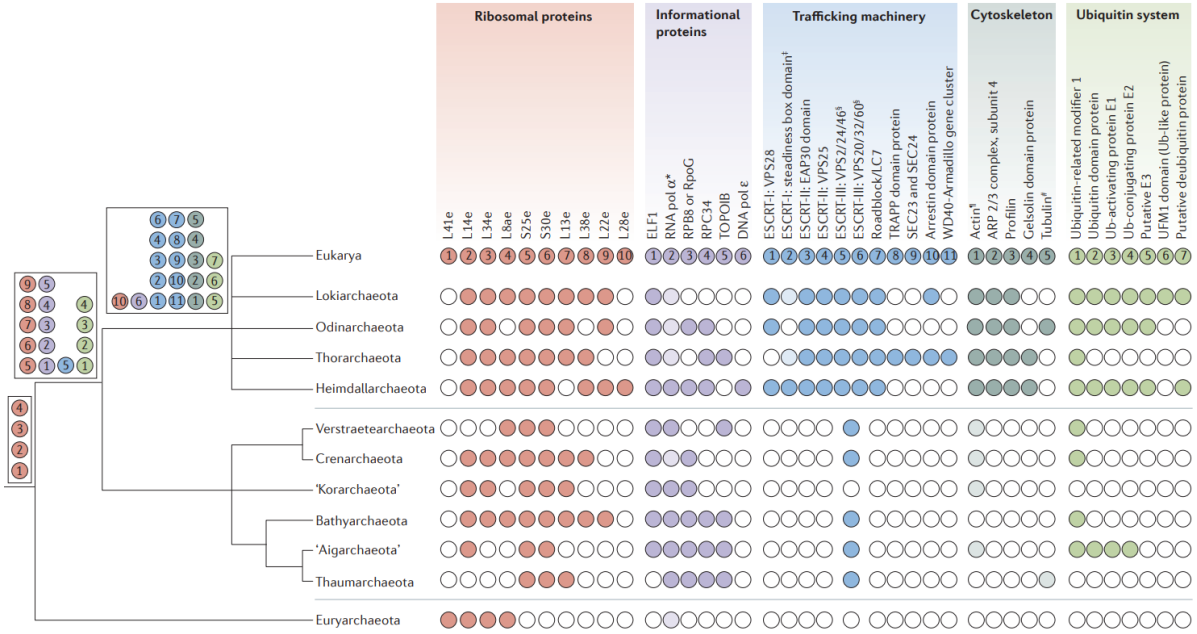


Figure 1.3: Presence and absence of eukaryotic signature protein homologues in archaeal lineages. Filled circles represent the presence of a homologue and the figure demonstrates the estimated time of emergence upon the phylogenetic tree on the left-hand side.

Figure taken from [13].

However, this is still disputed, and recent studies exist which provide support for the three-domain tree of life theory in which Archaea and Eukarya are independent monophyletic lineages[14-16]. These disputes may in part be due the lack of cultured species within the Asgard superphylum. In 2020, work by Imachi *et al*.[17] first demonstrated isolation and cultivation of an Asgard archaeon related to the Lokiarchaeota named "*Candidatus* Prometheoarchaeum syntrophicum". The archaeon's distinct metabolism and morphology allowed a hypothetical model of eukaryogenesis to be proposed named "the entangle-engulf-endogenize model" (Figure 1.4). Further cultivation of Asgard archaea may provide greater insight into the emergence of eukaryotes.

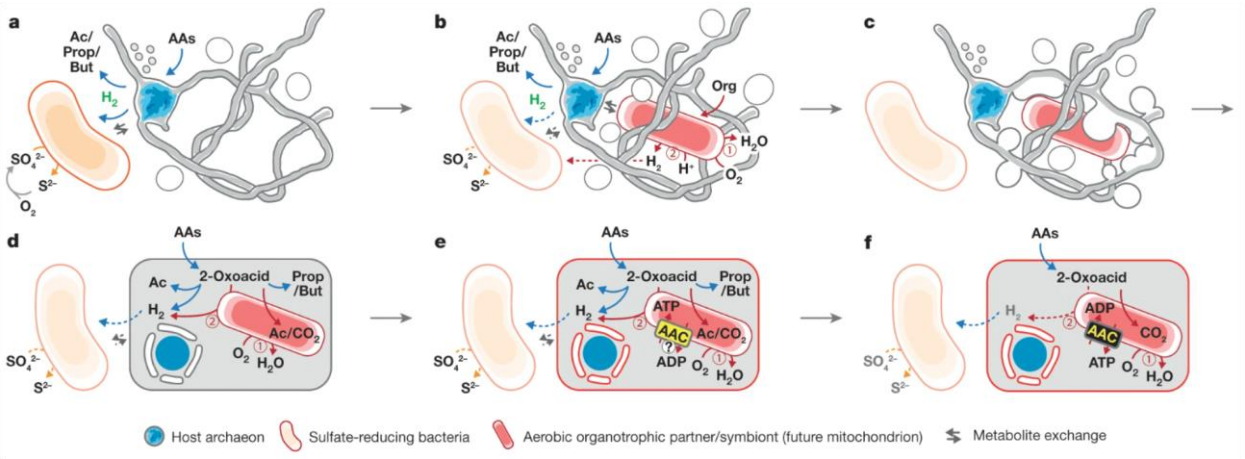


Figure 1.4: Stages in the "entangle–engulf–endogenize" model of eukaryogenesis hypothesized by Imachi *et al.*, 2020[17]. Figure taken from the same reference.

Characteristics of archaea

The Archaea comprise very diverse organisms and many archaeal species are able to survive in harsh environments including extremes in salinity, acidity, alkalinity, and temperature. This has led to archaea often being considered extremophiles which, although in some circumstances is true, is not always the case. Archaea can often be found in soil, fresh-water sediments, and the human gut[18,19].

Like bacteria, archaea are single-celled organisms with circular chromosomes. They do not have organelles and they do not possess nuclear envelopes. Both archaeal and bacterial genomes are normally 2-4 Mb in size and genes are organised into operons. Compared to eukaryotic and bacterial genomes, archaeal genomes appear to have a higher gene density[20]. Species in all three domains also have bidirectional origins of replication, with some archaeal species possessing more than one origin[20,21]. However, bacterial and archaeal cells membranes differ; bacterial cell membranes are protected with a peptidoglycan cell wall whereas most archaeal cell membranes consist of a single layer coated with a proteinaceous surface layer (S-layer) which is subject to glycosylation. Unlike the fatty acid side chains found in eukaryotic and bacterial cell membranes, archaeal cell membranes contain isoprenoid hydrocarbon side chains[22]. In bacteria and eukaryotes, fatty acid side chains are linked via and ester bond to the *sn*-glycerol-3-phosphate backbone whereas in archaea, isoprenoid hydrocarbon side chains are linked via an ether bond to the *sn*-glycerol-1-phosphate backbone.

The majority of archaeal species falling within the Euryarchaea, Ehaumarchaea, Eigarchaea, Eorarchaea, and Eanoarchaea package their genomes using histones. Histones can also be found in a minority of crenarchaeal genomes[23]. Archaeal histones are homologues of the H3 and H4 proteins found in eukaryotic histones, and bind to DNA as tetramers, each protecting around 60bp of DNA[24]. Unlike the

uniform discrete structure formed by eukaryotic histones, archaeal histones can form a range of histone-DNA complexes. *Haloferax volcanii* has tetrameric histones evenly spaced along the genome[25], whereas histones in *Thermococcus kodakarensis* can form multimers capable of protecting 500bp of DNA[26]. Due to this, the archaeal nucleosome can be flexible but may be less stable than the structures seen in eukaryotes. In the crenarchaea, Alba, a 10 kDa protein able to bind both DNA and RNA, is commonly found responsible for chromatin organisation. Alba normally acts as a homodimer and in addition to its role in genome compaction, it is thought to coat DNA as a filament to protect from damage by nucleases[27]. Other archaea-specific chromatin-binding proteins are used in some species[28].

In archaea, transcription and translation are coupled as seen in bacteria[29]. However, archaeal RNA polymerase (RNAP) shows more similarity to the eukaryotic homologue. Work using antibodies[2], sequence analysis[30], and X-ray crystallography (reviewed in [31]) demonstrated the similarities between archaeal and eukaryotic RNAP. RNAPs in all domains of life have a “crab-claw” shaped structure with the largest 2-3 subunits forming a cleft to bind dsDNA, the area around which is highly conserved. However, archaeal and eukaryotic RNAP have a protruding “stalk-like” structure which cannot be found in bacterial RNAP (Figure 1.5). In archaea, a TATA-box-binding protein (TBP) and transcription factor B (TBF) are required to initiate transcription, both of which have eukaryotic homologues[32]. Furthermore, archaeal translation is more similar to that in eukaryotes than in bacteria. Archaea have at least 5 translation initiation factors, all of which share homology with those found in eukaryotes[33].

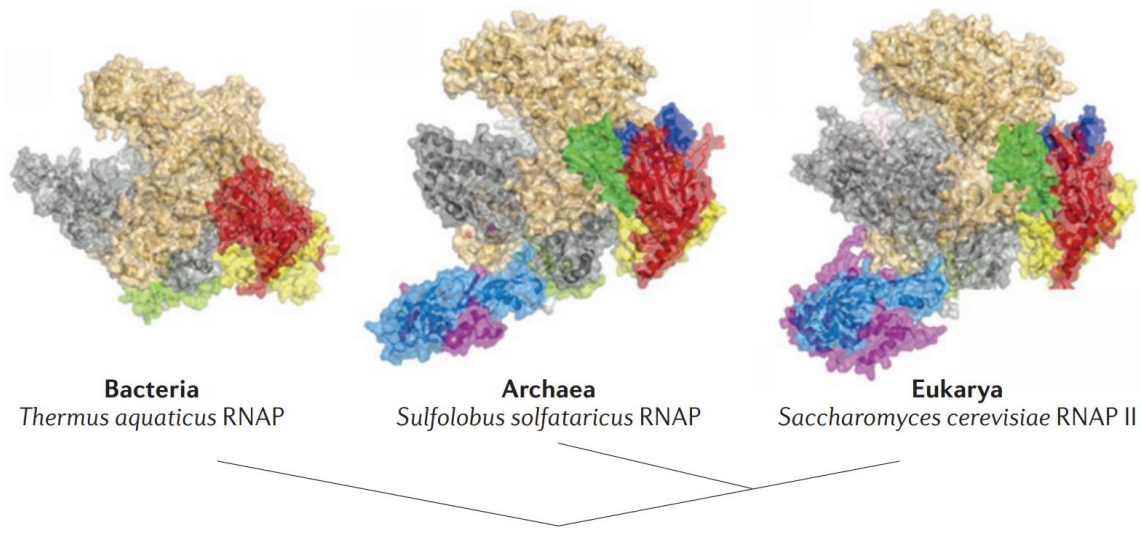


Figure 1.5: Structure of RNA polymerases (RNAPs) across the domains of life. Homologous subunits are shown in the same colour. Similarities in the structure of RNAPs across all domains of life can be seen, particularly those between archaeal and eukaryotic RNAP which have the protruding “stalk-like” structure not found in bacterial RNAP.

Figure taken from [34].

1.2- *Haloferax volcanii* as a model organism

Haloferax volcanii is a halophilic Euryarchaeon initially isolated from the Dead Sea in 1975[35]. It grows aerobically at 45°C in a concentration of 1.7-2.5 M NaCl upon both rich and minimal media. It has a generation time of 2-3 hours and forms colonies upon agar plates within around 4-5 days. *H. volcanii* cells are disc-shaped but become rod-shaped when mobile[36]. Growth occurs through elongation of the cell and reproduction occurs through binary fission[35]. Gene transfer between cells is also possible through natural mating via intracellular bridges[37-39]. *H. volcanii* uses a "salt-in" mechanism to survive in hypersaline environments[40]. The internal salt concentration is maintained to be the same concentration as the external environment and therefore, it is necessary for proteins to fold in the high ionic strength of the cytoplasm. Due to this, *H. volcanii* proteins contain a high number of acidic residues and are negatively charged on their surface, while hydrophobic residues folded into the protein's core (an example of which being the PCNA sliding clamp as described in[41]). This does mean, however, that *H. volcanii* proteins denature easily in low salt environments[42].

Shortly after its isolation, techniques were developed for the transformation and genetic manipulation of *H. volcanii*, allowing it to be used to study halophilic archaea[39,43,44]. More recently, additional genetic, molecular and biochemical tools have been established, allowing *H. volcanii* to become a well-established model for the study of DNA replication, repair, and homologous recombination[45-47]. Both auxotrophic and antibiotic selection, inducible gene expression using tryptophan-inducible and strong constitutive promotors, and the use of the native CRISPR system as a method of gene interference (CRISPRi) are examples of such tools, and a plethora of other techniques have also been developed, as mentioned in [45].

The wild-type *H. volcanii* genome (strain DS2) consists of a 2.85 Mb circular main chromosome, three circular mini chromosomes, namely pHV1 (85 kb), pHV3 (438 kb), and pHV4 (636 kb), and a 6.4 kb plasmid pHV2[48]. DS70, the ancestor of the *H. volcanii* laboratory strain H26 and its derivatives have been cured of pHV2[43]. In this process, unintentional stable integration of pHV4 onto the main chromosome occurred and is therefore also observed in all H26 derivatives[49]. This altogether results in a genome consisting of a 3.5 Mb main chromosome and mini chromosomes pHV1 and pHV3 (Figure 1.6). *H. volcanii* genome is GC rich (approximately 65%) and is highly polyploid with around 20 copies per cell [50,51]. *H. volcanii* is not thought to have a defined cell cycle and DNA replication profiling using deep sequencing has demonstrated non-synchronous origin firing [52]. Briefly, concurrent rounds of replication are carried out. An origin fires extending replication forks in opposite directions and may prime to fire again before the terminus has been replicated by the initial forks.

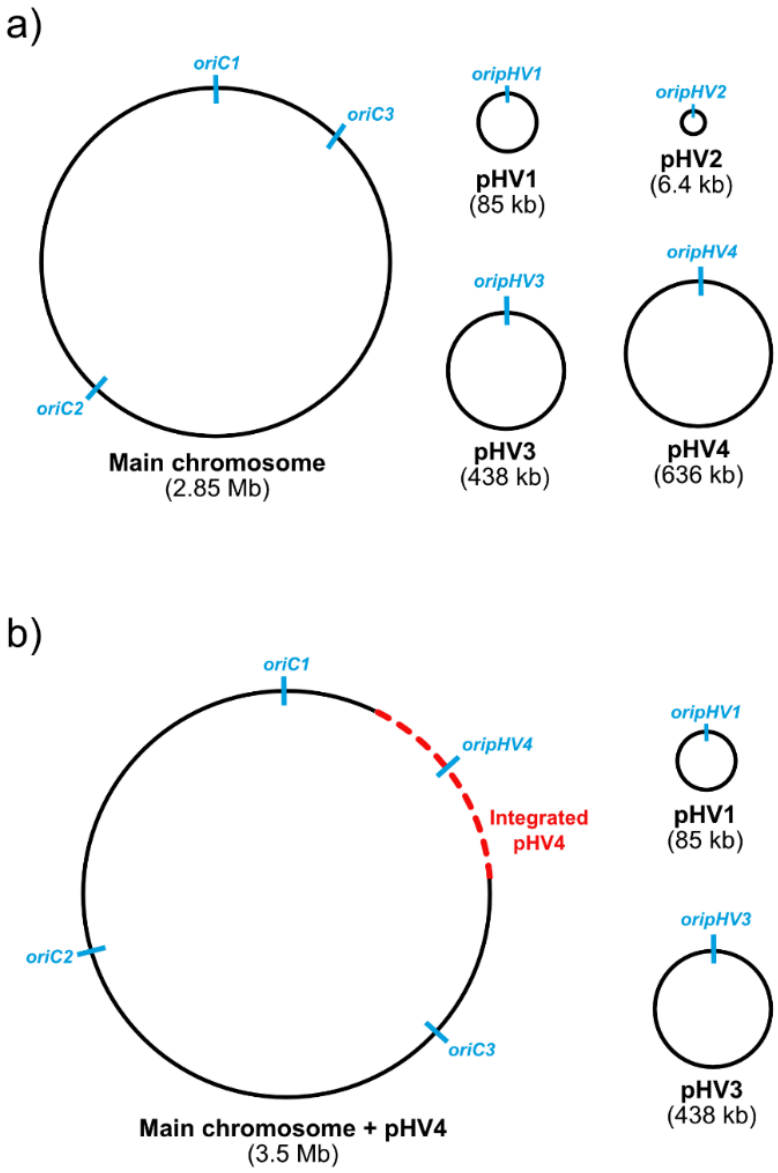


Figure 1.6: *Haloferax volcanii* genome architecture and origins of replication. a) Genome architecture of the *H. volcanii* DS2 wild isolate. b) Genome architecture of the H26 laboratory strain. pHV2 has been cured and pHV4 is integrated onto the main chromosome (red dotted line).

Six replication origins are present in the wild-type *H. volcanii* genome, each of which are adjacent to a gene encoding an Orc1/Cdc6 initiator protein[49,53]. In the laboratory strain, four active origins of replication are present on the main chromosome: *oriC1*, *oriC2*, *oriC3*, and *oripHV4.* Mini chromosomes pHV1 and pHV3 also have their own origins, namely *oripHV1* and *oripHV3*[48,53] (Figure 1.6). The origins have varying levels of activity with *oriC1* being the most active [49]. Deletion of *oriC1* and the adjacent initiator protein Orc1 results in reduced ploidy[54].

The *H. volcanii* genome architecture is notably fluid and is able to withstand large scale rearrangements. As mentioned previously, in the laboratory strain, pHV4 is found integrated onto the main chromosome. Integration occurred through recombination between homologous ISH18 elements[49]. Also, in an effort to delete *radA* in a strain in which *oriC1, oriC2 and oriC3* had previously been deleted, inversion of a region of integrated pHV4 was observed, again as a result of recombination between ISH18 elements[49]. A similar stable genomic rearrangement has been observed in which recombination between two identical *sod1* and *sod2* genes (with non-identical upstream and downstream regions) formed a new mini-chromosome[55].

1.3- Canonical DNA replication

Accurate DNA replication is a fundamental cellular process required to ensure that when a cell divides, each daughter cell contains a complete and correct copy of the genome. Therefore, diverse mechanisms have developed to ensure the process is well co-ordinated. Across all domains of life, DNA replication can be divided into three stages: initiation, elongation, and termination.

Initiation of DNA replication begins with an initiator protein binding to an origin of replication: a specific *cis*-acting sequence containing one or more AT-rich regions, termed a duplex-unwinding elements (DUEs), with weak hydrogen bonding to facilitate unwinding of the duplex (reviewed in [56]). A helicase is then recruited which upon activation, unwinds the DNA duplex further. Single-stranded DNA (ssDNA) is protected by single-stranded DNA binding proteins (SSBs) which later aid in the recruitment of further elements of the replisome complex. Bidirectional DNA synthesis is initiated once the replisome complex has formed. DNA polymerases prime synthesis on the leading strand in the 5'-3' direction from short RNA fragments generated by primases. On the lagging strand, replication is discontinuous; Okazaki fragments are formed which are later joined together to create a continuous strand. Clamp loader proteins, key elements of the replisome, load sliding clamp proteins which act as processivity factors for DNA polymerase and as molecular toolbelts. Replication forks meet and are resolved in the termination stage. During termination, remaining gaps are filled and ligated, the replication machinery is disassembled, and DNA catenanes are resolved[57]. Bacterial genomes contain termination (*Ter*) sites which block the replication machinery and halt replication. However, eukaryotes and archaea do not have genetic sequences determining replication termination. Instead, termination occurs stochastically upon the meeting of two replication forks [49,57,58]. Upon completion of replication, chromosomes can segregate.

Origins of replication

Regulation of DNA replication through origins ensures the ploidy of an organism is maintained. In eukaryotes, it is essential that the genome is replicated only once per cell cycle as otherwise, aneuploidy, genome instability and genome rearrangements may occur. Additionally, in bacteria, increased double strand breaks (DSBs) can occur due to over-replication of the genome, the consequences of which could be catastrophic to the cell[59-61].

In general, bacteria initiate replication from a single origin whereas eukaryotes have multiple origins per chromosome. The number of replication origins in archaea varies between species: *Pyrococcus abyssi* has one origin of replication but *Haloferax volcanii* has multiple[53,62]. Due the strict regulation of replication by origins and the consequences known to arise when correct ploidy is not maintained, it was thought that replication in their absence would not be possible. However, replication in the absence of origins is possible, and can result in faster growth than seen when origins are present. In *H. volcanii,* all origins on the main chromosome can be deleted and cells grow 7.5% faster than wild-type (discussed further in Section 1.4.3- Recombination-dependent replication)[49]. Additionally, origin deletion in *E. coli* is possible in cells with a mutated form of the gene encoding RNase HI but, unlike in *H. volcanii,* cell growth and viability is reduced[60]. Origin-independent replication is discussed later in this chapter.

Initiation of DNA replication in bacteria

In bacteria, DNA replication is initiated through the binding of the initiator protein, DnaA, to the origin, *oriC,* at DnaA box sequences (motifs with TT(A/T)TNCACA consensus sequences)[63]. DnaA is a member of the AAA+ (ATPases Associated with diverse cellular Activities) family of proteins[64]. The binding of DnaA opens the DNA duplex and creates a ssDNA bubble to which SSB binds. The formation of the ssDNA bubble by DnaA allows DnaC, a helicase loader also of the AAA+ protein family, to bind. DnaC acts as a chaperone for DnaB, a replicative helicase, which subsequently binds to the lagging strand. A single hexamer of DnaB is loaded per replication fork[6]. The activation of DnaB is reliant upon DnaC: if ADP is bound to DnaC, DnaB will be active, but if ATP is instead bound to DnaC, DnaB will be inactive. Although DnaC-ATP inhibits helicase action of DnaB, it is necessary for DnaC to bind ssDNA initially. DnaC-ATP is not necessary in the recruitment of DnaB, and ssDNA and DnaB act as cofactors to trigger the hydrolysis of ATP. Therefore, it is thought that DnaC acts as a dual-switch protein as both the ATP and ADP bound forms are required for initiation of replication sequentially[65,66].

Once activated, DnaB can further unwind the DNA duplex and additional components of the replisome can be recruited including primase, sliding clamp b-clamp and DNA polymerase. As replication from an origin is bidirectional, two replication forks are formed moving away from each other, and therefore two replisomes are assembled at the origin.

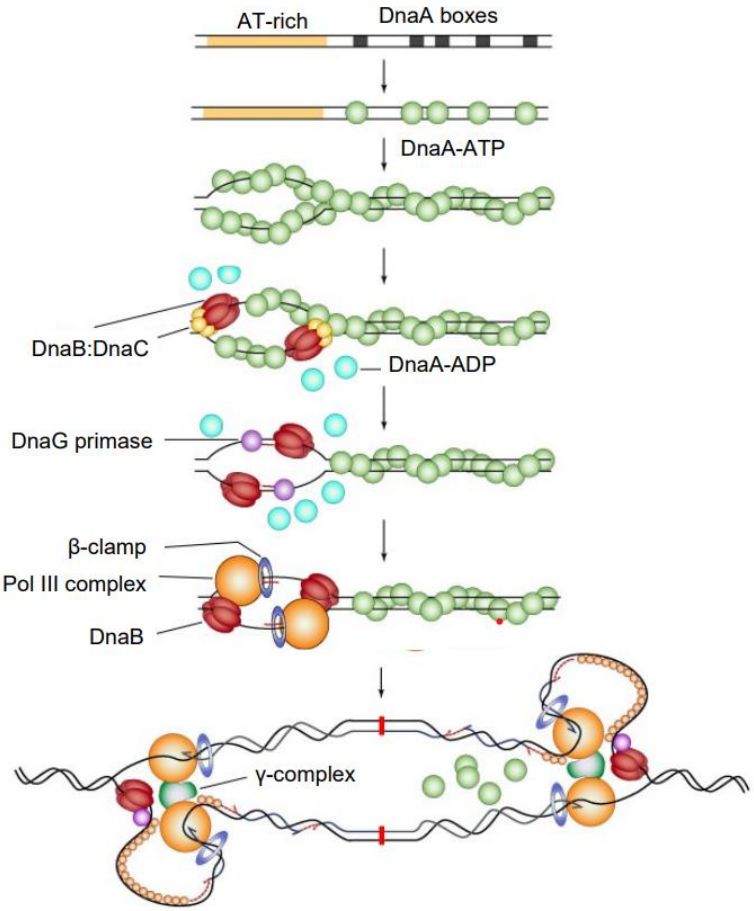


Figure 1.7: Replication initiation in *Escherichia coli*. DnaA binds at DnaA boxes, the duplex begins to unwind, and further proteins are recruited leading to replisome formation.

Figure created by Thorsten Allers.

Initiation of DNA replication eukaryotes

Eukaryotic DNA replication differs from and is more complex than DNA replication in bacteria. Eukaryotic DNA replication is initiated through binding of an origin recognition complex (ORC) to an origin of replication, many of which can be found along the length of linear eukaryotic chromosomes. The larger genome size and slower progression of replication forks in eukaryotes compared to bacteria mean that multiple origins of replication are necessary. The ORC complex comprises six origin recognition proteins (Orc1-Orc6) with proteins Orc1-5 facilitating binding to the origin through winged-helix domains in an ATP-dependent manner during G1-phase[67,68]. During S-phase, the ORC, regulator cell division cycle 6 protein (Cdc6), and licensing factor Cdc10-dependent transcript 1 protein (Cdt1) load the replicative helicase mini-chromosome maintenance (MCM2-7) to form the pre-replicative complex (pre-RC). MCM is a heterohexameric helicase encoded by six paralogous genes (MCM2-7). Like DnaB, the replicative helicase used in bacterial

DNA replication, each of the proteins forming the MCM are also part of the ATPase AAA+ family[69,70]. Many MCM proteins can associate with the replisome whereas in bacteria, a single helicase is loaded per replication fork. For its protection, ssDNA becomes bound by replication protein A (RPA), the eukaryotic SSB protein. When recruited to the pre-RC, MCM is inactive. ORC, Cdc6, and Cdt1 dissociate once MCM has been loaded. Activation of the pre-RC into a replisome, via activation of MCM is coordinated by cyclin-dependent kinase (CDK) and Dbf4-dependent kinase (DDK) which leads to recruitment of proteins Cdc45 and GINS. Together with MCM, these proteins form the CMG complex (Cdc45-MCM-GINS). Once the CMG complex is assembled, MCM switches from binding dsDNA to ssDNA and becomes activated[71-73]. Further elements of the replisome are recruited including DNA polymerases, primases and clamp proteins, and once fully formed, the replisome initiates replication in a 5' to 3' direction away from the origin. -

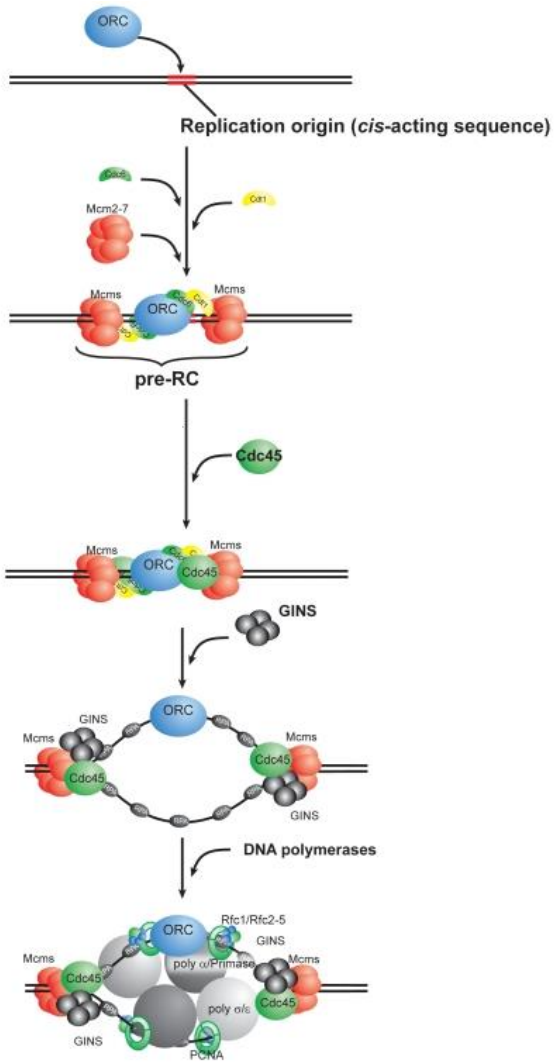


Figure 1.8: Replication initiation in eukaryotes. The ORC binds the origin of replication and subsequent binding of further proteins leads to the formation of a replisome.

Figure modified from [74].

Initiation of DNA replication in archaea

Archaeal DNA replication initiation is more similar to eukaryotic than bacterial DNA replication. Archaea use single or multiple origins, dependent upon the species, and Orc1/Cdc6-like proteins (referred to as Orc proteins)[21]. However, *Sulfolobus* species are an exception in that they use an alternative protein, WhiP, for replication initiation from *oriC3*[75]*.* 2D-gels, whole-genome microarrays, and marker frequency analysis (MFA) have led to the identification of origins of replication in over 20 species of archaea, and have demonstrated that across these species, the number of origins and Orc proteins is highly variable[76]. Despite this, in almost all cases, a gene encoding an Orc protein is found adjacent to the origin to which it binds[77].

Like bacterial and eukaryotic origins, archaeal origins are AT-rich. However, origin recognition box (ORB) sequences can be found either side of the origin, often in inverted pairs, to direct the binding of Orc proteins onto the DNA. ORB sequences contain a string or guanine nucleotides and are found on the minor groove of DNA[78,79]. The N-terminal AAA+ domain of the Orc protein binds to the DNA and the affinity with which the protein binds is determined by a winged-helix C-terminal domain[80,81]. MCM helicase is recruited by the Orc protein (when it is bound by ATP) which, in archaea, is a single polypeptide homologous to MCM2-7 in eukaryotes. Archaeal MCM is loaded onto the leading DNA strand as a double homohexameric ring[21,77,79]. As in eukaryotes, it is thought that a CMG complex forms after MCM helicase has been recruited, but unlike in eukaryotes, archaeal MCM has helicase activity when it is not bound to Cdc45 or GINS[82]. Archaeal GINS is a dimer of dimers: either a homodimer of GINS51 or a heterodimer of GINS51 and GINS23[83]. There are no direct archaeal Cdc45 homologue, but RecJ proteins are considered orthologues[83]. Once the CMG complex forms, it is thought that further proteins are recruited to create a fully formed replisome[6].

As mentioned previously, the laboratory strain of *H. volcanii* has six origins of replication, each with an associated Orc protein. Origin and Orc protein associations are as follows: *oriC1* with Orc1, *oriC2* with Orc5, *oriC3* with Orc2*, oripHV1* with Orc10*, oripHV3* with Orc6*,* and *oripHV4* with Orc3. However, the *H. volcanii* genome encodes 16 Orc proteins in total: 10 of which are not associated with origins (Table 1.1). Instead, associations of these proteins vary between native and non-native genes involving glycerol and DNA metabolism, protein glycosylation, and transposases . With the exception of orc10, all orc genes found adjacent to origins are native. It is thought that Orc9 plays a role in DNA repair[84].

Table 1.1: *orc* genes in *H. volcanii*: origin-associated; native but not origin-associated; non-native, not origin-associated. *orc* genes may be classified by synonymous codon usage (SCU) and are either native or non-native (>10% rare codons). SCU data generated by Thorsten Allers (unpublished).

orc	Location	SCU	Features
1	Chr		***oriC1***, main chromosomal origin.
2	Chr		***oriC3***
3	pHV4		***ori-pHV4***, in native region of pHV4.
4	Chr	23% rare	Associated with non-native protein glycosylation genes.
5	Chr		***oriC2***
6	pHV3		***ori-pHV3***
7	pHV4	27%	Both associated with (non-native) transposases.
8	pHV1	21%	
9	Chr		**Not adjacent to origin**.
10	pHV1	**33%**	***ori-pHV1***. Dominant in origin screen, difficult to delete.
11	Chr	35%	Both in (non-native) prophage region.
14	Chr	33%	
12	pHV4	14%	Both associated with non-native DNA metabolism genes.
13	pHV4	34%	
15	Chr		Both associated with native glycerol metabolism genes[85].
16	Chr		

1.4- Non-canonical DNA replication

Displacement loops arising from either ssDNA or RNA invading a homologous dsDNA duplex (named D-loops and R-loops, respectively) can be used to prime replication instead of typical initiation of replication at origins[86]. D-loops and R-loops can be remodelled to create a replication fork and from this, DNA can be synthesised. Replication from both D-loops and R-loops has been demonstrated to require RecA-family recombinases[87] and it is therefore thought that DNA replication in the absence of origins requires homologous recombination (recombination-dependent replication (RDR)).

1.4.1- Homologous recombination

Homologous recombination (HR) involves the exchange of genetic material between identical or near-identical DNA sequences and is considered to be a high-fidelity DNA repair mechanism. HR is used in the repair of double strand breaks, single strand breaks/gaps, and interstrand DNA crosslinks. Additionally, HR is involved in the restart of stalled DNA replication forks, and in chromosomal pairing and exchange during meiosis[88,89]. Mutations in genes in HR pathways are known to result in genomic instability and predisposition to some types of cancer[90].

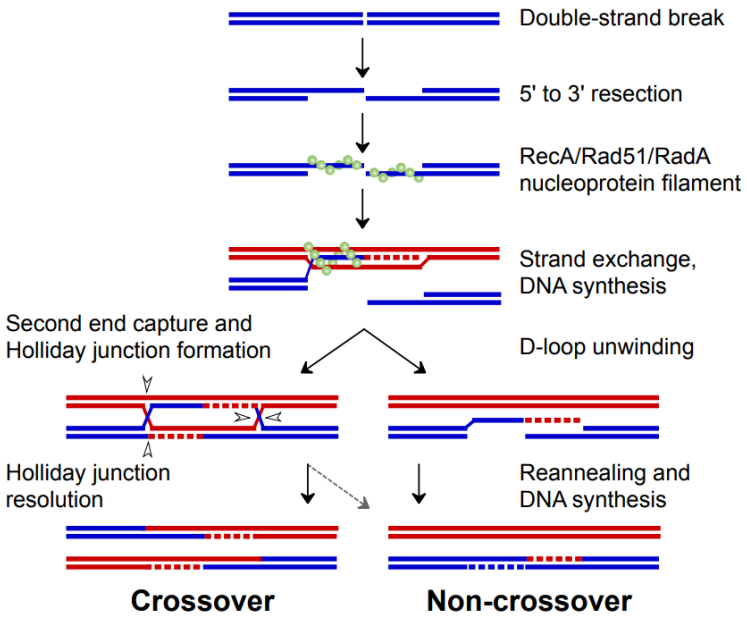


Figure 1.9: Homologous recombination in double-strand break repair. Ends of a double-strand break are resected 5' to 3' and are bound, depending upon the domain of life, by either RecA, Rad51 or RadA. The nucleoprotein filament catalyses strand exchange, leading to D-loop formation. A double Holliday junction forms which can be resolved to form either a crossover or non-crossover product.

Figure taken from [91].

The overall HR mechanism consists of end-resection, strand exchange, DNA synthesis and resolution. In brief, DNA is resected to generate 3' ssDNA overhangs which invade homologous regions of dsDNA eventually leading to the formation of a D-loop and a subsequent double Holliday junction (dHJ). DNA is synthesised using the invaded DNA duplex as a template and the dHJ is resolved. Two intact dsDNA duplexes are formed which, depending on the orientation of cleavage of the dHJ by resolvases, can either be crossover or non-crossover products (Figure 1.9). Homologous recombination is reviewed in [92,93].

1.4.1.1 - RecA-family proteins: RecA, Rad51, and RadA

Strand exchange during homologous recombination is catalysed by the RecA family of ATPases: the bacterial RecA, eukaryotic Rad51 and DMC1 (in repair/replication fork restart and meiosis respectively), and archaeal RadA. Proteins in this family form helical nucleoprotein filaments in an ATP-dependent manner upon the 3' ssDNA overhangs formed through end resection and invade homologous dsDNA. The invading strand binds to the complementary strand to form a heteroduplex and the other strand of the dsDNA duplex is displaced, creating a D-loop. Work by Clark and Margulies in 1965 into recombination-deficient *E. coli* mutants first identified the *recA* gene[94], and it was later discovered that homologous proteins able to form similar nucleoprotein filaments, Rad51 and DMC1, can be found in eukaryotic cells[95-98]. The discovery of archaeal RadA and its analogous ability to form nucleoprotein filaments[99,100] demonstrated that a recombination system involving filamentous recombinase structures spans all domains of life.

All recombinases share a RecA-fold ATPase domain consisting of Walker A and B motifs (for ATP binding and hydrolysis respectively), and motifs for binding ds- and ss-DNA[101-103]. Eukaryotic Rad51 and archaeal RadA share more amino acid identity and structural similarity with each other than with bacterial RecA. Archaeal RadA shares 40% amino acid identity with eukaryotic Rad51 compared to 20% with bacterial RecA[100]. RecA has a C-terminal domain that is not present in Rad51 or RadA which is responsible for mediating contact between dsDNA and ssDNA molecules within the filament[104]. Additionally, RecA has a much shorter N-terminal domain than Rad51 and RadA which in eukaryotes and archaea, is thought to play a role in activation of the nucleoprotein filament; a ~30° rotation of the N-terminal domain between active and inactive conformations has been observed[105]. Domain arrangement is demonstrated in Figure 1.10. An invariant phenylalanine residue can be found in the N-terminal domain of Rad51 and RadA which is inserted into a hydrophobic binding pocket found in the core domain of an adjacent monomer for polymerisation[102]. In RecA, an isoleucine residue is instead inserted into the adjacent monomer's binding pocket[106].

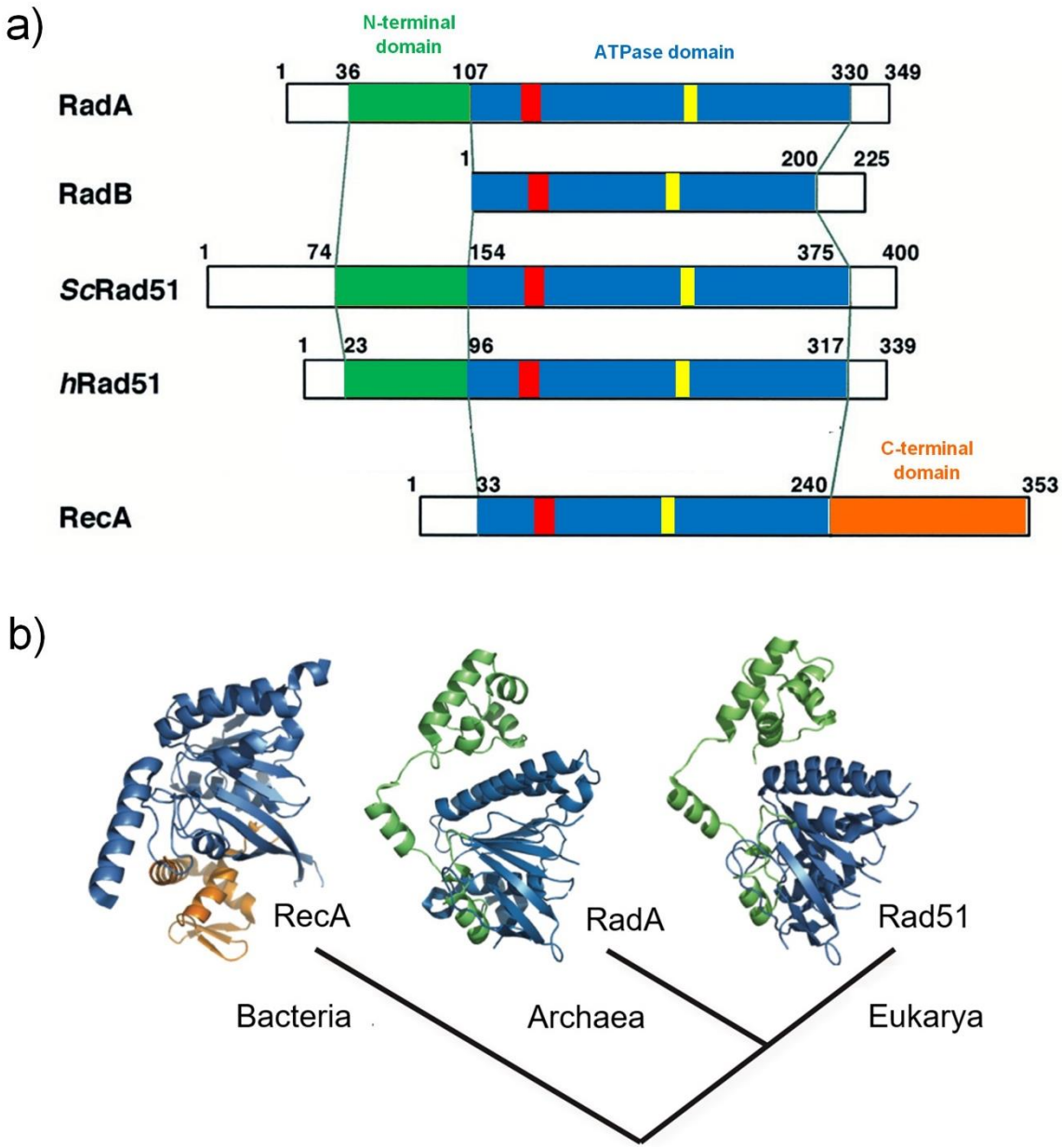


Figure 1.10: Domain arrangements and structural similarity of RecA-family recombinases. a) Comparison of domain arrangements of RadA and RadB from *Pyrococcus furiosus*, with Rad51 from *Saccharomyces cerevisiae* (ScRad51) and human (hRad51), and RecA from *Escherichia coli*. The N-terminal region specific to RadA and Rad51 is shown in green, the conserved ATPase domain is shown in blue, the RecA-specific C-terminal domain is shown in orange, and Walker A and B domains are shown in red and yellow, respectively. **b)** Structural similarity between RadA and Rad51 in comparison to RecA can be seen. Colours are the same as in a). RecA is from *E. coli*, RadA is from *P. furiosus*, and Rad51 is from *S. cerevisiae*. Crystal structures were accessed from the RCSB Protein Data Bank (2REB, 1PZN, 1SZP).

Figure 1.10a modified from [107]. Figure 1.10b modified from [108].

Nucleoprotein filament formation

Polymerisation of recombinases onto ssDNA occurs in three stages: nucleation, extension, and disassembly.

The number of RecA monomers necessary for nucleation is disputed with studies stating that either dimers or five monomers are required[109,110]. Each monomer binds three nucleotides and DNA is underwound up to 1.5-fold to aid in the search for homologous DNA sequence[111-113]. The structure of a RecA nucleoprotein filament can be seen in Figure 1.11. The ATP binding motif is formed at the interface between two monomers[114]. Between two and five monomers are required for Rad51 nucleation and like RecA, 3 nucleotides are bound per monomer[97,115-117]. Nucleation of RadA has not been studied directly but given the structural similarity between RadA and Rad51, it may function in the same manner.

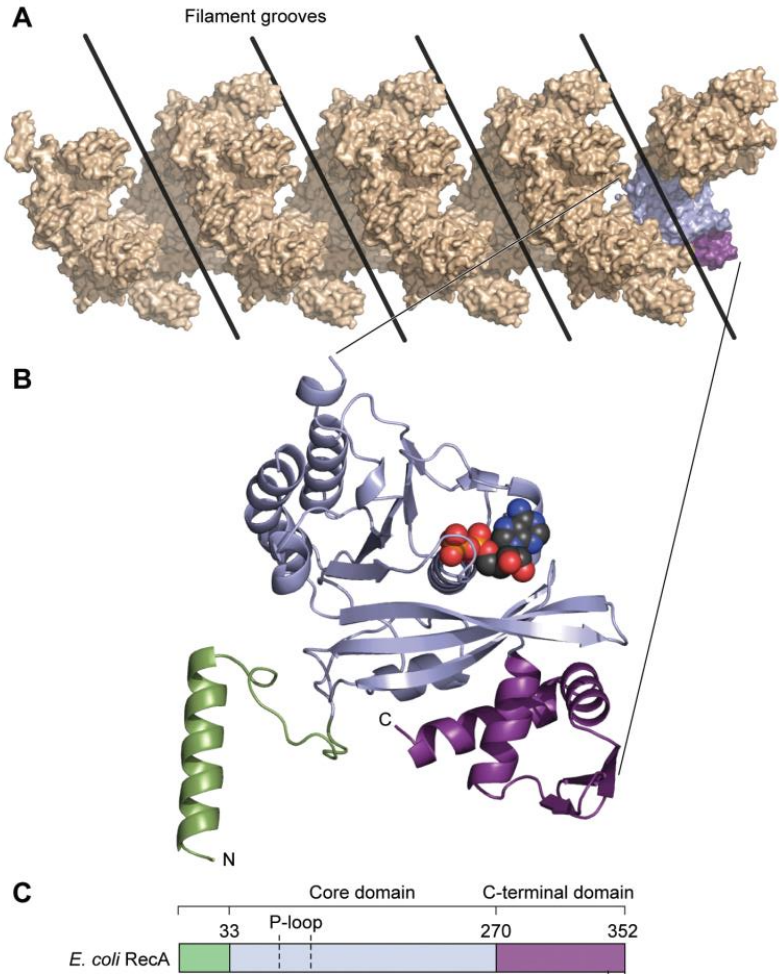


Figure 1.11: RecA structure. a) An *Escherichia coli* RecA nucleoprotein filament with four helical turns. 24 sub-units are shown there are six RecA monomers per helical turn of the filament with each turn corresponding to 18bp of DNA in extended conformation. **b**) One monomer of the nucleoprotein filament. Colours are as denoted in c). **c**) Domains of a RecA monomer.

Figure modified from [118].

Once nucleation has occurred, filaments can extend and polymerise upon ssDNA. There is a higher rate of RecA polymerisation at the 3' end of the filament than at the 5' end meaning that the filament shows net 5' to 3' growth[109]. However, Rad51 filaments have opposite polarity to RecA, and net growth is in the 3' to 5' direction[119,120].

ATP-binding and hydrolysis play an important role in the assembly and disassembly of nucleoprotein filaments: ATP-binding is required for filament assembly and ATP-hydrolysis is required for filament disassembly. RecA filaments can exist in an active form (when ATP-bound) able to carry out strand exchange, and an inactive form (when ADP-bound) thought to be a storage state[121,122]. Filaments elongate as ATP is bound and contract as ATP is hydrolysed[112]. ATPase deficient Rad51 monomers are able to form filaments but are not able to disassemble, but ATP hydrolysis is not necessary for disassociation from DNA[116,123].

The nucleoprotein filament promotes pairing of the ssDNA with dsDNA. Within the filament, ssDNA and dsDNA molecules align and underwinding of the DNA (RecA filaments extend DNA 1.5-fold relative to B-form) and initiates switching of base pairs. Base pair switching displaces one strand of the duplex and a D-loop is generated (Figure 1.12).

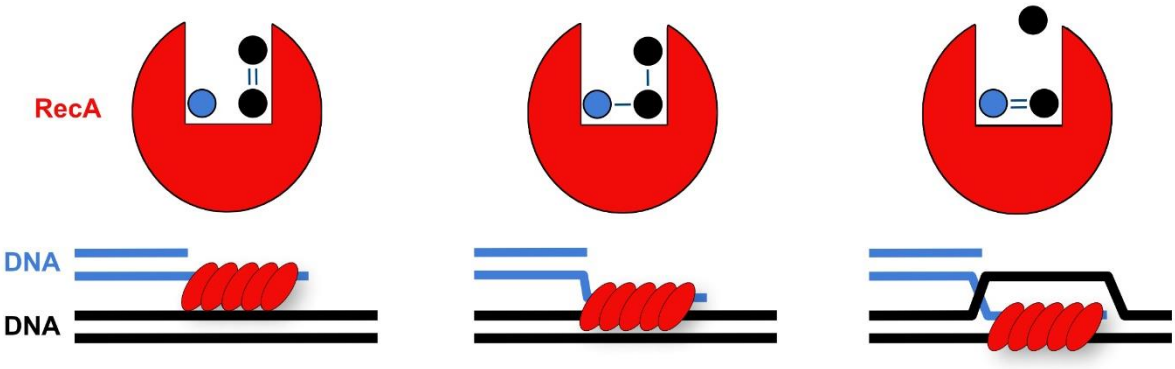


Figure 1.12: Strand exchange catalysed by RecA. The RecA filament promotes pairing with dsDNA. ssDNA and dsDNA molecules align (left-hand image). DNA underwinding initiates switching of base pairs (middle image) which displaces one strand of the DNA duplex (right-hand image).

1.4.2- R-loops

An R-loop is a three-stranded RNA: DNA hybrid structure first described by Davies R.W in 1976 when it was discovered that in the presence of formamide, RNA and DNA hybridise stably[124]. Their presence in vivo and co-transcriptional formation in bacteria was later demonstrated in 1995[125]. A nascent RNA strand invades double stranded-DNA and displaces the non-template strand as ssDNA. Due to the open structure created and presence of an RNA strand to be used as a primer, it is possible that the 3' OH end of the RNA strand within the R-loop is capable of priming DNA replication. However, R-loops leading to genome instability through transcription: replication conflicts are well documented[126-129], and the DNA breaks formed due to this may instead be used to prime DNA replication through homologous recombination.

Recent technological developments involving DNA:RNA immunoprecipitation (DRIP) using the monoclonal S9.6 antibody to bind non-sequence specific DNA:RNA hybrids[130,131], followed by high-throughput DNA sequencing (DRIP-seq), and variations upon this including DNA:RNA in vitro enrichment (DRIVE-seq)[132,133], tilling array hybridisation (DRIP-chip)[134], and reverse transcription of RNA into cDNA (DRIPc-seq)[135], have provided insights into the great frequency at which R-loops arise, and the previously questioned biological relevance of this method to replicate DNA. These techniques have allowed direct detection and subsequent genome-wide mapping of R-loops in a range of organisms including plant[136], yeast[134,137,138], and mammalian genomes[133], revealing that R-loops are both widespread and functional genomic structures.

R-loop formation

It was initially thought that the short DNA:RNA hybrid formed during transcription could be extended, leading to R-loops. However, the discovery that DNA and RNA exit through distinct RNAP II channels upon analysis of the enzyme's crystal structure[139] does not support this and instead, current research supports the "thread-back" model of R-loop formation. During transcription, RNA polymerase and the nascent RNA protruding from it must rotate relative to the DNA. Resistance to the rotational motion can lead to positive supercoiling of the DNA ahead of RNAP and negative supercoiling behind[140]. Unwinding of the DNA behind RNAP as a result of this may allow the nascent RNA to form an initial DNA:RNA interaction with the template strand, and the rest of the nascent RNA can then thread back to form a longer DNA:RNA hybrid[141]. For this to occur, the nascent RNA must remain single stranded and bind to the DNA template before the two DNA strands re-anneal. Additionally, R-loops can form *in trans* when transcripts hybridise to DNA far from the transcribed sequence. Examples of this include non-coding RNA R-loops and guide RNA (gRNA) R-loops arising when a gRNA associated with a CRISPR/Cas9 binds to its target sequence [142] (Figure 1.13).

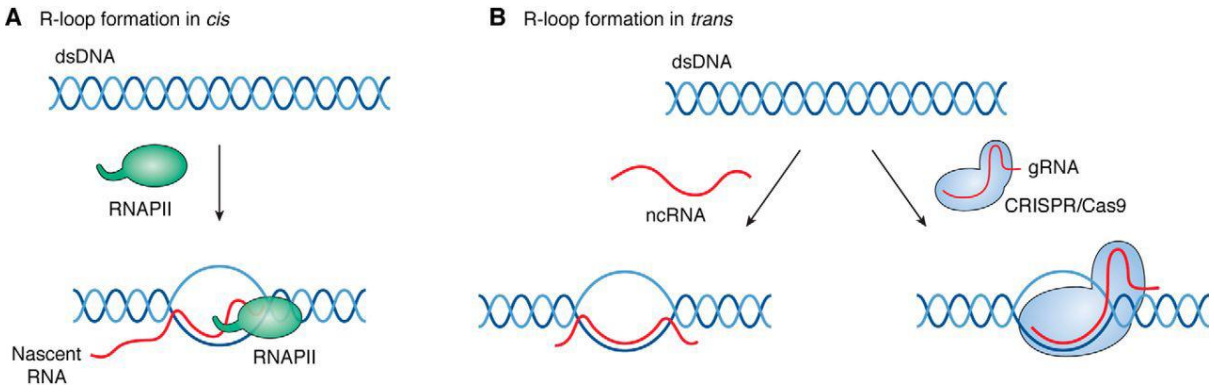


Figure 1.13: Possible mechanisms of R-loop formation. A) *In cis* formation in which the nascent RNA binds to the template. **B)** *In trans* formation in which a transcript anneals with DNA a distance from the template DNA sequence.

Figure taken from [142].

Topoisomerase I interacts with RNAP to relax transcription induced supercoiling[143], and has been found to inhibit R-loop formation in vitro[144-146]. Additionally, DNA:RNA hybrids are elevated in mutants unable to elongate the nascent RNA strand during transcription, presumed to be due to the increase in time the nascent RNA and negatively supercoiled DNA spend in close proximity due to RNAP stalling, and also in mutants deficient in splicing as the reduction in RNA masking may facilitate DNA:RNA hybridisation[147-150]. The culmination of these findings provides support for the "thread-back" model of R-loop formation. Other factors such as DNA nicks and polyA-tracts have been found to promote R-loop formation[138,151], and a correlation between GC content and DNA:RNA hybrid formation has been discovered in *Saccharomyces cerevisiae*[134,137].

Additionally, the role of RecA-family recombinases in the formation of R-loops has been investigated. It has been found that RecA is capable of catalysing invasion of an RNA transcript into a homologous DNA duplex in *E.coli* in vitro[152]. Furthermore, an inverse DNA:RNA strand exchange mechanism in *E.coli*, in which RecA coats dsDNA and promotes strand exchange with ssRNA, is also thought to be responsible for R-loop formation[153]. Additionally, RecA-dependent hybrid formation can form post-transcriptionally in vitro[152,153]. In *S. cerevisiae*, mutants defective in transcription repression and RNA degradation require Rad51 and Rad52 for R-loop formation and interestingly, Rad51 and Rad52 were found to promote hybridisation of transcripts to chromosomal regions distinct from their synthesis site[154]. The involvement of RecA-family recombinases in R-loop formation provide support for a recombinase-dependent DNA replication model initiating from R-loops[155,156].

1.4.3- Recombination-dependent replication

The T4 bacteriophage provided an early example of RDR, primed by both D-loops and R-loops depending upon the phase of replication[157]. However, its replication is not originless. T4 bacteriophage replication origins consist of a middle-mode promotor and a downstream AT-rich DUE[158]. Transcription factor MotA binds to the promotor at the origin, transcription is initiated, and R-loops persist within the DUE region. The R-loop forces the DNA duplex open, allowing elements of the replisome to bind. RNase H processes the RNA creating a 3' hydroxyl group which can be extended by a DNA polymerase, and replication begins[158-160].

Work by Tokio Kogoma demonstrated that D- and R-loops can be used to prime replication in *E. coli*. Regulated replication in bacteria is initiated through the protein DnaA binding to the origin *oriC* but investigation into E. coli mutants able to synthesise DNA after protein synthesis inhibition revealed that in *Escherichia coli*, *rnhA* (RNase HI) mutants are able to replicate their genomes without requirement for DnaA or *oriC* in a process termed constitutive stable DNA replication (cDSR)[161,162]. It was suggested that a role of RNase HI was therefore to degrade RNA in R-loops and that R-loops formed during transcription would remain present long enough, in the absence of degradation by RNase HI, to be used to prime replication. Additionally, topoisomerase I has been found to suppress replication initiation on plasmids lacking *oriC* which are not dependent on DnaA[163], acting as a specificity factor for regulated DNA replication, and its role involves inhibition of R-loop formation through relaxation of transcription induced supercoiling[125,145]. Mutations in both *rnhA* and *topA* genes would permit the use of R-loops to prime DNA replication.

Break-induced replication

Break-induced replication (BIR) is a form of RDR. When a double strand break (DSB) occurs in DNA, as in HR, ends are resected to form 3' ssDNA overhangs. However, unlike in HR whereby both 3' ssDNA overhangs invade the homologue, a single 3' ssDNA overhang invades dsDNA and a D-loop is created. The invading strand can be used to prime one-way replication[164,165]. BIR is associated with DNA damage and genome instability including genome rearrangements and copy number variation. It is thought that due to asynchronous leading and lagging strand synthesis seen during BIR, ssDNA tails extending from the migrating D-loop during replication are likely to sustain damage[164]. In eukaryotes, BIR can occur independently of Rad51 and Rad52 is instead required, unlike the strict requirement seen for recombinases in RDR in *E. coli* and *H. volcanii*[166,167].

RDR in the archaea

In *H. volcanii,* DNA replication in the absence of origins on the main chromosome is possible (Figure 1.14). When deleted for *oriC1, oriC2, oriC3* and *oripHV4,* cells grow faster than wild-type but are now critically dependent upon RadA[49]. Previous work has found that deleting RadA inhibits cells from carrying out homologous recombination[168]. It was not possible to delete the gene encoding RadA (*radA*) in a *ΔoriC1,2,3,pHV4* background (0 of 455 slow-growing colonies screened were *ΔradA*), and *radA* deletion in a *ΔoriC1,2,3* background recovered only one isolate (1 of 70 screened) in which a chromosomal rearrangement had occurred (inversion of a region of pHV4[49], as mentioned in Section 1.2- *Haloferax volcanii* as a model organism). As no mutations were seen in RNase HI in this study, it was thought that D-loops, not R-loops, were being used to prime replication. However, this was not proven experimentally.

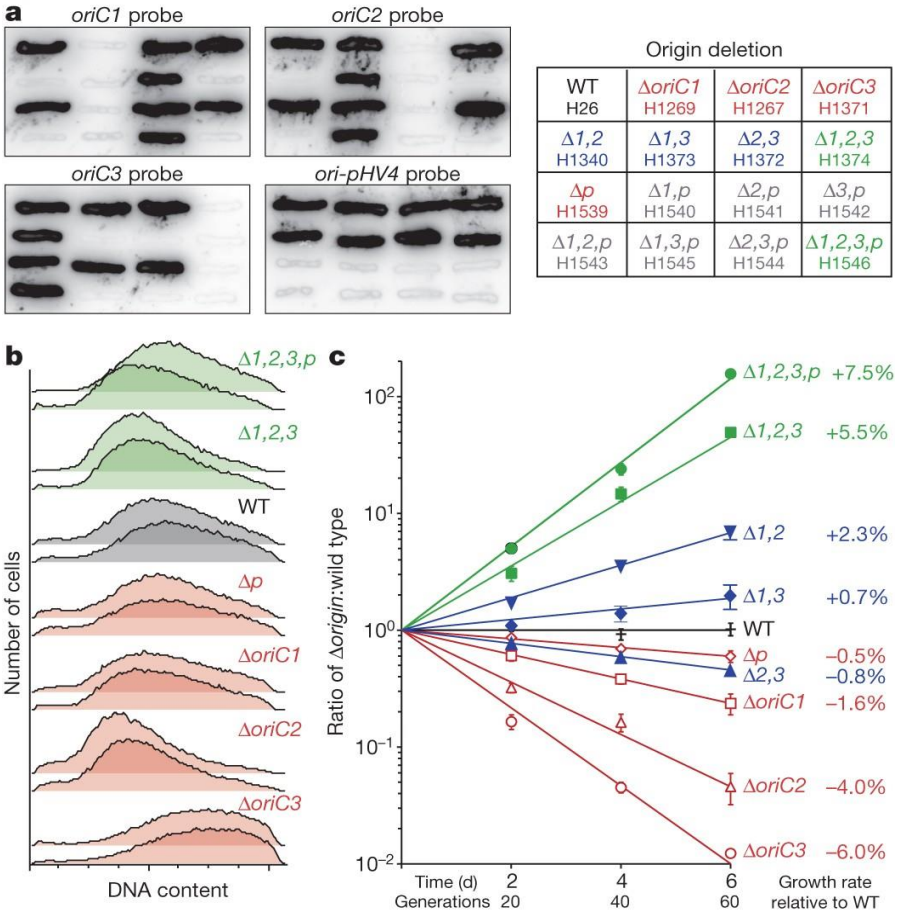


Figure 1.14: Origin deletion in *Haloferax volcanii*. a) Colony hybridisation with origin-specific probes confirms origins deletions. **b**) DNA content of cells was measured with flow cytometry. Two biological replicates are shown per strain. **c**) Pairwise growth competition assay demonstrating growth rates of strains with origin deletions in comparison to wild-type. In particular, a 7.5% increase growth rate in a *ΔoriC1,2,3pHV4* strain compared to wild-type is shown.

Figure taken from [49] .

Since this finding was reported, replication in the absence of origins has been demonstrated in other polyploid archaeal species including *Thermococcus barophilus*[169] and *Thermococcus kodakarensis* [170]. Although not archaeal, cyanobacterial species with multiple genome copies have recently been shown to replicate their DNA from many discrete sites across the chromosome, independent of DnaA, and as in *H. volcanii* strains deleted for all origins on the main chromosome, MFA profiles for these species do not demonstrate significant peaks indicative of origins of replication [171]. Additionally, when the gene encoding DnaA was deleted, ploidy increased. The combination of these findings may demonstrate that higher genome copy number permits successful origin-independent replication; replication in the absence of origins has been shown to depend upon RecA-family recombinases and more copies of the genome means that more opportunities are available for the strand invasion stage of homologous recombination.

Despite this, there exist other polyploid Euryarchaea, such as two close relatives of *H. volcanii,* namely *Haloarcula hispanica* and *Haloferax mediterranei*, that are not capable of originless replication. Deletion of all three chromosomal origins in *H. mediterranei* activates a dormant origin acquired through lateral gene transfer[172] and in *H. hispanica,* the two origins of replication cannot be deleted at the same time[173]. Therefore, further work is necessary to determine the requirements for origin independent replication across different species.

1.5- The role of transcription and R-loops in genomic instability

Transcription and the presence or R-loops have been shown to play a role in genome instability (reviewed in[129,174-178]). During transcription, chromatin structure and DNA topology undergo change and it is possible that the transcriptional machinery may collide with DNA replication forks. Factors such as these can lead to the destabilisation of the genome through transcription-associated mutation (TAM) and transcription-associated recombination (TAR)[179].

Transcription-associated mutation (TAM)

Examples of TAM noted in *E. coli* include increase in the mutation rate of the β-galactosidase gene upon transcription and when transcription is activated, the ICR-191 mutagen reverts lac$^-$ mutations more frequently[180,181]. Examples also exist in the T7 phage[182] and in yeast [183]. Additionally, the dependence upon transcription in somatic hypermutation (SHM) and class switch recombination (CSR) during a B-cells' response to a foreign antigen has been subject of much research[184,185]. During SHM, point mutations are introduced and during CSR, recombination events occur in the constant heavy chain immunoglobulin genes (meaning that CSR would fall under the bracket of TAM, not TAR).

In the formation of R-loops, a ssDNA strand is displaced upon invasion of the nascent RNA. G-quadruplex secondary structures can form within the displaced ssDNA strand forming sites susceptible to nuclease action[186], and DNA nicks and lesions arise through deaminases, DNA repair enzymes, and DNA damaging agents using the displaced ssDNA as a substrate[178,179,187].

Transcription-associated recombination (TAR)

Additionally, transcription has been found to strongly induce recombination. Possible reasons for this include conflict between transcription and replication forks and R-loop formation (Section 1.4.2- R-loops)[188].

The first example of TAR was reported in λ-phage studies in which recombination occurring in transcribed regions of DNA was found to be dependent upon *E. coli* RNAP activity but not upon RecA[189,190]. In *Saccharomyces cerevisiae*, the recombinant DNA sequence *HOT1* is a hot spot of recombination. It was found that the ability of *HOT1* to stimulate ectopic recombination of non-recombinant DNA sequences depends upon the activity of RNAPI[191,192]. Other examples of TAR include work on phage transduction[193], RNAPII-mediated recombination between DNA repeats in *S. cerevisiae* and rodent cells[194,195] and illegitimate recombination in bacterial plasmids[196]. Further and more recent examples of TAR are reviewed in [197].

It is widely accepted that stalling of replication forks due to collisions with transcriptional machinery is a major source of TAR[188]. However, how recombination events occur after stalling of the replication fork remains ambiguous. Accumulation of negatively supercoiled DNA during transcription may mean that damage via endogenous genotoxic agents is more likely. This can result in DNA lesions, preventing progression of the replication fork. A stretch of ssDNA would arise in front of either the leading or lagging strand polymerase. Recombinational template switching may then act to fill the un-replicated gap[188]. Another possibility is that head-on collisions between replication and transcription forks can lead to DSBs, positive supercoiling between the machineries, or creation of a “chicken-foot” structure as a result of extrusion of the replication fork backwards, all of which may lead to recombination in their resolution[188,198-200]. Additionally, codirectional collisions between replication and transcription forks may result in replication fork stalling. A transcription fork progressing ahead of a replication fork may be blocked or be required to backtrack due to the presence of a DNA lesion in the template may cause stalling of the replication fork behind[201]. Similarly, formation of an R-loop in a situation such as this would prevent the replication fork from progressing. In order to restart replication, DNA breakage may occur and recombination-dependent replication may arise or the R-loop structure may instead be used to prime replication[188].

Genomic instability in the form of recombination arising due to the presence of R-loops in particular was first demonstrated in *S. cerevisiae.* A hyper-recombination phenotype was observed in cells with mutant messenger RNA particle biogenesis and export factor THO which could be suppressed by mutation in RNAPII to reduce transcription[149,202-205]. The hyper-recombination phenotype was also reliant on the ability of the nascent RNA molecule to form an R-loop[149]. More recently, the relevance of R-loops in transcription-dependent instability has been demonstrated though R-loop formation in DT40 chicken and HeLa cells with inactive splicing factor ASF/SF2 in addition to chromosomal rearrangements, γ-H2AX foci, and hyper-recombination in a range of organisms containing mutations in factors involved in RNA metabolism (reviewed in [188]). An association between areas of the genome able to form non-B-form DNA structures and genome instability has also been noted[206,207].

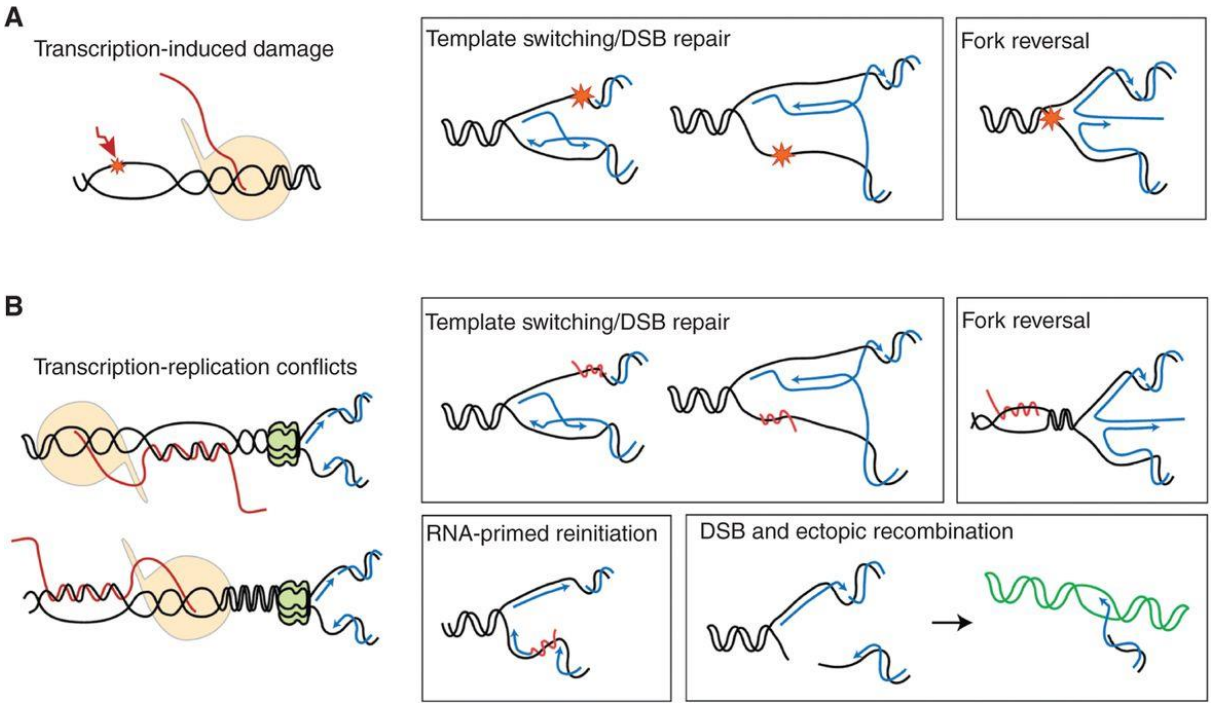


Figure 1.15: Possible mechanisms of transcription-associated recombination. A) ssDNA is exposed behind the transcription machinery and is susceptible to damage. Transcription-dependent DNA lesions may prevent replication forks from progressing, leading to template switching, DSB repair, or replication fork reversal. Other methods to overcome such scenarios exist, including translesion synthesis. **B)** Both codirectional or head-on collisions between transcription and replication forks may arise and require recombination to be resolved in the ways shown.

Figure taken from [188].

1.6- Aims of the study

Initially, this study had two over-arching goals: to determine RadA localisation on DNA in origin-independent replication, and to investigate the dependence upon RadA in replication initiation from R-loops. However, information discovered when investigating the latter (as detailed in Chapter 4 of this study) meant that *H. volcanii* genome instability was instead investigated as the initial aim was no longer possible in the time-period available.

In order to determine RadA localisation on DNA in origin-independent replication, aims were as follows:

- To create strains of *H. volcanii* expressing FLAG-tagged RadA in both wild-type and origin-deleted backgrounds and verify that FLAG-RadA was functional and could be detected through Western blotting
- To carry out ChIP-Seq to determine where FLAG-RadA localises upon DNA

To investigate dependence upon RadA in replication initiation from R-loops, this study aimed to undertake phenotypic and quantitative growth analysis of strains constructed prior to this work (Laura Mitchell, unpublished data). During this investigation, it was revealed that the genomic architecture of these strains was not as expected. The aims were adjusted to reflect this and comprised:

- Pulsed field gel electrophoresis to investigate integration of pHV3 onto the main chromosome and the frequency at which this occurs
- Determining the loci at which pHV3 became integrated onto the main chromosome through Oxford Nanopore MinION sequencing
- Investigation into the unexpected genotype of a strain discovered upon analysis of the Oxford Nanopore MinION sequencing data

Chapter 2- Materials and Methods

2.1- Materials

2.1.1- Strains

***Haloferax volcanii* strains**

Table 2.1: *H. volcanii* strains used in this study. Red text denotes discoveries made concerning the genotypes and parents of strains in Chapter 4. For strains with red text, the black text denotes the presumed genotype/parent of the strain upon commencement of the study.

Strain	Parent*	Genotype	Notes
H26	DS70	Δ*pyrE2*	Constructed by TA[46]
H53	H26	Δ*pyrE2* Δ*trpA*	Constructed by TA[46]
H298	H53	Δ*pyrE2* Δ*trpA* Δ*radA::trpA+*	Constructed by TA
H1637	H98	Δ*pyrE2* Δ*hdrB* *p.tnaA-radA+::hdrB+*	Constructed by MH[49]
H1642	H1608	Δ*pyrE2* *p.tnaA-radA+::hdrB+* Δ*oriC2* Δ*oriC1* Δ*oriC3* Δ*hdrB* Δ*ori-pHV4-2*	Constructed by MH[49]
H1804	H1546	Δ*pyrE2* Δ*trpA* Δ*oriC2* Δ*oriC1* Δ*oriC3* Δ*ori-pHV4-2*	Constructed by KP
H1965	H1544	Δ*pyrE2* Δ*trpA* Δ*oriC2* Δ*oriC3* Δ*ori-pHV4-2*	Constructed by HM
H2461	H2163 x H2240***	Δ*pyrE2* *leuB+::[pHV3]*	Constructed by HM[208]
H2184	H1948	Δ*pyrE2* Δ*trpA* Δ*{pHV1}*	Constructed by HM[208]
H4113	H53	Δ*pyrE2* Δ*trpA* *p.syn::adh2+*	Constructed by LM
H4318**	H4113	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* *oripHV3+::[ΔoripHV3::trpA+ pyrE2+]*	Constructed by LM
H4319**	H4113	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* *orc6+::[Δorc6::trpA+ pyrE2+]*	Constructed by LM
H4320**	H4113	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* *orc6* *oripHV3+::[Δorc6 ΔoripHV3::trpA+ pyrE2+]*	Constructed by LM
H4344	H4113	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* Δ*orc6* Δ*oripHV3::trpA+* (*pHV3* integrated onto the main chromosome)	Constructed by LM

H4345	H4113	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* Δ*orc6* Δ*oriHV3::trpA+* (*pHV3* integrated onto the main chromosome)	Constructed by LM
H4346	H4113	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* Δ*orc6* Δ*oriHV3::trpA+* (*pHV3* integrated onto the main chromosome)	Constructed by LM
H4347	H4113	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* Δ*orc6* Δ*oriHV3::trpA+* (*pHV3* integrated onto the main chromosome)	Constructed by LM
H4396	H4113	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* Δ*oriHV3::trpA+* (plasmid used in construction failed to pop out)	Constructed by LM
H4397	H4113	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* Δ*oriHV3::trpA+* (plasmid used in construction failed to pop out)	Constructed by LM
H4398	H4113	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* Δ*oriHV3::trpA+* (plasmid used in construction failed to pop out)	Constructed by LM
H4462**	H4113	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* *radA+::[ΔradA::trpA+ pyrE2+]*	Constructed by LM
H4481**	H4462	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* *radA+::[ΔradA::trpA+ pyrE2+]* *{radA+ pyrE2+ MevR}*	Constructed by LM
H4482**	H4462	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* *radA+::[ΔradA::trpA+ pyrE2+]* *{radA+ pyrE2+ MevR}*	Constructed by LM
H4492	H4344	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* Δ*orc6* Δ*oriHV3* (*pHV3* integrated onto the main chromosome)	Constructed by LM
H4722	H4113 (unknown)	Δ*pyrE2* *p.syn::adh2+* (*pHV3* integrated onto the main chromosome, Δ*pHV1*, Δ*oriHV3::trpA+*, Δ*mrr*, Δ*leuB*, *does not contain p.syn::adh2+)*	Constructed by LM
H4723	H4113	Δ*pyrE2* *p.syn::adh2+*	Constructed by LM
H4724	H4492	Δ*pyrE2* *p.syn::adh2+* Δ*orc6* Δ*oriHV3* (*pHV3* integrated onto the main chromosome)	Constructed by LM
H4780	H4722	Δ*pyrE2* *p.syn::adh2+* Δ*hdrB* (Derived from H4722, see above)	Constructed by LM
H4782	H4724	Δ*pyrE2* *p.syn::adh2+* Δ*orc6* Δ*oriHV3* Δ*hdrB*	Constructed by LM

		(*pHV3* integrated onto the main chromosome)	
H4943	H4780	Δ*pyrE2* *p.syn::adh2+* Δ*hdrB* *p.tnaA-radA+::hdrB+* (Derived from H4780, see above)	Constructed by LM
H4945	H4782	Δ*pyrE2* *p.syn::adh2+* Δ*orc6* Δ*oripHV3* Δ*hdrB* *p.tnaA-radA+::hdrB+* (*pHV3* integrated onto the main chromosome)	Constructed by LM
H5464**	H298	Δ*pyrE2* Δ*trpA* Δ*radA::trpA+::[FLAG-radA pyrE2+]*	Constructed in this study
H5465**	H1965	Δ*pyrE2* Δ*trpA* Δ*oriC2* Δ*oriC3* Δ*ori-pHV4-2* *radA+::[ΔradA::trpA+ pyrE2+]*	Constructed in this study
H5471**	H5465	Δ*pyrE2* Δ*trpA* Δ*oriC2* Δ*oriC3* Δ*ori-pHV4-2* *radA+::[ΔradA::trpA+ pyrE2+]* *{radA+ pyrE2+ MevR}*	Constructed in this study
H5472**	H5465	Δ*pyrE2* Δ*trpA* Δ*oriC2* Δ*oriC3* Δ*ori-pHV4-2* *radA+::[ΔradA::trpA+ pyrE2+]* *{radA+ pyrE2+ MevR}*	Constructed in this study
H5501	H298	Δ*pyrE2* Δ*trpA* *FLAG-radA*	Constructed in this study
H5502	H298	Δ*pyrE2* Δ*trpA* *FLAG-radA*	Constructed in this study
H5505	H298	Δ*pyrE2* Δ*trpA* *FLAG-radA*	Constructed in this study
H5532**	H1804	Δ*pyrE2* Δ*trpA* Δ*oriC2* Δ*oriC1* Δ*oriC3* Δ*ori-pHV4-2* *radA+::[FLAG-radA pyrE2+]*	Constructed in this study
H5547	H1804	Δ*pyrE2* Δ*trpA* Δ*oriC2* Δ*oriC1* Δ*oriC3* Δ*ori-pHV4-2* *FLAG-radA*	Constructed in this study
H5548	H1804	Δ*pyrE2* Δ*trpA* Δ*oriC2* Δ*oriC1* Δ*oriC3* Δ*ori-pHV4-2* *FLAG-radA*	Constructed in this study
H5549	H1804	Δ*pyrE2* Δ*trpA* Δ*oriC2* Δ*oriC1* Δ*oriC3* Δ*ori-pHV4-2* *FLAG-radA*	Constructed in this study

* Strains in this column are those that were transformed with a plasmid leading to the creation of the strains in the furthest left-hand column. This means that for pop-out strains (all strains not marked with **), an intermediate pop-in strain is not shown.

** Pop-in strain.

*** These strains were mated prior to this study to create the strain in the furthest left-hand column.

HM= Hannah Marriot, KP= Katarzyna Ptasinska, LM= Laura Mitchell, MH= Michelle Hawkins, TA= Thorsten Allers.

2.1.2- Plasmids

pTA29 (*radA* genomic clone)

pTA29 contains the *H. volcanii radA* open reading frame and was used in this study to create a *radA* Southern hybridisation probe. Sourced from Steve Sandler via Mike Dyall-Smith[100,168].

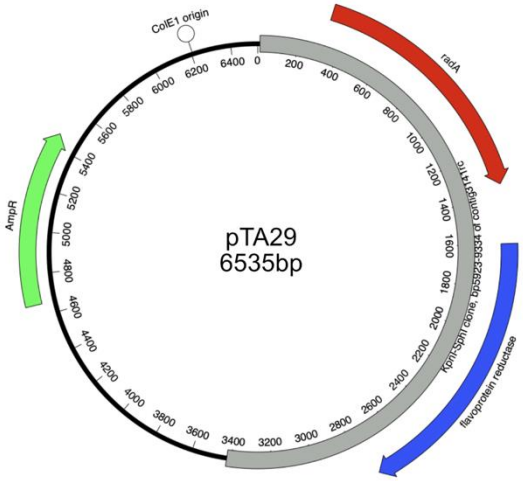


Figure 2.1: pTA29. Plasmid containing a *radA* genomic clone.

pTA637 (*radA in trans*-expression construct)

pTA637 is a dual-resistance shuttle vector for *in trans* expression of *radA*. The plasmid contains the pHV2 origin allowing it to remain as an episome when transformed into *H. volcanii*. Constructed by Thorsten Allers in 2006.

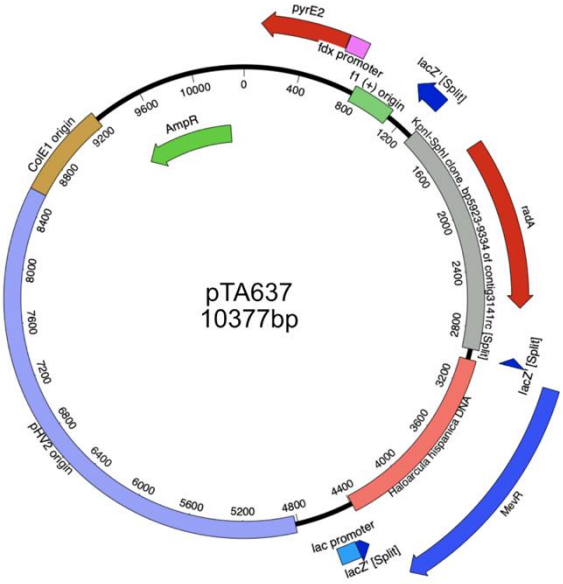


Figure 2.2: pTA637. *radA in trans*-expression construct.

pTA1651 (*radA* deletion construct)

pTA1651 is a *radA* deletion construct for use in a Δ*pyrE2* Δ*trpA* background. The deletion is flanked by regions homologous to the upstream and downstream regions of the *H. volcanii radA* gene and is marked with *trpA*. Constructed by Thorsten Allers in 2015.

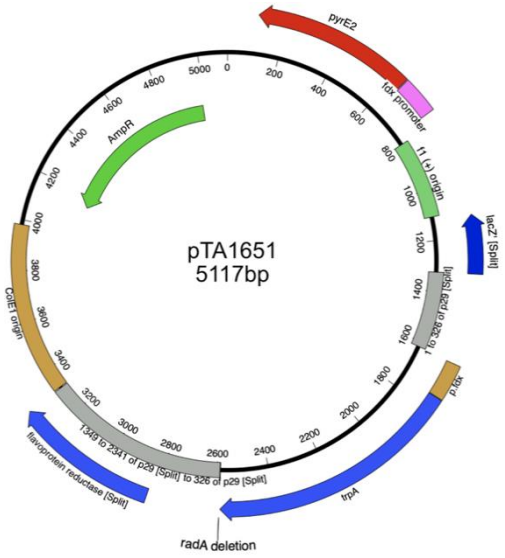


Figure 2.3: pTA1651. *radA* deletion construct.

pTA2424 (wild-type *radA or* Δ*radA* to FLAG-*radA* gene replacement construct)

pTA2424 contains a the *radA* gene 5'-tagged with a FLAG-tag (DYKDDDDK) and upstream and downstream regions. The FLAG-tag was introduced via PCR. Constructed by Ambika Dattani in 2019.

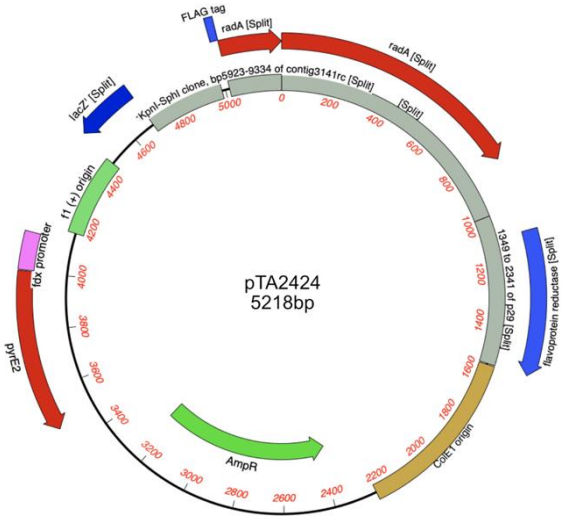


Figure 2.4: pTA2424. Wild-type *radA or* Δ*radA* to FLAG-radA gene replacement construct.

2.1.3- Oligonucleotides

Table 2.2: Oligonucleotides used in this study.

Name	Sequence (5'-3')	Description
o247	CGAGAATCGAAACGCT TATAAGTGCCCCCCGG CTAGAGAGAT	Binds within *p.syn*. Used in RT-PCR (Chapter 4).
o331	GACGATACGCTTGTCG CCC	To amplify a region within the *radA* gene. Used with o331 for DNase treatment verification (Chapter 4).
o332	GTCCGCCTCCTCTGTG TTGAC	See o331.
o1413	CGAGGGCGACGAGAG CTACTGGTACCCGC	Binds upstream of the *radA* gene. Used to screen pop-ins and pop-outs via colony PCR in conjunction with o2190 (Chapter 3).
o1827	CAACGACGACTTTCGA GGCGACTGGTTCG	Binds within a gene encoding a hypothetical protein on pHV1. Used with o1828 to create a colony hybridisation probe and in colony PCR (Chapter 4).
o1828	TCGAACAGGTAGTCCT CGAACCGCTTCCG	See o1827.
o1891	CGGCCGCGACTACAT GTACTGGAACAACG	Used with o1892 to amplify a section of the *narH* gene to form a Southern hybridisation probe for pHV3 detection (Chapter 4).
o1892	GGTGTCGGTCAGCATA TCCAGCAGTTCGC	See o1891.
o1948	CTCGTGACCGCGGTA GAACCAGAGG	Binds within *adh2*. Used in RT-PCR (Chapter 4).
o1976	CGTTCGCATCGAGGAA AAACCAGAACCGG	Binds within *adh2*. Used in RT-PCR (Chapter 4).
o2132	CACGTCATCCGCGAGA ACA	Used with o2133 to amplify *rpl10*. Used in RT-PCR (Chapter 4).
o2133	GACCTTCCCGAACGAC TGG	See o2132.
o2136	GAGTACGCCGACCAG ATTCC	Used with o2137 to amplify *polB1*. Used in RT-PCR (Chapter 4).
o2137	GGTCTGTTCCTGTCCG CTTT	See o2136.
o2190	CATGGAACCGGTCTCG AAGCCGCCCACGTCG G	Binds within the *radA* gene. Used to screen pop-ins and pop-outs via colony PCR in conjunction with o1413.
o2438	ACCGACTTCCTCGACG ATACCCATC	Binds within *adh2*. Used in RT-PCR (Chapter 4).
o2439	AACCAGAACCGGAGCT TCAACAACC	Binds within *adh2*. Used in RT-PCR (Chapter 4).

2.1.4- Chemicals and enzymes

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

2.1.5- Media

H. volcanii **media**

Media were sterilised by autoclaving for 15 minutes at 121°c. Liquid media were stored at room temperature in the dark and solid media (as poured agar plates) were stored in sealed bags at 4°c. All plates were dried for a minimum of 30 minutes prior to use.

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

18% saltwater (SW): 30% SW, 3mM $CaCl_2$ (added after autoclaving)**.**

Trace elements: 1.82mM $MnCl_2.4H_2O$, 1.53mM $ZnSO_4.7H_2O$, 8.3mM $FeSO_4.7H_2O$, 200µM $CuSO_4.5H_2O$. Filter sterilised and stored at 4°C.

Hv-Min salts: 0.4M NH_4Cl, 0.25M $CaCl_2$, 8% v/v of trace elements. Stored at 4°C.

Hv-Min carbon source: 10% DL-lactic acid Na_2 salt, 8% succinic acid Na_2salt·$6H_2O$, 2% glycerol, pH to 7.0 with NaOH. Filter sterilised.

10 x YPC: 5% yeast extract (Difco), 1% peptone (Oxoid), 1% casamino acids, 17.6mM KOH. Not autoclaved, used immediately.

10 x Ca: 5% casamino acids, 17.6mM KOH. Not autoclaved, used immediately.

Hv-Ca Salts: 362mM $CaCl_2$, 8.3% v/v trace elements, 615µg/ml thiamine, 77µg/ml biotin.

KPO_4 Buffer: 308mM K_2HPO_4, 192mM KH_2PO_4, pH to 7.0 with NaOH.

Hv-YPC agar: 1.6% Agar (Bacto), 18% SW, 1 x YPC, 3mM $CaCl_2$. Microwaved before adding 10x YPC and $CaCl_2$ to dissolve agar. 10 x YPC added and 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 before addition of 10x Ca, Hv-Ca Salts, and KPO_4 Buffer to dissolve agar. 10 x Ca added then autoclaved. Hv-Ca Salts and KPO_4 Buffer added prior to pouring.

Hv-YPC broth: 18% SW, 1 x YPC, 3mM $CaCl_2$. $CaCl_2$ added after autoclaving when cool.

Hv-Ca broth: 18% SW, 1 x Ca, 30 mM Tris.HCl pH 7.0, 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), 444nM biotin, 2.5µM thiamine. 18% SW, 1 x Ca, 30 mM Tris.HCl pH 7.0 autoclaved and all other components added once cool.

H. volcanii **media supplements**

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

Table 2.3: Media supplements used with *H. volcanii*.

Supplements	Final concentration (unless stated otherwise)
Tryptophan (Trp)	50 µg/ml
Uracil (Ura)	50 µg/ml
Thymidine (Thy)	50 µg/ml (+ 50 µg/ml hypoxanthine in Hv-Ca)
5-Fluoroorotic acid (5-FOA)	50 µg/ml (+ 10 µg/ml uracil)
Mevinolin (Mev)	6 µg/ml

Table 2.4: Growth of auxotrophic *H. volcanii* mutants on different media.

Genotype	Hv-YPC	Hv-Ca	Biosynthetic pathway disrupted
Δ*pyrE2*	+	-	Uracil
ΔtrpA	+	-	Tryptophan
ΔhdrB	-	-	Thymidine

+ indicates growth, - indicates no growth (supplement must be added to media)

2.1.6- Other solutions

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

Sodium acetate: 3M NaAc pH5.2, filter sterilised.

2.2- Methods

2.2.1- General *H. volcanii* microbiology

Growth and storage

Cultures on solid media were grown in a static incubator (LEEC) at 45°c for 5-10 days in plastic bags to prevent drying of plates. 5ml liquid cultures were grown overnight (approximately 16 hours) in a static incubator at 45°c with 8rpm rotation. Larger liquid cultures (50-333ml) were grown overnight (again approximately 16 hours) in an Innova 4330 floor-standing incubator (New Brunswick Scientific) at 45°c with 120 rpm shaking. For short term storage, cultures on solid media were stored at room temperature in plastic bags for up to 1 month. For long term storage, glycerol (80% glycerol in 6% SW) was added (20% v/v) to liquid cultures which were snap frozen on dry ice and stored at -80°c.

Transformation of *H. volcanii* using PEG600

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% PEG 600: 150 µl PEG 600 and 100 µl unbuffered spheroplasting solution.

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

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

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

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

Transformation of *H. volcanii* was carried out according to [209] with modifications detailed in [168]. In brief, plasmids or linear fragments of genomic DNA can be taken up by spheroplasts in the presence of polyethylene glycol. Plasmids must be passaged through a Dam methylase mutant *E. coli* strain (*dam*-) prior to transformation as the *H. volcanii* genome contains an *mrr* gene encoding a restriction enzyme which cuts methylated 5'-GATC-3' sequences[210]. Alternatively, *dam*+ plasmid DNA can be transformed into Δ*mrr H. volcanii* strains.

2.2.2- Nucleic acid extraction from cells

Genomic DNA extraction from *H. volcanii* by spooling

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

Lysis buffer: 100mM EDTA pH 8.0, 0.2% SDS

5ml Hv-YPC was inoculated with *H. volcanii* and was grown overnight at 45°c with 8 rpm rotation until A_{650}= 0.6-0.8 was reached. 1ml of culture was transferred to a 2ml round-bottomed tube and centrifuged at 3300 x *g* for 8 minutes at room temperature to pellet cells. The supernatant was removed, and the pellet was resuspended in 200µl ST buffer. 200µl lysis buffer was added, the tube was mixed by inversion, and 1ml 100% EtOH was added on top of the cell lysate. DNA at the interface was spooled with a capillary tip until the contents of the tube was homogeneous and clear. The spool was washed twice in 100% EtOH and the excess allowed to drain. The DNA was air-dried and resuspended in 500µl TE. 400µl isopropanol and 50µl sodium acetate were added to precipitate DNA. Sampled were centrifuged at 11,000 x *g* for 5 minutes at 4°c, and the pellets washed in 1ml 70% EtOH before air-drying. When the EtOH had evaporated, the pellet was resuspended in 100µl TE and stored at 4°c.

When genomic DNA was to be sequenced using Oxford Nanopore MinION, the following amendments to the above method were made:

2ml of culture was initially centrifuged and quantities of ST buffer and lysis buffer were both increased to 500µl. After spooling the DNA at the interface, washing with 100% EtOH and resuspending in 500µl TE, RNase A was added to a final concentration of 0.03mg/ml and was incubated for 1 hour at 37°c. Proteinase K was then added to a final concentration of 0.04 mg/ml and was left shaking gently for 1 hour at 37°c. DNA was overlayed with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) mix and was shaken. Samples were then centrifuged 11,000 x *g* for 5 minutes at at 4°c. The upper aqueous layer was removed and transferred to a fresh tube. An equal volume of chloroform:isoamyl alcohol (24:1) mix was added and samples were again centrifuged at 11,000 x *g* for 5 minutes at 4°c. 1/10 volume sodium acetate was added and overlayed with 2 volumes 100% EtOH. DNA was spooled at the interface as previously, swirled in 70% EtOH twice, excess drained, and resuspended in 200µl TE.

RNA extraction from *H. volcanii*

Eppendorf Biopur Safe-Lock tubes and RNase-free reagents were used throughout the RNA extraction protocol. Surfaces and equipment were cleaned with diethyl pyrocarbonate (DEPC) water prior before use.

5ml Hv-YPC broth was inoculated with *H. volcanii* and grown overnight at 45°c with 8rpm rotation. This culture was used to inoculate 5ml fresh Hv-YPC broth and was again grown overnight at 45°c with 8rpm rotation. The second overnight culture was used to inoculate 50ml Hv-YPC broth and was grown at 45°c with 120rpm shaking until A_{650} =0.2 was reached. The culture was transferred to a cold 50ml Falcon tube and was centrifuged at 4629 x *g* for 8 minutes at 4°c to pellet the cells. The pellet was thoroughly resuspended in 1ml Trizol, pipetted up and down to lyse cells, and was transferred to a 1.5ml Eppendorf Biopur Safe-Lock tube. The tube was then incubated at room temperature for 5 minutes. 0.2ml chloroform was added and the tube was shaken vigorously by hand before being allowed to stand at room temperature for 5 minutes. The tube was centrifuged at 12,000 x *g* for 15 minutes at 4°c. The upper aqueous phase was removed and transferred to a fresh tube ensuring that the DNA containing interphase was not disturbed. RNA was precipitated from the aqueous phase by adding 0.5ml isopropanol. The sample was mixed gently through inversion and incubated at room temperature for 15 minutes. The sample was then centrifuged at 12,000 x *g* for 10 minutes at 4°c. The supernatant was removed, and the pellet washed with 1.2ml cold 75% ethanol. The sample was gently mixed by inversion before being centrifuged at 7,500 x *g* for 5 minutes at 4°c. Washing with 1.2ml cold ethanol was repeated. The supernatant was thoroughly removed and the pellet air dried at room temperature for 10-15 minutes (ensuring the pellet did not dry completely). The pellet was resuspended in 50µl RNase-free water.

5µl RNA was run on a 1% agarose gel in 1xTAE to assess RNA quality at 60V for 45-60 minutes.

<u>DNase treatment</u>

10µl RNA was transferred to a different 1.5ml Eppendorf Biopur Safe-Lock tube to keep as a pre-DNase treated sample. To degrade contaminating DNA, RNA samples were DNase treated using a TURBO DNA-*free* Kit (Thermo Fisher Scientific), according to the manufacturer's instructions.

2.2.3- Nucleic acid manipulation

Restriction digests

Restriction digests were carried out according to the manufacturer's instructions (NEB). Plasmid DNA was digested for at least 1 hour. Genomic DNA and PCR products were digested for 16 hours. CutSmart buffer (NEB) was used for all digests in this study, in which each enzyme had 100% activity.

PCR

DNA amplification was carried out using either Q5 Hot Start or OneTaq Hot Start polymerase (NEB). Q5 Hot Start was used for reactions requiring high fidelity and OneTaq Hot Start was used for diagnostic amplifications. Reaction conditions are shown below. Reactions were carried out in a Prime Thermal Cycler (Techne).

Table 2.5: PCR components

Q5 Hot Start	OneTaq Hot Start
200µM of each dNTP	200 µM of each dNTP
0.5 µM of each primer	0.5 µM of each primer
1 ng –1 µg genomic DNA	10 ng of template DNA
1 x Q5Reaction Buffer	1 x OneTaq GC Buffer
1 x Q5 High GC Enhancer	-
0.02 U/µl Q5 Hotstart	0.025 U/µl OneTaq

Table 2.6: PCR reaction conditions

Step	Q5 Hot Start	OneTaq Hot Start	
Initial denaturation	98°c, 30 seconds	94°c, 30 seconds	
Denaturation	98°c, 10 seconds	94°c, 30 seconds	30 cycles
Annealing	Tm°c, 20 seconds	Tm°c, 30 seconds	
Extension	72°c, 30 seconds per 1kb	68°c, 1 minute per 1kb	
Final extension	72°c, 10 minutes	68°c, 5 minutes	

Annealing temperatures for primers (Tm°c) were calculated using the following equation[211]:

$$\boldsymbol{Annealing\ temperature}\ (°c) =$$

$$(\mathbf{81.5} + (\mathbf{16.6} * \boldsymbol{log10}[\boldsymbol{Na}+]) + (\mathbf{0.41} * \%\boldsymbol{GC}) - (\mathbf{100} - \%\boldsymbol{homology}) - (\frac{\mathbf{600}}{\boldsymbol{length}})$$

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

Purification of PCR products

PCR products were purified using a Macherey-Nagel DNA purification kit, according to the manufacturer's instructions. In brief, DNA binds pH dependently to a silica membrane and washing with ethanol separates DNA from contaminants. DNA was eluted in 30µl of the provided elution buffer.

Colony PCR

H. volcanii patches growing on solid media were gently touched with a sterile 200µl pipette tip to pick up a small number of cells which were then resuspended in 100µl of sterilized distilled water by pipetting up and down. Cells were lysed through heating to 100°c for 30 minutes and were then placed on ice for 10 minutes. 1µl of the resulting crude DNA was used in PCR with OneTaq Hot Start polymerase (NEB).

Oligonucleotide synthesis

All oligonucleotides were synthesised by Eurofins MWG, Germany.

Nucleic acid quantification

To determine concentration and purity of DNA and RNA samples, absorbance at 260nm and the 260:280nm absorbance ratio were measured by spectrophotometer (Thermo Scientific NanoDrop 2000 Spectrophotometer).

Agarose gel electrophoresis

Agarose gels were cast using agarose powder and either TBE or TAE buffer; TBE was used as standard practice and TAE used when running RNA gels or when Southern blotting was required (with the exception of pulsed field gel electrophoresis). SYBR Safe (Invitrogen) was added prior to casting, diluted to a 1:25,000 strain to buffer ratio. Gel loading dye was added to samples to a final concentration of 1 x. Molecular markers, either 1kb ladder (NEB) or 100bp (NEB), were loaded onto the gel alongside samples. TBE gels (10cm) were run at 100V for ~1 hour and TAE gels as detailed in specific methods.

Agarose gel extraction and purification of DNA

Bands were excised from agarose gels using Dark Reader transilluminator (Clare Chemical Research) for visualisation so as to not expose the DNA to UV irradiation. DNA was purified using the Macherey-Nagel DNA purification kit, according to the manufacturer's instructions.

cDNA synthesis

cDNA was synthesised using SuperScript III Reverse Transcriptase (Invitrogen) from RNA extracted and DNase treated as described earlier in this chapter, according to the manufacturer's instructions. The enzyme used is an engineered version of Moloney murine leukemia virus reverse transcriptase with reduced RNase H activity. cDNA was stored at -20°c.

RT-PCR

PCR using OneTaq Hot Start polymerase (as previously detailed in this chapter) was carried out upon cDNA samples, all diluted to 100ng/µl prior to the PCR reaction.

Pulsed field gel electrophoresis

Lysis solution: 20mM Tris.HCl pH 8.8, 500mM EDTA, 1% N-lauroylsarcosine, 1 mg/ml proteinase K (as necessary), 10 µg/ml RNase A (Ribonuclease A from bovine pancreas, as necessary).

Wash solution: 25mM Tris.HCl pH 7.5, 100mM EDTA, 1mM PMSF (Phenylmethylsulfonyl fluoride, as necessary).

TBE: 89mM Tris, 89mM boric acid, 2mM EDTA pH 8.0.

5ml Hv-YPC or Hv-Ca broth (+Thy and +Mev as necessary) was inoculated with *H. volcanii* and incubated overnight at 45°c with 8 rpm rotation until A_{650} =0.8-1.0 was reached. 1ml of cells were transferred to a 2ml round-bottomed tube and pelleted at 3300 x *g* for 10 minutes. After removing the supernatant, the pellet was resuspended in 100µl 18% SW. 8µl of resuspended cells was transferred to a fresh tube and incubated at 42°c for 10 minutes. 320µl 1% low-melt agarose (SeaPlaque GTG, Lonza) in 18% SW, also at 42°c, was then mixed gently with the cells, and 80µl of the mixture was pipetted into plug molds (BioRad) which were then left at 4°c for 30 minutes to set. Plugs were transferred to a 2 ml tube containing 2ml lysis solution and 1 mg/ml proteinase K and were incubated at 45°c with gentle shaking for 3 hours. The lysis solution was removed and 2ml fresh lysis solution containing 1 mg/ml proteinase K and 10 µg/ml RNase A was added before incubating overnight at 45°c with gentle shaking. The lysis solution was removed, replaced with 2ml wash solution, and incubated at 37°c with gentle shaking for 30 minutes. Agarose plugs were transferred to fresh 2ml tubes containing 2ml wash solution with PMSF and were incubated at 37°c with gentle shaking for 1 hour. This wash step was repeated twice, once with PMSF and once without. Plugs were transferred to a fresh 2ml tube containing 500µl wash solution and were gamma irradiated with an exposure of approximately 35 Gy (Gammacell 1000 Elite Model ^{137}Cs source). This was

determined to be sufficient to linearise the 3483 kb *H. volcanii* main chromosome with integrated pHV4 (calculated as in [212]). Plugs were then washed twice in 2ml 0.5x TBE at 37°c with gentle shaking for 30 minutes. For short term storage, plugs were stored at 4°c in 1ml wash solution.

A 1.2% agarose (SeaKem, Lonza) gel was prepared with 0.5 x TBE and allowed to set. Plugs were sealed onto the gel comb using an aliquot of the same molten agarose prior to pouring the gel. Lambda PFG ladder was loaded after removal of the comb and wells were sealed with molten agarose. The gel was run in a CHEF Mapper (BioRad) with 2.2 L 0.5 x TBE at 14°c for 20 hours 46 minutes. The running programme consisted of a two-state included angle of 120°, voltage gradient of 6 V/cm with 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 for visualisation and Southern blotted.

2.2.4- Genetic manipulation of *H. volcanii*

Gene deletion/replacement using the pop-in/pop-out method

∆pyrE2 *H. volcanii* strains were transformed with gene deletion or replacement constructs (*pyrE2*+) which integrate through recombination at the gene locus. Transformations were plated on Hv-Ca agar (plus required additives) to select for the *pyrE2* marker (pop-in). Pop-in colonies were picked and grown without *pyrE2* selection in 5ml of the appropriate broth (as detailed for each transformation in subsequent chapters) overnight at 45°C with 8rpm rotation until A_{650}= 1.0 was reached. This culture was diluted 1/250 into fresh broth, and growth and dilution were repeated (the number of successive overnights again detailed for each transformation in subsequent chapters). Relieving *pyrE2* selection allows recombination to remove the plasmid and native gene from the genome. The final culture was plated on Hv-Ca + 5FOA (+additives as required) to select for loss of *pryE2* and therefore loss of the plasmid (pop-out). Recombination to remove the plasmid can result in either wild-type cells or deletion/replacement mutants. If the gene is replaced with a selectable marker (*trpA* in this study), mutants can be selected for directly. This method of gene deletion/replacement was developed in [213] and is illustrated in Figure 2.5. Colonies were then re-streaked onto appropriate media, and their genotypes verified through colony PCR and/or colony hybridisation or restriction digestion followed by Southern blotting.

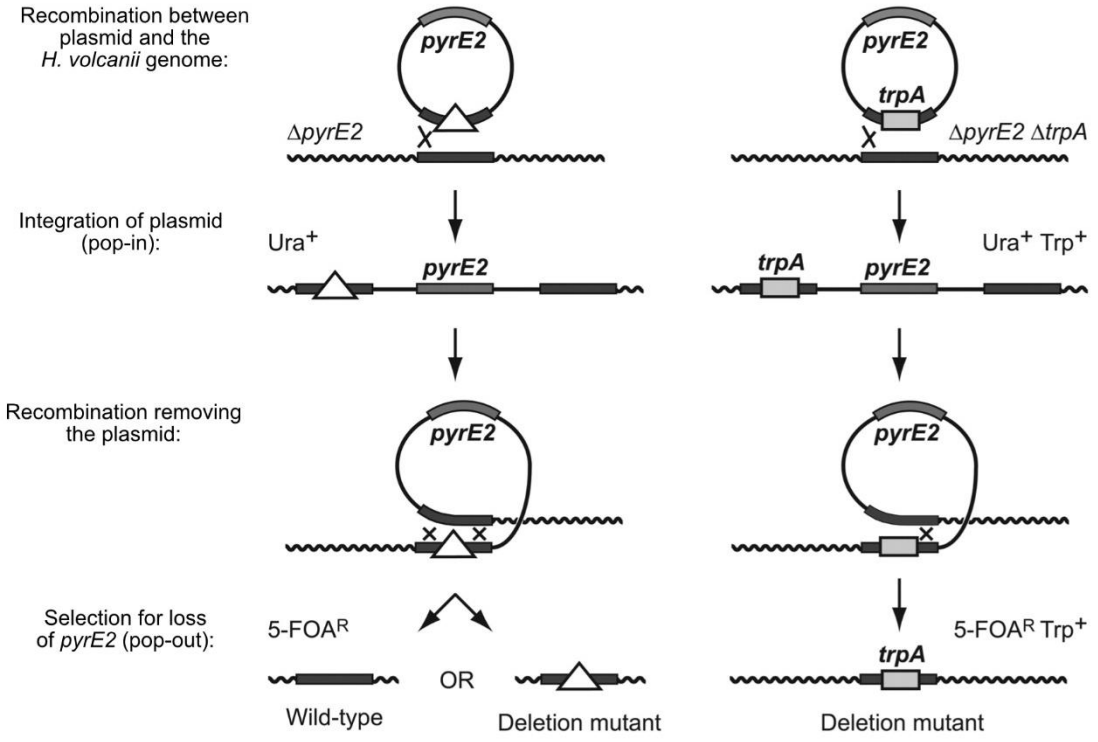


Figure 2.5: Gene deletion via the pop-in/pop-out method. The plasmid recombines with the genome and its integration is selected for using the pyrE2 marker, creating pop-in strains. A second recombination event is permitted when pyrE2 selection is removed and cultures are plated on 5-FOA for *pyrE2* counter-selection (creating pop-out strains). In the left of the figure, the second recombination event can result in either wild-type or mutant genotypes. In the right of the figure, *trpA* is used to directly select for mutants.

Figure modified from [46].

2.2.5- Genotype screening

Multiple methods can be used to screen presence and absence of a gene in *H. volcanii*. Strains with selectable phenotypes can be selected for through plating on selectable media (see Table 2.4). However, due to the polyploid *H. volcanii* genome, merodiploidy can occur meaning that a mixture of mutant and wild-type alleles may be present on different chromosome copies. To ensure all copies of the genome carry the same allele, colony hybridisation and Southern blotting were used. In both cases, DNA is denatured and transferred to a positively charged membrane through colony lift or vacuum transfer, respectively.

Colony lift

20 x SSPE: 3M NaCl, 230mM NaH_2PO_4, 32mM EDTA, pH7.4 (adjusted with NaOH)

Denaturing solution: 1.5M NaCl, 0.5M NaOH.

Neutralisation solution: 1.5m NaCl, 0.5M Tris.HCl, 1mM EDTA.

H. volcanii colonies were patched onto solid media using autoclaved wooden toothpicks and incubated at 45°C for ~4 days. A circular piece of Amersham Hybond – N^+ membrane was placed onto the surface of the media for 1 minute to lift the colonies. The membrane was placed colony side up onto Whatman paper soaked in 10% SDS for ≥5 minutes to lyse cells. The membrane was then transferred onto Whatman paper soaked in denaturing solution for ≥5 minutes, then Whatman paper soaked in neutralising solution twice, again for ≥5 minutes per transfer. The membrane was washed briefly (≤ 30 seconds) in 2 x SSPE and was allowed to air dry. DNA was crosslinked to the membrane with 3000 J/m^2 UV.

Southern blot vacuum transfer

20 x SSPE: 3M NaCl, 230mM NaH_2PO_4, 32mM EDTA, pH7.4 (adjusted with NaOH).

Denaturing solution: 1.5M NaCl, 0.5M NaOH.

H. volcanii genomic DNA was digested and resolved on a 200ml 0.75% TAE gel at 50V for 16 hours with buffer circulation. For visualisation, the gel was post-stained with 0.5µg/ml ethidium bromide for 30 minutes. The DNA within the gel was acid-nicked for 15 minutes in 0.25M HCl, washed with dH_2O for 10 minutes, and denatured in denaturing solution for 45 minutes. For pulsed field gels, timings were doubled for acid-nicking, washing in dH_2O, and denaturing. A 15 x 25cm GE Healthcare Amersham Hybond – XL membrane was soaked in dH_2O for 5 minutes

before equilibration in denaturing solution for 2 minutes. For pulsed field gels, a smaller membrane of 14 x 15cm was used. A Vacugene XL gel blotter and VacuGene pump (Pharmacia Biotech) were used for 1 hour (2 hours for pulsed field gels) at 40mBar for vacuum transfer. The membrane was washed briefly (≤ 30 seconds) in 2 x SSPE before crosslinking the DNA to the membrane with 3000 J/m^2 UV.

Hybridisation

100 x Denhardt's solution: 2% Ficoll 400, 2% polyvinyl pyrrolidone (PVP) 360, 2% BSA (bovine serum albumin, Fraction V).

20 x SSPE: 3M NaCl, 230mM NaH_2PO_4, 32mM EDTA, pH7.4 (adjusted with NaOH).

Pre-hybridisation solution: 6 x SSPE, 1% SDS, 5 x Denhardt's solution, 200 µg/ml salmon sperm DNA (Roche, boiled for 5 minutes before addition).

Hybridisation solution: 6 x SSPE, 1% SDS, 5% dextran sulphate.

Low-stringency wash solution: 2 x SSPE, 0.5% SDS.

High-stringency wash solution: 0.2% SSPE, 0.5% SDS.

Membranes were incubated in 40ml pre-hybridisation solution at 65°C for ≥ 3 hours with rotation in a Hybridiser HB-1D oven (Techne). To generate radiolabelled DNA probes, 50ng of DNA was denatured at 100°C for 5 minutes, and then incubated with 0.74 MBq of [α-^{32}P] dCTP (Perkin Elmer) and HiPrime random priming mix (Roche) at 37°C for 15 minutes. A BioRad P-30 column was used for purification. The probe was added to 450µl of 10 mg/ml salmon sperm DNA and denatured at 100°C for 5 minutes, then placed on ice. For Southern blotting, 1µl of 1µg/ml 1 kb ladder or 1µg/ml Lambda ladder was added in the radiolabelling mix. The pre-hybridisation solution was removed and 30ml hybridisation solution followed by the DNA probe was added. Membranes were then incubated at 65°C overnight with rotation. Membranes were then washed 4 times: twice in 50ml low-stringency wash solution for 10 minutes and 30 minutes, and twice in 50ml high-stringency wash solution for 30 minutes. Whatman paper was used to blot off excess liquid from membranes before they were wrapped in cling film and exposed to a phosphorimager screen (Fujifilm BAS Cassette 2325) for ≥ 24 hours. A GE Healthcare Amersham™ Typhoon™ Biomolecular Imager was used to scan the screen.

2.2.6- Phenotyping of *H. volcanii*

Tryptophan gradient plates

A gradient of tryptophan across a plate was generated through pouring 17ml Hv-Ca +Trp agar of the desired tryptophan concentration on a 7° slant, forming a tapered wedge (Figure 2.6). The plate was then placed flat once set and 50ml Hv-Ca agar was poured on top of the tapered wedge[49,214].

5ml cultures of the required H. volcanii strains were grown overnight in Hv-Ca broth with added tryptophan at the desired concentration, with 8 rpm rotation at 45°C until an A_{650} of 0.6-0.8 was reached. These cultures were then diluted into fresh Hv-Ca broth with tryptophan added to the same concentration used prior and incubated at 45°C with 8 rpm rotation overnight until an A_{650} of 0.4 was reached. Serial dilutions of the cultures in 18% SW to 10^{-4} were prepared. For each strain, a fresh autoclaved paintbrush (The Range) was first wetted in 18% SW, dipped into the diluted culture, and then painted in one direction across the plate. The same paintbrush was then dipped back into the diluted culture and a second line was painted over the first in the opposite direction. The plates were allowed to dry and then incubated at 45°C for 4-5 days.

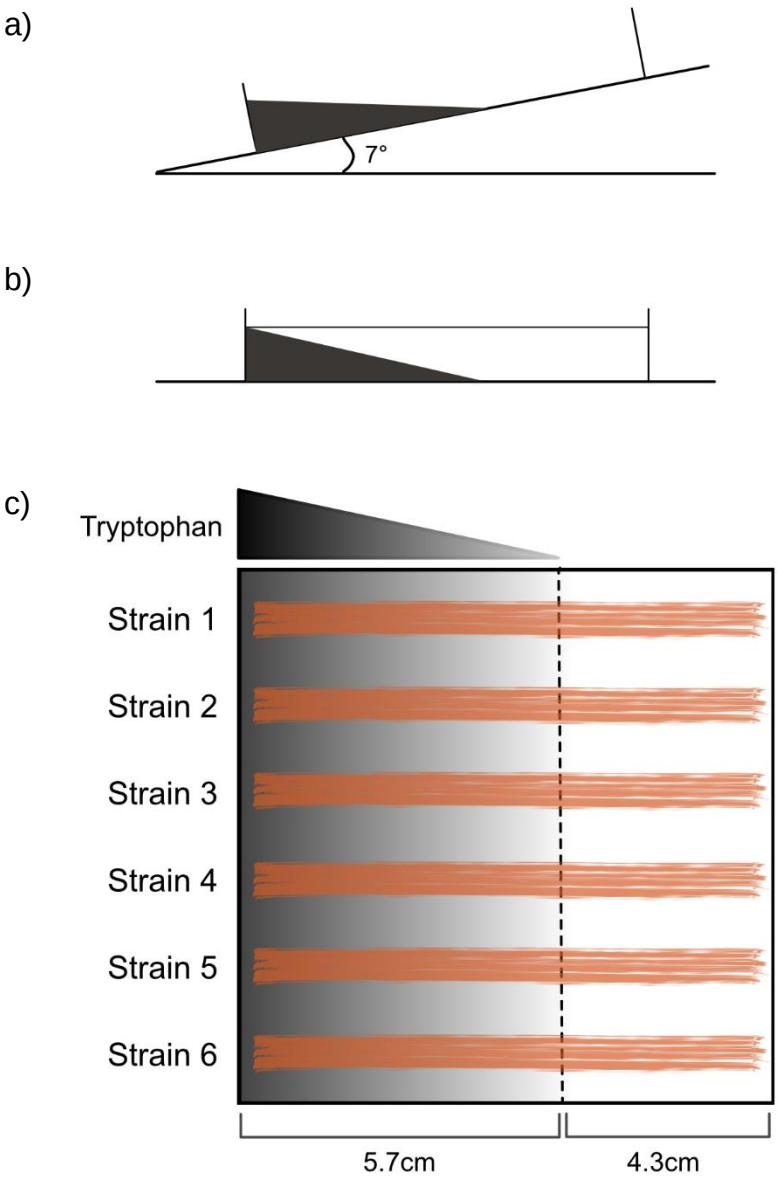


Figure 2.6: Tryptophan gradient plates. **a**) A 17ml wedge of Hv-Ca agar with tryptophan added to the desired concentration is poured with the plate at 7° angle, allowed to dry, and **b**) 50ml of Hv-Ca agar without tryptophan is poured on the top with the plate on a flat surface. **c**) Strains are then painted across the gradient.

Standard growth assay

5ml of appropriate broth (as specified per experiment) was inoculated with colonies grown on solid media and incubated at 45°C overnight with 8 rpm rotation. The culture was then diluted in 5ml of fresh broth and grown overnight. Cultures were diluted for a second time in fresh broth and grown overnight until A_{650}= 0.4 was reached. Cultures were diluted to 1 in 1000 before adding 150µl of culture (containing additions as necessary) and appropriate blanks to the wells of a 96 well

microtiter plate (Coring). The plate was sealed with 1.25cm microporous tape (Boots) and incubated in an Epoch 2 Microplate Spectrophotometer (BioTek) at 45°C with double-orbital shaking at 425 rpm for 72 hours. Readings were taken every 15 minutes at A_{650}. To calculate generation time, A_{650} values were plotted on a $\log_2$ axis and the time interval corresponding to a doubling in A_{650} value during logarithmic growth phase was found.

Generation times of strains vary between experiments when using this method to measure growth and consequently, this study compares only the growth of strains within the same experiment i.e., those that have been grown on the same microtiter plate with A_{650} measurements made simultaneously. For this reason, generation times are not absolute but the relationship between strains remains consistent.

DNA damage sensitivity assays

Ultraviolet (UV) irradiation sensitivity

Colonies grown on Hv-YPC agar were used to inoculate 5ml of Hv-YPC broth and were grown at 45°C overnight with 8 rpm rotation. This culture was then diluted in fresh Hv-YPC broth and again grown at 45°C overnight with 8 rpm rotation. The second overnight was then diluted again in fresh Hv-YPC broth to reach an A_{650} of ≈ 0.4 on the following day after growing at 45°C overnight 8 rpm rotation for the third time. When an A_{650} of ≈ 0.4 was reached, serial dilutions of the cultures from 10^0- 10^{-7} were made and 20µl of culture was spotted on to Hv-YPC agar in duplicate. Plates then dried at room temperature before either being exposed to UV or not exposed to UV radiation to form a control. Plates were incubated at 45°C until colonies became visible.

Gamma irradiation sensitivity

Colonies grown on Hv-YPC agar were used to inoculate 5ml of Hv-YPC broth and were grown at 45°C overnight with 8 rpm rotation. This culture was then diluted in fresh Hv-YPC broth and again grown at 45°C overnight with 8 rpm rotation. The second overnight was then diluted again in fresh Hv-YPC broth to reach an A_{650} of ≈ 0.4 on the following day after growing at 45°C overnight 8 rpm rotation for the third time. When an A_{650} of ≈ 0.4 had been reached, serial dilutions of the cultures from 10^0- 10^{-8} were made and 20µl of culture was spotted on to Hv-YPC agar plates in duplicate. Plates then dried at room temperature before either being exposed to a ^{137}Cs gamma radiation source or not exposed to gamma radiation to form a control. Exposure varied between 3.02 and 5.55 Gy/min across the barrel in which the plates were placed for irradiation. The centre of the barrel was exposed at a rate of 4.44 Gy/min and is the rate used to calculate exposure in subsequent chapters of this study. Plates were incubated at 45°C until colonies became visible.

Once spotted, all plates in the DNA damage assays described above were stored in black plastic bags to shield from visible light to prevent DNA repair through photoreactivation catalysed by photolyases[215]. Colonies were counted at the lowest dilution at which single colonies could be seen to minimise pipetting error. Two repeats per dilution were plated per exposure dosage (technical replicates) and experiments were carried out three times (biological replicates). The proportion of cells seen at varying exposures in comparison to those surviving in the unirradiated control in the same experiment were calculated, resulting in a total of six values across the three independent experiments, each with two technical replicates. The mean of these six values was calculated and is denoted as “fraction of cells surviving” in this study.

2.2.7- Proteomic analysis

Whole cell lysate protein extraction

5ml Hv-YPC broth was inoculated with *H. volcanii* grown on solid media and grown overnight at 45°C. The culture was then used to inoculate 50ml Hv-YPC broth and grown overnight at 45°C until an A_{650} of ≈ 0.8 was reached. Cells were centrifuged at 6000 *x g* for 12 minutes at room temperature in a 50ml falcon tube. The supernatant was removed, cells were resuspended in 2ml 18% SW, and centrifuged again. The supernatant was again removed, the cells were resuspended in 1ml 18% SW, and transferred to a 5ml Eppendorf tube. Samples were placed on ice and sonicated at ≤ 8µm amplitude for 5 x 20 second cycles (or repeated until the sample became clear). The sonicated culture was transferred to a 1.5ml Eppendorf tube and centrifuged at 12,000 *x g* for 30 minutes at room temperature. The supernatant (clarified cell lysate) was transferred to a new 1.5ml Eppendorf tube and was stored at -80°C.

SDS- Polyacrylamide Gel Electrophoresis (SDS-PAGE)

12.5% SDS-PAGE gel (resolving): 12.5% acrylamide/bisacrylamide Protogel, 0.38M Tris pH 8.8 with 0.5% 2,2,2-trichloroethanol (TCE; Sigma T54801), 0.1% SDS, 0.05% AMPS (ammonium persulfate), 0.05% TEMED (tetramethylethylenediamine).

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

10x SDS-PAGE running buffer: 0.25M Tris, 1.92M glycine, 1% SDS.

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

Gels were cast in 1.0 mm Novex casettes (ThermoFisher Scientific). ~6 ml 12.5% resolving gel was poured into the cassette and overlayed with 200µl isopropanol. Once set, the isopropanol was removed, and the resolving layer was washed with dH_2O. ~2ml 3% stacking gel was poured on top of the resolving layer and an appropriate comb was inserted. Laemmlli buffer was added to the protein samples at a 1x concentration which were then boiled at 94°C. Proteins were resolved at 200V, 36mA for 90 minutes in 1x SDS-PAGE running buffer alongside either Blue Protein Standard, Broad Range Ladder (NEB) or ColorPlus™ Prestained Protein Ladder, Broad Range (NEB). For protein visualisation, gels were incubated with SimplyBlue™ SafeStain (ThermoFisher Scientific) overnight and de-stained in dH_2O.

Protein identification by Western Blot

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

TBST: 1 x TBS, 0.1% Tween20

Freshly run SDS-PAGE gels were removed from the cassettes and the stacking gel and foot were removed. Proteins were transferred onto a Midi-Size PVDF membrane, using the Bio-Rad Trans-Blot Transfer System and 1.5mm gel Bio-Rad pre-programmed transfer protocol according to the manufacturer's instructions.

When transfer was complete, the membrane was transferred to a 50ml Falcon tube. To block the membrane, 10ml 5% milk in TBST (w/v) was added and was placed on a tube roller to rotate at room temperature for 1 hour. The solution was removed and replaced with 5ml 1% milk in TBST (w/v) and primary antibody in a 1:2000 dilution, then left to rotate on a tube roller at 4°C overnight. The membrane was washed 4 times with 10ml TBST at room temperature before incubation with the secondary antibody. 5ml TBST containing the secondary antibody (peroxidase conjugate) in a 1:5000 dilution was added, and the membrane incubated at room temperature on a tube roller for 2 hours. The membrane was then washed 4 times with 10ml TBST at room temperature.

Western blots were imaged with Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare). Luminol enhancer and hydrogen peroxide mixed 1:1 was placed evenly across the membrane and the excess removed. Chemiluminescence was detected using a Vilber Fusion FX.

2.2.8- ChIP-Seq

Protocol adapted from [216,217] by the Schmid lab, Duke University, Durham, NC, 27708 USA. Obtained via personal communication further adapted by the Allers lab, according to resources available.

Lysis buffer: 1mM EDTA, 0.1% Na-deoxycholate, 1 x protease inhibitor cocktail, 140mM NaCl, 50mM HEPES-KOH pH 7.5, 1% Triton X-100. Prepared fresh at 4°C.

PBS/BSA solution: 50ml PBS, 250mg BSA. Prepared fresh at 4°C.

Wash buffer: 1mM EDTA, 10mM Tris-HCl pH 8.0, 0.5% NP40, 250mM LiCl, 0.5% Na-deoxycholate. Storage at 4°C.

Elution buffer: 50mM Tris-HCl pH 8.0, 10mM EDTA, 1% SDS.

TE/SDS: 1mM EDTA, 10mM Tris-HCl pH 8.0, 1% SDS.

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

Crosslinking of samples

3 x 5ml of Hv-YPC broth was inoculated with *H. volcanii* and grown overnight. The culture was diluted in fresh media and grown overnight twice before being used to inoculate 100ml and 1 litre Hv-YPC broth. The 100ml cultures were grown to A_{650}= 0.5 and the 1 litre cultures to A_{650}= 0.05. Cultures were centrifuged at 5,000 x *g* and resuspended in 40ml Hv-YPC broth to reduce the volume of formaldehyde used. This results in approximately the same final cell density when the cells are resuspended in 40ml Hv-YPC. Formaldehyde was added to all cultures to a final concentration of 1% and incubated on a rocking platform for 20 minutes at room temperature. Glycine was then added to a final concentration of 0.125mM. Cells were pelleted at 2880 x *g* for 10 minutes at 4°C. Each cross-linked culture was washed 3 times with 10ml cold basal salts buffer, pelleting as previously. After the final wash, all liquid was removed by tapping the mouths of the tubes on a paper towel. Pellets were stored at -80°C.

Bead preparation

50µl Dynabeads (protein A, Invitrogen) per sample plus an extra 50µl were placed into a 15ml falcon tube. Beads were centrifuged at 716g for 5 minutes at 4°C three times: the first to remove buffer, and the second and third to wash with 10mL cold PBS/BSA. Beads were then centrifuged for a further 1 minute to spin down liquid from the side of the tube. Beads were resuspended in 250µl cold PBS/BSA plus 1µg antibody per sample and were incubated for at least 2 hours at 4°C on a rotating

platform. After incubation, beads and antibodies were centrifuged at 711 x *g* for 5 minutes at 4°C three times: the first to remove the current PBS/BSA, and the second and third to again wash with 10ml cold PBS/BSA. Beads were then centrifuged for a further 1 minute to spin down liquid from the side of the tube. Beads were resuspended in 30µl PBS/BSA per sample and stored briefly at 4°C until lysates had been sonicated.

Sonication of lysates

Crosslinked pellets were thawed on ice, resuspended in 800µl lysis buffer, and transferred 5ml Eppendorf tubes. Samples were sonicated at ≤ 8µm amplitude for 8 x 20 second on/off cycles and 14 x 20 second on/off cycles for samples grown to A_{650}≈ 0.05 and A_{650}≈ 0.5 prior to crosslinking respectively. 5µl of each sonicated sample was run on a 50ml 1.2% agarose gel to ensure that DNA was a smear between 200-800bp with the highest concentration of fragments at around 500bp. Lysates were transferred to pre-chilled 2ml tubes and centrifuged at 20,817 x *g* for 10 minutes at 4°C. 500µl supernatant (cleared cell lysate) was transferred to a new 2ml tube. The antibody-bound beads were very briefly vortexed and 30µl was added to 500µl cleared cell lysate for each sample. These immunoprecipitation reactions were incubated overnight at 4°C on a rotating platform. The remaining cleared cell lysate was transferred to another 2 ml tube and stored at -20°C as a pre-immunoprecipitation control.

Bead washing and crosslink reversal

The immunoprecipitation reactions were placed onto a GE Healthcare MagRack 6 magnet and supernatant was removed. The beads were washed twice with 1ml lysis buffer, twice with 1ml wash buffer, and once with 1ml TE. They were then centrifuged at 956 x *g* for 3 minutes at 4°C, and residual buffer was removed. 50µl elution buffer was added to beads which were then incubated for 10 minutes in a water bath at 65°C and tapped gently every 2 minutes to maintain the beads in suspension. The bead suspension was centrifuged at 20,238 x *g* for 30 seconds at room temperature and then placed back onto the magnet. 40µl and 10µl supernatant was transferred to new, separate 2ml tubes to be used in the crosslink reversal step and in Western blotting respectively. The pre-immunoprecipitation cleared cell lysate samples were thawed on ice. 5µl and 10µl of this was transferred to new, separate 2ml tubes to again be used in the crosslink reversal step and in Western blotting respectively. 160µl TE/SDS was added to each tube containing 40µl supernatant from the elution step to raise the total volume to 200µl, and 95µl TE/SDS was added to each tube containing 5µl pre-immunoprecipitation cleared cell lysate to raise the total volume to 100µl. These samples were incubated in a water bath at 65°C overnight to reverse crosslinks. Samples for Western blotting were stored at -20°C.

Purification of DNA

Samples were removed from the water bath and diluted 1:1 in TE. RNase A was added to a final concentration of 0.2 mg/ml. Samples were mixed by inversion, spun down briefly, and incubated at 37°C for 2 hours. Proteinase K was added to a final concentration of 0.2 mg/ml. Samples were mixed by inversion, spun down briefly, and incubated in a water bath at 55°C for 2 hours. 200µl TE was added to the pre-immunoprecipitation cleared cell lysate samples to raise their total volume to 400µl. 400µl phenol:chloroform:isoamyl alcohol (25:24:1) was to each sample and organic and aqueous phases were mixed thoroughly. Samples were centrifuged at 11,000 x *g* for 5 minutes at 4°C to separate phases. The upper nucleic acid containing phase was pipetted off and transferred to a fresh 2ml tube. 1.5µl 20 µg/µl glycogen, NaCl to a final concentration of 200mM, and 800µl cold 80% EtOH was added. Samples were incubated overnight at -80°C. After incubation, samples were centrifuged at 20,817 x *g* for 30 minutes at 4°C to pellet DNA. Supernatant was removed, and pellets were washed with 500µl cold EtOH and centrifuged at 20,817 x *g* for 10 minutes at 4°C. Pellets were dried until all ethanol had evaporated, then resuspended in 50µl sterilized distilled water and stored at -20°C.

2.2.9- - DNA sequencing

All DNA sequencing was carried out by the Deep Seq facility, University of Nottingham. Analysis of ChIP-Seq data was also carried out by the Deep Seq facility. Analysis of the MinION seq data was carried out in this study, as described in Chapter 4. Methodologies written by the Deep Seq facility can be found in the appendices.

Chapter 3- ChIP-Seq to determine RadA localisation

3.1- Background

Previous data from the laboratory suggests that in *H. volcanii,* replication initiates from origins during early log phase growth (A_{650}= 0.1 or lower) but during mid/late log phase growth (A_{650}= 0.2 or higher), the use of origins is reduced, and recombination-dependent replication is used instead. In MFA plots of DNA extracted from cultures at between A_{650} = 0.05 and 0.1, peaks corresponding to origins of replication are well defined and show a maximum copy number between 2.2 and 2.0[49]; a copy number higher than 2.0 indicates that an origin has fired for a second time before the replication forks from the first firing have reached the terminus. However, MFA plots of DNA extracted from cultures grown to A600= 0.2 show less-defined peaks at origins of replication and the maximum copy number is reduced to 1.1 (Darya Ausiannikava, unpublished data), indicating that origin firing is significantly reduced. Nevertheless, the growth rate and genome copy number of cultures in mid/late log phase growth is not reduced, indicating that DNA replication is maintained at the same level as in early log phase. This suggests that cells have switched from origin-dependent to recombination-dependent replication, and therefore RadA may play a greater role during mid/late log phase growth.

RadA

As described in Section 1.4.1.1- RecA-family proteins: RecA, Rad51, and RadA, RadA is an archaeal RecA-family recombinase demonstrating greater structural and sequence homology to eukaryotic Rad51 than bacterial RecA. RecA-family recombinases have been found to be essential in origin-independent replication[86,87]. RadA is essential when all origins of replication found upon it are deleted from the *H. volcanii* main chromosome, presumably due to its role in recombination-dependent replication[49].

In *Pyrococcus furiosus*, RadA was split into two parts to better understand the link between the structure and function of the protein *in vitro*[218]. RadA was separated into the N-terminal third of the protein (RadA-n) containing the N-terminal domain only found in eukaryotic and archaeal homologues, and the C-terminal two thirds of the protein (RadA-c) containing the core domain found in all RecA-family homologues. RadA-n and RadA-c were assayed for ATPase activity and DNA pairing and strand

exchange activity. It was concluded that the core domain of RadA was necessary in D-loop formation and the N-terminal domain acts to enhance the efficiency of the reaction.

Recombination mediator proteins also exist in many organisms and are often paralogues of RecA-family recombinases. For example, Rad52 and Rad55-57 in yeast displace RPA (eukaryotic and archaeal SSB) and aid in loading Rad51 onto the DNA[219-221]. Rad55-Rad57 are Rad51 paralogues. In euryarchaeal species, RadB is a recombination mediator and is a paralogue of RadA. RadB consists of only the core ATPase domain found in this family of proteins and does not contain the N-terminal domain found in Rad51 and RadA[218]. In *Pyrococcus furiosus,* RadA and RadB have been shown to interact *in vitro*[218] and in *Haloferax volcanii*, their interaction was demonstrated *in vivo*[222]. In *H. volcanii*, RadB was found to be necessary for efficient homologous recombination and survival in cells subjected to DNA-damaging agents[222]. However, Δ*radB* phenotype could be suppressed by point mutations *radA-S101P* or *radA-A196V*. The *radA-S101P* point mutation is found at an “elbow-joint” between the core domain and the N-terminal domain of RadA. It was hypothesised that this could alter the orientation of the N-terminal domain and as this overcomes the Δ*radB* phenotype, a role of RadB could be to induce conformational chape change in RadA. The *radA-A196V* point mutation is within the hydrophobic binding pocket implicated in RadA polymerisation. As the hydrophobicity of the binding pocket would increase with an alanine to valine substitution, it was thought that the interactions between the RadA monomers would be stronger and that if a role of RadB is to stabilise RadA filaments, this would be overcome by the *radA-A196V* point mutation[222].

Chromatin Immunoprecipitation Sequencing (ChIP-Seq)

Chromatin immunoprecipitation followed by next generation sequencing allows interaction between proteins and DNA to be assayed *in vivo*. Formaldehyde is used to crosslink proteins to DNA, and chromatin is sheared via sonication into small fragments, typically between 200 and 600bp. Antibodies are used to immunoprecipitate the protein of interest and bound DNA fragments. Crosslinks between the protein and DNA are reversed, and sequencing of the DNA fragments can determine where in the genome the protein of interest was bound. ChIP-Seq can identify targets for transcription factors and enhancers, and specific sequence motifs[223,224]. Some of the first studies using ChIP-Seq concerned binding of the STAT1 transcription factor in humans[225], profiling of histone methylation in human T cells[226], and chromatin state changes in the progression of cells from immature to adult states[227].

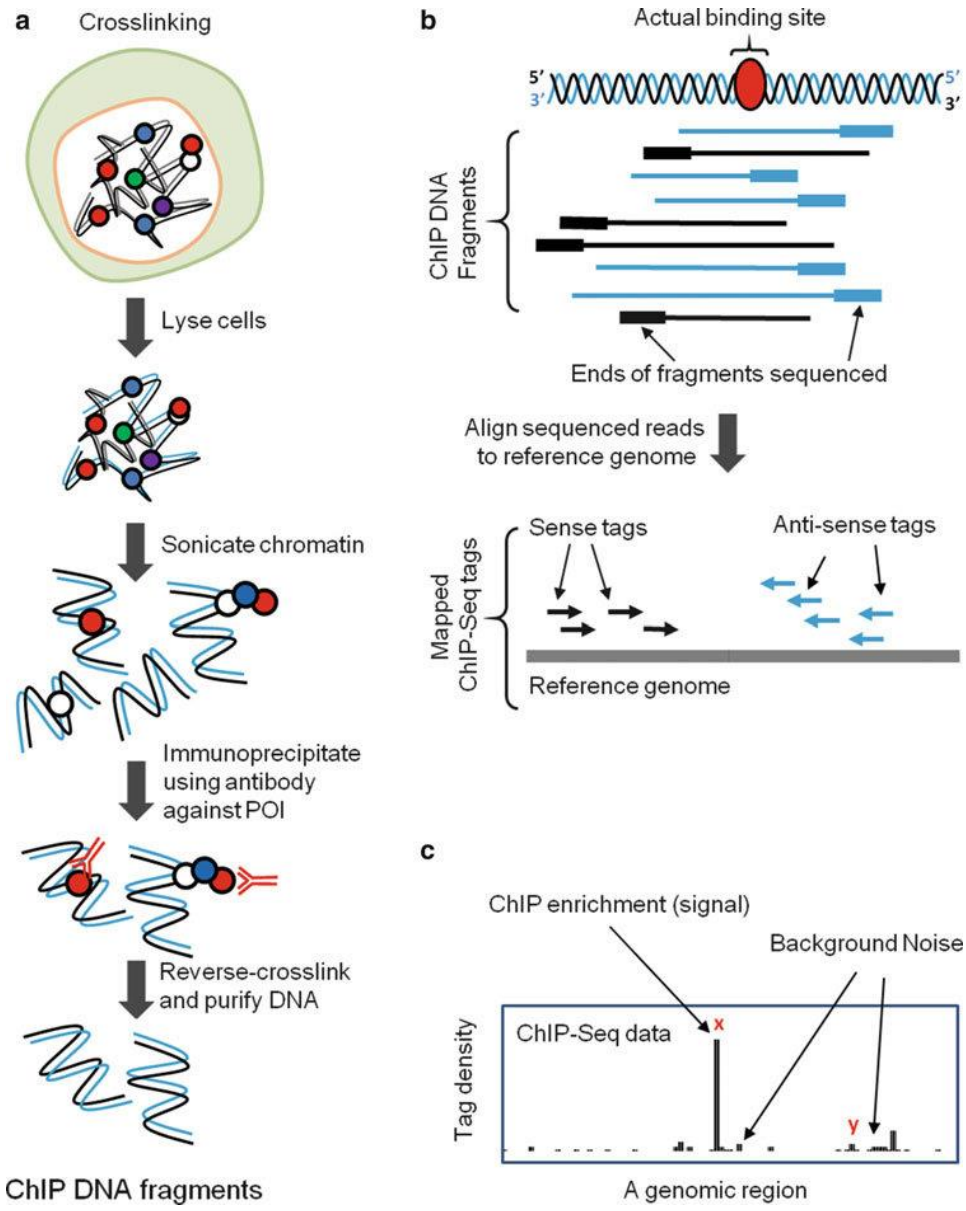


Figure 3.1: ChIP-Seq sample preparation and sequencing. a) Steps in the preparation of ChIP DNA. Cultures are crosslinked and cells are lysed. The lysate is sonicated, immunoprecipitation against the protein of interest (POI) is undertaken, and crosslinks between proteins and DNA are reversed. Finally, DNA is purified. **b)** Fragments obtained are sequenced and aligned to the reference genome. **c)** Histograms are created from sequencing data demonstrating the density of fragments mapping to areas of the reference genome. x and y demonstrate signal and noise, respectively.

Figure taken from [228].

Recently, the utility of ChIP-Seq to investigate protein-DNA interactions in *H. volcanii* was demonstrated. A conserved gene (*cdrS*) encoding a small protein involved in multiple gene networks was found to be highly transcribed. ChIP-Seq identified 18

binding sites of the protein encoded by *cdrS*, including one upstream of a promotor for a cell division gene, and one upstream of the essential gene *dacZ* involved in c-di-AMP signalling[229]. These results indicated that CdrS is a transcription factor involved in the regulatory network controlling cell division and metabolism. Another study was able to couple ChIP-Seq with quantitative real-time PCR and determined the function and binding motifs of a thiol-based transcription factor OxsR[230]. ChIP-Seq has also been used in other model archaeal species. Binding sites of the histone HypA were determined in halophile *Halobacterium salinarum*[231], and also of the BarR Lrp family transcription factor in hyperthermoacidophile *Sulfolobus acidocaldarius*[232].

Antibody quality is a key component in ChIP-Seq experimental design. The antibody must be sufficiently sensitive and specific to enrich for the protein over the background and thereby detect binding events. It is for this reason that validation of antibody binding via Western blotting should undertaken prior to carrying out ChIP-Seq, to confirm that it is sufficiently sensitive and specific.

RecA family recombinase localisation

RecA

ChIP-Seq against RecA in *E. coli* has been used to demonstrate Chi-dependent loading of RecA in DNA double-strand break (DSB) repair[233]. Additionally, DSB-independent RecA binding was revealed at rRNA genes, tRNA genes, and ribosomal protein genes – all loci not associated with the positions of Chi sites[233]. In *E. coli,* repair of DSBs occurs through homologous recombination. In brief, the RecBCD complex acts to unwind the DNA double helix at a DSB and recognises Chi sequences (5′-GCTGGTGG-3′) within the 3’ ssDNA strand. When a Chi sequence is reached, RecBCD is thought to undergo conformational shape change and the nuclease domain of RecB loads RecA onto a ssDNA loop created when unwinding the duplex (reviewed in [234]). Since ChIP-Seq found RecA binding at loci not associated with Chi sequences, it is suggested that RecA binding at these sites is not RecBCD-dependent.

Rad51

ChIP-Seq has also been used to determine Rad51 localisation. Recently, ChIP-Seq was used to assess regulatory roles of Rad51 in cancer cell lines and revealed that in all cell lines tested, promotors of autophagy pathway-related genes were occupied by Rad51 more often than in non-cancerous cell lines[235]. Therefore, it was possible to predict a non-canonical role of Rad51 in the regulation of autophagy-related genes.

FLAG-tag

A FLAG-tag is a peptide tag commonly used in chromatin immunoprecipitation. The tag consists of the amino acid sequence DYKDDDDK and is approximately 1 kDa. Single, multiple or tandem FLAG-tags can be added to a protein at either the N- or C-terminus. The FLAG-tag is hydrophilic and is regarded as a highly specific protein tag with a dissociation constant of 100nM[236-238]. Due to this specificity, FLAG-tags are often used in ChIP-Seq (examples include [239-241]) and methods to attach a FLAG-tag to proteins using CRISPR for ChIP-Seq have also been developed[242,243].

3.2- Aims

The goal of this chapter was to determine RadA localisation via ChIP-Seq in *H. volcanii* strains using either origin-dependent or origin-independent replication. To do this, aims were as follows:

- Generate strains expressing FLAG-RadA in a wild-type background (therefore retaining all origins)
- Generate strains expressing FLAG-RadA with origins deleted from the main chromosome
- Ensure antibody detection of FLAG-RadA is sufficiently sensitive and specific for ChIP-Seq analysis
- Measure the sensitivity of the FLAG-RadA strains to UV- and gamma-irradiation to ensure that the tagged RadA protein behaves in the same way as the wild-type protein
- Carry out ChIP-Seq to determine RadA localisation

3.3- Results

Prior to this study, plasmid pTA2424 was generated (Chapter 2: Materials and Methods), in which *radA* is FLAG-tagged at the 5' end (and therefore the N-terminus of the protein). In the crystal structure of RadA from *Pyrococcus furiosus*, the C-terminus appears to be buried (accessed through the RCSB Protein Data Bank, entry ID 1PZN[102]) and previous work demonstrated that the use of a His-tag at the N-terminal protein did not affect the function of the protein[244].

3.3.1- Strain generation and validation

FLAG-radA **in an** ***ori*****+ background**

In *H. volcanii,* when gene replacements are made, instead of replacing the gene directly, it is preferable to delete the wild-type allele leaving a tryptophan-marked deletion construct (the *trpA* gene necessary for tryptophan biosynthesis flanked by the upstream and downstream regions of the gene to be deleted) in its place, and then replace the deletion construct with the allele to be inserted. Therefore, ability to grow upon media lacking tryptophan should indicate successful deletions, and tryptophan selection can be used again when replacing the deletion construct with the allele to be inserted. This is of particular importance due to the polyploid nature of the *H. volcanii* genome. Colonies able to grow upon media containing tryptophan but not able to grow upon media lacking tryptophan should have replaced the deletion construct with the new allele and should not be merodiploid (if some copies of the genome retained the *trpA*+ deletion construct, colonies would be able to grow upon media lacking tryptophan).

H. volcanii strain H298 (Δ*radA::trpA*+, Δ*pyrE2*) was transformed with plasmid pTA2424 (FLAG-*radA*+, *pyrE2*+) to create pop-in strain H5464. For the pop-out step, 3 successive overnight cultures of H5464 were carried out in Hv-YPC broth to provide time for double crossover events to occur which would leave FLAG-*radA* in place of the Δ*radA::trpA*+ construct and remove the *pryE2* gene. Pop-outs were plated onto Hv-Ca with additions of 5-FOA and tryptophan to select for loss of *pyrE2* and to permit growth in loss of the Δ*radA::trpA*+ construct. Colonies from these plates were patched onto tryptophan+ and tryptophan- plates and those which only grew in the presence of tryptophan, and had therefore lost the Δ*radA::trpA*+ construct, were re-streaked and screened via Southern blotting (Figure 3.2). 3 colonies showing the correct genotype (H5501, H5502, and H5505) and were taken forward for further analysis.

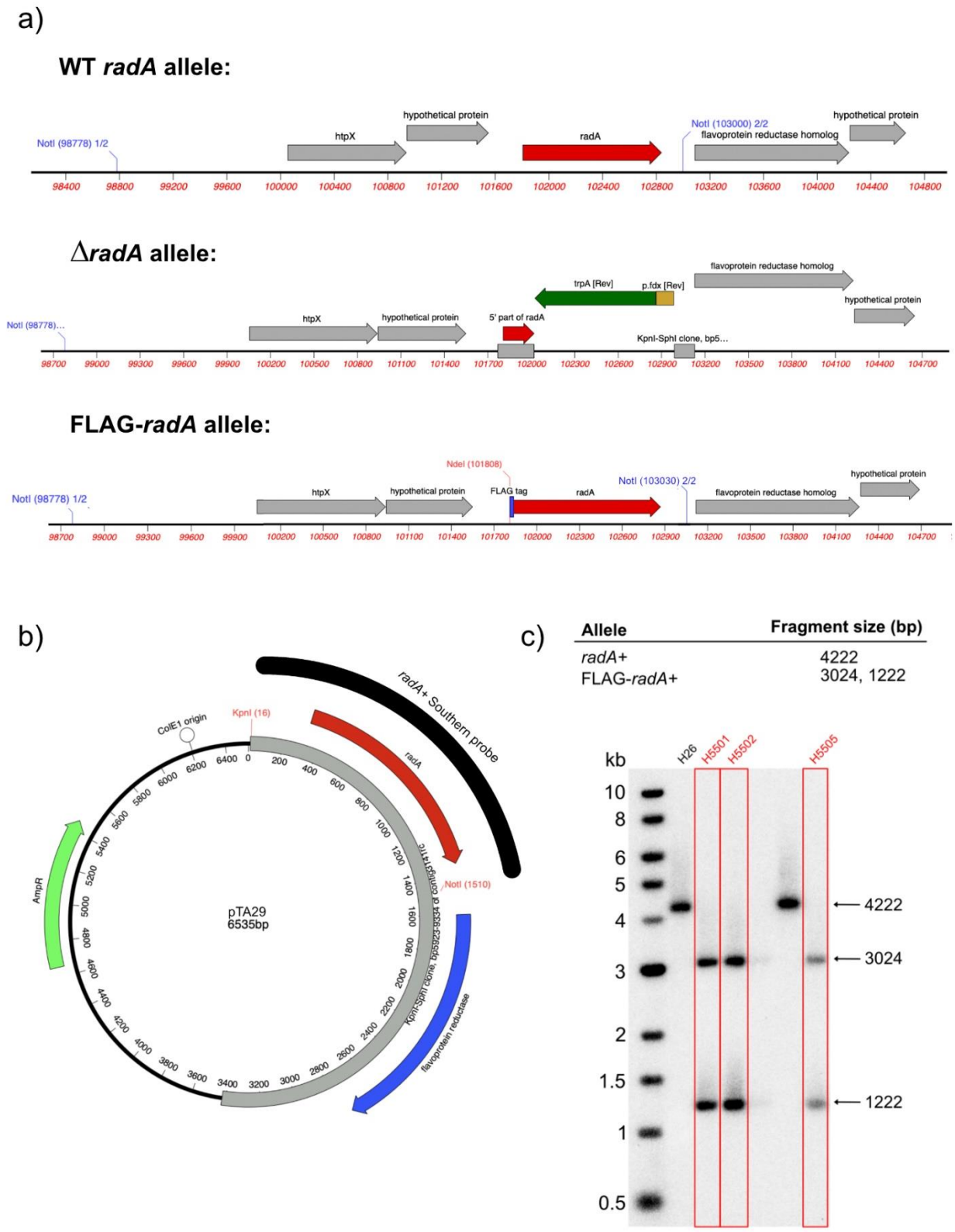


Figure 3.2: Southern blotting to screen FLAG-*radA* candidate strains. a) 3 alleles: *radA*+, Δ*radA::trpA*+, and FLAG-*radA*+. The *NotI* and *NdeI* restriction sites used to digest genomic DNA are shown. **b**) A 1494bp *radA*+ Southern hybridisation probe was created through digesting pTA29 with *KpnI* and *NotI*. **c**) Size of predicted fragments and a Southern blot showing the wild-type radA+ allele (strain H26) and three FLAG-*radA*+ strains (H5501, H5502, and H5505). Digestion of a Δ*radA::trpA*+ allele with *NotI* and *NdeI* would result in a fragment >10,000bp and would therefore not be visible on the Southern blot so was not used as a control.

FLAG-radA in a Δ*oriC1,* Δ*oriC2,* Δ*oriC3,* Δ*oripHV4-2* background

The initial aim was to delete *radA* in a strain with only one origin remaining on the main chromosome, namely *oriC1*, then insert the FLAG-*radA* gene and delete the final origin. *radA* was to be deleted and replaced with a tryptophan-marked deletion construct, as is the general case for gene replacements in *H. volcanii* as described previously. Previous work found that it is not possible to delete *radA* in a strain in which all origins are deleted on the main chromosome and therefore, it was necessary to replace *radA* with FLAG-*radA* prior to deletion of *oriC1*.

Strain H1965 (Δ*pyrE2,* Δ*trpA,* Δ*oriC2,* Δ*oriC3,* Δ*ori-pHV4*-2) was transformed with the *radA::trpA*+ deletion construct pTA1651 (Δ*radA::trpA*+*, pyrE2*+) and was plated on Hv-Ca agar using uracil and tryptophan biosynthesis as selection to create the pop-in strain H5465. H5465 was then transformed with p637 (*radA*+*, MevR+)* for *in trans* complementation of RadA and plated on Hv-Ca + Mev agar to select for uptake of the plasmid and to maintain tryptophan and uracil biosynthesis selection (to ensure pTA1651 remained integrated). RadA is necessary in recombination to allow the plasmid to pop-out of the genome and therefore when aiming to delete it, expression from an episome is necessary. For the pop-out step, two successful transformants (H5471 and H5472) were then grown for four successive overnights, the first two of which in Hv-Ca broth +Ura +Mev to select for p637 to provide the RadA necessary for recombination, and the second two in Hv-Ca +Ura to cure the strains of the plasmid. Pop-outs were then plated on Hv-Ca + 5-FOA to select for loss of *pyrE2*. To check for *radA* deletions, colonies were patched on Hv-YPC agar and screened through colony hybridisation using a *radA*+ probe (Figure 3.3a). All colonies retained the *radA* gene. An alteration was then made to the pop-out process: five successive overnight cultures were grown, the first two in Hv-Ca broth + Ura + Mev and the final 3 in Hv-Ca broth + Ura, and the final culture was plated on Hv-Ca +5-FOA agar. Once grown, pop-out colonies were again then patched on Hv-YPC agar. The screening was repeated and again, all colonies retained the *radA* gene (Figure 3.3b).

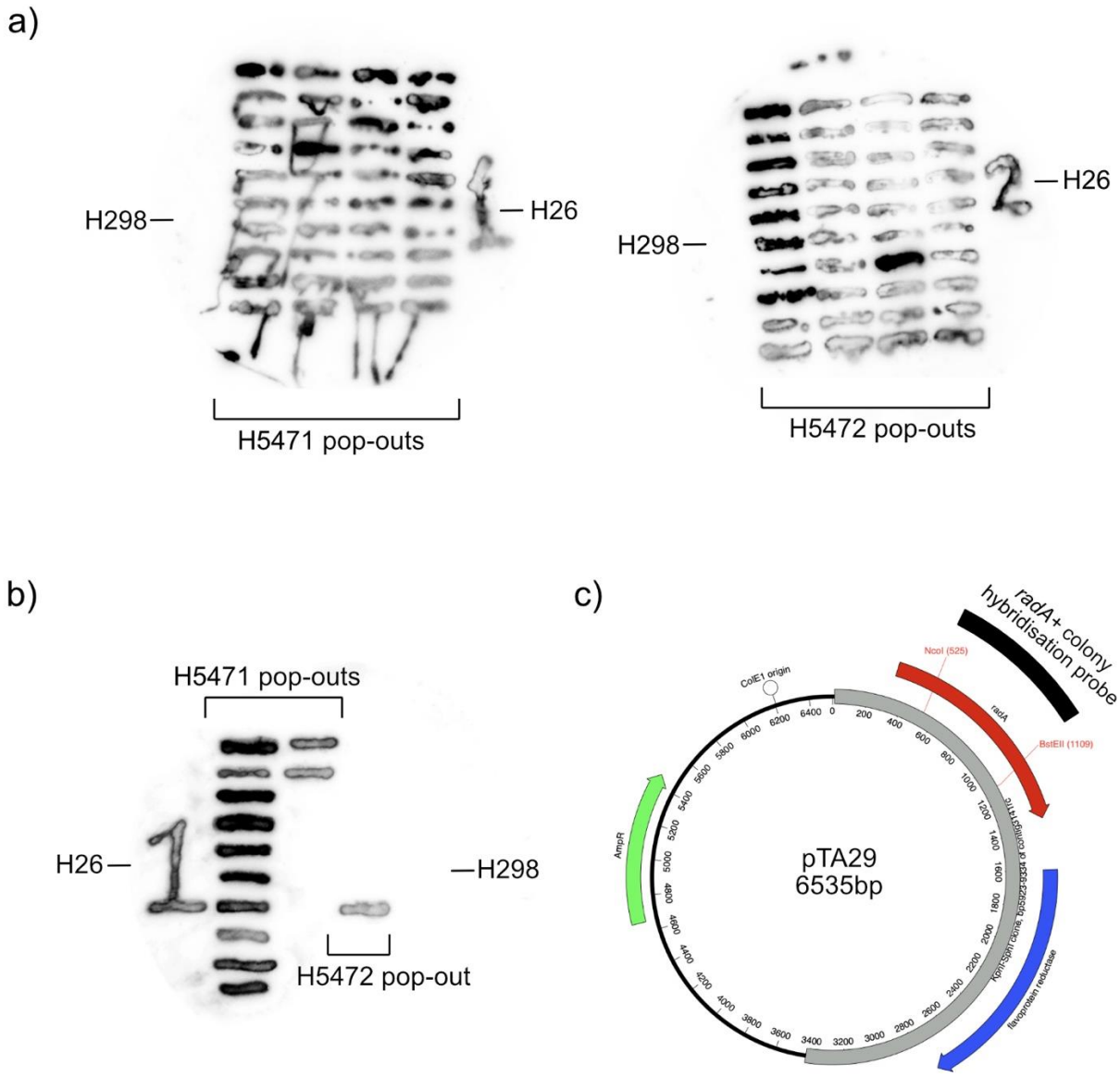


Figure 3.3: Colony hybridisation to check Δ*radA* colonies. a) Colony hybridisation of H5471 and H5471 pop-out colonies. As expected, the *radA*+ probe did not bind to H298 (Δ*radA*) and therefore, these patches cannot be seen. The positions of patched H298 on the membranes are labelled. All patched colonies retained *radA*. **b**) Colony hybridisation of H5471 and H5472 pop-out colonies after modification to the pop-out process. Fewer colonies were present after screening for 5-FOA resistance than previously (one colony was present in the case of H5472). Again, all colonies retained *radA*. **c**) The 584bp colony hybridisation probe used in a) and b) was created through digestion of pTA29 with *NcoI* and *BstEII*.

To instead replace wild-type *radA* with FLAG-*radA* directly, H1804 (Δ*pyrE2,* Δ*trpA,* Δ*oriC2,* Δ*oriC1,* Δ*oriC3,* Δ*ori-pHV4-2*) was transformed with pTA2424 (FLAG-*radA, pyrE2*+) and plated on Hv-Ca +Trp agar.

The region of homology between H1804 and pTA2424 upstream of the *radA* gene is 301bp, whereas the region of homology consisting of the *radA* gene and the downstream region is 1759bp (Figure 3.4). Due to this, integration of pTA2424 onto the main chromosome of H1804 is more likely to occur through recombination within the *radA* gene and downstream region than with the upstream region. Subsequent recombination during the pop-out step is again more likely to occur within the *radA* gene and the downstream region. However, the FLAG-tag is found at the 5' end of *radA*. In order for the FLAG-tag to remain in the genome after a downstream pop-out, pTA2424 must initially recombine with the region upstream of the *radA* gene (as shown in Figure 3.6a). Therefore, selection for recombination between H1804 and pTA2424 upstream of the *radA* gene will increase the likelihood of the FLAG-tag remaining in the genome after the pop-out step (Figure 3.5 demonstrates pop-in and pop-out stages, and Figure 3.6 and Figure 3.7 demonstrate selection for upstream pop-ins and downstream pop-outs, respectively).

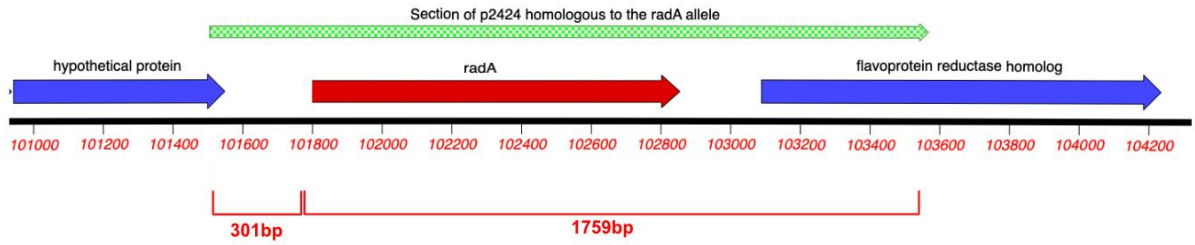


Figure 3.4: Homology between p2424 and the *radA* locus. The region upstream of *radA* with homology to p2424 is 301bp and the region consisting of *radA* and the downstream region with homology to p2424 is 1759bp. Therefore, the plasmid is more likely to recombine with the *radA* gene plus downstream region than the upstream region.

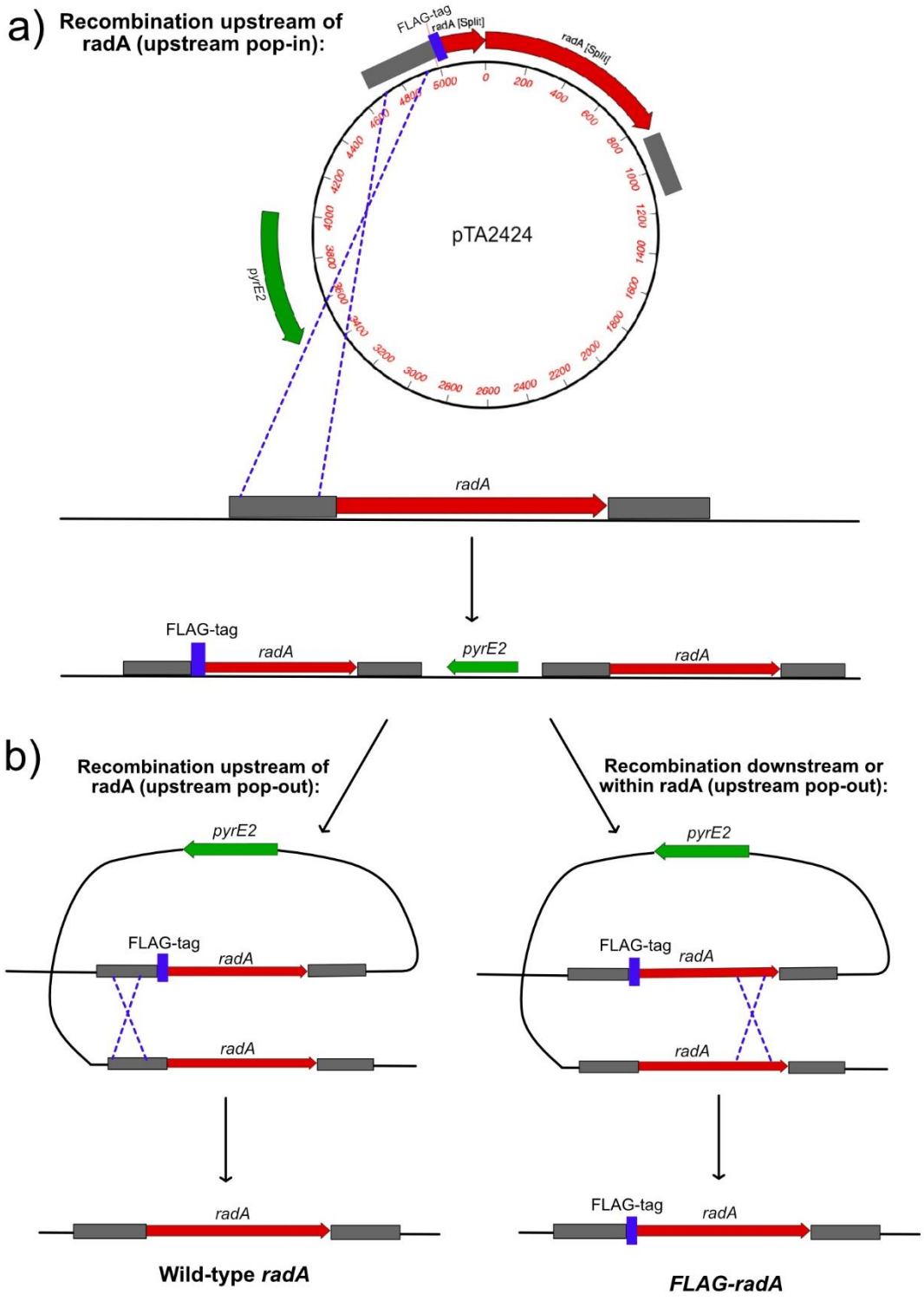


Figure 3.5: Pop-in/pop-out stages for strain H1804 transformed with pTA2424. a) Recombination between the region of homology upstream of *radA* of pTA2424 and H1804 leading to popping-in of the plasmid. Recombination between the *radA* gene or the downstream region of homology is also possible but upstream pop-ins were selected (for reasons described previously). **b)** Subsequent recombination in the popping-out step can result in either the wild-type *radA* allele (if recombination occurs between the regions of homology upstream of the FLAG-tag), or the FLAG-*radA* allele (if recombination occurs downstream of the FLAG-tag).

Transformants were screened via colony PCR using oligonucleotides o1413 and o2190, and subsequent restriction digestion with *NdeI* to select for upstream pop-ins (Figure 3.6). o1413 binds upstream of the *radA* gene within the H1804 genome and does not bind within pTA2424. o2190 binds within the *radA* gene. An *NdeI* cut site is found within FLAG-*radA* and is not present in the wild-type allele.

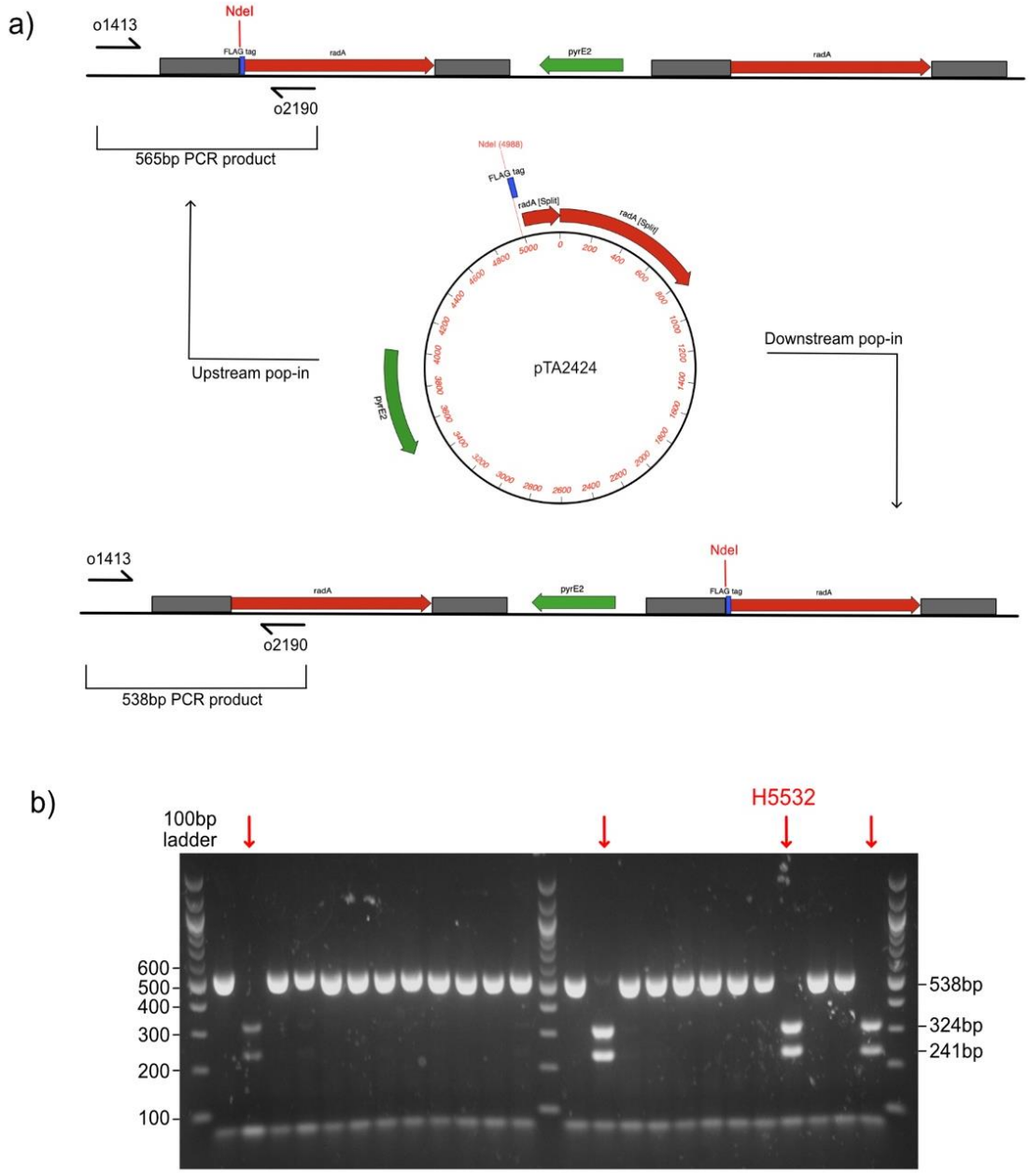


Figure 3.6: Selection for an upstream p2424 pop-in in H1804. a) Upstream and downstream pop-ins of p2424 at the *radA* locus in H1804. Colony PCR with oligonucleotides o1413 and o2190 generates a 565bp product containing an *NdeI* restriction site when p2424 has integrated into the genome upstream of *radA*, and a 538bp product without an *NdeI* site when p2424 has integrated downstream of *radA*. **b**) 24 pop-in colonies were screened via colony PCR and the products digested with *NdeI*. Digestion with *NdeI* results in the formation of two fragments of sizes 324bp and 241bp from the upstream pop-in colony PCR product. Red arrows indicate upstream pop-ins.

Strain H5532 (as seen in Figure 3.6b) was selected as a successful upstream pop-in and then underwent the pop-out step consisting of 3 consecutive overnight cultures with dilution in Hv-YPC broth. The final overnight culture was plated on Hv-Ca + 5-FOA + Trp agar to select for loss of *pyrE2*. Pop-out colonies were patched on Hv-YPC agar in preparation for downstream pop-out screening.

As explained previously, when pTA2424 has recombined upstream of the *radA* gene, recombination in the region downstream of the FLAG-tag is necessary for the tagged allele to remain in the genome. Screening for downstream pop-outs was carried out through colony PCR using oligonucleotides o1413 and o2190, and digestion of the PCR product with restriction enzyme NdeI (Figure 3.7) , exactly as carried out for selection of an upstream pop-in (Figure 3.6). 12 colonies were re-streaked and screened via Southern blotting (Figure 3.8). 3 colonies showing the correct genotype (H5547, H5548, and H5549) and were taken forward for further analysis.

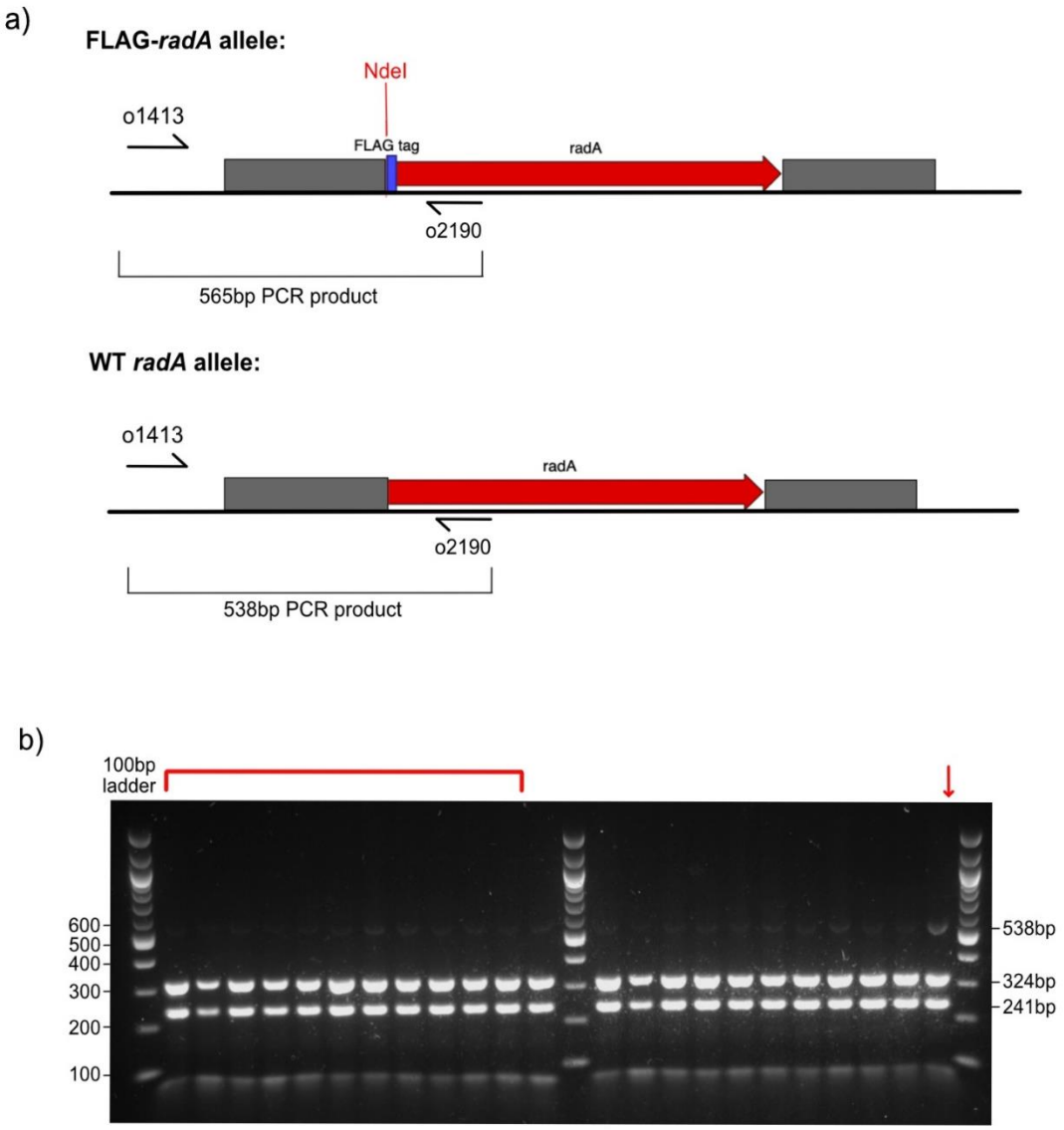


Figure 3.7: Colony PCR to screen H5532 pop-outs. **a**) Two genotypes are possible after the pop-out stage: FLAG-*radA* and wild-type *radA*. PCR amplification of the FLAG-*radA* allele using oligonucleotides o1413 and o2190 results in a 565bp fragment containing an *Nde1* restriction site, and PCR amplification of the wild-type *radA* allele results in a 538bp fragment without an Nde1 restriction site. **b**) 23 H5332 pop-out colonies were screened through colony PCR using o1413 and 2190 and digested with restriction enzyme Nde1. Digestion with *NdeI* results in the formation of two fragments of sizes 324bp and 241bp. Two bands at the expected sizes are present for every colony tested meaning that all 23 contain the FLAG-*radA* allele. However, faint bands for all colonies can be seen at 538bp which could be indicative of either incomplete digestion or a merodiploid state. Colonies 1 to 11 (indicated via the red bracket) and colony 23 (indicated via the red arrow) were re-streaked and screened again via Southern blotting. Colony 23 was selected as the band present at 538bp can be seen more clearly in this lane than in any other and Southern blotting would reveal whether this strain was merodiploid. In this case it would act as a control to compare to other strains (those not expected to be merodiploid).

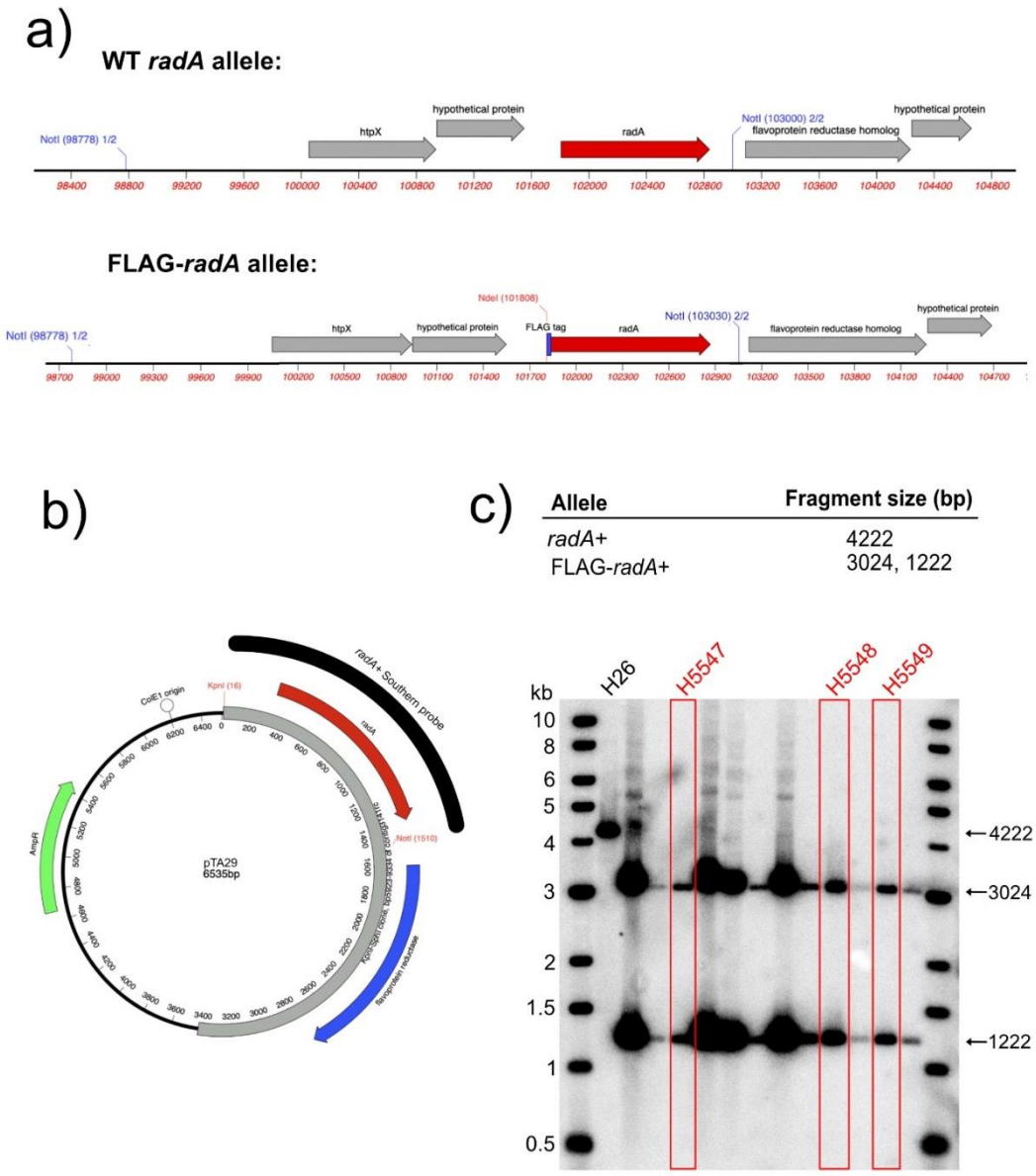


Figure 3.8: Southern blotting to screen FLAG-*radA* candidate strains. **a**) 2 alleles: wild-type *radA*+ and FLAG-*radA*+. The *NotI* and *NdeI* restriction sites used to digest genomic DNA are shown. **b**) A 1494bp *radA*+ Southern hybridisation probe was created through digesting pTA29 with *KpnI* and *NotI*. **c**) Size of predicted fragments and a Southern blot showing the wild-type radA+ allele (strain H26) and three FLAG-*radA*+ strains (H5547, H5548, and H5549). For these strains, only bands seen in each lane are those of the expected sizes for the FLAG-*radA* allele and therefore, they are not merodiploid.

FLAG-RadA expression and detection with the anti-DDDDK antibody

Whole cell lysate was prepared from the 3 strains showing the correct genotype in the Southern blot in Figure 3.2 (H5501, H5502, and H5505), and the 3 strains showing the correct genotype in the Southern blot in Figure 3.8 (H5547, H5548, and H5549). SDS-PAGE and Western blotting against the FLAG-tag was carried out to validate expression of FLAG-RadA and to ensure that a singular FLAG-tag was sufficient for anti-DDDDK antibody binding and protein detection. The Western blot confirmed that a single FLAG-tag was deemed sufficient for detection of the protein and suitable for use in ChIP-Seq. Two bands were observed in each lane for every *FLAG-radA* strain, in both the wild-type and origin-deleted backgrounds, and in both Western blots, no band was present for H26, the wild-type control (Figure 3.9). RadA has a molecular weight of 37.2kDa[222]. Protein bands from *H. volcanii* appear at a higher molecular weight due to the abundance of acidic residues in halophilic proteins[41,42]. Both bands seen were around the correct molecular weight at which *H. volcanii* RadA would be expected, as seen in [222]. The presence of 2 bands per lane for each of the FLAG-RadA strain cell lysates could be indicative of post-translational modification or truncation at the C-terminus of the protein.

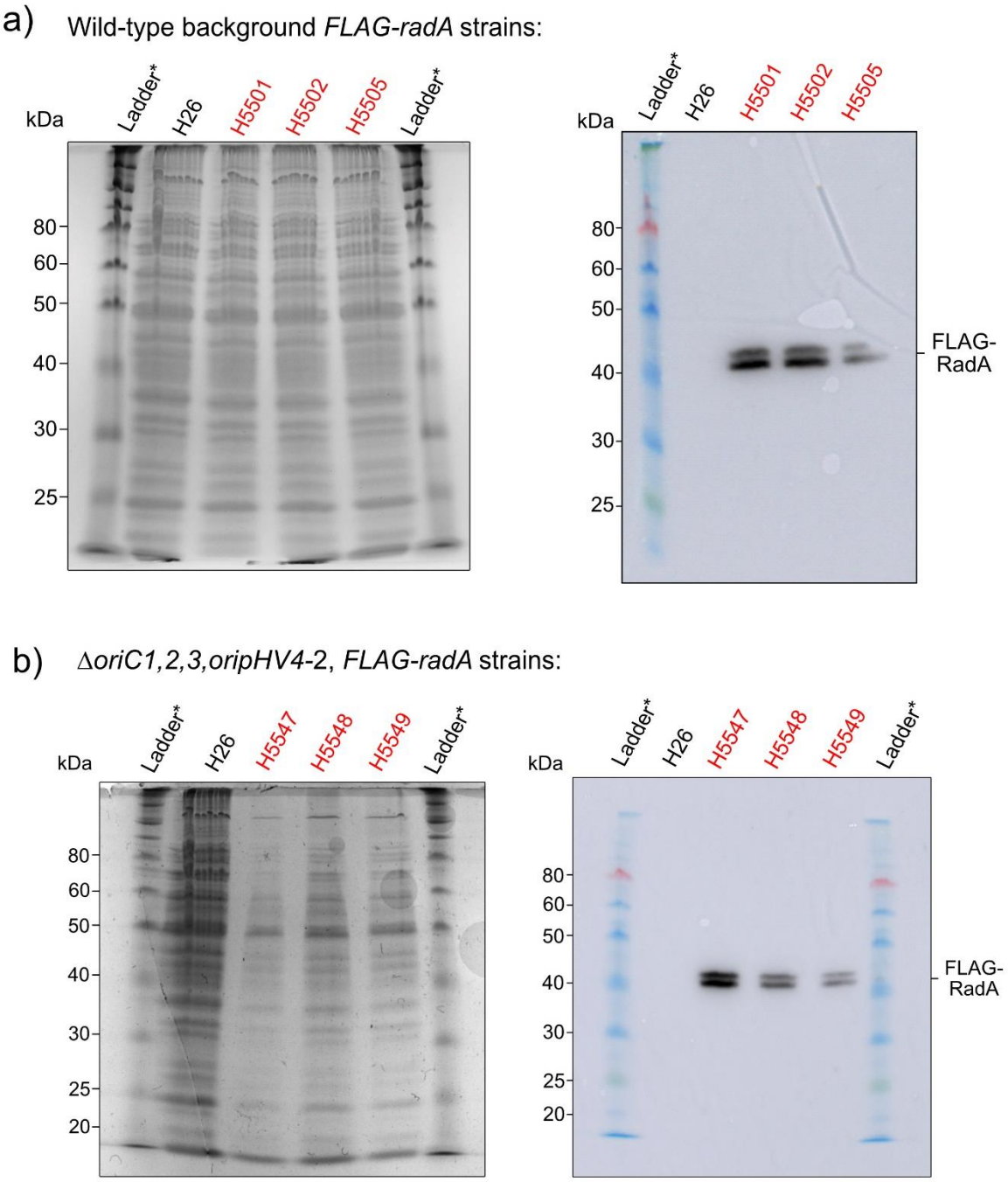


Figure 3.9: FLAG-RadA expression in both wild-type and origin-deleted background strains. **a**) Left: An SDS-PAGE gel stained in SimplyBlue™ SafeStain showing protein expression in wild-type and wild-type background FLAG-*radA* strains. Right: A Western blot against FLAG-RadA using a primary monoclonal anti-DDDDK tag antibody produced in mouse (Sigma) and a secondary anti-mouse IgG peroxidase conjugate (Sigma). FLAG-RadA is seen at the expected size. **b**) Left: An SDS-PAGE gel stained in SimplyBlue™ SafeStain showing protein expression in wild-type and Δ*oriC1,2,3,oripHV4-2* FLAG-*radA* strains. Right: A Western blot against FLAG-RadA using the same antibodies as used in a). FLAG-RadA is again seen at the expected size. *ColorPlus Prestained Protein Ladder, Broad Range (10-230 kDa) (NEB) was used.

Growth rates

Standard growth rates in liquid media were determined for the FLAG-RadA strains in comparison to strains expressing wild-type RadA and a *radA* deletion strain. Strains were grown in Hv-YPC broth, and standard growth assays were carried out according to *Chapter 2: Materials and Methods*.

The FLAG-RadA *ori*+ strains (H5501, H5502, and H5505) all grew more quickly than the Δ*radA* strain (H298) with generation times of 2.5 hours, 2.75 hours, and 2.25 hours respectively, compared to 4.75 hours (Figure 3.10). This indicates that FLAG-RadA is functional within these strains. H5501, H5502 and H5505 all also show a decreased generation time in comparison to wild-type (H26) with a generation time of 3 hours. Possible explanations for this are explained in the discussion of this chapter.

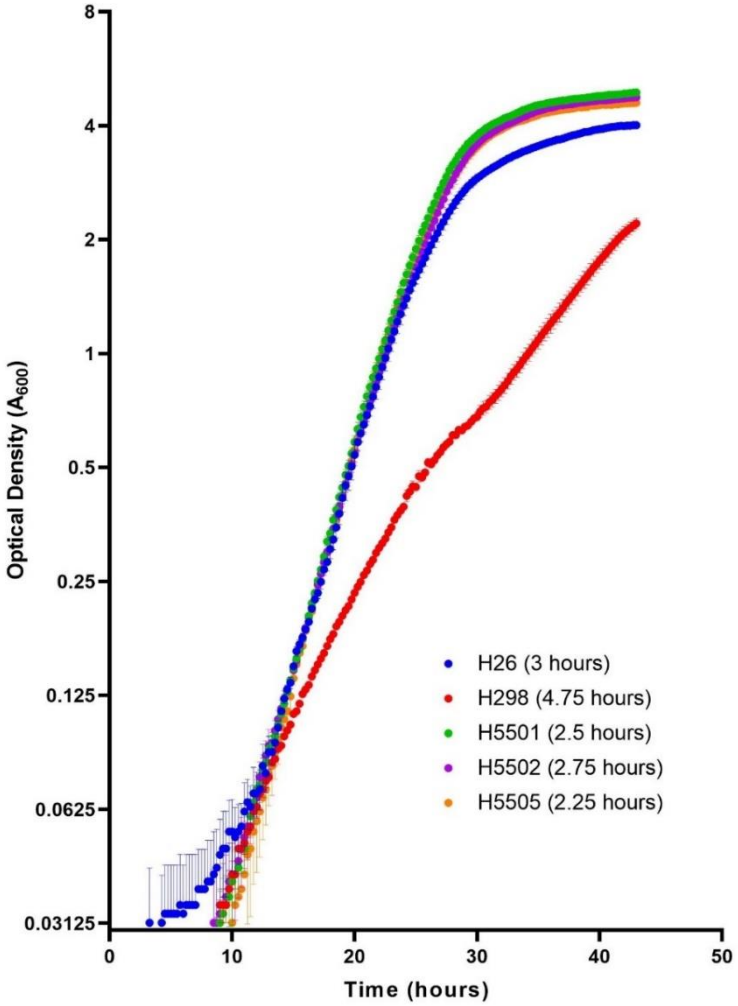


Figure 3.10: Growth of FLAG-RadA *ori*+ strains. H26 (wild-type), H298 (Δ*radA*), and H5501, H5502, H5505 (all *FLAG-radA*). Generation times during logarithmic phase are shown in the key. Each data point is the mean of four technical repeats and error bars display standard error of difference.

The FLAG-RadA ΔoriC1,2,3,oripHV4-2 strains (H5547, H5549, and H5549) all grew more quickly than the Δ*radA* strain (H298) with generation times of 2 hours, 2 hours, and 2.125 hours respectively, compared to 5.375 hours (Figure 3.11). This indicates that FLAG-RadA is functional within these strains. H5547, H5548 and H5549 all also show a decreased generation time in comparison to H1804 (Δ*oriC1,2,3,oripHV4-2*, wild-type *radA*) with a generation time of 2.25 hours.

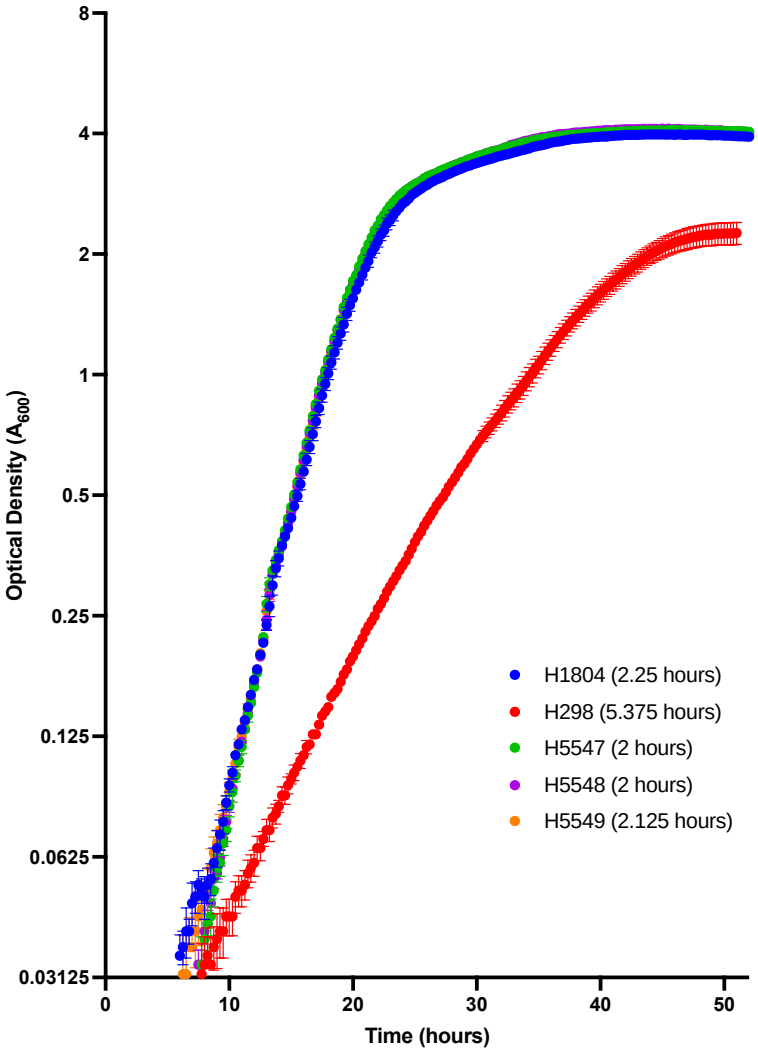


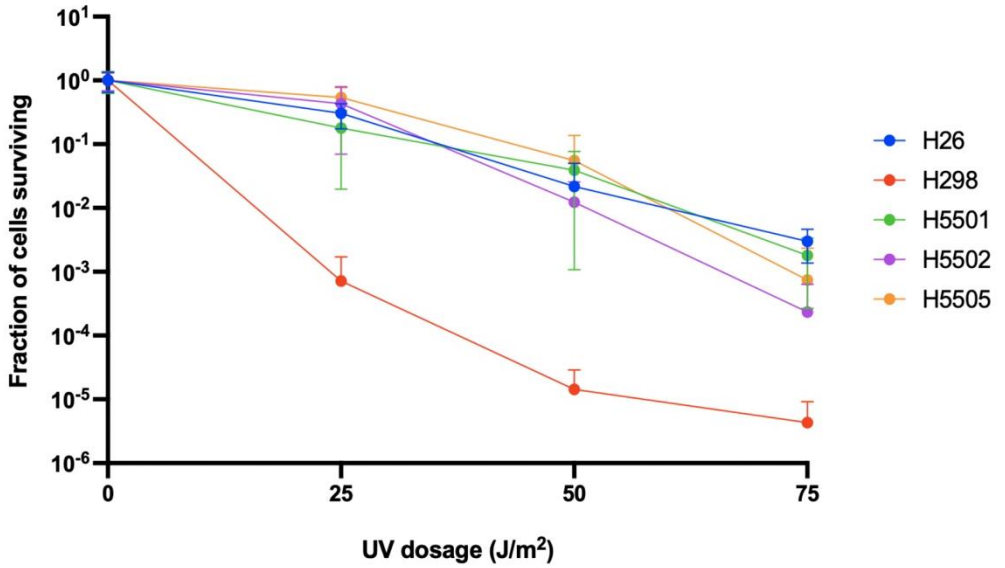
Figure 3.11: Exponential growth of FLAG-RadA origin-deleted background strains. H1804 (Δ*oriC1,2,3,oripHV4*-2, wild-type *radA*), H298 (Δ*radA*), and H5547, H5548, H5549 (*FLAG-radA, ΔoriC1,2,3,oripHV4*-2). Generation times during logarithmic phase are shown in the key. Each data point is the mean of four technical repeats and error bars display standard error of difference.

UV-irradiation resistance

UV radiation can lead to many types of DNA lesions including pyrimidine dimers, 6-4 photoproducts, and single/double-stranded DNA breaks. “Bulky” lesions such as pyrimidine dimers and 6-4 photoproducts can be repaired by photolyases [245] or through nucleotide excision repair [246]. UV-irradiation can also result in reactive oxygen species and free radicals formed can induce ssDNA breaks. ssDNA breaks can lead to dsDNA breaks (DSBs) if two occur in close proximity or if encountered by a replication fork. In *H. volcanii,* repair of DSBs can occur through homologous recombination or microhomology-mediated end joining. Repair through homologous recombination requires RadA, and Δ*radA* strains show severe growth defects after UV-irradiation compared to wild-type [168]. Therefore, subjecting strains expressing FLAG-RadA to UV-irradiation and comparing their survival to that of wild-type and Δ*radA* strains will demonstrate the functionality of FLAG-RadA.

FLAG-radA strains were subjected to a range of doses of UV radiation (See Chapter 2: Materials and Methods) and their ability to repair DNA damage, and therefore the functionality of FLAG-RadA in these strains, was compared to that in H26 (wild-type) for *ori*+ background strains, H1804 (Δ*oriC1,2,3,oripHV4*-2, wild-type *radA*) for Δ*oriC1,2,3,oripHV4*-2 strains, and H298 (Δ*radA*). All strains expressing FLAG-RadA showed a similar resistance to UV-irradiation to H26 and H1804 for *ori*+ background strains and Δ*oriC1,2,3,oripHV4*-2 strains respectively, and were all more resistant than H298 (Figure 3.12). This indicates that in these strains, FLAG-RadA is functional. The resistance of strains H5501 and H5548 most closely resembled that of H26 and H1804 respectively and were taken forward to be tested for resistance to gamma irradiation.

a)



b)

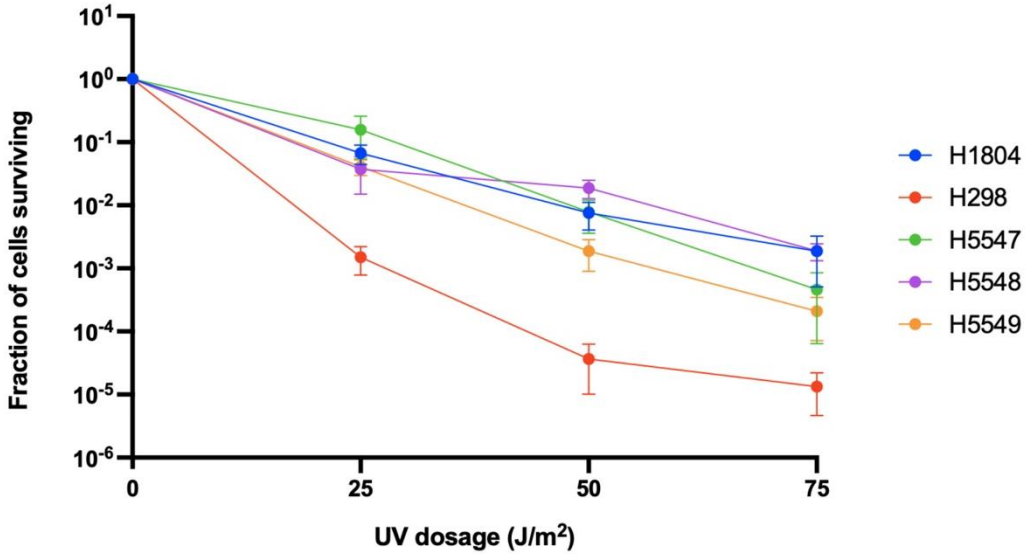


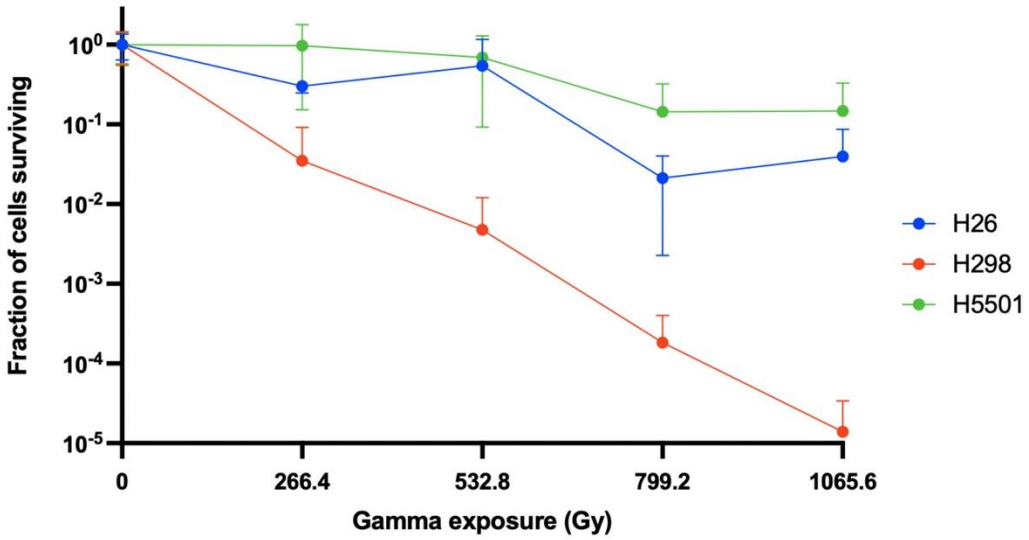
Figure 3.12: Fraction of cells surviving after UV exposure. a) Strains H5501, H5502 and H5505 (FLAG-radA) all show greater resistance to UV-irradiation than H298 (ΔradA) and show similar resistance to H26 (wild-type). The resistance of strain H5501 most closely resembles that of H26 in comparison to the other FLAG-*radA* strains. **b**) Strains H5547, H5548 and H5549 (all FLAG-radA, Δ*oriC1,2,3,oripHV4*-2) all show greater resistance to UV than H298 and show similar resistance to H1804 (wild-type *radA*, Δ*oriC1,2,3,oripHV4*-2). The resistance of strain H5548 most closely resembles that of H1804 in comparison to the other FLAG-*radA* Δ*oriC1,2,3,oripHV4*-2 strains. In both graphs, each data point demonstrates the mean of 3 independent experiments, each consisting of 2 technical repeats. Error bars represent standard error of the mean. In some cases, lower error bars are not shown as a data point of zero contributed to the mean and therefore, this cannot be plotted on a logarithmic axis.

Gamma-irradiation resistance

Gamma radiation induces oxidative damage which leads to DNA lesions such as ssDNA and dsDNA breaks, DNA-protein crosslinks, oxidised bases, and abasic sites[247,248]. Repair of gamma-irradiation induced damage is often through the base excision repair pathway[249]. However, repair of dsDNA breaks requires either homologous recombination or microhomology mediated end joining. Repair through homologous recombination requires RadA and therefore, gamma-irradiation of strains expressing FLAG-RadA and comparison of their survival with Δ*radA* strains and wild-type, as carried out previously to determine UV-irradiation resistance, will further demonstrate the functionality of FLAG-RadA in the strains created.

Strains H5501 and H5548 were subjected to a range of doses of gamma radiation (*See Chapter 2: Materials and Methods*). The same control strains were used as in the UV-irradiation resistance assay described previously, namely H26 (wild-type), H1804 (Δ*oriC1,2,3,oripHV4*-2, wild-type *radA*) and H298 (Δ*radA*). H5501 showed a resistance to gamma-irradiation resembling that of H26 and was more resistant than H298 (Figure 3.13a). H5548 showed a resistance to gamma-irradiation resembling that of H1804 and like H5501, was also more resistant than H298 (Figure 3.13b). This result mirrored that seen in the UV-irradiation resistance assay and again demonstrated that FLAG-RadA is functional in these strains.

a)



b)

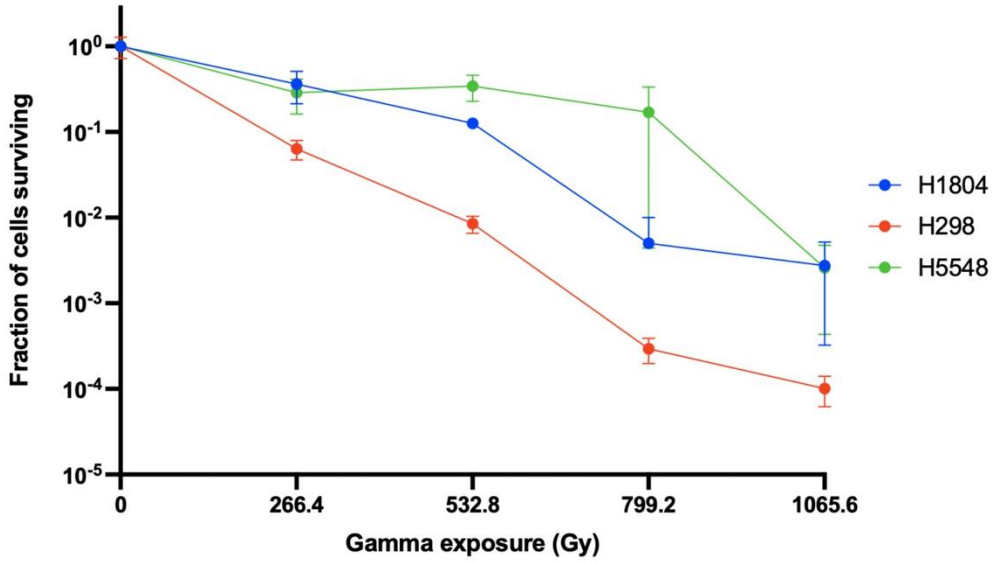


Figure 3.13: Fraction of cells surviving after exposure to a gamma source. **a**) H5501 (*FLAG-radA*) shows similar gamma-irradiation resistance as H26 (wild-type) and is more resistant than H298 (Δ*radA*). **b**) H5548 (Δ*oriC1,2,3,oripHV4*-2, *FLAG-radA*) shows similar gamma-irradiation resistance to H1804 (Δ*oriC1,2,3,oripHV4*-2, wild-type *radA*) and is more resistant than H298. In both graphs (excluding the H1804 data point at 799.2 Gy), each data point demonstrates the mean of 3 independent experiments, each consisting of 2 technical repeats. The H1804 799.2 Gy datapoint consists of one experiment with two technical repeats as in two of the three experiments, no colonies were present at the dilutions plated. Error bars represent standard error of the mean.

3.3.2- ChIP-Seq

Sample preparation and interpretation of ChIP-Seq data are detailed in this section. Library preparation, sequencing, and bioinformatic analysis were undertaken by the Deep Seq facility, University of Nottingham.

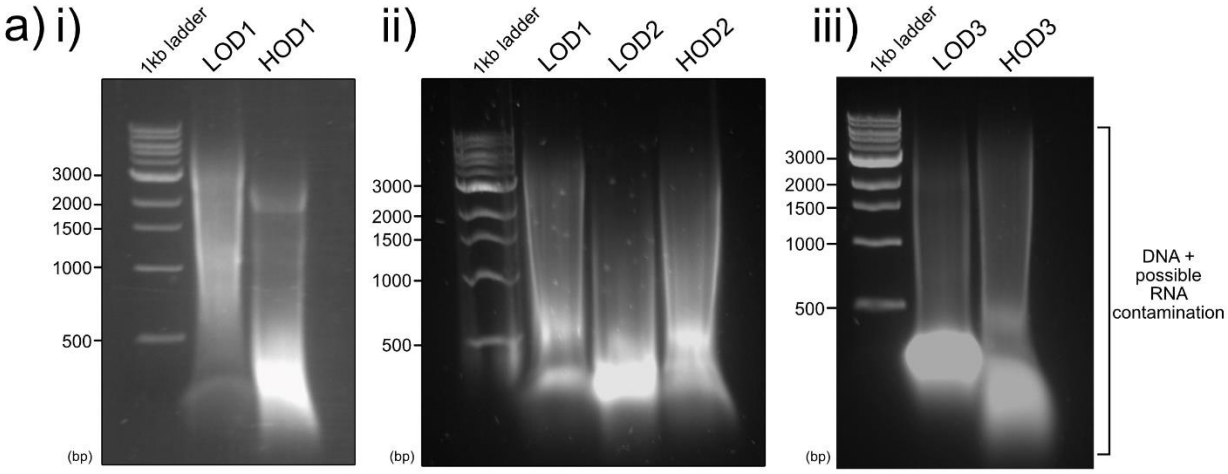
Due to time constraints and dependence upon external sources to process samples, it was only possible to carry out ChIP-Seq on strain H5501 (FLAG-RadA, *ori*+ background). Localisation of FLAG-RadA in an origin-deleted background was not undertaken but FLAG-*radA* ∆*oriC1,2,3,pHV4* strains were created and validated meaning that it is possible for ChIP-Seq can be carried out in the future. ChIP-Seq against FLAG-RadA in an *ori*+ background should be able to identify recombinational hotspots and areas of the genome prone to double-strand breaks from which recombination-dependent replication may occur.

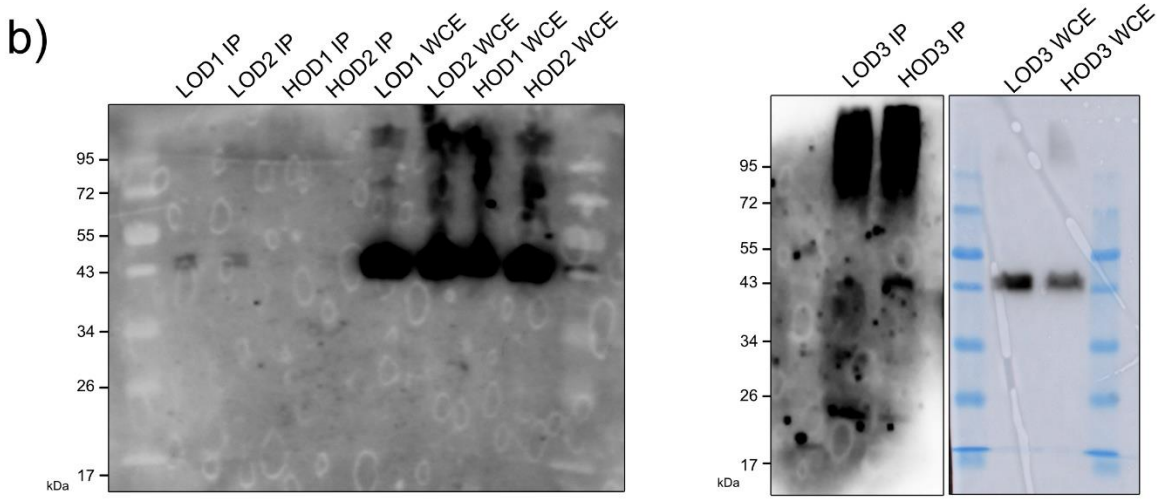
3.3.2.1 - Sample preparation (FLAG-RadA in an ori+ background)

As described in the background for this chapter, the use of origins in *H. volcanii* appears to be reduced between early and mid/late log phase. In early log phase, MFA plots show well defined peaks at origins of replication with a copy number between 2.2 and 2.0, but in mid/late log phase, peaks become less well defined, and the copy number reduces to approximately 1.1[49]. It was hypothesised that in mid/late log phase, recombination-dependent replication (instead of origin-dependent replication) may have been responsible for the flattening of peaks. It was of interest to determine RadA localisation in both early and mid/late log phase, therefore cells at A650= 0.05 and 0.5 were chosen, corresponding to early and mid/late log phase, respectively.

3 x 5ml Hv-YPC broth cultures of H5501 (∆*pyrE2,* ∆*trpA, FLAG-radA*) were grown over successive overnights with dilution and used to inoculate 3 x 100ml cultures which were then grown to A_{650}= 0.5 (referred to as High OD samples HOD1, HOD2 and HOD3), and 3 x 1 litre cultures grown to A650= 0.05 (referred to as Low OD samples LOD1, LOD2 and LOD3) as described in Chapter 2: Materials and Methods, ChIP-Seq. The same 5ml overnight was used to inoculate one 100ml high OD culture and one 1 litre low OD culture meaning that in total, there were three biological replicates, each with a low OD and a high OD sample. Sample preparation was carried out according to the protocol described in Chapter 2: Materials and Methods; Western blots using the anti-DDDK antibody were undertaken using aliquots of samples removed prior to crosslink reversal, to ensure FLAG-RadA was present.

Samples LOD3 and HOD3 were prepared after the other samples and therefore, in Figure 3.14, the agarose gel and Western blot images are separate.





c)

Sample	DNA concentration (ng/µl)
LOD1 IP-X	37.2
LOD2 IP-X	29.4
LOD3 IP-X	16.9
HOD1 IP-X	19.9
HOD2 IP-X	29.9
HOD3 IP-X	26.9
LOD1 WCE-X	99.8
LOD2 WCE-X	138.9
LOD3 WCE-X	184.6
HOD1 WCE-X	145.2
HOD2 WCE-X	105.0
HOD3 WCE-X	145.9

Figure 3.14: Preparation of ChIP-Seq samples to determine RadA localisation in a wild-type background. a) Size of DNA fragments after sonication. Sample LOD1 was initially sonicated for 4 x 20 second on/off cycles leading to the size of fragments seen in **i)** and was then sonicated for a further 4 x 20 second on/off cycles to generate the size of fragments seen in **ii)**. Samples LOD2 and LOD 3 were sonicated for 8 x 20 second on/off cycles and samples HOD1, HOD2, and HOD3 for 14 x 20 second on/off cycles. For all samples, most fragments fell between 200 and 800bp as required. However, average fragment length determined by the Deep Seq facility when samples were run on both a D1000 and D5000 ScreenTape was between 1000 and 2000bp, possibly due to RNA contamination (as described in section 3.3.2.2- Library preparation and sequencing. **b)** Western blots against FLAG-RadA. IP denotes samples which have undergone the immunoprecipitation reaction and WCE denotes whole cell extract samples which were not subjected to immunoprecipitation (an aliquot of clarified cell lysate was removed prior to incubation with antibody-bound beads). 2 faint bands can be seen per IP sample lane at the size expected and seen in Chapter 3, Figure 3.9. WCE samples show a greater concentration of FLAG-RadA which is to be expected due to loss of protein in the immunoprecipitation reaction. The two halves of the LOD3 and HOD3 Western blot (right hand image) were imaged separately with the aim of increasing the visibility of the IP bands. **c)** Concentration of DNA after purification. X denotes samples which have undergone crosslink reversal. IP samples all have a lower DNA concentration than WCE samples as expected.

Figure 3.14a shows fragments of DNA between 200 and 800bp after sonication, with the highest concentration slightly below 500bp. Bands can be seen at the expected size for RadA in all lanes in the Western blots in Figure 3.14b. Bands were very faint for samples HOD1 IP, HOD2 IP, and LOD3 IP but upon close inspection, were present. It was thought that this was due to loss of protein in the immunoprecipitation reaction. After crosslink reversal, the concentration of DNA in each of the samples was measured and all samples which underwent the immunoprecipitation reaction had a lower concentration of DNA than the whole cell extract samples, again as expected. Concentrations are shown in Figure 3.14c.

3.3.2.2 - Library preparation and sequencing

Following sample preparation, samples were sequenced, and data were analysed by the Deep Seq facility, University of Nottingham. Methodologies followed and Tapestation profiles of final library preparations and can be found in the appendices of this study. For each condition (low OD and high OD), two WCE samples and three IP samples were sequenced.

In the process of creating libraries to sequence, the Deep Seq facility ran the samples on both a D1000 and D5000 ScreenTape to check fragment-length distribution. Despite having run the samples on an agarose gel to ensure fragments were of the correct size (as seen in Figure 3.14), the average fragment length was between 1000 and 2000bp. This is longer than the ideal length to create ChIP-Seq libraries (100-500bp). Large fragments make peak detection more difficult and reduce the chance of detecting specific binding motifs. It was not expected that specific motifs would be detected in the same way may be seen with transcription

factors due to the nature and function of the RadA protein. However, it remained important that peaks, even broad ones, remained detectable. To narrow the fragment size, the Deep Seq facility performed a double-sided SPRI size selection after the adapter ligation and before the final library amplification step to target 200-700bp fragments. This narrowed the fragment size distributions, but some libraries still had an average fragment size of ~900bp. These libraries were sequenced and data TapeStation profiles can be found in the Appendices.

The discrepancy seen between the lengths of fragments on the agarose gels and from the D1000 and D5000 ScreenTape data could be explained by the presence of RNA. RNase treatment of samples occurs during the final purification step in the ChIP-Seq sample preparation protocol (see Section 2.2.8- ChIP-Seq, Materials and Methods) but the samples run on the agarose gel were taken directly from sonicated samples, before any further stages were undertaken, and therefore still contained RNA. RNA may be responsible for the bright band of small fragments seen.

3.3.2.3 - Analysis and interpretation of ChIP-Seq data

The bioinformatic pipeline used by the Deep Seq facility can be found here [250]. In brief, sequencing adapters are removed, and paired end reads are aligned to the reference genome. Peaks are called (corresponding to regions of the genome seen more frequently in the IP samples compared to WCE samples), annotated, and quality control (QC) analysis is performed. The software used and further detail regarding the stages in data processing is as referenced.

The pipeline used creates outputs for both broad and narrow peaks. Due to the nature of RadA binding (as filaments spanning large stretches of DNA- see Section 1.4.1.1- Rec-A family proteins: RecA, Rad51, and RadA) and time constraints due to external factors, it was determined that data corresponding to broad peaks would be of most relevance and is the data discussed in the remainder of this chapter.

Samples previously referred to with notation, for example, “LOD1 IP-X” are from this point onwards in this chapter referred to with notation “LOD_IP_REP1”.

QC

This section discusses key elements of the MultiQC[251] report created by the Deep Seq facility.

Duplicate reads

In the library preparation step, 6 cycles of PCR were carried out on each of the samples. Bias in PCR (which may arise due to the high GC content of the *H. volcanii* genome) can result in duplicated reads being present in the library. It is important that duplicates are removed during bioinformatic analysis to reduce the PCR bias to the natural complexity of the sample. However, biological duplicates may also arise

in which the exact same sequence may be present multiple times within the same sample. In this case, it is not possible to differentiate between technical duplicates arising due to PCR bias and biological duplicates arising due to RadA binding at identical sequences. When duplicates are then removed, there may be discrepancy reads in the library compared to the sequences to which RadA binds and the extent to which it does so.

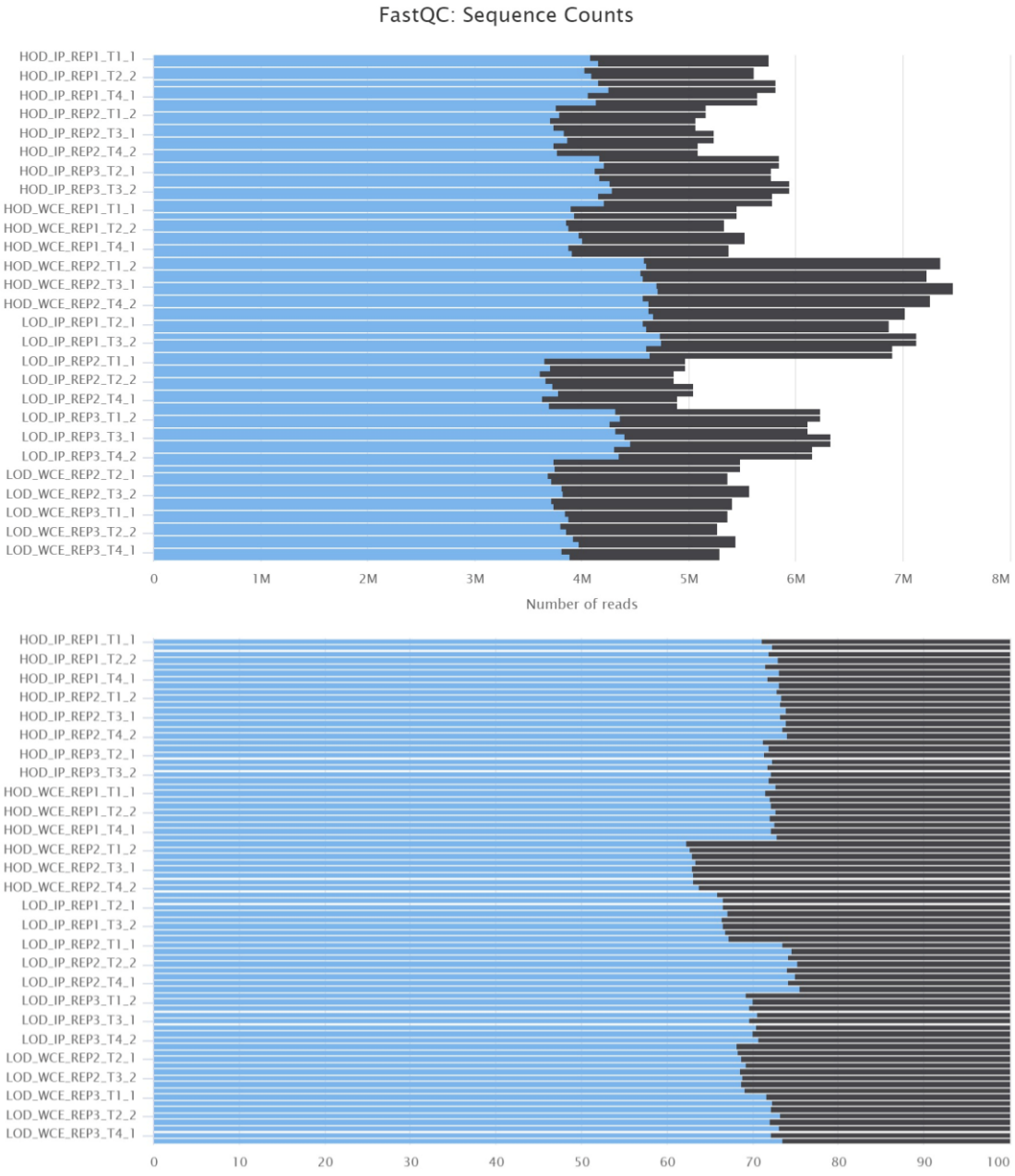


Figure 3.15: The number (upper image) and percentage (lower image) of unique and duplicate reads in ChIP-Seq samples. During analysis, multiple read files were created per sample corresponding to the multiple horizontal bars seen for each sample as labelled in the figure.

Figure created by the Deep Seq facility at the University of Nottingham using the MultiQC tool[251].

The proportion of unique reads to duplicate reads in all the samples was above 60%, with the average proportion falling around 70% (Figure 3.15). It is recommended that 80% of reads should be unique in ChIP-Seq experiments[252] which is not met in the preparation of libraries in this experiment. Also, it would be expected that WCE samples have a greater proportion of unique reads than IP samples as the IP should pull down specific areas of the genome (those that the protein of interest is bound to), whereas in whole cell extract samples, all DNA sequences previously bound to any protein in within cells should be present therefore leading to more unique reads. This is not seen in this experiment. In Figure 3.15, percentage of unique reads is similar across all samples with some WCE samples even showing fewer unique reads than IP samples. Therefore, the lower-than-expected number of unique reads compared to duplicate reads in this study may indicate that PCR bias has arisen due to the high GC content of the *H. volcanii* genome, or that biological duplicates were present which have now been removed, and that the reads present in the library may not be a true representation of all of the sequences to which RadA was bound in the samples.

Principal component analysis

Principal component analysis (PCA) was used to demonstrate the level of variability among the IP samples (Figure 3.16). Figure 3.16 was created using the MultiQC tool[251] used by the Deep Seq facility with the DESeq2 R package[253]. PCA allows large data sets to be summarized by a smaller set of principal components through creating uncorrelated variables which reduce the dimensionality of the data. The maximum variance is explained by the first principal component and the second highest variance is explained by the second principal component. The first two principal components (PCA1 and PCA2) displayed 80% and 11% variance respectively across the samples.

It is expected that biological replicates would cluster together in a PCA plot and that there would be a greater variance between high OD samples and low OD samples than the variance found within either condition. However, in the PCA plot generated for samples in this study, this is not the case. Biological replicates do not appear to cluster together other than LOD_IP_REP2 and LOD_IP_REP3 which clustered more closely than any other samples. Biological replicates show as much variation as seen between LOD and HOD samples and therefore, it is difficult to draw conclusions about the significance of the data comparing the LOD and HOD conditions.

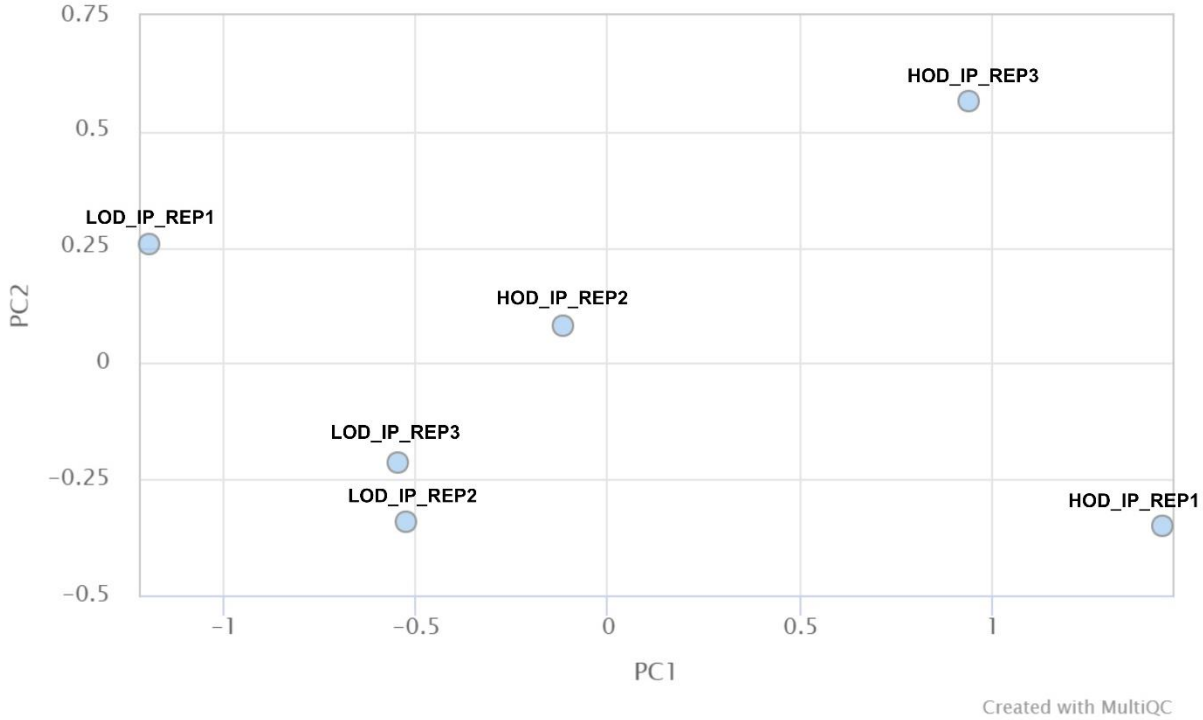


Figure 3.16: PCA of ChIP-Seq samples. PCA values (and plot co-ordinates) to three significant figures are as follows. LOD_IP_REP1: (-1.19, 0.259), LOD_IP_REP2: (-0.526. -0.341), LOP_IP_REP3: (-0.545, -0.213), HOD_IP_REP1: (1.44, -0.352), HOD_IP_REP2: (-0.116, 0.0812), HOD_IP_REP3: (0.937, 0.566). It is expected that biological replicates cluster together which is not the case here, other than for samples LOD_IP_REP2 and LOD_IP_REP3.

Figure created by the Deep Seq facility at the University of Nottingham using the MultiQC tool[251]. Annotations have been added for clarity.

Total peak count

Total peak count is calculated by MultiQC[251] from the total number of peaks called by MACS2 software[254].

The total number of peaks also varies greatly across the samples. Samples HOD_IP_REP2 and HOD_IP_REP3 show significantly more peaks than the other four samples (Figure 3.17) (390 and 314 peaks respectively compared to 17 peaks in HOD_IP_REP1, 36 peaks in LOD_IP_REP1, 3 peaks in LOD_IP_REP2, and 19 peaks in LOD_IP_REP3). If more peaks were seen across all three HOD samples compared to all three LOD samples, it might be possible to suggest that RadA binds more widely across the genome during later stages of growth (when it is thought that origins of replication might be more essential) and that during early stages of growth (when recombination-dependent replication is thought to play a greater role), RadA binds more specifically to a reduced number of sequences. However, only two of the

three HOD samples show a greater number of peaks meaning that conclusions such as this cannot be drawn and further optimization to the protocol is necessary.

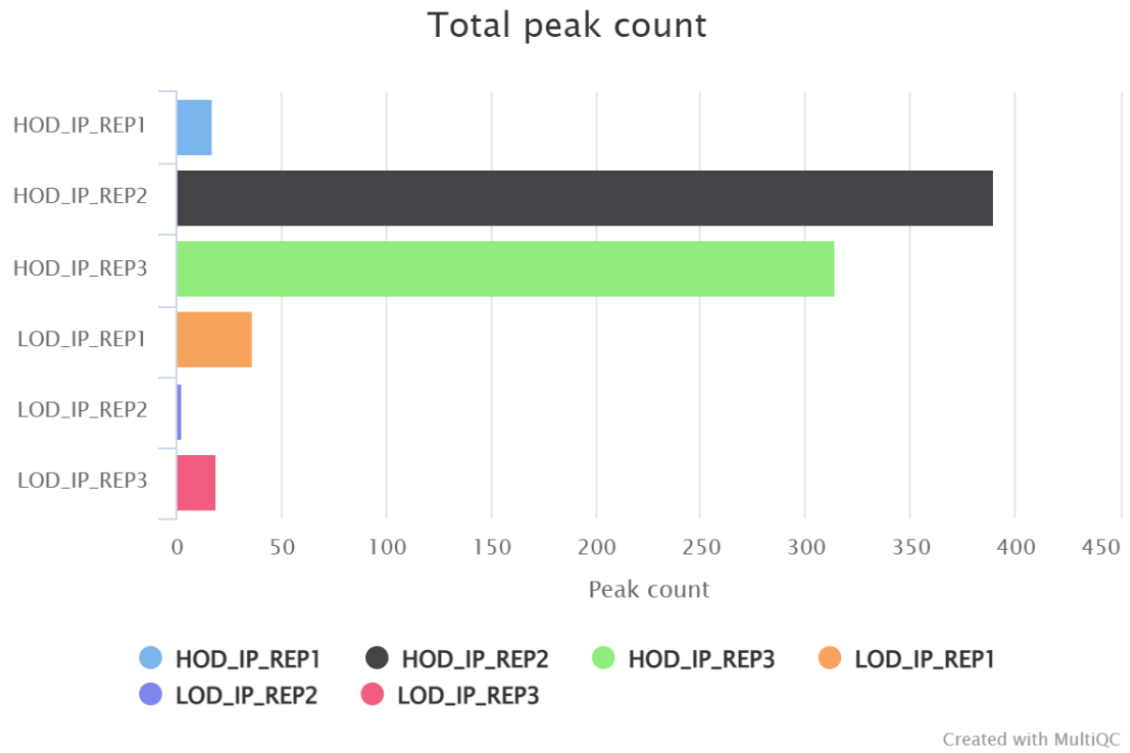


Figure 3.17: Total peak count of ChIP-Seq samples. The exact number of peaks per sample from top to bottom of the bar chart is as follows: HOD_IP_REP1= 17, HOD_IP_REP2= 390, HOD_IP_REP3= 314, LOD_IP_REP1= 36, LOD_IP_REP2= 3, LOD_IP_REP3= 19. Samples HOD_IP_REP2 and HOD_IP_REP3 show many more peaks than any other samples.

Figure created by the Deep Seq facility at the University of Nottingham using the MultiQC tool[251].

Fraction of reads falling within peaks (FRiP scores)

The fraction of all mapped reads falling within peaks (FRiP) was calculated for each sample. It is expected that a minority of reads will fall within peaks in a ChIP-Seq experiment, and the remainder are background[252]. FRiP therefore measures the success of the immunoprecipitation stage of the protocol. The majority of ChIP-Seq datasets from the ENCODE consortia have FRiP enrichment of 1% or greater. However, experiments with FRiP values greater than 1% can still be unsuccessful, such as for proteins which bind to many locations across the genome[252]. In these cases, many broad peaks can be present across a large percentage of the genome meaning that the likelihood of a mapped read falling within a peak is increased.

For FRiP scores calculated by MultiQC, a read must overlap a peak by at least 20% to be counted. Samples HOD_IP_REP2 and HOD_IP_REP3 had extremely high FRiP scores of 59% and 37%. Peaks from these samples can be seen to span a large proportion of the genome later in this chapter and therefore, it is expected that a large proportion of the reads would fall within them. Samples HOD_IP_REP1, LOD_IP_REP1, and LOD_IP_REP3 had scores above or equal to 1% (3%, 1% and 1%, respectively). Sample LOD_IP_REP2 had score <1% and therefore falls below the standard threshold. These scores correlate with the total peak count seen above as if more peaks are present, it is more likely that a larger proportion of the reads in the sample fall within a peak.

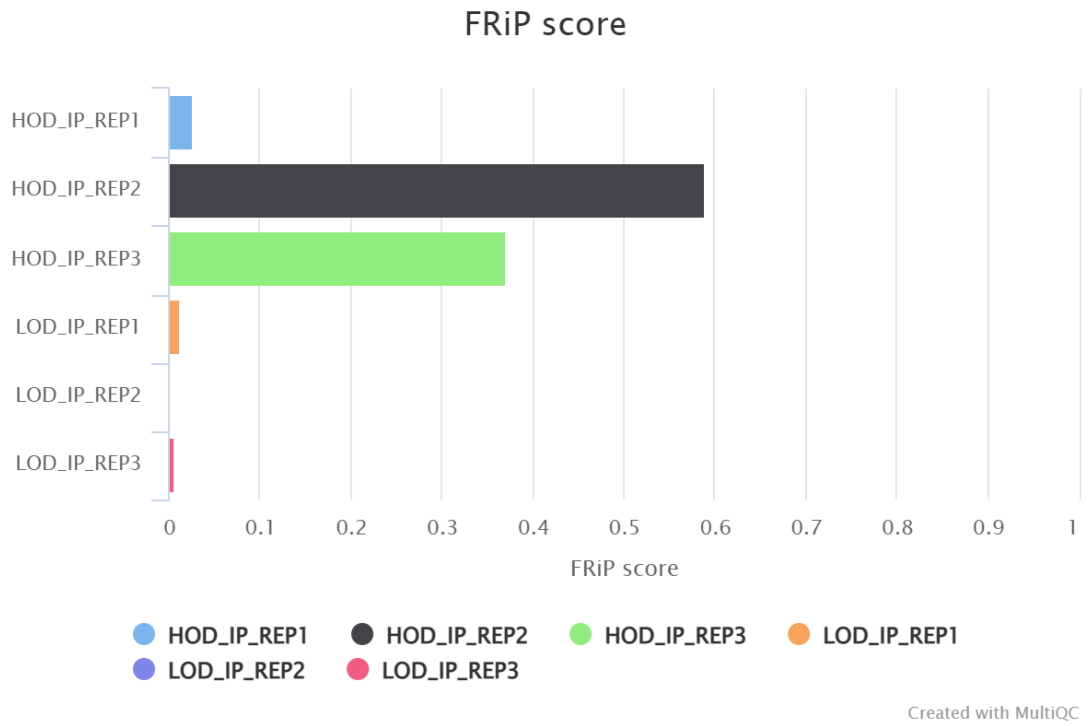


Figure 3.18: FRiP scores of ChIP-Seq samples. The exact FRiP score per sample from top to bottom of the bar chart is as follows: HOD_IP_REP1= 0.03, HOD_IP_REP2= 0.59, HOD_IP_REP3= 0.37, LOD_IP_REP1= 0.01, LOD_IP_REP2= 0.00, LOD_IP_REP3= 0.01. Samples HOD_IP_REP2 and HOD_IP_REP3 have a considerably higher FRiP score than the other four samples, correlating with the higher total peak count for these samples seen in Figure 3.17.

Figure created by the Deep Seq facility at the University of Nottingham using the MultiQC tool[251].

Overall, QC analysis demonstrates that further optimization of the protocol is necessary to allow reliable conclusions to be drawn. Variability among biological replicates can be seen in PCA, total read count, and FRiP scores. Additionally, one of the samples fell below the accepted FRiP score threshold. Proposed modifications to address this are described in the discussion of this chapter.

Read and peak mapping

Due to the elements of QC described above, all conclusions drawn in this section are tentative.

Reads mapping to particular areas of the genome and peaks called by the MACS software were visualized using the Integrative Genomics Browser (IGV). Figure 3.19 demonstrates reads from each sample mapping to each component of the *H. volcanii* genomic architecture. The greater the depth of the blue regions seen in the figure, the greater the depth of read coverage of this area of the genome. Overall, this figure demonstrates lack of specific pull-down. WCE samples closely resemble IP samples and therefore, IP samples do not immediately show enrichment for particular areas of the genome. In the majority of cases, the majority of the genome is spanned in both WCE and IP samples. Had all elements of the QC been deemed acceptable, this would indicate that RadA binds non-specifically all over the genome. However, large fragment sizes like those seen for these samples mean that large areas of the genome can be spanned by reads, even if the protein of interest binds at very specific regions, leading to results like those seen here.

It is worth noting that in Figure 3.19, the LOD samples seem to show a greater depth of coverage across pHV1 than seen in the LOD WCE sample which would indicate that there is read enrichment in this region. However, data later in this chapter reveals very few peaks (none of which are common among all three LOD samples) in this region and therefore the seemingly increased depth of coverage here does not demonstrate significant enrichment. It is possible that RadA binds widely across pHV1 (meaning that no peaks are seen) which may play a role in the formation of pHV1 concatemers demonstrated in [53] and [208] (explained further in the discussion of this chapter).

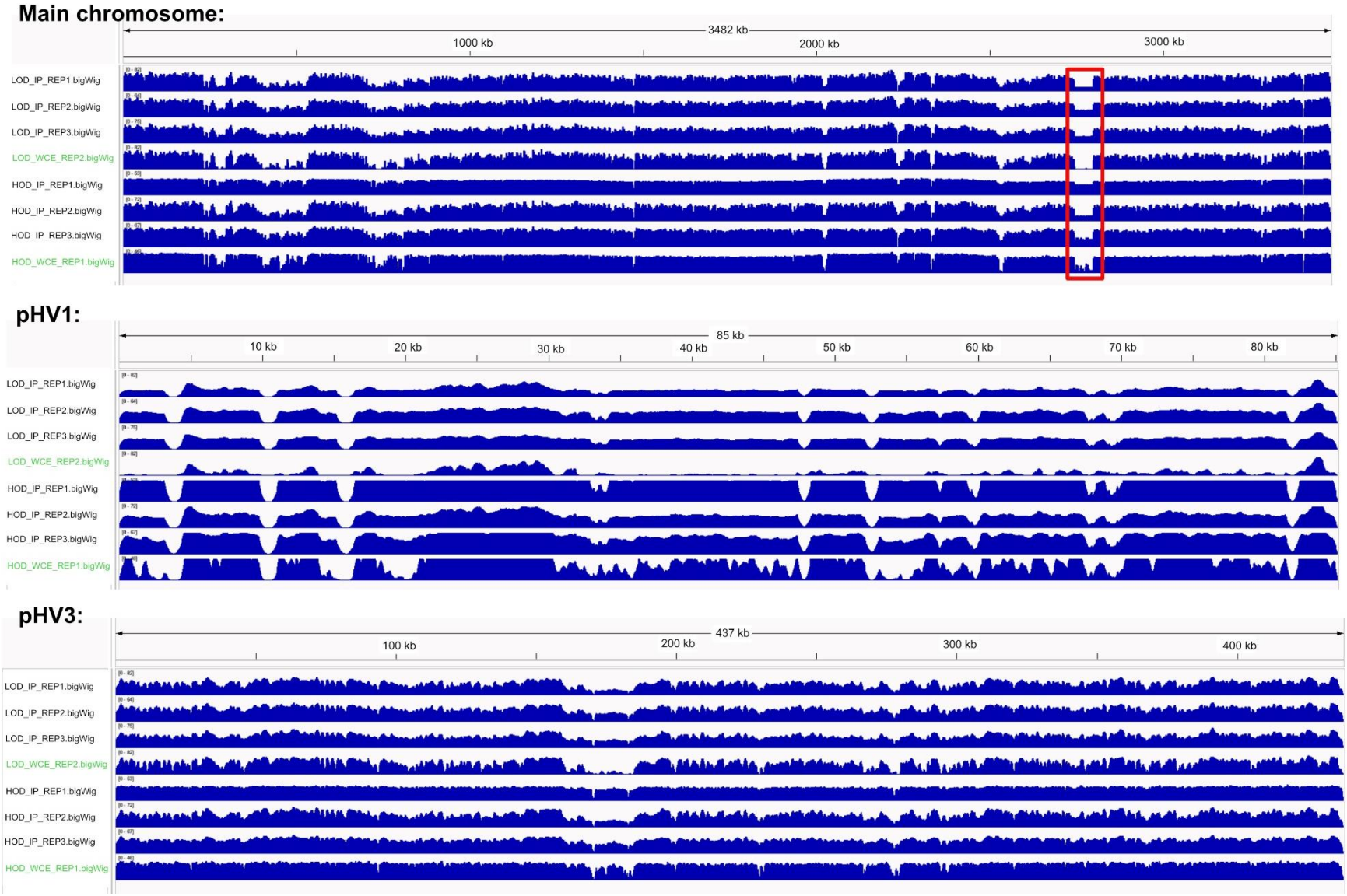


Figure 3.19: Read coverage of ChIP-Seq samples. Green text signifies WCE samples. The red box indicates a putative prophage region.

Additionally, the region contained in the red box in Figure 3.19 shows a lesser depth of coverage than the regions adjacent to it in all IP and WCE samples. As for pHV1, LOD samples seemingly appear to show read enrichment in this area (Figure 3.19) alone, but no peaks were called corresponding to this region in any of the LOD samples. The area maps exactly to a putative prophage region discovered in [48]. Previously, MFA sequencing undertaken in [49] had also found under-representation of this region; the MFA sequence used ABI SOLiD as opposed to Illumina sequencing used in this study, arguing against technical reasons for under-representation. This region has 50% GC content (lower than the remainder of the *H. volcanii* genome, which is on average 65% GC) but is not repetitive, therefore sequencing should not be inhibited for these reasons.

Figure 3.20 shows depth of coverage data across the main chromosome for representative IP and WCE samples from this study, MFA sequencing data from [49], and the GC content of the main chromosome. Correlation can be seen between the GC content in a particular region and the copy number (MFA) or read coverage (ChIP samples): areas with lower GC content show a reduction in copy number and read depth. When the MFA sequencing data was previously analysed in [49], sequence composition was found to influence depth of coverage (a reduction in mapped sequence reads was found at high GC content) and GC-bias was corrected for. The putative prophage region has a lower GC content than average for the *H. volcanii* genome, therefore GC-bias should lead to an increase in depth of coverage of this area; the Deep Seq facility did not correct for GC-bias during their analysis. Due to this, it is not thought that the lower coverage of the region seen in MFA and ChIP Seq is an artifact of the sequencing process or analysis, and instead it may hold biological significance. Possible explanations for this are explored in the discussion of this chapter.

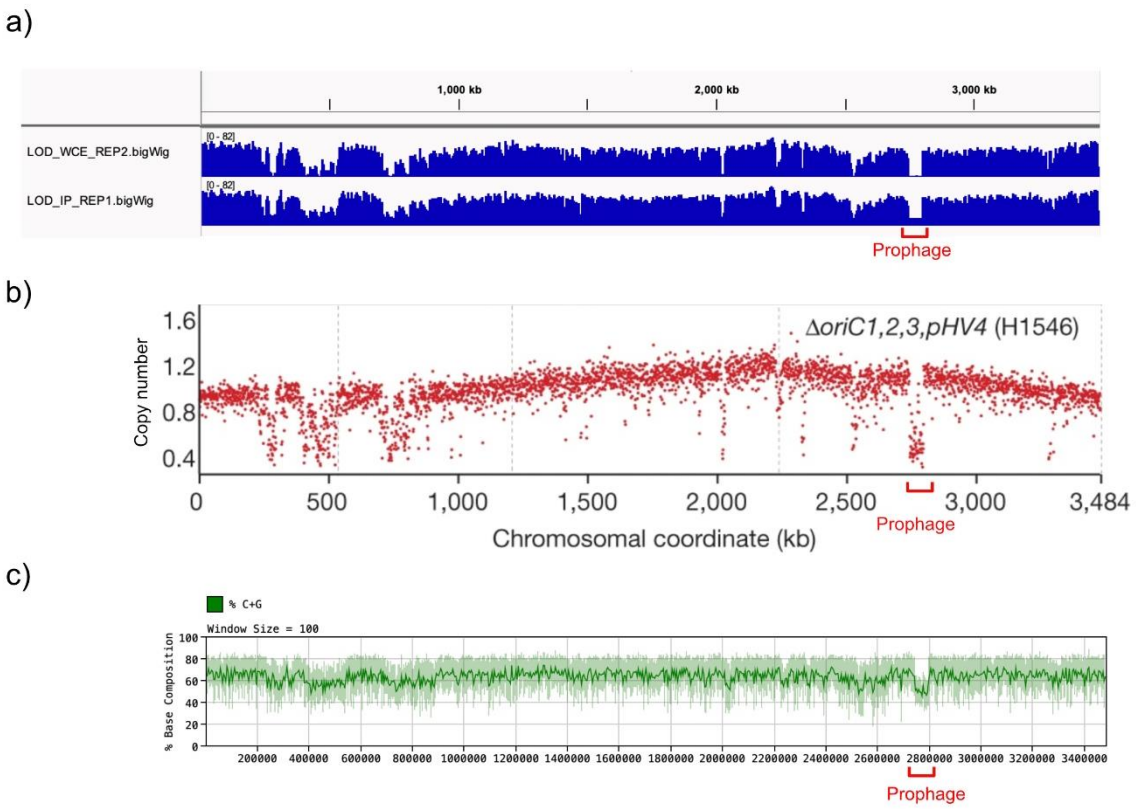


Figure 3.20: ChIP-Seq and MFA sequencing data compared to the GC content of the H. volcanii genome. a) Main chromosome read coverage of two representative ChIP-Seq samples. **b)** MFA sequencing demonstrating copy number across the main chromosome. **c)** GC content of the main chromosome.

Figure 3.20b taken from [49].

Consensus peaks

The bioinformatic pipeline used by the Deep Seq facility determines intervals (regions of the genome) in which RadA enrichment can be seen over the background. Intervals may contain multiple peaks and are of varying sizes. Combined analysis of LOD and HOD samples allows general recombination hotspots to be identified (not those specific to LOD or HOD samples, if such regions exist). The maximum number of IP samples for which enrichment is seen within a given interval is four out of the six, meaning that for a given interval, two out of the six samples did not demonstrate significant read enrichment in this region of the genome. Out of 389 intervals identified in total, 8 intervals were enriched for in four samples, and 35 were enriched for in three samples. Therefore, in each case two or

more of the samples did not show enrichment for RadA in each interval identified. This result is not unexpected due to the variability across samples seen in QC.

Information concerning each of the intervals demonstrating RadA enrichment in across all of the samples can be found in the appendices Table 7.2. Figure 3.21 provides information for the 8 intervals in which four samples show enrichment for RadA binding.

The table in Figure 3.21 demonstrates the p-values corresponding to the peaks falling within a particular interval for each sample. None of the peaks have a value equal to or lower than the generally accepted 0.05. Instead, values as high as 38 can be seen. This further reinforces the necessity of protocol optimisation before reliable conclusions can be drawn.

Chromosome	Start	End	Peaks	Samples	HOD_IP_REP1 p-value	HOD_IP_REP2 p-value	HOD_IP_REP3 p-value	LOD_IP_REP1 p-value	LOD_IP_REP2 p-value	LOD_IP_REP3 p-value	Annotation	GC content	Distance to TSS	Nearest promoter	Nearest gene	Gene name	Description
M.C.	541794	567045	8	4	NA	20.4938	4.68562; 4.00772; 11.244; 3.98815; 10.3673	5.59648	NA	7.16036	TTS (HVO_A0619)	68,9%	461	HVO_A0618	HVO_A0618	-	ferredoxin-NAD+ reductase
M.C.	641615	646225	4	4	NA	38.0442	13.331	4.48588	NA	6.60607	promoter-TSS (HVO_A0524)	70.8%	-763	HVO_A0524	HVO_A0524	*3-hydroxyacyl-CoA dehydrogenase*	
M.C.	898407	915128	7	4	NA	17.4065; 12.5885	5.37078	4.72091; 3.57163; 5.53931	NA	5.27401	TTS (HVO_0301)	68.1%	742	HVO_0302	HVO_0302	-	hypothetical protein (ABC-type multidrug transport system, permease component)
M.C.	1299695	1337026	8	4	3.61193	40.219	16.6471; 3.85494; 7.86105	NA	NA	6.77394; 4.49177; 5.27579	promoter-TSS (HVO_0761)	70.5%	-276	HVO_0761	HVO_0761	-	hypothetical protein (Uncharacterized conserved protein)
M.C.	1679351	1683190	4	4	NA	22.6726	11.8996	4.64046	NA	4.29122	promoter-TSS (HVO_1143)	69.1%	436	HVO_1145	HVO_1145	*rps3aR*	30S ribosomal protein S3a.eR
M.C.	2660736	2687823	9	4	NA	11.9013; 26.9585; 5.25171; 27.4476	13.909	4.12497; 4.5317; 6.20704	NA	7.57215	TTS (HVO_2167)	68.8%	574	HVO_2168	HVO_2168	*moxR3*	methanol dehydrogenase regulatory protein
M.C.	2975710	2992724	4	4	NA	19.4962	9.53598	3.81337	NA	4.36235	promoter-TSS (HVO_2479)	69.6%	308	HVO_2481	HVO_2481	-	phosphohydrolase
M.C.	3052114	3072434	4	4	3.39782	25.4962	12.2666	NA	NA	5.02916	TTS (HVO_2576)	68.7%	288	HVO_2575	HVO_2575	-	translation elongation factor aEF-1 subunit alpha-like protein

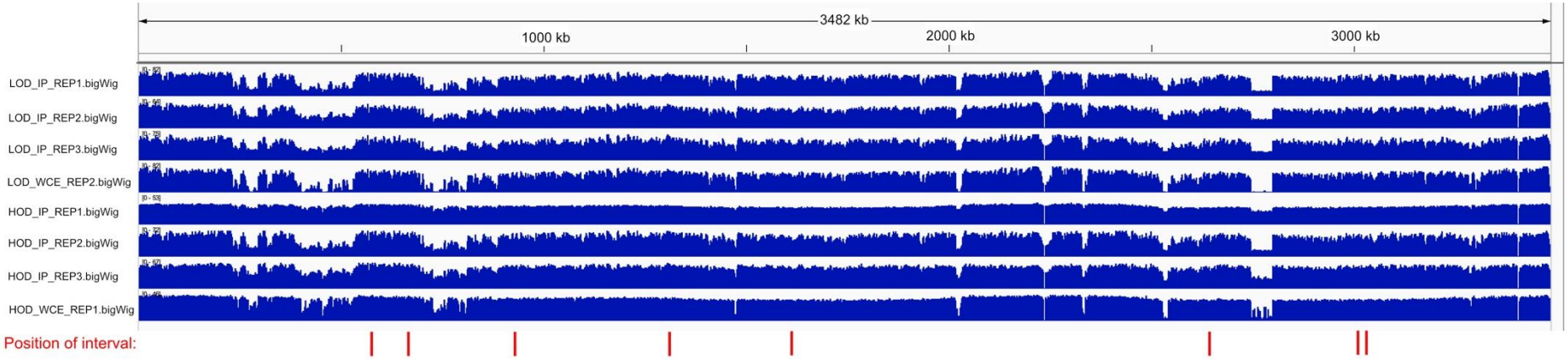


Figure 3.21: Intervals identified as demonstrating enrichment. a) Table of intervals. M.C. denotes main chromosome and where text is red, the interval is found upon pHV4 (integrated onto the main chromosome in all laboratory strains). “Start” and “End” denote the position on the main chromosome at which the interval starts and ends. “Peaks” denotes the number of peaks within the interval and “Samples” denotes the number of samples for which the interval was identified. Where NA is present in p-value columns, the interval was not identified as being enriched in this sample. “TSS” denotes transcription start site. **b)** ChIP-Seq read coverage data with the positions of intervals from a) marked.

Peak to feature proportion

In all samples, a greater number of peaks map to promotor-transcription start site regions (promotor-TSS) than to either transcription termination sites (TSSs) or within exons (Figure 3.22). Additionally, a greater proportion of peaks map to TSSs than within exons.

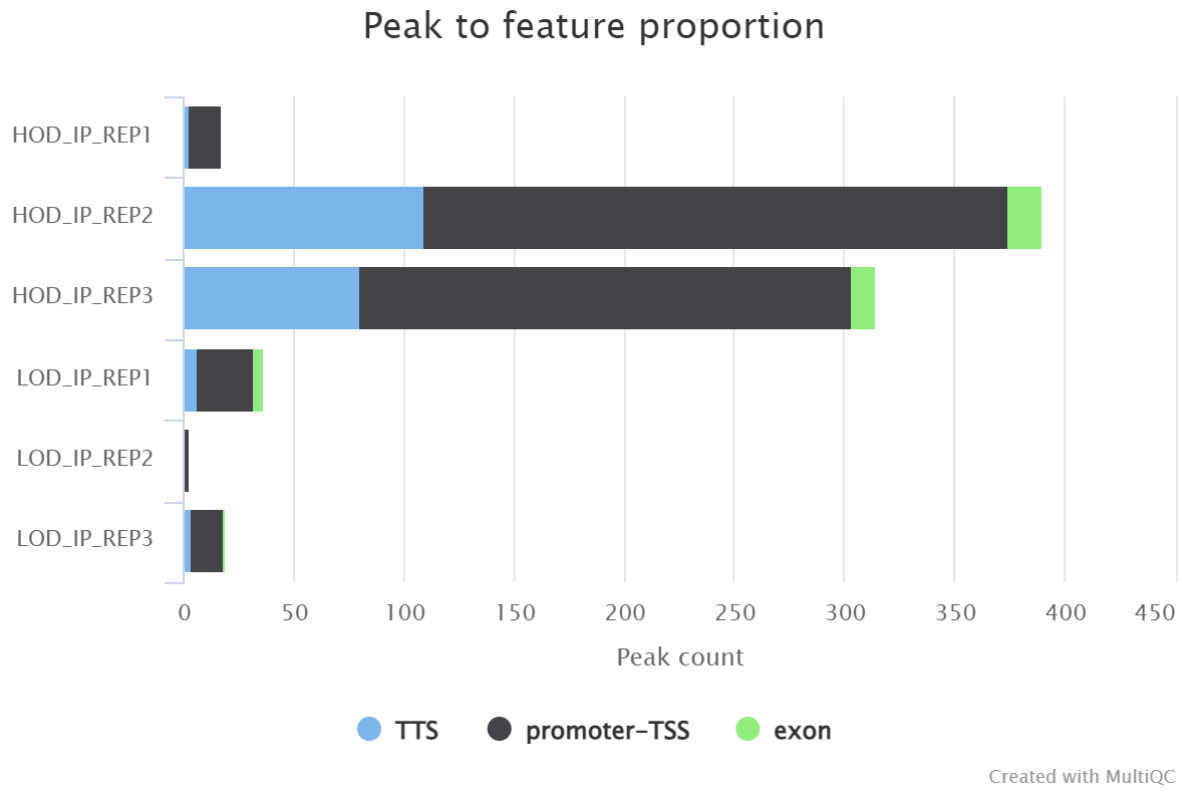


Figure 3.22: Peak to feature proportion of ChIP-Seq samples. A greater proportion of peaks map to promotor-TSS regions for all samples than at TSSs or within exons.

Figure created by the Deep Seq facility at the University of Nottingham using the MultiQC tool [251].

Several potential explanations may be offered as to why a greater number of peaks may be seen at promotor-TSS regions, including abortive transcription initiation leading to R-loops via strand invasion requiring RadA, and an increase in transcription-mediated DNA damage using RadA in its repair.

Peak visualisation

To investigate the differences in RadA binding between early and mid/late exponential phase growth (when differential use of origins for replication initiation may occur), peaks were visualised using IGV.

The distribution of peaks for LOD samples across the *H. volcanii* genome can be seen in Figure 3.23, and distribution of peaks for HOD samples can be seen in Figure 3.24.

In Figure 3.23, very few peaks can be seen for sample LOD_IP_REP2 which is expected as the FRiP score (fraction of reads mapping to peaks) for this sample was less than 1% (described earlier in this chapter). The distribution of peaks on the main chromosome varies greatly between samples LOD_IP_REP1 and LOD_IP_REP3 (no peaks are present on the main chromosome in sample LOD_IP_REP2), which is again as expected due to the variability between samples seen in PCA. No peaks at all can be seen mapping to pHV3 for any of the LOD samples.

Deletion of the origin found upon pHV3 has been attempted several times in the laboratory and has yet to be completed successfully unless pHV3 is found integrated onto the main chromosome (Laura Mitchell and Hannah Marriott, unpublished data). The apparent lack of RadA binding upon the pHV3 mini-chromosome might indicate that recombination-dependent replication is not initiated from pHV3 during early exponential phase and may explain why origin deletion is seemingly impossible. It has also been found that the level of transcription occurring from pHV3 is notably low (Darya Ausiannikava- unpublished data, [255]) which may mean that the level of transcription related DNA damage and R-loop formation is reduced, and therefore RadA binding is reduced. Further interpretation of this finding can be found in the discussion of this chapter and the relationship between transcription and recombination-dependent replication is explored in Chapter 4. However, all conclusions drawn from this data are tentative due to the issues in QC described previously.

In Figure 3.24, sample HOD_IP_REP1 shows fewer peaks upon the main chromosome than samples HOD_IP_REP2 and HOD_IP_REP3, and no peaks are present at all for this sample on pHV3. However, upon pHV1 very broad peaks can be seen for this sample, spanning more of pHV1 than seen for any other of the samples. Many peaks can be seen for HOD_IP_REP2 and HOD_IP_REP3 on the main chromosome, but HOD_IP_REP2 has no peaks on pHV1 and HOD_IP_REP3 shows fewer peaks than HOD_IP_REP2 on pHV3. The lack of consensus among peaks within biological replicates such as this again demonstrates the necessity for protocol optimisation. Changes to be made to rectify such issues are described in the discussion of this chapter.

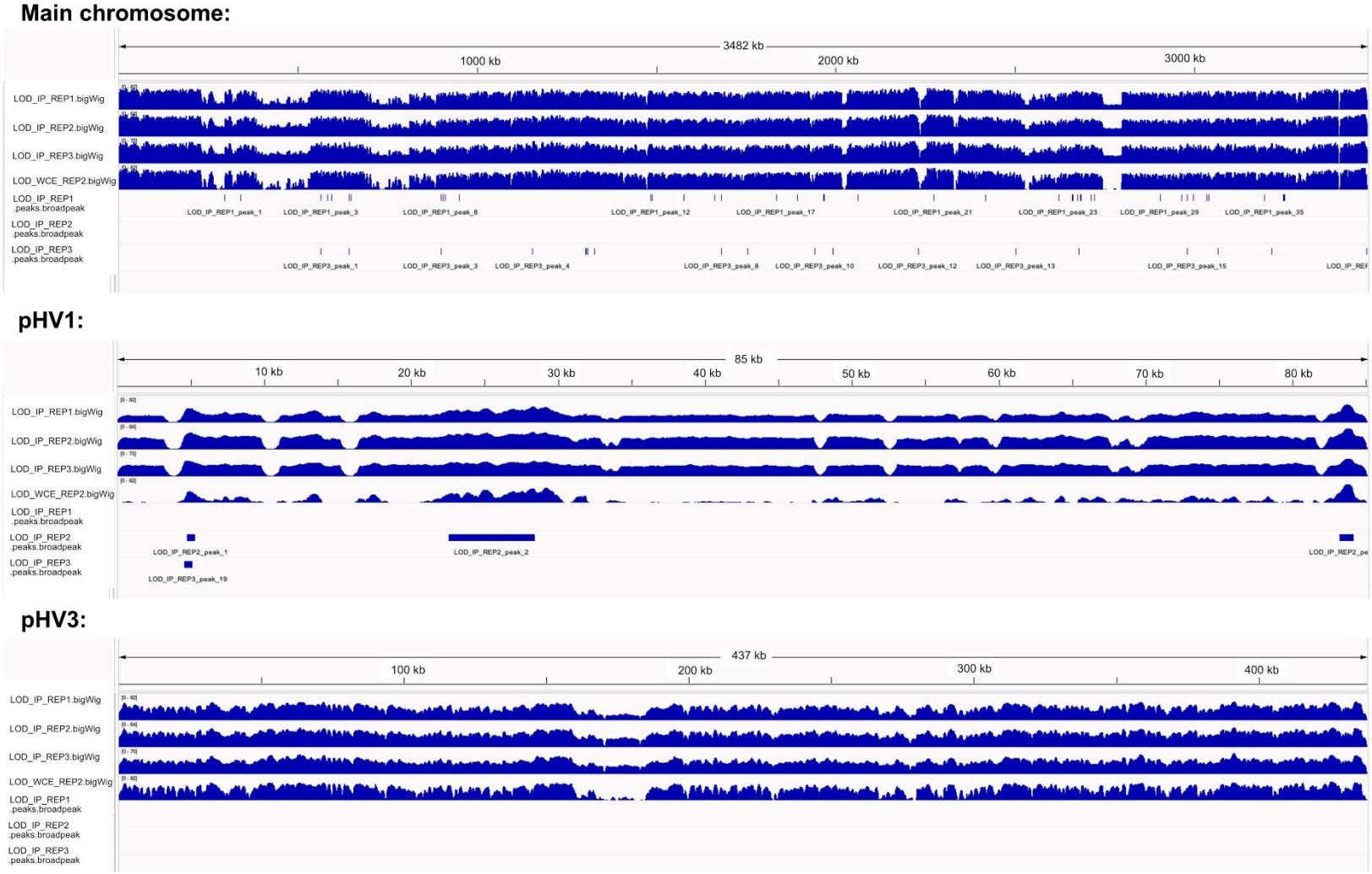


Figure 3.23: Depth of coverage and peaks called in LOD samples. The upper four lanes per component of the *H. volcanii* genomic architecture represent depth of read coverage in the order LOD_IP_REP1, LOD_IP_REP2, LOD_IP_REP2, and LOD_WCE_REP2 (as shown in the figure). The lower three lanes per component represent peaks corresponding to regions of the genome enriched in IP samples compared to the background (in the order LOD_IP_REP1, LOD_IP_REP2 and LOD_IP_REP3, again as shown in the figure).

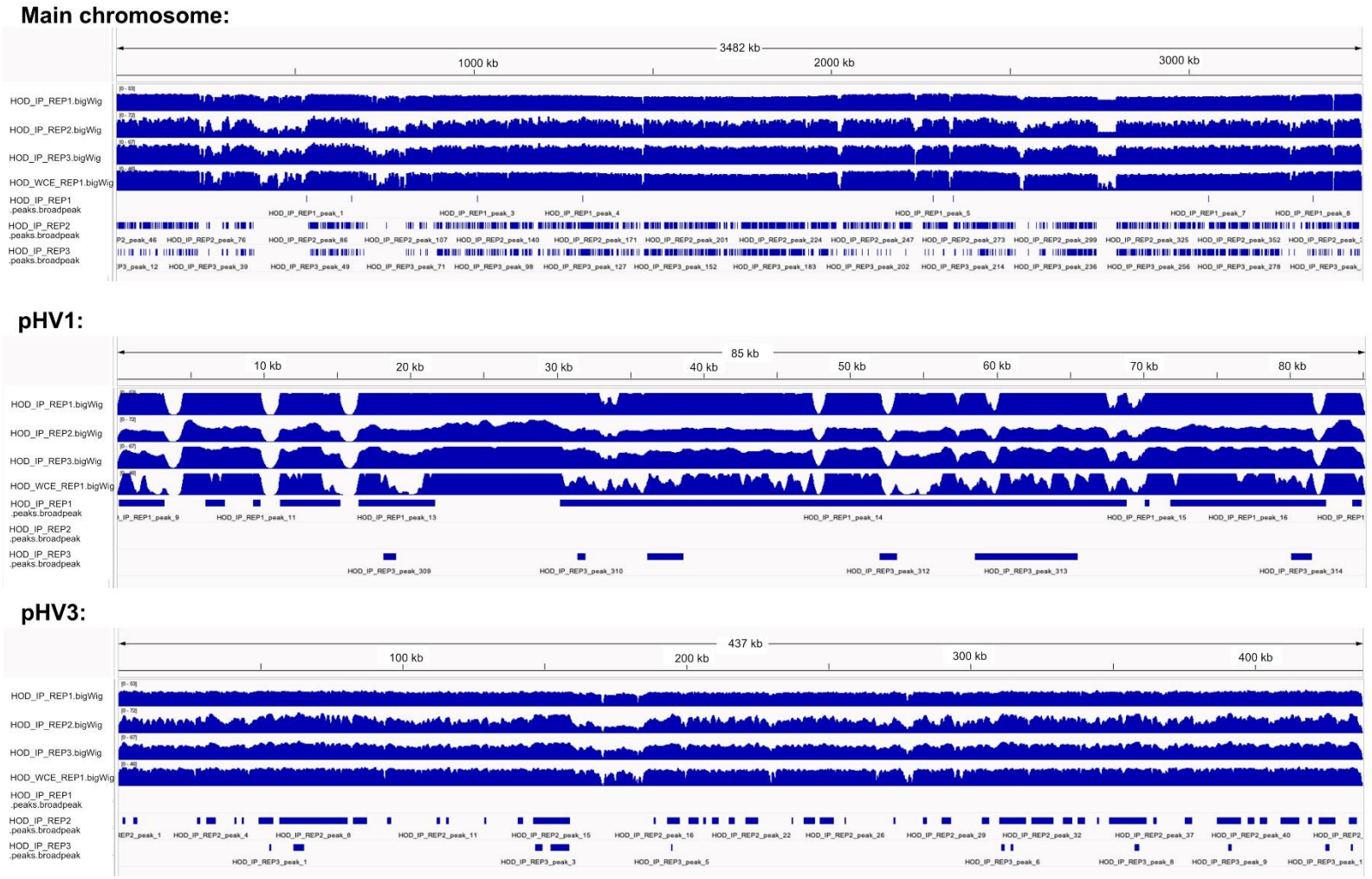


Figure 3.24: Depth of coverage and peaks called in HOD samples. The upper four lanes per component of the *H. volcanii* genomic architecture represent depth of read coverage in the order HOD_IP_REP1, HOD_IP_REP2, HOD_IP_REP3, and HOD_WCE_REP1 (as shown in the figure). The lower three lanes per component represent peaks corresponding to regions of the genome enriched in IP samples compared to the background (in the order HOD_IP_REP1, HOD_IP_REP2 and HOD_IP_REP3, again as shown in the figure).

Initially, this chapter aimed to investigate potential differentiation in RadA binding between early and mid/late log phase samples, based upon the hypothesised dependence upon origins for replication initiation. As described previously, MFA sequencing data from the laboratory prior to this study appeared to show differential use of origins to initiate replication, depending on the growth phase; in early exponential growth (during which LOD samples were taken) it is thought that cells rely heavily upon origins for replication initiation, and in mid/late exponential phase (during which HOD samples were taken) cells instead use recombination-dependent replication. However, difficulties arose in this comparison due to the variability across the samples as described throughout this section, meaning that it was not possible for this to be completed by the Deep Seq facility due to time constraints.

3.4- Discussion

3.4.1- Evaluation of FLAG-RadA strain construction

This chapter aimed to generate *H. volcanii* strains that express FLAG-RadA for use in ChIP-Seq, to investigate RadA localisation in origin-independent replication. It was necessary that RadA remained functional in these strains and that FLAG-RadA could be detected reliably with the anti-DDDDK antibody, so that the same antibody could be used in the immunoprecipitation stage of ChIP-Seq sample preparation.

Southern blotting confirmed the genotypes of 3 candidate FLAG-RadA strains with both wild-type and origin-deleted backgrounds. The expression of FLAG-RadA in each of these strains was verified through Western blotting and a single FLAG-tag was sufficient for detection with the anti-DDDDK antibody. The growth rate of FLAG-*radA* strains, and resistance to both UV- and gamma-irradiation were significantly increased in comparison to a strain deleted for *radA*, and largely matched those of a wild-type *radA*$^{+}$ strain, demonstrating that FLAG-*radA* is functional. Overall, this chapter documents successful creation of strains expressing FLAG-RadA in wild-type and origin-deleted backgrounds, suitable for ChIP-Seq.

In the standard growth assays undertaken for both wild-type background and origin-deleted background FLAG-*radA* strains, the generation time of the FLAG-*radA* strains was slightly less than that of the respective control expressing wild-type RadA. In *H. volcanii,* RadA polymerises through insertion of an invariant phenylalanine (Phe-96) found in the N-terminal domain of the protein into the hydrophobic binding pocket of an adjacent protein [102]. The FLAG-tag is also located at the N-terminal domain, and it is therefore possible that the presence of the tag plays a role in the polymerisation and function of the protein, leading to strains with faster growth rates than strains expressing the wild-type protein. Additionally, interaction with RadB may be altered[222].

Investigation into the presence of two bands per lane seen when targeting FLAG-RadA through Western blotting would be of interest. Post-translational modification contributes to the complexity of the *H. volcanii* proteome (as reviewed in[256]) and could be responsible for the two distinct bands formed by FLAG-RadA. In this case, modifications could be identified through treating samples with enzymes suitable for the removal of the modification such as alkaline phosphatase for phosphorylation or Peptide:N-glycosidase F (PNGase F) for N-glycosylation, and attempting further Western blotting to look for disappearance of the second band. Additionally, truncation of the protein at the C-terminus could also be responsible for the two bands seen. Both post-translational modification and truncation could be investigated through mass spectrometry, and the extent of the truncation could be identified.

Initially, this chapter aimed to delete *radA* in a strain with only one origin remaining on the main chromosome- *oriC1*. However, colony hybridisation showed that all

candidate Δ*radA* pop-out colonies retained the *radA* gene (Figure 3.3). When all origins found upon the *H. volcanii* main chromosome are deleted, strains become critically dependent on RadA[49]. When origins are deleted, the percentage of successful *radA* deletion transformations decreases. Therefore, in agreement with published work, deletion of *radA* in strains with only one remaining origin proves difficult. However, direct replacement of wild-type *radA* with FLAG-*radA* was successful, meaning that deletion of *radA* to then replace the deletion construct was no longer necessary. Direct replacement was not initially attempted as the *radA* deletion construct was *trpA*+ and the strain into which the plasmid was transformed was Δ*trpA*. When replacing *radA* with FLAG-*radA,* it was not possible to use tryptophan selection for FLAG-*radA* over wild-type *radA*. Instead, upstream pop-ins and downstream pop-outs were selected for and the Southern blot in Figure 3.8 showed that the strains taken forward were not merodiploid.

3.4.2- ChIP-Seq conclusions

Optimisation of the protocol

The ChIP-Seq data in this chapter demonstrated a need for further optimisation of the protocol. PCA demonstrated a variability across biological replicates and lengths of fragments were significantly greater than optimal for ChIP-Seq experiments.

It was hypothesised that the discrepancy between size of fragments when running sonicated samples on an agarose gel compared to the data the Deep Seq facility generated using the D1000 and D5000 ScreenTape was due to the presence of RNA when the fragment length was checked through pulsed-field gel electrophoresis. To test this and to optimise the protocol, the sonicated samples should be subjected to RNase treatment prior to running them on an agarose gel.

In addition to immunoprecipitation against RadA, a positive control could be implemented in which immunoprecipitation and ChIP-Seq against a different protein known to bind to specific parts of the genome (for example, an Orc1 homologue, which binds to origins). This would demonstrate whether any further issues in sample preparation have arisen and a conclusive result in the positive control would provide greater confidence in conclusions drawn for RadA binding. Alternatively, a positive control identical to that seen in published work such as [229] could be implemented.

Findings

Tentative conclusions were drawn from the data, but key elements of the QC compromise their significance and reliability.

IP samples showed greater depth of coverage across pHV1 than WCE samples, although few peaks were observed upon this mini chromosome. This could indicate that RadA binds extensively across the entirety of pHV1, which may play a role in pHV1's replication mechanism[53,208]. In a PFG undertaken in [53], a probe hybridising to the origin found upon pHV1 (*oripHV1*) bound at both ~85kb and at a similar size to pHV4 (636kb). However, since then, it has been discovered that pHV1 exists in a ~6-fold concatemeric form, which is of a similar size to pHV4[208] and explains the result seen in [53]. In a Δ*radA* strain, the monomeric form of pHV1 is the predominant form, suggesting that the 6-fold concatemer requires RadA (and by implication, recombination-dependent replication) for its formation[208]. The dependence upon RadA for concatemerisation provides a possible explanation for the greater depth of pHV1 coverage seen in IP samples than in WCE samples.

A lack of peaks corresponding to RadA binding in LOD samples mapping to pHV3 was observed. Furthermore, the proportion of peaks mapping to promotor-transcription start site regions was greater than that mapping to transcription termination sites and exons. The low level of transcription occurring from pHV3 may result in a reduction in transcription-mediated DNA damage and therefore the requirement for RadA is reduced. Deletion of the origin found upon pHV3 is not possible unless pHV3 is integrated onto the main chromosome. When all origins on the main chromosome are deleted, RadA becomes essential, and it is therefore thought that recombination-dependent replication is utilised. The lack of peaks on pHV3 may demonstrate a low level of recombination on this mini-chromosome (which may be as a result of the lack of transcription and formation of R-loops). In this case, it would be difficult to carry out recombination-dependent replication and therefore would also be difficult to delete the pHV3 origin, as shown previously (Laura Mitchell and Hannah Marriott, unpublished data). Therefore, there may be a correlation between transcription, RadA binding, and ease of origin deletion.

To further investigate the link between transcription and recombination, a comparison between RNA sequencing data and the intervals identified as being enriched for RadA binding in this study would be of interest. It is expected that regions of the genome enriched for RadA binding would be highly transcribed.

Across all ChIP samples, a reduction in coverage can be seen for a putative prophage region. It could be hypothesized that within this region, recombination-dependent replication is suppressed. If so, under-representation in MFA sequencing would be expected due to the reduction in replication initiation, and RadA would therefore not bind meaning that read enrichment for this region would not be seen in ChIP-Seq. However, this does not explain why fewer reads are seen mapping to this region in WCE samples. Recombination and chromosome rearrangement events are

seen regularly within the *H. volcanii* genome[49,55] and due to its polyploid nature, it could be possible that the prophage region is lost in this manner in some of the copies of the genome but remains in others, meaning that the copy number for this region would be lower than for the remainder of the genome in MFA, and also that fewer reads are present in ChIP-Seq. Reduction in reads could be due to transient phage excision in log phase samples which reintegrate onto the chromosome in stationary phase and therefore are not lost. Further work would be necessary to determine the biological significance of this result.

Chapter 4- Replication initiation from R-loops and genomic rearrangements

4.1- Background

In *H. volcanii*, all origins of replication found on the main chromosome can be deleted[49]. Deletion of the origin found on pHV1 is also possible[53], and the origin found on pHV4 can be deleted both when pHV4 is integrated onto the main chromosome[49] and when pHV4 is found as an episome (Darya Ausiannikava, unpublished data). However, deletion of the origin on pHV3 is not possible, nor can pHV3 be lost due to the essential genes found upon it (Hannah Marriott, unpublished data, [257]). It is possible to delete the origin when pHV3 is integrated onto the main chromosome (Hannah Marriott, unpublished data), showing that there is nothing inherently essential about the origin itself.

Transcription on pHV3 appears to occur at a very low level (Darya Ausiannikava-unpublished data, [255]). When transcription levels were dramatically increased on pHV3 through insertion of a strong constitutive promotor (*p.syn*- a synthetic 43bp sequence based upon the *H. volcanii* consensus tRNA promotor[258,259]), it was seemingly possible to delete the origin (Laura Mitchell, unpublished data). The promotor drives high levels of native *adh2* and over-expression of this gene does not have any ill effects on the cell[258]. This suggested that origin-independent replication of pHV3 involves R-loops as the absence of highly transcribed genes on pHV3 would provide fewer opportunities for their formation, and DNA replication in the absence of origins would be impossible on pHV3 if initiated from R-loops.

Previous work into replication initiation from R-loops in both bacteria and eukaryotes found homologous recombination to be essential for this process. In *rnhA E. coli* mutants, RecA was found to be essential for cSDR[260]. A mutation which inactivates the recombinase activity of RecA but has no effect on its coprotease activity was found to prevent DNA replication. Additionally, a mutation causing great decrease in coprotease activity but little effect on recombinase activity shows only slight decrease in DNA replication, suggesting that the recombinase activity of RecA specifically is necessary for replication in this manner [260,261]. A similar recombinase dependence seems to also be present in eukaryotes. Investigation into R-loops priming origin-independent replication at the ribosomal DNA (rDNA) locus in *Saccharomyces cerevisiae* finds an increase in Rad52-YFP foci in cells with RnaseH1 and RnaseH2 mutations [262]. The homologous DNA strand annealing mediator function of Rad52 and its role in DNA repair combined with its heightened accumulation at the highly transcribed rDNA locus indicates an increase in DNA damage and increased dependence upon homologous recombination.

Additionally, the requirement upon RadA in replication from R-loops may depend upon the structure of DNA polymerase D.

DNA polymerases can be separated into seven different families based upon sequence homology: A, B, C, D, X, Y, and RT (reverse transcriptase) [263-265]. Bacterial DNA replication relies primarily upon the C-family polymerase Pol-III [264]. A-family polymerases can also be found in bacteria which are involved in removal of RNA primers and in Okazaki fragment processing[266]. In eukaryotes, Pol-α, Pol-ε, and Pol-δ are the main replicative DNA polymerases and all fall within the B-family [267]. All archaeal species possess B-family polymerases, and archaeal B-family polymerases and the catalytic subunit of eukaryotic B-family replicative polymerases show structural homology[268,269]. However, a phylogenetic divide within the domain becomes evident when investigating D-family polymerases. This family is specific to one of the two archaeal phyla; D-family polymerases with a small proofreading exonuclease subunit DP1 and a polymerase catalytic subunit DP2 are present in Euryarchaeota but are absent in Crenarchaeota[270,271]. It is hypothesised that B-family polymerases in Crenarchaeota may have gained specialised functions to compensate for the D-family polymerase absence as multiple different PolBs are present within a species. *H. volcanii* is a member of the Euryarchaeota with two B-family polymerase homologues, PolB1 and PolB2, and a D-family polymerase[272].

Investigation into the crystal structure of DP1 and DP2 DNA polymerase D subunits in the hyperthermophilic archaeon *Pyrococcus abyssi* revealed a catalytic core more closely resembling that found in the RNAP superfamily than in other known DNAPs [273]. The same “double-psi β-barrel” structure found in RNAPs is seen in the DP2 catalytic core. Interestingly, MFA analysis and origin deletion studies in *Thermococcus kodakarensis*, another hyperthermophilic archaeon, demonstrated no specific origin usage and instead, replication appears to be initiated at many sites across the genome [170]. *T. kodakarensis* is critically dependent upon RadA and RadB and it was therefore presumed that a recombination-dependent DNA replication mechanism is employed. Additionally, *T. kodakarensis* does not require DNA polymerase B for replication but is instead dependent upon DNA polymerase D [274]. Would the structural similarity between archaeal D family DNA polymerases and the RNA polymerase superfamily permit the direct use of R-loops to initiate origin-independent DNA replication? Furthermore, if use of R-loops is direct, would RadA be required? (Figure 4.1)

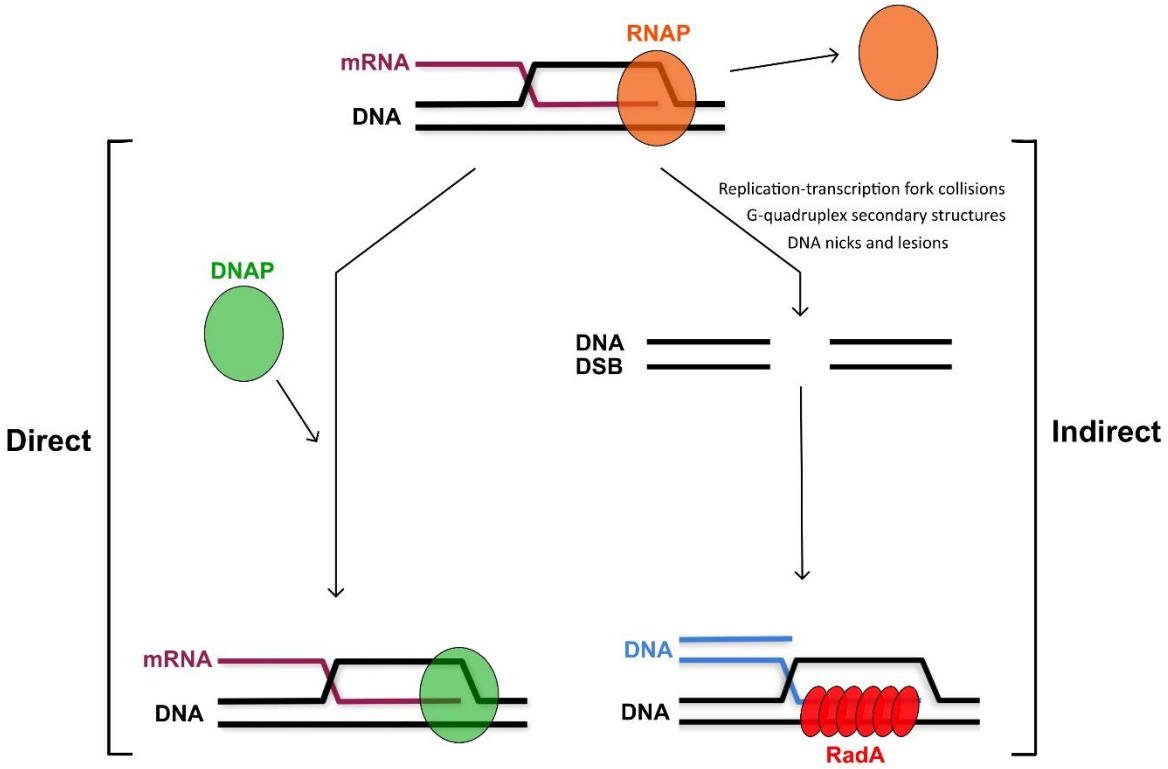


Figure 4.1: Direct vs indirect use of R-loops to prime replication. The question exists as to whether the structural similarity between euryarchaeal DNA polymerase D and RNA polymerase would permit the direct use of R-loops to prime replication (left of figure), or whether use of R-loops is indirect and dsDNA breaks arise which are processed by the homologous recombination machinery and that replication is initiated in a RadA-dependent manner (right of figure).

DNA damage and genomic instability as a consequence of transcription is detailed in Section 1.5- The role of transcription and R-loops in genomic instability. Additionally, the *H. volcanii* genome is notably fluid. Integration of the mini-chromosome pHV4 onto the main chromosome[49] and spontaneous formation of a new mini-chromosome[55] have both been observed as a result of homologous recombination.

4.2- Aims

The initial objective of this chapter was to explore the role of homologous recombination in the initiation of replication from R-loops in *H. volcanii*. Upon investigation into this, aims were expanded to explore the integration of pHV3 onto the main chromosome. This was carried out through:

- Phenotypic assessment of RadA-dependence in origin-independent replication of pHV3 using strains in which *radA* is under the control of a tryptophan inducible promotor (*p.tnaA*) (constructed prior to this study) on tryptophan gradient plates
- Quantitative assessment of growth rate in the same strains in ranging concentrations of tryptophan
- RT-PCR to investigate the level of *adh2* transcription in strains which were thought to contain the *p.syn::adh2*+ construct (constructed prior to this study)
- Pulsed field gel electrophoresis to explore the genomic architecture of the strains used in the abovementioned RT-PCR and assessment of RadA-dependence in origin-independent replication
- Investigation into the frequency of pHV3 integration onto the main chromosome in strains containing the *p.syn::adh2*+ construct (all constructed prior to this study)
- Determining the loci of pHV3 integration onto the main chromosome using Oxford Nanopore MinION sequencing and subsequent analysis
- Investigation into loss of pHV1 seen through the Oxford Nanopore MinION data

4.3- Results

4.3.1- Analysis of RadA-dependence in origin-independent replication of pHV3

Strains H4945 (supposed genotype: Δ*pyrE2, p.syn::adh2*+, Δ*orc6* Δ*oripHV3,* Δ*hdrB, p.tnaA-radA*+*::hdrB*+) and H4943 (supposed genotype: Δ*pyrE2, p.syn::adh2*+, Δ*hdrB, p.tnaA-radA*+*::hdrB*+) were constructed prior to this study. Upon commencement of this study, it was though that in strain H4945, *radA* expression was under the control of a tryptophan-inducible promotor (*p.tnaA-radA*+), and on pHV3, the level of transcription was thought to be dramatically increased through insertion of a strong synthetic promotor controlling expression of an alcohol dehydrogenase gene (*p.syn::adh2*+). *adh2* overexpression has no detrimental effect on strains[258]. The origin and corresponding Orc initiator protein on pHV3 had been

deleted (Δ*oripHV3,* Δ*orc6*). In common with other laboratory strains of *H. volcanii*, H4945 is deficient in uracil biosynthesis (Δ *pyrE2*). In strain H4943, the origin and Orc initiator protein on pHV3 were thought to be present but other than this, H4943 has the same genotype as H4945. Therefore, analysis of these strains in varying concentrations of tryptophan should provide insight into the dependence upon RadA for replication initiation in the absence of origins on pHV3. Further analysis of these strains in this chapter demonstrates that these supposed genotypes are incorrect. This section (Chapter 4.3.1) details experiments and the conclusions drawn from them according to the genotypes of strains as they were thought to be at the time the experiments were undertaken.

Phenotypic analysis

The following strains were grown for two consecutive overnights in Hv-Ca broth +Ura +0.1mM Trp and painted onto tryptophan gradient plates. The plates comprised bottom tapered wedge of Hv-Ca +Ura+ 0.1mM Trp agar overlayed with a converse wedge of Hv-Ca agar +Ura (see Materials and Methods).

Table 4.1: Strains plated upon the tryptophan gradient plate in Figure 4.2.

Strain	Genotype	Construction
H26	Δ *pyrE2*	Constructed by TA
H53	Δ *pyrE2* Δ*trpA*	Constructed by TA
H1637	Δ*pyrE2* Δ*hdrB* *p.tnaA-radA+::hdrB+*	Constructed by MH
H1642	Δ*pyrE2* *p.tnaA-radA+::hdrB+* Δ*oriC2* Δ*oriC1* Δ*oriC3* Δ*hdrB* Δ*ori-pHV4-2*	Constructed by MH
H4943	Δ*pyrE2* *p.syn::adh2+* Δ*hdrB* *p.tnaA-radA+::hdrB+*	Constructed by LM
H4945	Δ*pyrE2* *p.syn::adh2+* Δ*orc6* Δ*oripHV3* Δ*hdrB* *p.tnaA-radA+::hdrB+*	Constructed by LM

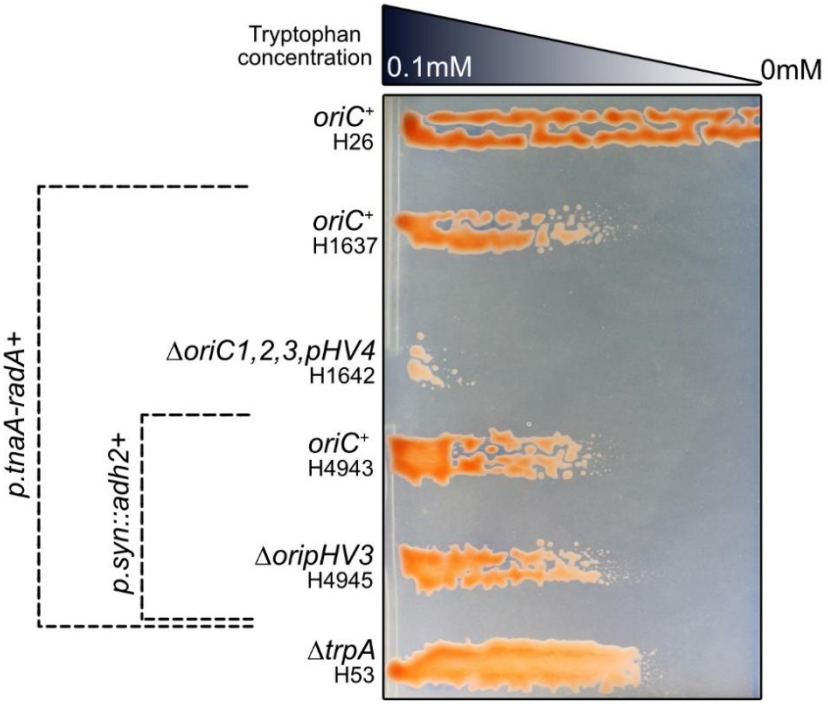


Figure 4.2: RadA is not essential in H4945. *radA* was placed under the control of a tryptophan-inducible promotor (*p.tnaA*) and transcription levels on pHV3 were thought to be increased by a strong constitutive synthetic promotor (*p.syn*) in Δ*oripHV3* and *oriC*$^{+}$ strains. The Δ*oripHV3* strain is also Δ*Orc6* (the Orc initiator protein associated with *oripHV3*).

Strains H26 and H53 are trpA+ and ΔtrpA respectively and were used to demonstrate the tryptophan gradient across the plate. The dependence upon RadA in origin-independent replication from the main chromosome is shown through H1637 and H1642; all origins remain in H1637, all origins found on the main chromosome have been deleted in H1642 (including the integrated pHV4), and expression of *radA* in both strains is under the control of a tryptophan inducible promotor. The transcription levels on pHV3 are increased by through insertion of a strong synthetic promotor to drive expression of an alcohol dehydrogenase gene (*p.syn::adh2+*). Strains H4943 and H4945 are as described previously.

In Figure 4.2, growth of H4945 closely resembles that of both H1637 and H4943. As H1637 does not have any origin deletions, it is not dependent upon RadA for initiation of replication. In H1642, all origins on the main chromosome have been deleted and reflecting previous work[49], absence of RadA is lethal as growth on the plate can only be seen at the highest concentration of tryptophan. If RadA was necessary in the initiation of replication of pHV3 from R-loops, growth of H4945 would resemble that of H1642. However, we do not see this and instead, growth of H4945 mirrors H1637 and H4943 – strains which do not require RadA for initiation of replication.

Quantitative analysis

Growth rates of the same strains seen in Figure 4.2 were quantified in liquid media. A standard growth assay was carried out according to Chapter 2: Materials and Methods. Strains were grown in Hv-Ca broth +Ura +0.1mM Trp for the 3 successive overnight cultures, and then diluted into Hv-Ca broth +Ura and Hv-Ca broth +Ura + 0.1mM Trp immediately before preparing the 96-well plate for differential RadA expression during A_{650} measurements.

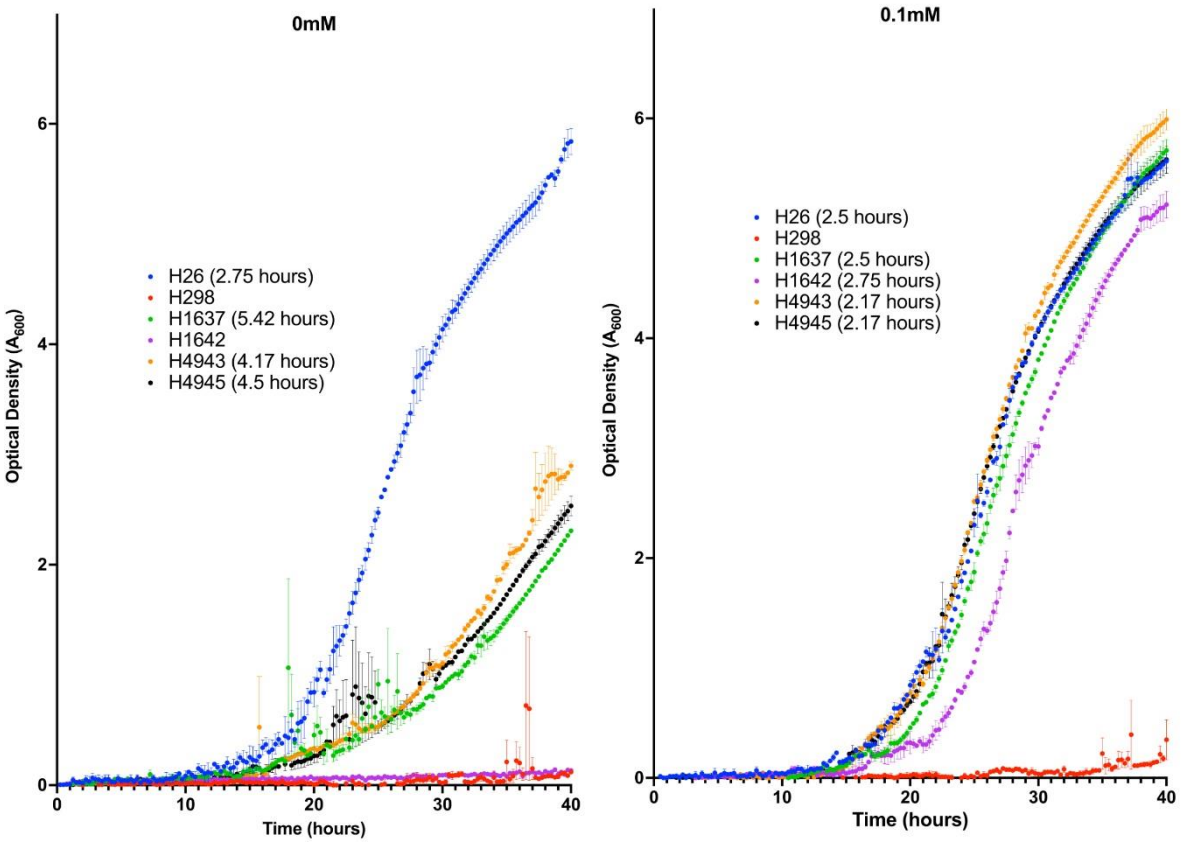


Figure 4.3: Exponential growth of strains to investigate RadA dependence in origin-independent replication. H26 (wild-type), H298 (ΔradA), H1637 (p.tnaA-radA+), H1642 (ΔoriC1,2,3,pHV4, p.tnaA-radA+), H4943 (p.syn::adh2+, p.tnaA-radA+), H4945 (p.syn::*adh2+, p.tnaA-radA+, Δorc6 ΔoripHV3*). Generation times during logarithmic phase are shown in the keys. Each data point is the mean of two technical repeats and error bars display standard error of difference. **a**) Growth of strains in media at 0mM tryptophan. **b**) Growth of strains in media at 0.1mM tryptophan.

At 0mM tryptophan (Figure 4.3), strains H298 (Δ*radA*) and H1642 (Δ*oriC1,2,3,pHV4*, *p.tnaA-radA+*) show no growth. H1637 (*p.tnaA-radA+)*, H4943 (*p.syn::adh2+, p.tnaA-radA+*), and H4945 (*p.syn::adh2+, p.tnaA-radA+, ΔOrc6 ΔoripHV3*) have similar growth rates but are all reduced in comparison to H26 (wild-type). At 0.1mM tryptophan (Figure 4.3), H298 still shows no growth. However, H1642 now grows well with a growth rate almost that of H26 as *radA* is being expressed in the presence of tryptophan and is essential for origin-independent replication on the main chromosome. H1637 has a generation time equal to that of H26 (2.5 hours) and both H4943 and H4945 grow slightly faster than H26 (2.17 hours). If genotypes were as thought, the growth rate of H4945 showing an increase in comparison to H1642 and similar to that of H1637 and H4945 at 0mM tryptophan would demonstrate that origin-independent replication of pHV3 is not dependent upon RadA.

adh2 expression

RT-PCR was carried out upon H26 (wild-type), H4943 (*p.syn::adh2+, p.tnaA-radA+*), and H4945 (*p.syn::adh2+, p.tnaA-radA+, Δorc6 ΔoripHV3*) to confirm that an increase in *adh2* transcripts can be seen. Both H4943 and H4945 have the *p.syn::adh2+* construct and it was when this construct was introduced that deletion of the origin on pHV3 seemingly became possible. Confirmation that an increase in *adh2* mRNA transcripts in H4943 and H4945 were present in comparison to wild-type was necessary to determine whether deletion of oripHV3 was made possible due to the increased level of transcription occurring upon the pHV3 mini-chromosome and that R-loop dependent replication had the potential to occur.

RNA was extracted from strains H26, H4943 and H4945. The RNA samples were DNase treated and cDNA was synthesised. RNA samples pre- and post-DNase treatment were PCR amplified to assess whether any contaminating DNA had been degraded (Figure 7.4, appendices). Next, PCR was undertaken on the cDNA for each strain with OneTaq HotStart polymerase. Figure 4.4a demonstrates PCR amplification of cDNA corresponding to *adh2*, *rpl10*, and *polB1* transcripts (all diluted to the same concentration) using oligonucleotides o2438 and o2439, o2132 and o2133, and o2136 and o2137 respectively. Each oligonucleotide binds within the target gene and would therefore amplify cDNA created from the mRNA transcripts. If the level of *adh2* transcription was increased in H4943 and H4945, as expected in strains with the *p.syn::adh2+* construct, the band seen through PCR amplification for *adh2* cDNA would be brighter than the band seen in H26 as there would be a greater proportion of *adh2* mRNA transcripts in the RNA sample initially. Genes *rpl10* and *polB1* form controls as these genes had not been modified in strains H26, H4943 and H4945 and so their transcription would therefore be expected to be equal.

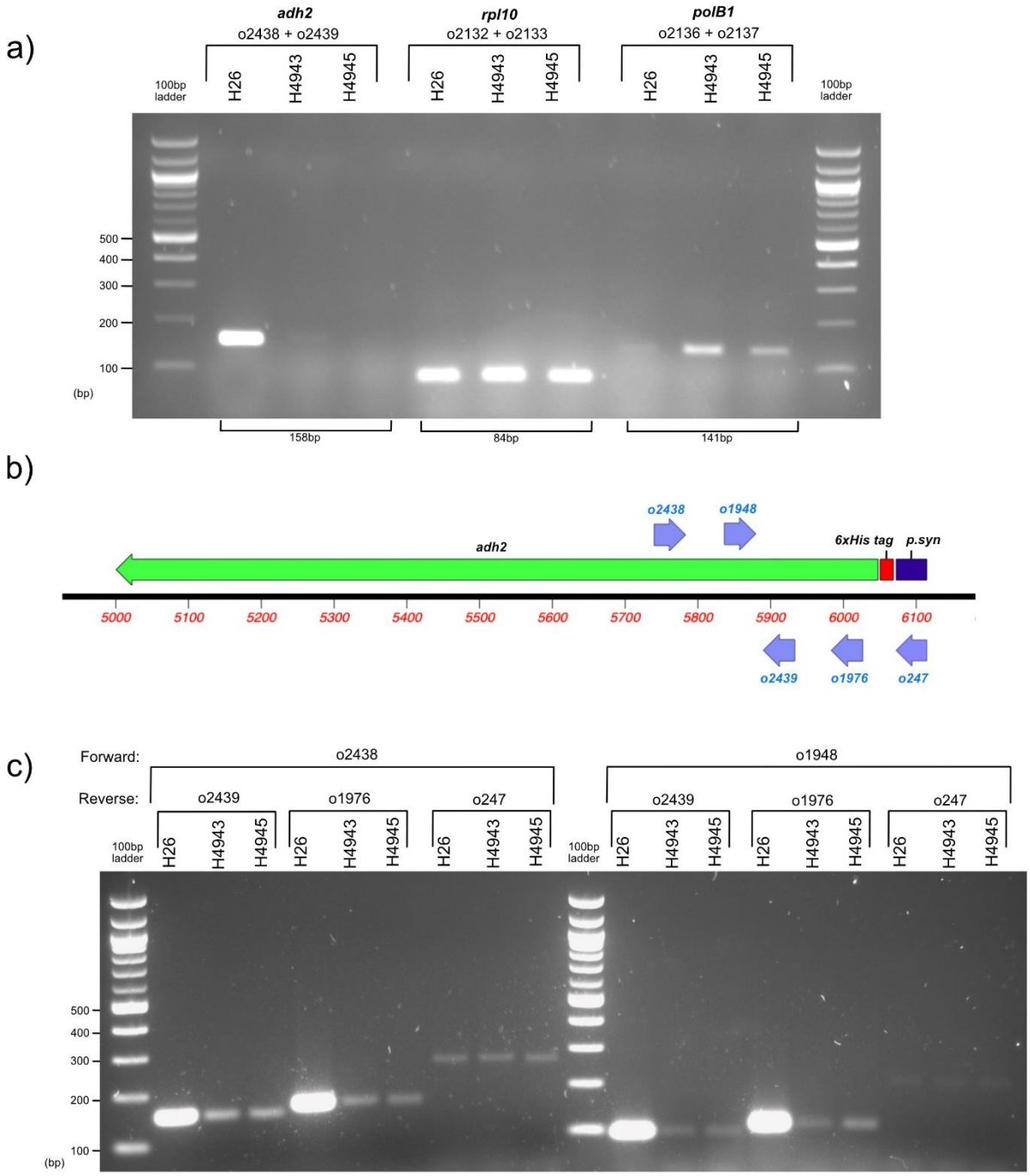


Figure 4.4: PCR amplification of cDNA from wild-type strains and strains thought to contain the *p.syn::adh2*+ construct. **a**) PCR amplification of *adh2*, *rpl10*, and *polB1* cDNA run on an agarose gel. The band seen for amplification of H26 (wild-type) *adh2* cDNA is brighter than that seen in H4943 and H4945 (both supposedly *p.syn::adh2*+). *rpl10* and *polB1* formed controls. **b**) Location of oligonucleotide binding within the *adh2* gene. **c**) PCR amplification of *adh2* cDNA using all six possible combinations of the two forward and three reverse primers seen in b) run on an agarose gel. For all four combinations of oligonucleotides binding within the *adh2* gene, bands corresponding amplification of H26 cDNA are brighter than those for H4943 and H4945 cDNA. Amplification with o247 demonstrates DNA contamination as o247 binds within *p.syn* which has not been engineered into H26 and as *p.syn* is a promotor, would not be expected to be found in mRNA transcripts.

In Figure 4.4a, the band corresponding to PCR amplification of H26 *adh2* cDNA is much brighter than the extremely faint bands corresponding to H4943 and H4945 *adh2* cDNA. Bands corresponding to PCR amplification of *rpl10* cDNA are of equal brightness in each of the three strains. Bands corresponding to *polB1* cDNA are of varying brightness but are all brighter than those seen for *adh2* cDNA in H4943 and H4945. All bands seen were of the expected size.

Further PCR amplification of *adh2* cDNA was undertaken using all possible combinations of the oligonucleotides shown in Figure 4.4b. PCR products were again run on an agarose gel (Figure 4.4). For all four of the possible combinations of the four oligonucleotides binding within the *adh2* gene (forward: o2438 and o1948, reverse: o2439 and o1976) and therefore within the cDNA, the bands seen for H26 is brighter than the band seen for H4943 and H4945. Oligonucleotide o247 binds within *p.syn* and bands seen for PCR amplification using this create faint bands of equal brightness for each of the strains. *p.syn* has not been engineered into H26 and is a promotor, therefore, amplification cDNA using o247 is indicative of DNA contamination.

Overall, the results shown in Figure 4.4 demonstrate that the level of *adh2* transcription in H4943 and H4945, strains expected to express *adh2* at an extremely high level which was thought to permit deletion of oripHV3, have a significantly reduced levels of *adh2* transcription compared to the wild-type.

The results of these experiments are revisited later in this chapter.

4.3.2- Integration of pHV3 onto the main chromosome

Integration of pHV3 onto the main chromosome in H4943 and H4945 lineages

Genome instability resulting from high levels of transcription is well documented (see Section 1.5- The role of transcription and R-loops in genomic instability). It was hypothesised that a possible explanation for the low level of *adh2* expression could be genomic rearrangement, such as integration of pHV3 onto the main chromosome, as a result of the high level of transcription initiated at *p.syn*.

A pulsed field gel followed by Southern blotting was carried out on genomic DNA from H4943, H4945, and all strains in their lineages leading back to the lab wild-type strain H26 (Figure 4.6). H2461 was used as a positive control as pHV3 had purposely been integrated onto the main chromosome, through recombination between mutant *leuB* genes in this strain prior to this study (Hannah Marriott, unpublished data).

To create H2461, two *leuB* (necessary for leucine biosynthesis) alleles were created: one missing the 3' end of the gene (*leuB△3'Bb*) and the other missing the 5' end (*leuB△5'Dd*). Both of these mutations render the protein non-functional. Two different strains were created; in one, *leuB△5'Dd* replaced *adh2* on pHV3 and in the other, *leuB△3'Bb* replaced wild-type *leuB* gene on the main chromosome. These strains were mated and cells capable of leucine biosynthesis were selected for. The only way in which a functional *leuB* gene can be created is through recombination between the two mutant *leuB* alleles, leading to forced integration of pHV3 onto the main chromosome. When H2461 was created, it was decided that two strains would be made, each with a different mutant *leuB* allele, which would then be mated instead of two sequential transformations with the same strain. For example, if a plasmid carrying the *leuB△5'Dd* allele was transformed into a strain carrying the chromosomal *leuB△3'Bb* allele, recombination could have occurred between the *leuB△5'Dd* allele on the plasmid and the *leuB△3'Bb* allele on the main chromosome, instead of recombination between the plasmid and pHV3 to insert *leuB△5'Dd* at the *adh2* locus. This could have resulted in a fully functional form of the *leuB* gene on the main chromosome without integration of pHV3.

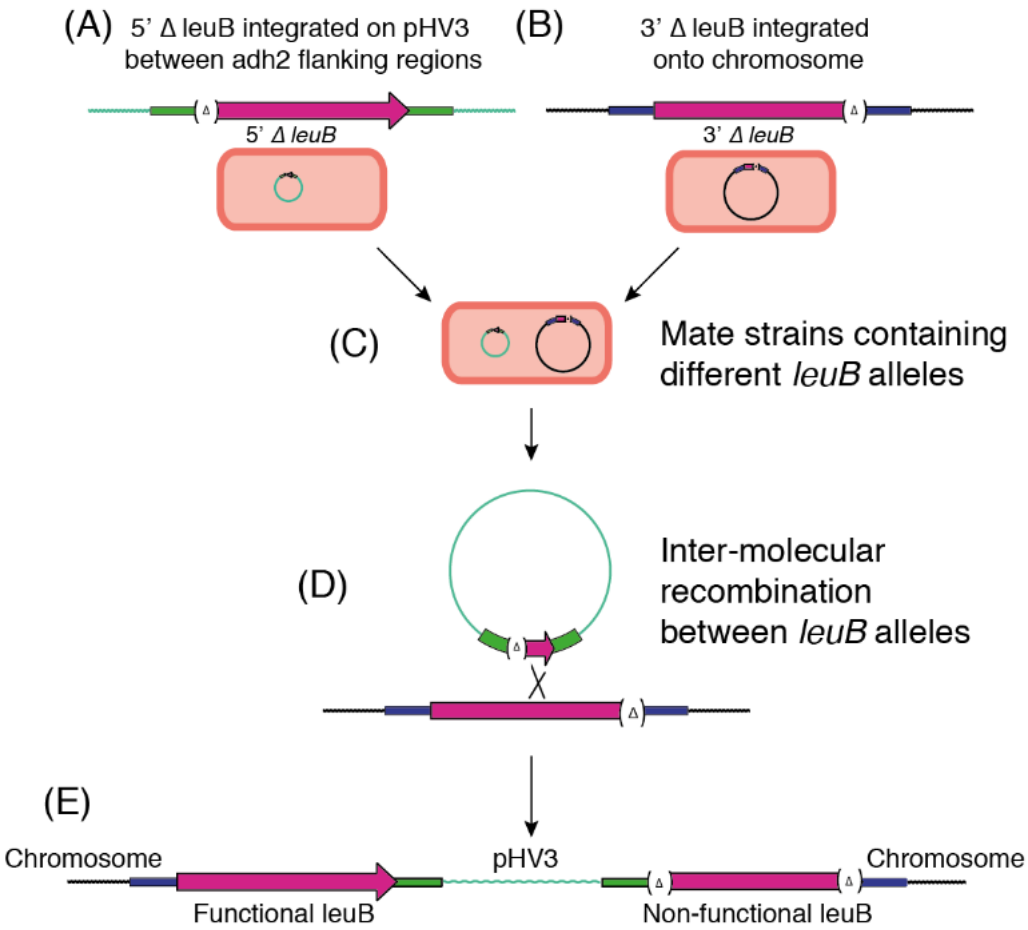
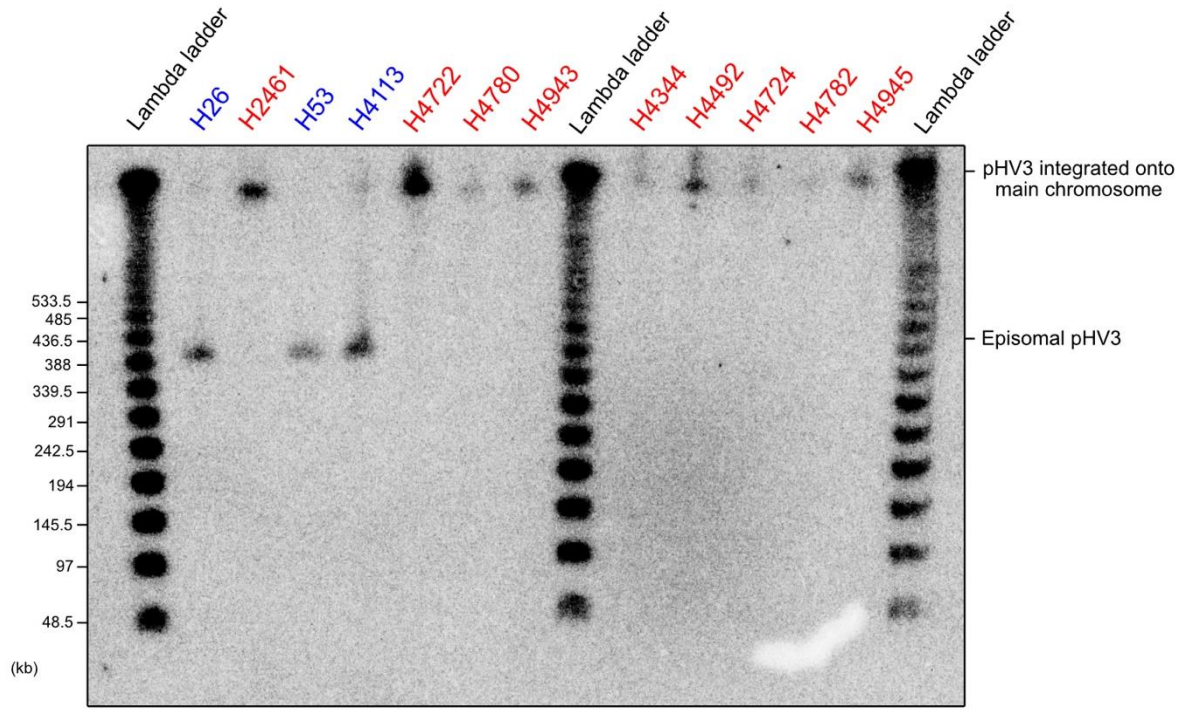


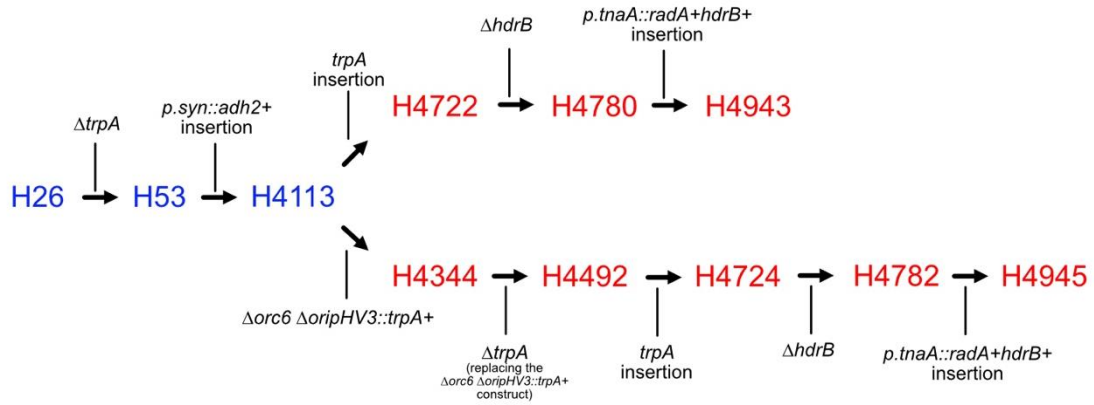
Figure 4.5: Forced integration of pHV3 onto the main chromosome through recombination between mutant *leuB* alleles. A) A strain was created containing the *leuBΔ5'Dd* allele on pHV3. B) A strain was created containing the *leuBΔ3'Bb* allele on the main chromosome. C) The strains from A) and B) were mated. D0 Intermolecular recombination between pHV3 and the main chromosome which creates E) a strain (H2461) with pHV3 integrated onto the main chromosome flanked by a functional copy of the *leuB* gene and a copy of the gene missing both the 5' and 3' ends.

Figure created by Hannah Marriot[208].

a)



b)



H2461: Forced pHV3 integration at the *leuB* locus (control)

c)

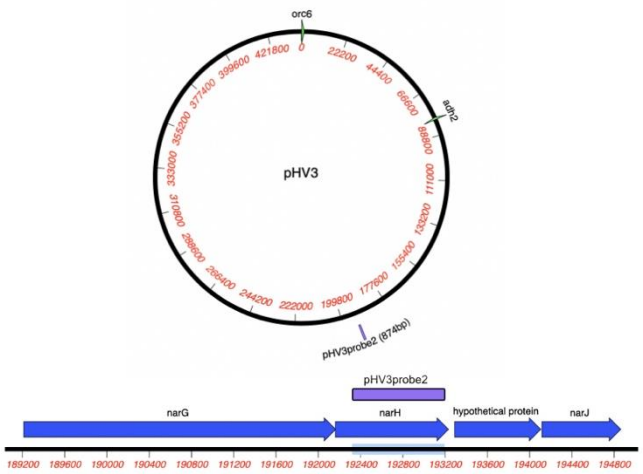


Figure 4.6: Pulsed field gel electrophoresis and Southern blotting demonstrating integration of pHV3 onto the main chromosome in the lineages of strains H4943 and H4945. a) A Southern blot demonstrating integration of pHV3 onto the main chromosome in strains from which H4943 and H4945 descended, as shown in b). pHV3 was forced onto the main chromosome in H2461 and it therefore acts as a control. Blue and red strain numbers indicate those in which pHV3 remains as an episome and is found integrated, respectively. **b**) The supposed genetic modifications occurring in each strain's formation, culminating in H4943 and H4945. Further investigation detailed in this chapter revisits the genotype of strain H4722 and therefore the strains created from it. **c**) The 847bp Southern hybridisation probe used in a) was made via PCR using oligonucleotides o1891 and o1892, with H26 genomic DNA as the template.

Both H4943 and H4945 originated from H4113. After H4113, the lineages of H4943 and H4945 diverge. The Southern blot of the pulsed field gel (Figure 4.6a) shows that in H26, H53, and H4113, pHV3 remains as an episome; the pHV3 probe hybridises to ~400kb and pHV3 is 438kb in size. When loading the pulsed field gel, plugs with DNA from each of the strains are stuck onto the comb prior to pouring the gel but the ladder is loaded directly into the well once the comb has been removed. This leads to a slight discrepancy the ladder and the plug DNA, and accounts for the ~38kb difference between where the pHV3 probe should hybridise on the Southern blot (438kb) compared to the binding seen (~400kb). In every successive strain after H4113 culminating in strains H4943 and H4945 (portrayed in red in Figure 4.6a and b), pHV3 is seen integrated onto the main chromosome. The pHV3 probe hybridises higher up on the blot, where the main chromosome runs (the wild-type main chromosome is 2.8 Mb and when pHV3 is integrated, it becomes 3.2Mb). This indicates that two independent integration events occurred when genetically manipulating strain H4113: one when seemingly replacing the previously deleted *trpA* gene to create H4722, and another when deleting the origin and corresponding Orc initiator protein on pHV3 in H4344.

Previous work has already shown that oripHV3 deletion is possible when pHV3 is integrated onto the main chromosome (Hannah Marriot, unpublished data). Therefore, it can no longer be said that oripHV3 deletion is made possible through insertion of p.syn leading to increased transcription, unless significant genomic rearrangements occur. Additionally, previous work in this chapter into the role of RadA in replication initiation from R-loops on pHV3 was predicated upon pHV3 remaining as an episome, but in the strains used within this analysis, pHV3 is found integrated onto the main chromosome.

Later in this chapter, the genotype of H4722 was further interrogated and found to not be as expected. However, when undertaking these experiments, it appeared that integration of pHV3 onto the main chromosome had occurred twice independently: when deleting *oripHV3* and *orc6*, and after a linear transformation to re-introduce the *trpA* gene. However, the driving force of these integrations remained undetermined. In one case, pHV3 may have integrated as a result of deleting *oripHV3* and *orc6*,

and this mini-chromosome (which contains essential genes) cannot be maintained without its origin. This, however, does not explain integration when re-introducing *trpA*. The possibility that the dramatic increase in transcription resulting from *p.syn* causes the genome to become too unstable to sustain any further genetic modifications without significant genomic rearrangements is necessary to consider.

To investigate this, pulsed field gel electrophoresis was carried out on other strains resulting from transformations of H4113 (Figure 4.7), all constructed by Laura Mitchell prior to this study. Their genotypes can be seen in Table 4.2 and lineages in Figure 4.7b. Investigating the frequency of pHV3 integration onto the main chromosome provides insight into whether *p.syn* drives integration generally, or instead permits integration upon deletion of *oripHV3*. Integration of pHV3 onto the main chromosome had not previously been observed in the lab when deleting *oripHV3*. Therefore, it is possible that genomic instability due to the high level of transcription initiated at *p.syn* allows pHV3's integration onto the main chromosome upon the stress experienced by the cell when trying to delete *oripHV3*.

Table 4.2: Strains generated from H4113 (Δ*pyrE2,* Δ*trpA, p.syn::adh2+*) examined through pulsed field gel electrophoresis seen in Figure 4.7. All created prior to this study.

Strain	Genotype	Notes
H4318	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* *oripHV3+::[*Δ*oripHV3::trpA+ pyrE2+]*	Pop-in strain.
H4396	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* Δ*oripHV3::trpA+*	Pop-out strain from H4318. Plasmid pTA1801 did not pop-out successfully. Clone 1.
H4397	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* Δ*oripHV3::trpA+*	Pop-out strain from H4318. Plasmid pTA1801 did not pop-out successfully. Clone 2.
H4398	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* Δ*oripHV3::trpA+*	Pop-out strain from H4318. Plasmid pTA1801 did not pop-out successfully. Clone 3.
H4319	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* *orc6+::[*Δ*orc6::trpA+ pyrE2+]*	Pop-in strain.
H4320	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* *orc6* *oripHV3+::[*Δ*orc6* Δ*oripHV3::trpA+* *pyrE2+]*	Pop-in strain.
H4344	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* Δ*orc6* Δ*oripHV3::trpA+*	Pop-out strain from H4320. Clone 1.
H4345	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* Δ*orc6* Δ*oripHV3::trpA+*	Pop-out strain from H4320. Clone 2.
H4346	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* Δ*orc6* Δ*oripHV3::trpA+*	Pop-out strain from H4320. Clone 3.
H4347	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* Δ*orc6* Δ*oripHV3::trpA+*	Pop-out strain from H4320. Clone 4.
H4722*	Δ*pyrE2* *p.syn::adh2+*	Linear transformation to re-introduce *trpA*. Clone 1.
H4723	Δ*pyrE2* *p.syn::adh2+*	Linear transformation to re-introduce *trpA*. Clone 2.
H4462	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* *radA+::[*Δ*radA::trpA+ pyrE2+]*	Pop-in strain.
H4481	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* *radA+::[*Δ*radA::trpA+ pyrE2+]* *{radA+ pyrE2+ MevR}*	Pop-in strain.
H4482	Δ*pyrE2* Δ*trpA* *p.syn::adh2+* *radA+::[*Δ*radA::trpA+ pyrE2+]* *{radA+ pyrE2+ MevR}*	Pop-in strain.

*Upon further analysis later in this chapter, the genotype of this strain was found to not be as stated in this table. The table shows what was presumed to be the genotype at the time the experiment was undertaken.

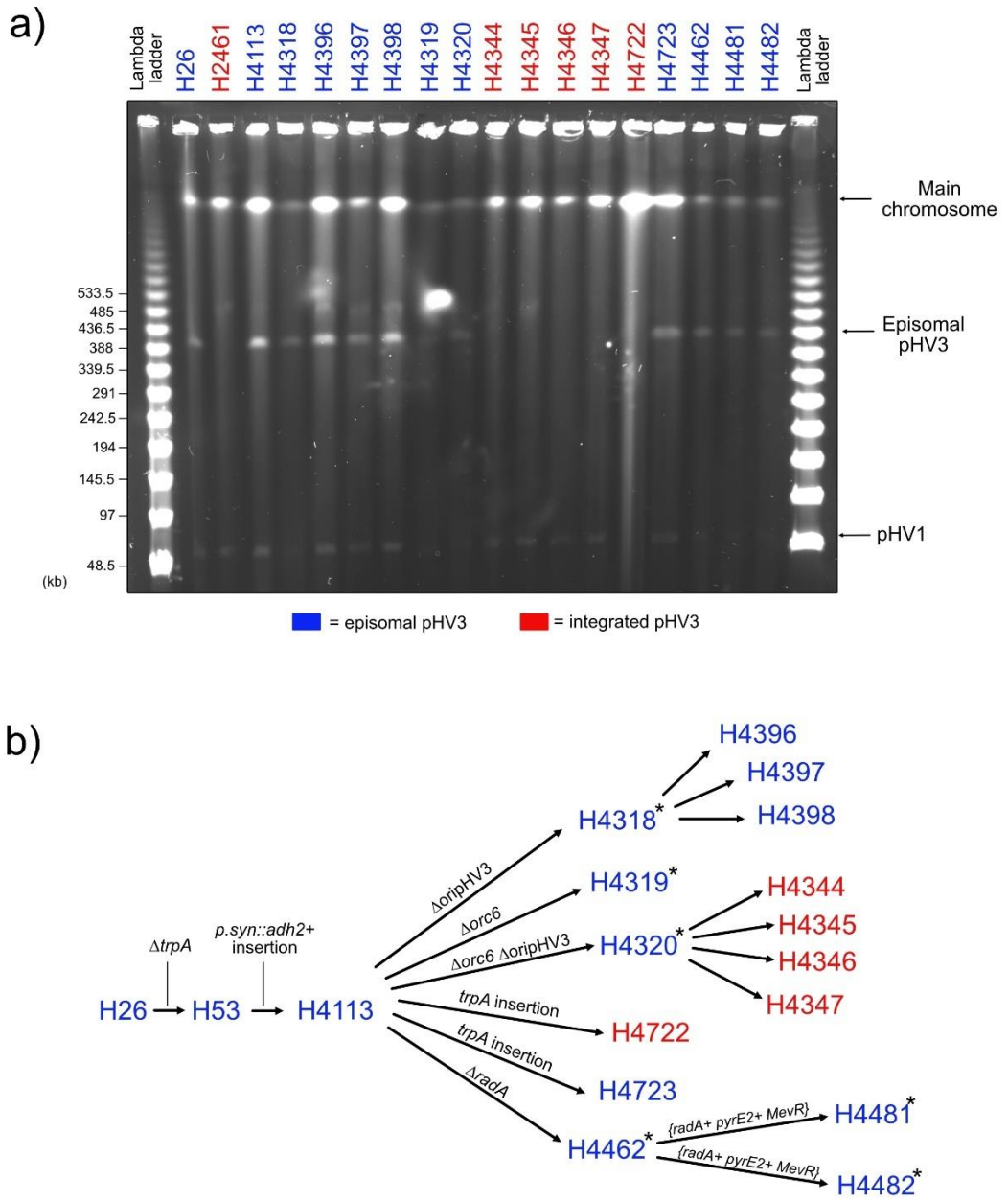


Figure 4.7: Pulsed field gel electrophoresis to investigate the status of pHV3 in strains generated after *p.syn* insertion. a) A pulsed field gel electrophoresis agarose gel showing the absence of a band corresponding to pHV3 in H2461 (a control strain in which pHV3 had previously been forced to integrate onto the main chromosome), H4344, H4345, H4346 and H4347. The H4722 lane contains a smear and so the status of pHV3 cannot be determined from this gel. However, the Southern blot in Figure 4.6a previously showed integration of pHV3 onto the main chromosome in this strain. A band corresponding to episomal pHV3 can be seen in all other strains, including H26, the wild-type control. **b**) A flow diagram demonstrating the lineages of strains from which DNA was extracted and used in a). As in a), blue denotes strains in which pHV3 remains as an episome and red denotes strains in which pHV3 has integrated onto the main chromosome. Strains with * are pop-in strains. Strains created from H4318 and H4320 (pop-in strains) are pop-out strains. {} denotes a plasmid. The plasmid within H4482 and H4483 remains as an episome.

In Figure 4.7a, a band can be seen at the size expected for pHV3 in H26 (wild-type) and H4113, and cannot be seen for H2461 in which pHV3 is known to be integrated onto the main chromosome. These strains form controls. Bands can be seen corresponding to episomal pHV3 in all strains investigated apart from H4344, H4345, H4346, H4347, and H4722. H4344, H4345, H4346 and H4347 are all clones created through the pop-out stage with pop-in strain H4320. A smear can be seen in the lane for strain H4722 meaning that the status of pHV3 cannot be determined from this gel. However, in the Southern blot from the previous pulsed field gel in this chapter (Figure 4.6), pHV3 was found to be integrated onto the main chromosome in this strain. Southern blotting with a probe hybridising to pHV3 was attempted several times but the vacuum transfer of the DNA within the gel to the membrane was not effective for sufficient pHV3 detection. Due to time constraints, it was not possible for the pulsed field gel to be run again and Southern blotting to be repeated. However, later in this chapter, Oxford Nanopore MinION sequencing data confirms that in all strains that do not have a band corresponding to pHV3 in this gel, pHV3 is found integrated onto the main chromosome.

In strains H4396, H4397 and H4398, *oripHV3* deletion was attempted but the plasmid used in the transformation was found not to have popped out (Laura Mitchell, unpublished data). pHV3 remains episomal in these strains. The only other strains in which *oripHV3* deletion was attempted on the gel are H4344 – H4347. In all of these strains, pHV3 is found integrated onto the main chromosome but deletion of *oripHV3* was successful. None of the other strains on the pulsed field gel were as a result of transformations to delete *oripHV3* and in all of them (excluding H4722), pHV3 remains as an episome.

Integration of pHV3 onto the main chromosome in H4722 was later found to have occurred independently of *p.syn*. H4722 was not created in a transformation with H4113 or any other strain containing the *p.syn* construct. Therefore, this strain should not be taken into consideration when investigating the role of *p.syn* in pHV3's integration onto the main chromosome.

Therefore, the only strains created directly from H4113, in which pHV3 is seen integrated onto the main chromosome, are those in which *oripHV3* and *orc6* deletions were made (Figure 4.8).

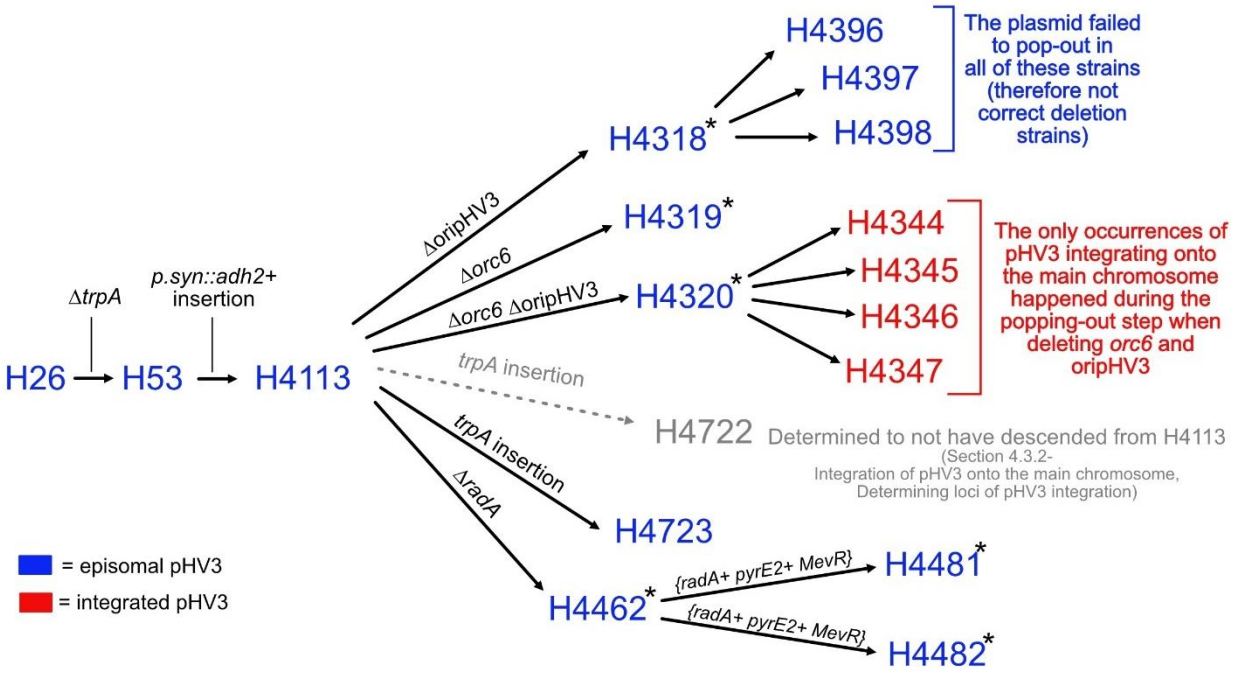


Figure 4.8: The only instances observed of pHV3 integrated onto the main chromosome all occurred upon the deletion of *orc6* and *oripHV3*. Integration occurred during the popping out step as pHV3 is found as an episome in H4320 (the pop-in strain and pre-cursor to H4344, H4345, H4346, and H4347. Later in this study, it was found that strain H4722 was not derived from H4113 and is therefore denoted in grey in this flow diagram. As in Figure 4.7, strains with * are pop-in strains and {} denotes a plasmid. The plasmid within H4482 and H4483 remains as an episome.

Given these findings, is seems probable that upon deletion of *oripHV3*, the presence of *p.syn* permitted integration of pHV3 onto the main chromosome and therefore, deletion of the origin. In previous attempts to delete *oripHV3*, when *p.syn* had not been inserted onto pHV3, integration of pHV3 onto the main chromosome was never observed. Instead, pop-out colonies were either not obtained, or all pop-out colonies retained the origin (Laura Mitchell and Hannah Marriott, unpublished data). In all other strains tested that were generated directly from H4113, pHV3 remained as an episome. All of these strains contain the *p.syn::adh2+* construct. This indicates that the presence of *p.syn* alone does not lead to such genomic instability that pHV3 is forced to integrate onto the main chromosome.

Determining loci of pHV3 integration

Large scale genomic rearrangements have previously been observed in *H. volcanii* [55], including the integration of pHV4 onto the main chromosome[49]. The locus of pHV4's integration is a *transposase (ISH18)* sequence found on both the main chromosome and pHV4. The paper concluded that recombination occurred between the two ISH19 elements resulting in large scale rearrangements of the genomic architecture.

To determine the loci of pHV3 integration onto the main chromosome, gDNA was extracted from strains shown in Table 4.3, and Oxford Nanopore MinION sequencing was undertaken to locate the loci of integration of pHV3 upon the main chromosome. H2461 was constructed prior to this study (described earlier in this chapter) and it was expected that pHV3 would be found integrated at the *leuB* locus in this strain. H4113, H4396 and H4397 formed controls as pulsed field gel electrophoresis had already demonstrated that pHV3 remained as an episome in these strains. However, it was of interest to investigate the region of the genome in which the plasmid used in the transformation to make H4396 and H4397 had failed to pop-out. Strains H4344, H4345, H4346 and H4347 were created upon *oripHV3 orc6* deletion and the locus of integration was unknown. At the time, H4722 was thought to have been created in a linear transformation to re-introduce the *trpA* gene and the locus of integration was again unknown in this strain.

MinION sequencing is unrestricted in the length of reads generated up to 4Mb[275] and is therefore ideal for *de novo* genome assembly. Long reads can be used to generate highly contiguous genome maps and can identify genomic rearrangements more easily than short read sequencing. For example, in the extension of the *Caenorhabditis elegans* reference genome, mean flow cell read lengths ranged from ~13 kb to ~20 kb and two complex rearrangements were noted in a wild-type strain[276]. Long reads gained through MinION sequencing would confirm integration of pHV3 onto the main chromosome and determine the loci at which the integration event occurred.

Fastq read files were uploaded to MacVector and consensus sequences were assembled with Flye (Expected genome size= 4Mb, Threads= 3, Flye polishing iterations= 1, Minimum overlap between reads= 1,000bp, Initial minimum coverage= 100). Pustell DNA matrices were created in MacVector (Window size= 30, Minimum % Score= 80%, Hash Value= 8) to locate homology between consensus sequences and the H26 reference genome where pHV4 is integrated onto the main chromosome (from [49]). The portion of the consensus sequence mapping to the non-contiguous regions of the Pustell matrices created were taken and plotted against pHV3 in a further Pustell DNA matrix using the same conditions. The locus upon the main chromosome at which pHV3 had integrated was determined using the initial Pustell DNA matrix and the locus upon pHV3 located using the second, as seen in

Figure 7.5 - Figure 7.13 (Appendices). The results of this analysis are shown in Table 4.3. For the strains in which pHV3 is found integrated onto the main chromosome, regions spanning the junctions between pHV3 and the main chromosome were auto-annotated in MacVector using sequence data and annotation from [48] (Minimum feature length= 10, Residues around point feature= 30, Maximum allowed mismatches= 1%, Maximum allowed gaps= 0.5%). The annotated sequences provide a visual representation of the regions of integration (Figure 7.14, Appendices).

Table 4.3: Loci of integration of pHV3 onto the main chromosome found through Oxford Nanopore MinION sequencing and analysis.

Strain	Locus of integration on the H26 main chromosome	Locus of integration on H26 pHV3
H2461	*leuB* (HVO_1502)	*adh2* (replaced with *leuB* in the creation of this strain) (HVO_B0071)
H4113	pHV3 not integrated	pHV3 not integrated
H4344	*transposase* (*ISH5*) (HVO_1835)	*transposase* (*ISH5*) (HVO_B0157)
H4345	*transposase* (*ISH5*) (HVO_1835)	*transposase* (*ISH5*) (HVO_B0157)
H4346	*transposase* (*ISH5*) (HVO_1835)	*transposase* (*ISH5*) (HVO_B0157)
H4347	*transposase* (*ISH5*) (HVO_1835)	*transposase* (*ISH5*) (HVO_B0157)
H4396	pHV3 not integrated	pHV3 not integrated
H4397	pHV3 not integrated	pHV3 not integrated
H4722	*leuB* (HVO_1502)	*adh2* (HVO_B0071)

In H2461, pHV3 is seen integrated at the *leuB* locus as anticipated. When the consensus sequences were compared to the H26 genome and the Pustell DNA matrices created, the locus identified on pHV3 was confirmed to be *adh2*. In the generation of this strain, *adh2* was replaced with a form of the *leuB* allele, as described earlier in this chapter, and this is therefore expected. For strains H4133, H4396 and H4397, two individual contigs were found, each mapping contiguously to the H26 main chromosome and pHV3, meaning that pHV3 is not integrated onto the main chromosome in these strains, again as expected.

In strains H4344, H4345, H4346 and H4347, the loci of integration on the main chromosome and on pHV3 share 96.17% sequence identity in an NCBI BLAST alignment with 98% query cover and E-value of 0 [277]. The locus identified on the main chromosome was a 1353bp *transposase* (*ISH5*) gene (HVO_1835) and on pHV3 was a 1353bp *transposase* (*ISH5*) gene (HVO_B0157) in all four strains. Each

of these strains is a clone derived from the pop-in strain H4320. Pulsed field gel electrophoresis, described earlier in this chapter, found that pHV3 remains as an episome in H4320. This indicates that integration onto the main chromosome occurred four times independently at these loci upon *oripHV3 orc6* deletion.

Regions ± 500bp of the *transposase* (*ISH5*) gene (HVO_1835) on the main chromosome and the *transposase* (*ISH5*) gene (HVO_B0157) on pHV3 were taken and the homology between them was investigated using a Pustell DNA matrix with the same conditions as previously (Figure 4.9). The matrix shows homology between the genes but no homology between the upstream and downstream regions. As in the case of pHV4's integration onto the main chromosome[49], it is likely that the homology between *transposase* genes allowed recombination between the main chromosome and pHV3, permitting their fusion.

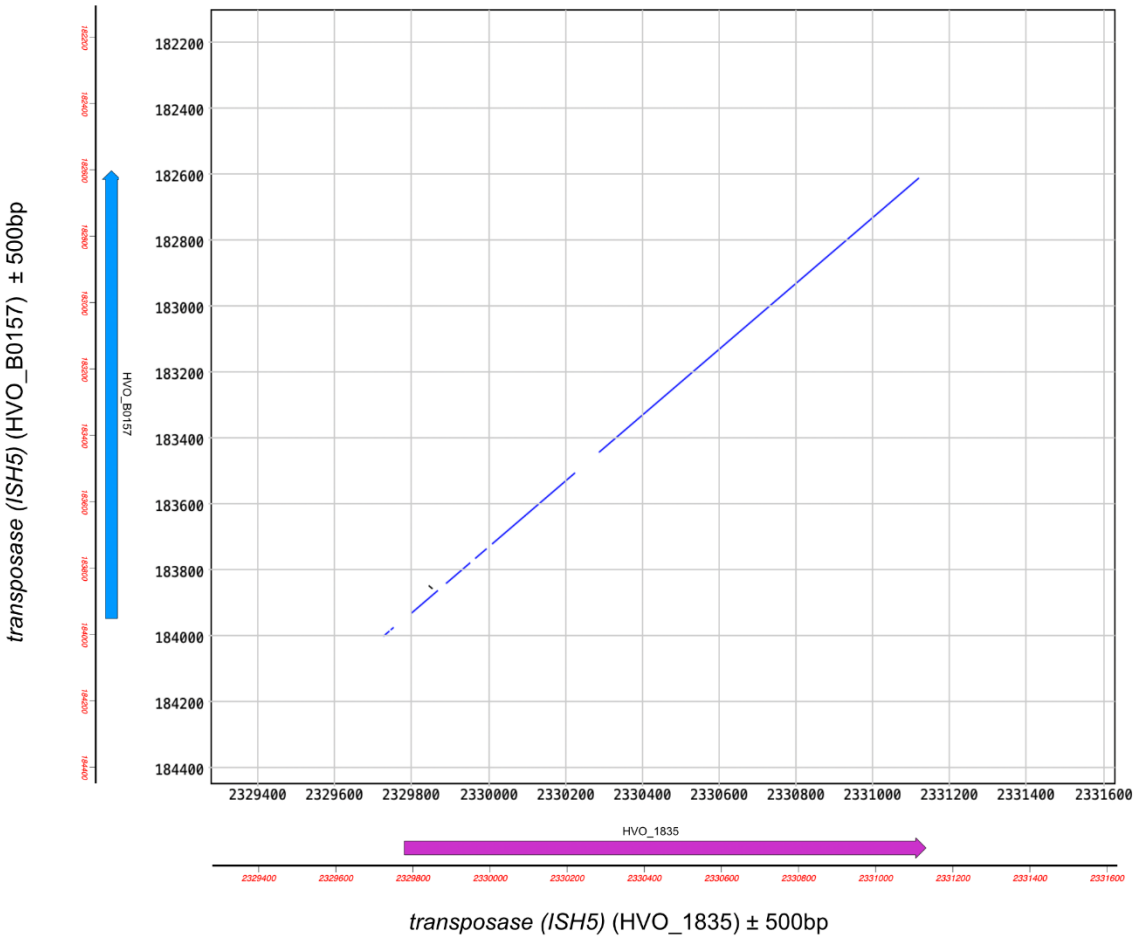


Figure 4.9: Homology between transposase (ISH5) genes HVO_1835 and HVO_B0157 on the main chromosome and pHV3 respectively. A Pustell DNA matrix was created to determine homology between HVO_1835 and HVO_B0157- the loci at which integration of pHV3 onto the main chromosome in strains H4344 - H4347 can be seen. Homology exists between the genes but does not exist between the regions upstream and downstream.

In strain H4722, integration of pHV3 onto the main chromosome occurred at the *leuB* locus on the main chromosome and the *adh2* locus on pHV3. In wild-type strains, *leuB* and *adh2* loci do not share homology (NCBI blastn finds no significant homology[277]). Additionally, when assembling contigs for this strain, a contig corresponding to the mini chromosome pHV1 was not present. However, there exist other strains in the laboratory that have this genotype and are derivatives of strain H2461 (the construction of which is detailed earlier in this chapter). In these strains, *adh2* gene was replaced with a version of the *leuB* allele prior to forced integration of pHV3 onto the main chromosome and pHV1 had been cured.

These findings were unexpected and required further analysis to determine which strain H4722 could be. As native *leuB* and *adh2* loci do not share homology to permit pHV3's integration onto the main chromosome and random curing of pHV1 has not been noted previously, it is most likely that this strain is a product of other work in the laboratory. The question remained as to whether the incorrect strain had been streaked out upon commencement of this study, or whether upon the linear transformation of H4113 to re-introduce the *trpA* gene which should have resulted in strain H4722 with the genotype it was initially thought to have, a glycerol stock of the incorrect strain was frozen.

In the initial pulsed field gel electrophoresis and Southern blot in this chapter (Figure 4.6), the lineages of strains H4943 and H4945 were examined. Strain H4722 gave rise to H4780 which in turn gave rise to H4943. In Figure 4.6, pHV3 is seen integrated onto the main chromosome in all three of these strains. If the incorrect strain had been streaked out at the start of the study and the H4722 used to create H4780 had the expected genotype, pHV3 must have integrated onto the main chromosome in an independent event- one not explained by forced integration between *leuB* loci. If the incorrect strain was frozen from the outset, all strains made using H4722 after this point (including H4780 and H4943) would contain pHV3 integrated onto the main chromosome and would not contain pHV1.

All strains included in this chapter were generated prior to this study.

Investigation into strain H4722

To confirm that pHV1 was not present in strain H4722, colony hybridisation with a pHV1 probe was undertaken (Figure 4.10). Strains H26 (wild-type) and H2184 (Δ*{pHV1}*) were used as controls.

In Figure 4.10a, the pHV1 probe hybridised to H26 but did not hybridise to H2184, as expected. The probe did not hybridise to H4722, confirming that pHV1 is not present in this strain, as shown through Oxford Nanopore MinION sequencing data analysis previously.

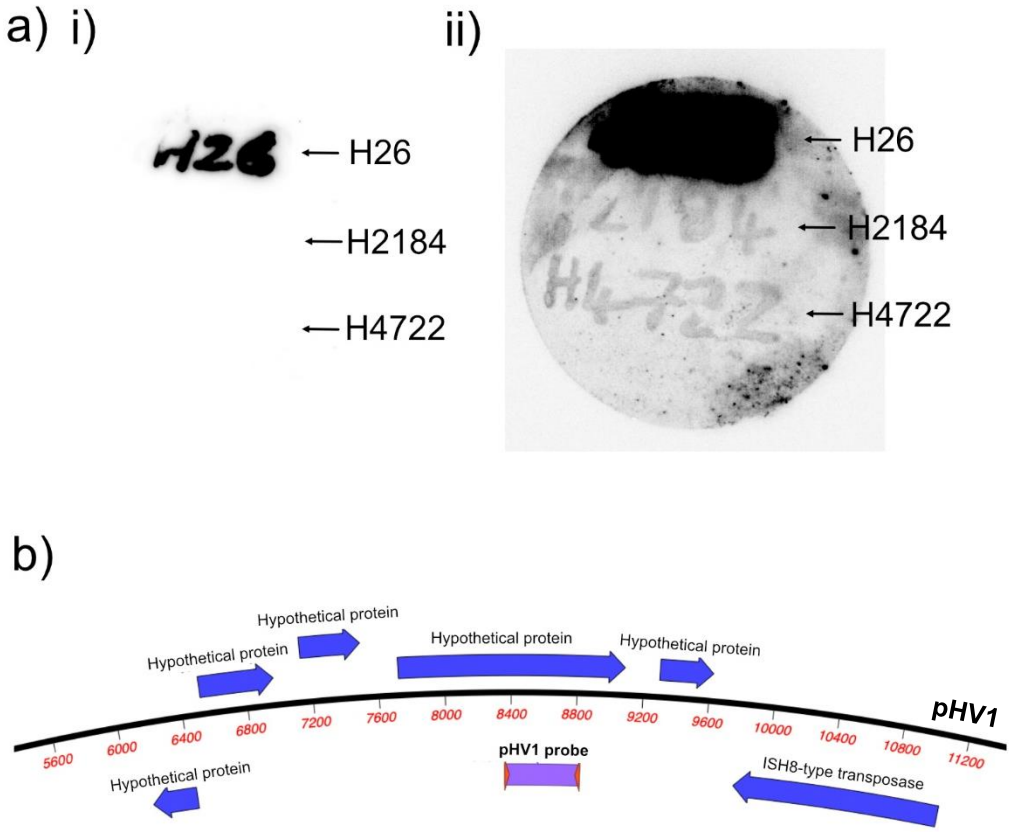


Figure 4.10: Colony hybridisation to determine presence/absence of pHV1 in H4722. **a**) Colony hybridisation using a pHV1 probe. Images i and ii are of the same membrane with different exposures. i) Hybridisation of the probe to H26 (wild-type) demonstrates that pHV1 is present. ii) The probe had not hybridised to H2481 (Δ*{pHV1}*) nor H4722 demonstrating that pHV1 is not present in these strains. **b**) PCR amplification of H26 gDNA using oligonucleotides o1827 and o1828 created the 471bp pHV1 probe used in a) shown in purple.

To determine whether the incorrect strain had been streaked out from the -80°c freezer at the start of the study or whether glycerol stock of the wrong strains had been frozen upon their creation, colony PCR to amplify a region of pHV1 was undertaken on the H4722 cells taken from the same plate as used for the previous work in this chapter, and on fresh plates of H4722 streaked out from both the normal and back-up -80°c freezers. The oligonucleotides used were the same as those used to create the pHV1 probe (o1827 and o1828) used in Figure 4.10. Figure 4.11 demonstrates the results.

In Figure 4.11, the band seen for H26 (wild-type) is at the expected size. A faint band is seen for H2184 and H4722* (cells taken from the plate used in the prior work in this chapter). This is indicative of contamination as the colony hybridisation in Figure 4.10 previously demonstrated the loss of pHV1 in these strains. Amplification seen for H4722** and H4722*** (freshly streaked colonies from both -80°c freezers) mirrors that seen in H2184 and H4722* and is fainter than that seen for H26. This indicates that pHV1 is not present in the glycerol stocks for these strains and prior to this study, the wrong strains were frozen.

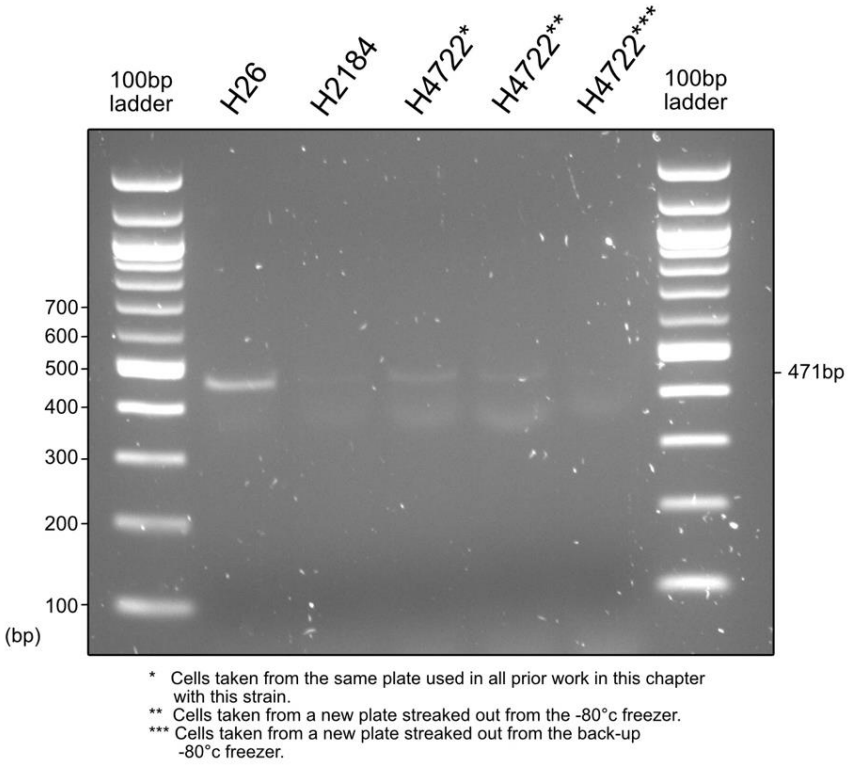


Figure 4.11: Colony PCR to amplify a region of pHV1. Oligonucleotides o1827 and o1828 were used to amplify a region of pHV1. The band corresponding to amplification of wild-type DNA is the correct size. Faint bands can be seen in all other lanes. It has been shown previously in this chapter that pHV1 is not present in strains H2184 and H4722* and amplification of DNA from these strains mirrors that for H4722** and H4722*** indicating that pHV1 is not present in these strains either.

To determine which other strain in the strain database may have mistakenly been frozen in place of the strain with the genotype H4722 was intended to have, the database was searched systematically using the information already known about H4722 and through searching for and aligning genes to the MinION sequencing contigs in MacVector (Align to Reference function) (these factors are shown in Table 4.4). It was known that H4722 contained pHV3 integrated at the *leuB* locus and that pHV1 was not present- modifications that are not known to happen spontaneously. Key genes to look for in the H4722 genome to narrow the search included *mrr*, *trpA, orc6* and *polB1*, and the origins of replication *oriC1* and *oripHV3*. Figure 7.15 (Appendices) shows alignments of genes to H4722.

Table 4.4: Elements of the H4722 genotype

Information discovered prior to aligning genes against the H4722 MinION sequencing contig	Information discovered from aligning sequences to the H4722 MinION sequencing contig
pHV3 integrated at the *leuB* locus	Δ*mrr*
ΔpHV1	Δ*oripHV3::trpA*+
	Wild-type *polB1*
	Wild-type *oriC1*
	Wild-type *orc6*

In brief, the *mrr* gene had been deleted and *trpA* was found in place of *oripHV3*, next to *orc6* (where the origin is natively found). When deletion of *oripHV3* is undertaken in the laboratory, it is replaced with a *trpA* to select for the loss of the origin. *polB1* and *oriC1* both remained in the genome as in the wild-type strain. No strains were identified when the database was searched using these criteria. Therefore, this is not a strain with a genotype documented by the laboratory.

Revisiting low levels of *adh2* transcription in H4943 and H4945

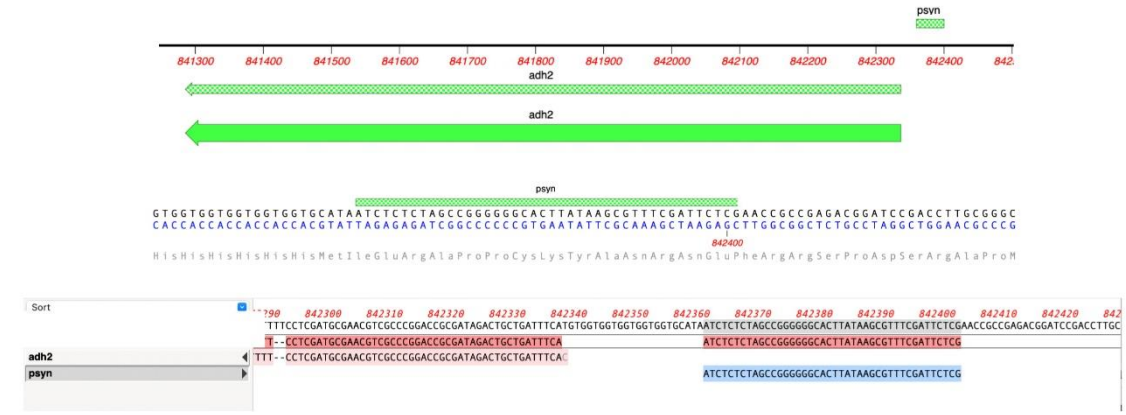
Earlier in this chapter, RT-PCR data showed unexpectedly low levels of transcription of *adh2,* despite its expression supposedly being driven by *p.syn*. However, work done in this chapter after undertaking RT-PCR provides some level of explanation.

Figure 4.6 demonstrated the lineages of strains H4943 and H4945- the strains in which low levels of *adh2* cDNA was shown in Figure 4.4 despite the supposed *p.syn::adh2*+ construct within their genomes. H4943 is descended from H4722 and H4945 is descended from H4344. As the entire genome sequences of strains H4344 and H4722 were obtained through Oxford Nanopore MinION sequencing, is was possible to align the sequence for the *p.syn::adh2*+ construct to them to determine whether this construct was present at this stage.

H4722 did not contain the *p.syn::adh2*+ construct. Instead, this sequence aligned to a gene encoding a formaldehyde dehydrogenase (Figure 4.12). The construct was inserted into H53 to create strain H4113. H4722 was thought to have been created from H4113 but as shown previously, the incorrect strain was frozen and therefore it is understandable that this strain does not contain the *p.syn::adh2*+ construct. Additionally, in strains in the laboratory in which pHV3 has been forced onto the main chromosome, the *adh2* gene is replaced by a modified version of the *leuB* allele and therefore is not present. It is for these reasons that the PCR product corresponding to amplification of *adh2* cDNA in strain H4943 is not as bright as the band for PCR amplification of H26 *adh2* cDNA (Figure 4.4). In fact, the band for H4943 is of equal brightness to the PCR products gained using o247 which should not have shown amplification and was indicative of contamination. This suggests that any amplification resulting from PCR of H4943 *adh2* cDNA also demonstrated contamination as the *adh2* gene is not present.

H4344 does contain the *p.syn::adh2*+ construct. *p.syn* is intact and is in the correct orientation to promote expression of *adh2* (Figure 4.12). This means that either the *p.syn::adh2*+ construct is still present in H4945, in which case high levels of *adh2* transcription should be observed through RT-PCR (which was not the case), the construct has been lost or damaged at some point in the lineage between H4344 and H4945, or something completely different is preventing *adh2* from being expressed at a high level such as abortive transcription. Due to time constraints, it was not possible to explore this further. Had this not have been the case, is would have been of interest to sequence the *p.syn::adh2*+ region of H4945 to determine whether the construct remained in the genome in order to explain the low level of *adh2* expression seen earlier in this chapter in RT-PCR.

adh2 and *p.syn* aligned to H4344:



p.syn::adh2+ aligned to H4722:

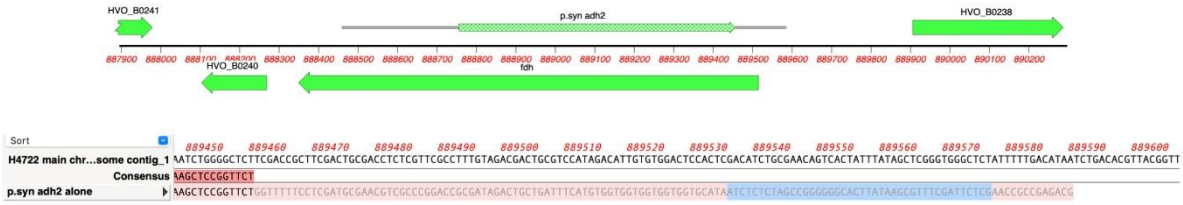


Figure 4.12: *p.syn* and *adh2* aligned to H4344 and H4722 contigs. Green checkerboard-filled arrows indicate sequences that have been aligned to the H4722 genome. Non-checkerboard green arrows are auto-annotated genes (detailed previously in this chapter). In H4344, *p.syn* and *adh2* map to the contig, and *p.syn* is intact and in the correct orientation to enable transcription of *adh2*. In H4722, the *p.syn::adh2* construct partially maps to a formaldehyde dehydrogenase gene (*fdh)*. The *p.syn* sequence does not align (grey lines in the figure indicate regions that do not align and contains *p.syn*). *p.syn* alone (not found as part of the *p.syn::adh2* construct) was also tested in this manner and did not align (not shown in figure).

4.4- Discussion

Initially, this chapter aimed to investigate the dependence upon RadA for replication initiation in the absence of origins on pHV3. It was thought that *oripHV3* deletion was made possible due to insertion of *p.syn* to drive *adh2* expression on pHV3, leading to increased levels of transcription upon the mini-chromosome. However, in the strains used for this analysis (H4943 and H4945), RT-PCR found that there was not an increase in *adh2* transcripts compared to wild-type. Pulsed field gel electrophoresis demonstrated the integration of pHV3 onto the main chromosome in strains in the lineages of H4943 and H4945, which had occurred during the genetic modification after insertion of the *p.syn::adh2* construct. Oxford Nanopore MinION sequencing revealed the loci of integration and also determined that strain H4722 did not have the genotype previously expected.

Overall, this chapter demonstrates the instability of the *H. volcanii* genome when deleting *oripHV3,* thought to be due to the high level of transcription initiated at *p.syn*. Previous attempts to delete *oripHV3* in the laboratory prior to this study had not noted pHV3 integration onto the main chromosome. Instead, deletion of *oripHV3* was not successful, and no colonies resulted after the pop-out stage, or as mentioned previously in the case of H4945, H4396 and H4397, the plasmid failed to pop-out completely (Laura Mitchell and Hannah Marriott, unpublished data). Through pulsed field gel electrophoresis and Oxford Nanopore MinION sequencing, this chapter discovered four independent integration events when deleting *oripHV3* and *orc6*, all of which occurred by recombination between the same *transposase* loci on pHV3 and the main chromosome.

It remains unknown whether the high level of transcription initiated at *p.syn* leads to small-scale genomic rearrangements in addition to the large-scale integrations seen (Figure 4.13). RT-PCR demonstrated the presence of fewer *adh2* transcripts in H4943 and H4945 compared to wild-type. This is consistent with MinION sequencing data for H4722 (from which H4943 is descended), as *adh2* is not present in H4722. Therefore, there would be no *adh2* transcripts in H4943 as the gene is not present in this strain. However, MinION sequencing data shows that in a strain in the lineage of H4945 (H4344), the *p.syn::adh2* construct is present and intact. It is possible that this construct was lost between H4344 and the creation of H4945, potentially as a result of instability resulting from increased transcription. Sequencing the *p.syn::adh2* region of H4945 would be of interest to provide an explanation for the lack of *adh2* transcripts seen through RT-PCR.

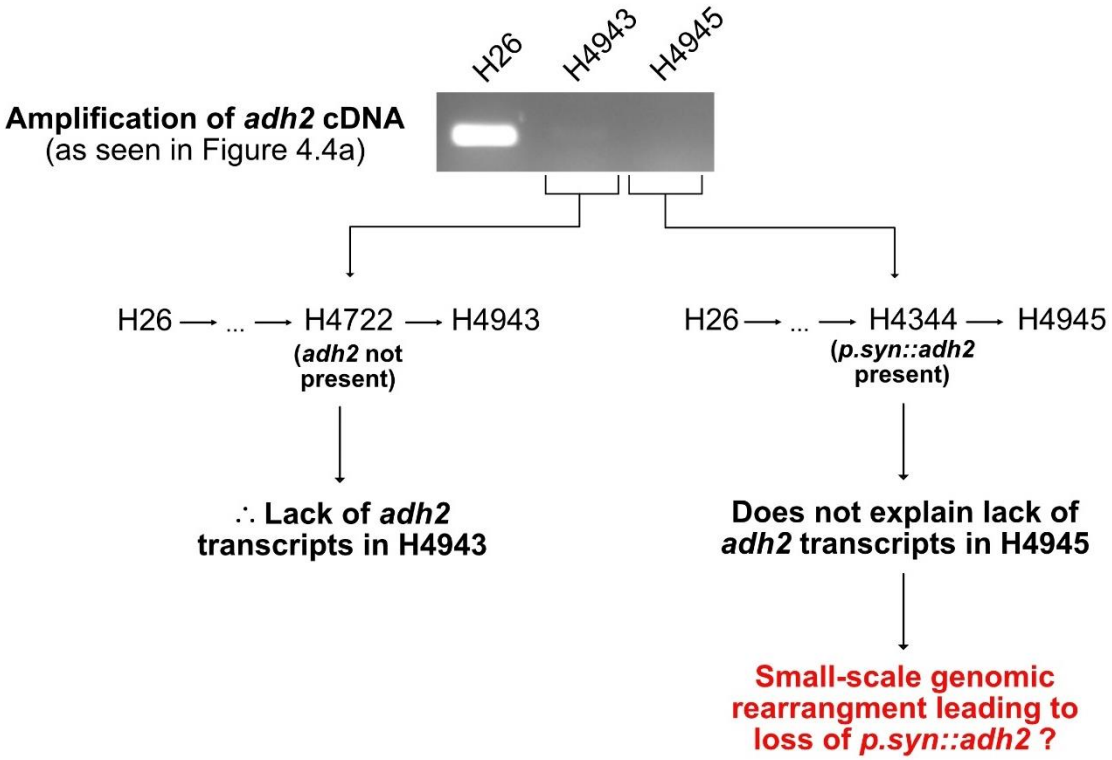


Figure 4.13: Interpretation of RT-PCR results shown in Figure 4.4. The agarose gel shown in this figure is directly taken from Figure 4.4a. *adh2* cDNA from H26, H4943 and H4945 was amplified using oligonucleotides o2438 and o2439. Greater amplification can be seen for wild-type (H26) than the strains in which *adh2* was thought to be under the control of *p.syn* (H4943 and H4945). … denotes successive genetic modifications have taken place to create H4722 and H4344 from H26 (not relevant to this figure). MinION sequencing revealed that *adh2* was not present in H4722. H4943 descended from H4722 meaning that *adh2* would also not be present in this strain, explaining the lack of amplification seen in RT-PCR corresponding to *adh2* transcripts. However, H4344 (from which H4945 descended) does contain the *p.syn::adh2* construct and it is therefore expected that there would be more adh2 transcripts in H4945 than H26, which is not seen (Figure 4.4). A possible explanation for this could be loss of *p.syn::adh2* (a small-scale genomic rearrangement) in generation of H4945 from H4722 due to the presence of *p.syn.*

This chapter also demonstrates the utility of Oxford Nanopore MinION sequencing in *H. volcanii* strain verification. Its use not only revealed the loci of pHV3 integration onto the main chromosome but showed that the genotype of strain H4722 was completely different to that previously thought. Loss of pHV1 was identified and alignments of genes to contigs assembled in MacVector meant that it was possible to search the database to try to identify which other strain has been incorrectly frozen in its place prior to this study. This is a powerful tool for verifying strains in future, especially considering the spontaneous large-scale genomic rearrangements seen in this chapter and elsewhere[49].

When this study began, it was thought that deletion of *oripHV3* and *orc6* was made possible due to the increased level of transcription initiated at *p.syn.* In the work presented here, integration of pHV3 onto the main chromosome is seen when deleting *oripHV3* and *orc6.* All other attempts to delete *oripHV3* in the laboratory prior to this were unsuccessful and either no colonies were present after the pop-out stage, or the plasmid did not pop-out completely, meaning that the origin remained. Therefore, it is not known as to whether deletion of oripHV3 is possible when the level of transcription occurring on pHV3 is increased and when pHV3 remain as an episome is possible. This gives rise to two possibilities (Figure 4.14). It could be that *oripHV3* was deleted prior to the integration of pHV3 onto the main chromosome, and that replication in its absence is indeed initiated from R-loops. Alternatively, the increase in transcriptional activity due to the insertion of *p.syn* may have led to transcription-associated recombination and integration of pHV3 onto the main chromosome occurred prior to the deletion of *oripHV3.* If the level of transcription initiated at *p.syn* leads to genomic instability and results in the integration of pHV3 onto the main chromosome when aiming to delete *oripHV3*, a weaker promotor could be inserted onto pHV3 which like *p.syn,* drives expression of *adh2 (*as its overexpression has no ill effects on the cell[258]). A weaker promotor may mean that the level of transcription is increased sufficiently to delete to the origin, but not increased so much that the genome becomes unstable leading to genomic rearrangements like those seen in these chapter. Additionally, an inducible promotor could be used. If the increased level of transcription does permit deletion of *oripHV3,* this should only be possible when the promotor is turned on. The strain itself would be an innate control.

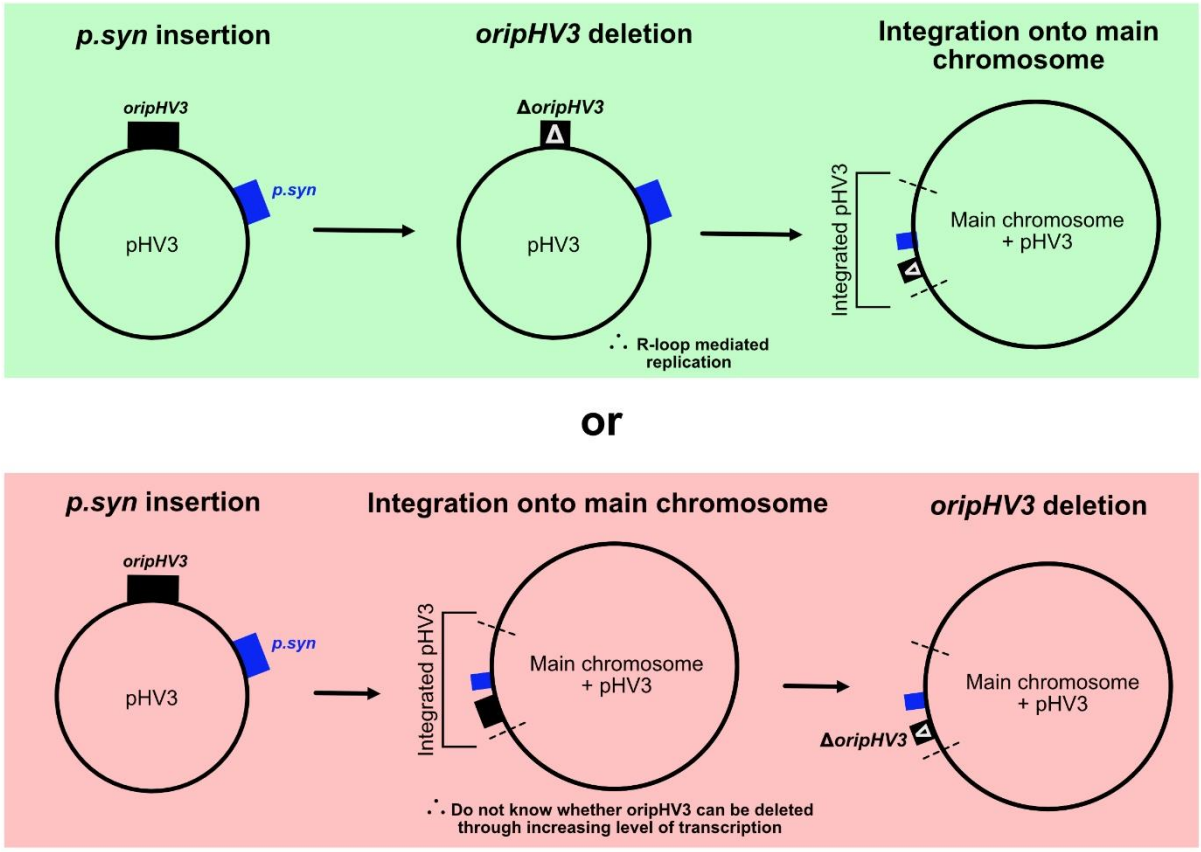


Figure 4.14: Possible orders of events leading to deletion of *oripHV3* in a strain in which pHV3 is integrated onto the main chromosome. The upper green panel demonstrates the possible scenario in which deletion of *oripHV3* occurs prior to integration of pHV3 onto the main chromosome, made possible by the insertion of *p.syn*. This would indicate that replication of pHV3 can occur via R-loop mediated replication and pHV3 can remain as episome. The lower red panel indicates a second possibility in which integration of pHV3 onto the main chromosome occurs prior to *oripHV3* deletion. In this case, it cannot be determined whether *p.syn* insertion enables deletion of *oripHV3* and pHV3 remain as an episome.

The finding that when the level of transcription upon pHV3 is increased, pHV3 is able to integrate onto the main chromosome directly relates to ChIP-Seq data for pHV3 seen in Chapter 3 of this study. In Chapter 3, a lack of peaks corresponding to RadA binding upon pHV3 was seen and the majority of peaks seen across all six IP samples mapped to promotor-transcription start site regions. Therefore, in combination, these findings could indicate that as the level of transcription upon pHV3 is increased by the presence of *p.syn*, recombination (and therefore RadA binding) also increases due to transcription mediated damage and R-loop formation. An increase in recombination may mean that pHV3 is now able to recombine with the main chromosome resulting in their fusion. This would not have previously been possible (and was not seen) when attempting to delete the origin found upon it prior to this study as in the strains in which this was attempted, transcription upon pHV3 was occurring at a native level.

Chapter 5- Future Perspectives

In Chapter 3, ChIP-Seq was carried out upon a FLAG-RadA strain in which all origins were present. Due to time constraints, ChIP-Seq was not carried out upon the other strain constructed in this chapter- FLAG-RadA with all origins deleted on the main chromosome. Future work to compare RadA binding in these strains would determine whether recombination-dependent replication initiates from specific sites within the genome.

Additionally, Chapter 3 highlighted a need for ChIP-Seq protocol optimisation. A positive control in which immunoprecipitation is undertaken against a protein known to bind at specific sites within the genome would be of use. It is also necessary that lengths of fragments are reduced by further sonication and through an additional RNase treatment step prior to running sonicated samples on an agarose gel to ensure bands seen when checking fragment size are not due to the presence of RNA in the samples.

Chapter 4 investigated genomic rearrangements that occurred when deleting the origin and corresponding Orc initiator protein on pHV3, potentially made possible by the increased level of transcription initiated at *p.syn*. As it was already known that *oripHV3* deletion is possible when pHV3 is integrated onto the main chromosome, it can no longer be said that an increase in the level of transcription and presence of R-loops also permits this deletion. As mentioned previously, alternative promotors could be used of varying strengths, and the use of an inducible promotor (such as the tryptophan-inducible promotor used to investigate the dependence upon RadA for replication initiation on pHV3 in this study) would be of interest. Furthermore, deletion of any regions of homology (such as other ISH elements) between pHV3 and the main chromosome that are found upon pHV3 (pHV3 is smaller and there would therefore be fewer regions to delete) may prevent recombination between the two and therefore also prevent integration events like those demonstrated in this study.

If it is possible to delete *oripHV3* when the level of transcription occurring on pHV3 is increased, marker frequency analysis (MFA) and DRIP-seq would be necessary to confirm the involvement of R-loops. The highly transcribed region would form a “pseudo-origin” if replication was initiated there. Therefore, in an actively replicating cell, there would be a higher number of DNA sequences corresponding to it than in areas of the genome which are not origins of replication. MFA would therefore be able to identify the highly transcribed region as a site of replication initiation given that the hypothesis was correct. DRIP-seq would be necessary to demonstrate the presence of R-loops at the highly transcribed region from which replication could be initiated.

Investigation into the roles of RNase H and topoisomerases in replication initiation from R-loops would also be of interest. Their roles in the prevention of replication from R-loops is detailed in Section 1.4.2- R-loops. In brief, in *E. coli,* mutations in genes encoding RNase HI and topoisomerase I mean that R-loops persist and can be used in replication in the absence of a *dnaA/oriC* system. If increasing the level of transcription occurring upon pHV3 is not sufficient to permit *oripHV3* deletion, the genes encoding the multiple RNase H and topoisomerases found in the *H. volcanii* genome could be deleted and *oripHV3* deletion repeated in these backgrounds.

This study begins to describe the relationship between transcription and genomic instability in *H. volcanii*. To further investigate this relationship, promotors of varying strengths could be inserted onto pHV3 and the frequency of pHV3's integration onto the main chromosome measured. As described in Chapter 4 in the creation of H2461 prior to this study, a 5'*leuB* mutant allele and 3'*leuB* mutant allele (or similar mutations in a different selectable marker) could be inserted onto the main chromosome and pHV3 respectively in strains with varying transcriptional activity occurring upon pHV3. All strains would be deficient in leucine biosynthesis as the mutations in the *leuB* allele render the protein non-functional. Native regions of homology between the main chromosome and pHV3, specifically the *transposase* loci, could be deleted meaning that the only region of homology remaining would be between the two mutant *leuB* alleles. Strains could be plated upon leu+ and leu- plates and only cells in which the mutant *leuB* alleles have recombined (resulting in integration of pHV3 onto the main chromosome) would have a functional *leuB* allele and would therefore be the only cells to survive on leu- plates. The fraction of cells which have undergone integration of pHV3 onto the main chromosome could be calculated for strains with each promotor to demonstrate the interplay between transcription and genome instability.

Hi-C could be used to investigate chromatin interactions between pHV3 and the main chromosome. Hi-C is based upon Chromosome Conformation Capture. Chromatin is crosslinked with formaldehyde, digested, and re-ligated incorporating a biotin-labelled nucleotide at the ligation junction. Chimeric DNA ligation junctions can then be purified and sequenced to determine which areas of the genome interact with each other[278]. Genome instability due to increased levels of transcription may lead to an increase in the interactions between pHV3 and the main chromosome. A speculative idea might be that as the native level of transcription occurring from pHV3 is very low, it may be packaged into a transcriptionally quiescent compartment[279,280]. When the level of transcription occurring upon it is increased through insertion of a strong promotor, this might no longer be the case meaning that pHV3 may be able to interact with the main chromosome (for example, permitting interactions between *transposase* (ISH5) genes seen in Chapter 4). Hi-C would demonstrate which areas of the genome are interacting, and whether there is an increase in interactions between pHV3 and the main chromosome when the level of transcription occurring upon pHV3 is increased.

Chapter 6- References

1 Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences* **74**, 5088-5090, doi:10.1073/pnas.74.11.5088 (1977).

2 Huet, J., Schnabel, R., Sentenac, A. & Zillig, W. Archaebacteria and eukaryotes possess DNA - dependent RNA polymerases of a common type. *The EMBO journal* **2**, 1291-1294, doi:10.1002/j.1460-2075.1983.tb01583.x (1983).

3 Kandler, O. & Konig, H. in *New Comprehensive Biochemistry* Vol. 26 223-259 (Elsevier, 1993).

4 Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences* **87**, 4576-4579, doi:10.1073/pnas.87.12.4576 (1990).

5 Grabowski, B. & Kelman, Z. Archaeal DNA replication: eukaryal proteins in a bacterial context. *Annual review of microbiology* **57**, 487, doi:10.1146/annurev.micro.57.030502.090709 (2003).

6 Barry, E. R. & Bell, S. D. DNA replication in the archaea. *Microbiology and Molecular Biology Reviews* **70**, 876-887, doi:10.1128/MMBR.00029-06 (2006).

7 Allers, T. & Mevarech, M. Archaeal genetics - the third way. *Nature Reviews Genetics* **6**, 58-73, doi:10.1038/nrg1504 (2005).

8 Winker, S. & Woese, C. A definition of the domains Archaea, Bacteria and Eucarya in terms of small subunit ribosomal RNA characteristics. *Systematic and applied microbiology* **14**, 305-310, doi:10.1016/S0723-2020(11)80303-6 (1991).

9 Guy, L. & Ettema, T. J. The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends in microbiology* **19**, 580-587, doi:10.1016/j.tim.2011.09.002 (2011).

10 Zaremba-Niedzwiedzka, K. *et al.* Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353-358, doi:10.1038/nature21031 (2017).

11 Spang, A. *et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173-179, doi:10.1038/nature14447 (2015).

12 Seitz, K. W., Lazar, C. S., Hinrichs, K. U., Teske, A. P. & Baker, B. J. Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. *The ISME journal* **10**, 1696-1705, doi:10.1038/ismej.2015.233 (2016).

13 Eme, L., Spang, A., Lombard, J., Stairs, C. W. & Ettema, T. J. G. Archaea and the origin of eukaryotes. *Nature Reviews Microbiology* **15**, 711-723, doi:10.1038/nrmicro.2017.133 (2017).

14 van der Gulik, P. T. S., Hoff, W. D. & Speijer, D. In defence of the three-domains of life paradigm. *BMC Evolutionary Biology* **17**, 218, doi:10.1186/s12862-017-1059-z (2017).

15 Harris, J. K., Kelley, S. T., Spiegelman, G. B. & Pace, N. R. The genetic core of the universal ancestor. *Genome research* **13**, 407-412, doi:10.1101/gr.652803 (2003).

16 Ciccarelli, F. D. *et al.* Toward automatic reconstruction of a highly resolved tree of life. *science* **311**, 1283-1287, doi:10.1126/science.1123061 (2006).

17 Imachi, H. *et al.* Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* **577**, 519-525, doi:10.1038/s41586-019-1916-6 (2020).

18 Khelaifia, S. & Raoult, D. Haloferax massiliensis sp. nov., the first human-associated halophilic archaea. *New microbes and new infections* **12**, 96-98, doi:10.1016/j.nmni.2016.05.007 (2016).
19 Chaban, B., Ng, S. Y. & Jarrell, K. F. Archaeal habitats--from the extreme to the ordinary. *Canadian journal of microbiology* **52**, 73-116, doi:10.1139/w05-147 (2006).
20 Koonin, E. V. & Wolf, Y. I. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic acids research* **36**, 6688-6719, doi:10.1093/nar/gkn668 (2008).
21 Wu, Z., Liu, J., Yang, H. & Xiang, H. DNA replication origins in archaea. *Frontiers in microbiology* **5**, 179, doi:10.3389/fmicb.2014.00179 (2014).
22 Albers, S. V. & Meyer, B. H. The archaeal cell envelope. *Nature Reviews Microbiology* **9**, 414-426, doi:10.1038/nrmicro2576 (2011).
23 White, M. F. & Bell, S. D. Holding it together: chromatin in the Archaea. *Trends in Genetics* **18**, 621-626, doi:10.1016/s0168-9525(02)02808-1 (2002).
24 Reeve, J. N., Sandman, K. & Daniels, C. J. Archaeal histones, nucleosomes, and transcription initiation. *Cell* **89**, 999-1002, doi:10.1016/S0092-8674(00)80286-X (1997).
25 Ammar, R. *et al.* Chromatin is an ancient innovation conserved between Archaea and Eukarya. *Elife* **1**, e00078, doi:10.7554/eLife.00078 (2012).
26 Mattiroli, F. *et al.* Structure of histone-based chromatin in Archaea. *Science* **357**, 609-612, doi:10.1126/science.aaj1849 (2017).
27 Peeters, E., Driessen, R. P., Werner, F. & Dame, R. T. The interplay between nucleoid organization and transcription in archaeal genomes. *Nature Reviews Microbiology* **13**, 333-341, doi:10.1038/nrmicro3467 (2015).
28 Reeve, J. N. Archaeal chromatin and transcription. *Molecular microbiology* **48**, 587-598, doi:10.1046/j.1365-2958.2003.03439.x (2003).
29 French, S. L., Santangelo, T. J., Beyer, A. L. & Reeve, J. N. Transcription and translation are coupled in Archaea. *Molecular biology and evolution* **24**, 893-895, doi:10.1093/molbev/msm007 (2007).
30 Langer, D., Hain, J., Thuriaux, P. & Zillig, W. Transcription in archaea: similarity to that in eucarya. *Proceedings of the National Academy of Sciences* **92**, 5768-5772, doi:10.1073/pnas.92.13.5768 (1995).
31 Jun, S. H., Reichlen, M. J., Tajiri, M. & Murakami, K. S. Archaeal RNA polymerase and transcription regulation. *Critical reviews in biochemistry and molecular biology* **46**, 27-40, doi:10.3109/10409238.2010.538662 (2011).
32 Bell, S. D. & Jackson, S. P. Mechanism and regulation of transcription in archaea. *Current opinion in microbiology* **4**, 208-213, doi:10.1016/S1369-5274(00)00190-9 (2001).
33 Benelli, D. & Londei, P. Translation initiation in Archaea: conserved and domain-specific features. *Biochemical Society Transactions* **39**, 89-93, doi:10.1042/BST0390089 (2011).
34 Albers, S. V., Forterre, P., Prangishvili, D. & Schleper, C. The legacy of Carl Woese and Wolfram Zillig: from phylogeny to landmark discoveries. *Nature Reviews Microbiology* **11**, 713-719, doi:10.1038/nrmicro3124 (2013).

35 Mullakhanbhai, M. F. & Larsen, H. Halobacterium volcanii spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Archives of microbiology* **104**, 207-214, doi:10.1007/BF00447326 (1975).
36 Duggin, I. G. *et al.* CetZ tubulin-like proteins control archaeal cell shape. *Nature* **519**, 362-365, doi:10.1038/nature13983 (2015).
37 Shalev, Y., Turgeman-Grott, I., Tamir, A., Eichler, J. & Gophna, U. Cell Surface Glycosylation Is Required for Efficient Mating of Haloferax volcanii. *Frontiers in microbiology* **8**, 1253, doi:10.3389/fmicb.2017.01253 (2017).
38 Rosenshine, I., Tchelet, R. & Mevarech, M. The mechanism of DNA transfer in the mating system of an archaebacterium. *Science* **245**, 1387-1389, doi:10.1126/science.2818746 (1989).
39 Mevarech, M. & Werczberger, R. Genetic transfer in Halobacterium volcanii. *Journal of Bacteriology* **162**, 461-462, doi:10.1128/jb.162.1.461-462.1985 (1985).
40 Siglioccolo, A., Paiardini, A., Piscitelli, M. & Pascarella, S. Structural adaptation of extreme halophilic proteins through decrease of conserved hydrophobic contact surface. *BMC structural biology* **11**, 1-12, doi:10.1186/1472-6807-11-50 (2011).
41 Winter, J. A., Christofi, P., Morroll, S. & Bunting, K. A. The crystal structure of Haloferax volcanii proliferating cell nuclear antigen reveals unique surface charge characteristics due to halophilic adaptation. *BMC Structural Biology* **9**, 55, doi:10.1186/1472-6807-9-55 (2009).
42 Oren, A. Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Systems* **4**, 2, doi:10.1186/1746-1448-4-2 (2008).
43 Wendoloski, D., Ferrer, C. & Dyall-Smith, M. L. A new simvastatin (mevinolin)-resistance marker from Haloarcula hispanica and a new *Haloferax volcanii* strain cured of plasmid pHV2. *Microbiology* **147**, 959-964, doi:10.1099/00221287-147-4-959 (2001).
44 Cline, S. W., Lam, W. L., Charlebois, R. L., Schalkwyk, L. C. & Doolittle, W. F. Transformation methods for halophilic archaebacteria. *Canadian journal of microbiology* **35**, 148-152, doi:10.1139/m89-022 (1989).
45 Perez-Arnaiz, P., Dattani, A., Smith, V. & Allers, T. Haloferax volcanii-a model archaeon for studying DNA replication and repair. *Open Biology* **10**, 200293, doi:10.1098/rsob.200293 (2020).
46 Allers, T., Ngo, H. P., Mevarech, M. & Lloyd, R. G. Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. *Applied and environmental microbiology* **70**, 943-953, doi:10.1128/AEM.70.2.943-953.2004 (2004).
47 Allers, T. & Ngo, H. P. Genetic analysis of homologous recombination in Archaea: Haloferax volcanii as a model organism. *Biochemical Society Transactions* **31**, 706-710, doi:10.1042/bst0310706 (2003).
48 Hartman, A. L. *et al.* The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS One* **5**, e9605, doi:10.1371/journal.pone.0009605 (2010).
49 Hawkins, M., Malla, S., Blythe, M. J., Nieduszynski, C. A. & Allers, T. Accelerated growth in the absence of DNA replication origins. *Nature* **503**, 544-547, doi:10.1038/nature12650 (2013).

50 Ludt, K. & Soppa, J. Polyploidy in halophilic archaea: regulation, evolutionary advantages, and gene conversion. *Biochemical Society Transactions* **47**, 933-944, doi:10.1042/BST20190256 (2019).
51 Breuert, S., Allers, T., Spohn, G. & Soppa, J. Regulated polyploidy in halophilic archaea. *PLoS One* **1**, e92, doi:10.1371/journal.pone.0000092 (2006).
52 Hawkins, M. *et al.* High-resolution replication profiles define the stochastic nature of genome replication initiation and termination. *Cell reports* **5**, 1132-1141, doi:10.1016/j.celrep.2013.10.014 (2013).
53 Norais, C. *et al.* Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS genetics* **3**, e77, doi:10.1371/journal.pgen.0030077 (2007).
54 Maurer, S., Ludt, K. & Soppa, J. Characterization of Copy Number Control of Two Haloferax volcanii Replication Origins Using Deletion Mutants and Haloarchaeal Artificial Chromosomes. *Journal of bacteriology* **200**, doi:10.1128/JB.00517-17 (2018).
55 Ausiannikava, D. *et al.* Evolution of Genome Architecture in Archaea: Spontaneous Generation of a New Chromosome in Haloferax volcanii. *Molecular Biology and Evolution* **35**, 1855-1868, doi:10.1093/molbev/msy075 (2018).
56 Leonard, A. C. & Mechali, M. DNA replication origins. *Cold Spring Harbor perspectives in biology* **5**, a010116, doi:10.1101/cshperspect.a010116 (2013).
57 Dewar, J. M. & Walter, J. C. Mechanisms of DNA replication termination. *Nature Reviews Molecular Cell Biology* **18**, 507-516, doi:10.1038/nrm.2017.42 (2017).
58 Duggin, I. G., Dubarry, N. & Bell, S. D. Replication termination and chromosome dimer resolution in the archaeon Sulfolobus solfataricus. *EMBO J* **30**, 145-153, doi:10.1038/emboj.2010.301 (2011).
59 Michel, B. & Bernander, R. Chromosome replication origins: do we really need them? *Bioessays* **36**, 585-590, doi:10.1002/bies.201400003 (2014).
60 Kogoma, T. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiology and Molecular Biology Reviews* **61**, 212-238, doi:10.1128/mmbr.61.2.212-238.1997 (1977).
61 DePamphilis, M. L. & Bell, S. D. *Genome duplication*. (Garland Science, 2010).
62 Myllykallio, H., & Forterre, P. Mapping of a chromosome replication origin in an archaeon: response. *Trends in Microbiology* **8**, 537-539, doi:10.1016/S0966-842X(00)01881-3 (2000).
63 Schaper, S. & Messer, W. Interaction of the initiator protein DnaA of Escherichia coli with its DNA target. *Journal of Biological Chemistry* **270**, 17622-17626, doi:10.1074/jbc.270.29.17622 (1995).
64 Messer, W. The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS microbiology reviews* **26**, 355-374, doi:10.1111/j.1574-6976.2002.tb00620.x (2002).
65 Mott, M. L., Erzberger, J. P., Coons, M. M. & Berger, J. M. Structural synergy and molecular crosstalk between bacterial helicase loaders and replication initiators. *Cell* **135**, 623-634, doi:10.1016/j.cell.2008.09.058 (2008).
66 Davey, M. J., Fang, L., McInerney, P., Georgescu, R. E. & O'Donnell, M. The DnaC helicase loader is a dual ATP/ADP switch protein. *The EMBO journal* **21**, 3148-3159, doi:0.1093/emboj/cdf308 (2002).
67 Li, H. & Stillman, B. The origin recognition complex: a biochemical and structural view. *Sub-cellular biochemistry* **62**, 37-58, doi:10.1007/978-94-007-4572-8_3 (2012).

68 Bleichert, F., Botchan, M. R. & Berger, J. M. Crystal structure of the eukaryotic origin recognition complex. *Nature* **519**, 321-326, doi:10.1038/nature14239 (2015).
69 Makarova, K. S. & Koonin, E. V. Archaeology of eukaryotic DNA replication. *Cold Spring Harbor Perspectives in Biology* **5**, a012963, doi:10.1101/cshperspect.a012963 (2013).
70 Bell, S. P. & Dutta, A. DNA replication in eukaryotic cells. *Annual review of biochemistry* **71**, 333-374, doi:10.1146/annurev.biochem.71.110601.135425 (2002).
71 Li, H. & O'Donnell, M. E. The Eukaryotic CMG Helicase at the Replication Fork: Emerging Architecture Reveals an Unexpected Mechanism. *Bioessays* **40**, doi:10.1002/bies.201700208 (2018).
72 Tanaka, S. & Araki, H. Helicase activation and establishment of replication forks at chromosomal origins of replication. *Cold Spring Harbor perspectives in biology* **5**, a010371, doi:10.1101/cshperspect.a010371 (2013).
73 Kelman, L. M., O'Dell, W. B. & Kelman, Z. Unwinding 20 years of the archaeal minichromosome maintenance helicase. *Journal of Bacteriology* **202**, doi:10.1128/JB.00729-19 (2020).
74 Sawa, M. & Masai, H. Drug design with Cdc7 kinase: a potential novel cancer therapy target. *Drug design, development and therapy* **2**, 255, doi:10.2147/dddt.s4303 (2008).
75 Robinson, N. P. & Bell, S. D. Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proceedings of the National Academy of Sciences* **104**, 5806-5811, doi:10.1073/pnas.0700206104 (2007).
76 Ausiannikava, D. & Allers, T. Diversity of DNA Replication in the Archaea. *Genes* **8**, doi:10.3390/genes8020056 (2017).
77 Samson, R. Y. *et al.* Specificity and function of archaeal DNA replication initiator proteins. *Cell reports* **3**, 485-496, doi:10.1016/j.celrep.2013.01.002 (2013).
78 Wu, Z., Liu, H., Liu, J., Liu, X. & Xiang, H. Diversity and evolution of multiple orc/cdc6-adjacent replication origins in haloarchaea. *BMC genomics* **13**, 1-16, doi:10.1186/1471-2164-13-478 (2012).
79 Samson, R. Y., Abeyrathne, P. D. & Bell, S. D. Mechanism of Archaeal MCM Helicase Recruitment to DNA Replication Origins. *Molecular cell* **61**, 287-296, doi:10.1016/j.molcel.2015.12.005 (2016).
80 Kawakami, H., Ohashi, E., Kanamoto, S., Tsurimoto, T. & Katayama, T. Specific binding of eukaryotic ORC to DNA replication origins depends on highly conserved basic residues. *Scientific reports* **5**, 14929, doi:10.1038/srep14929 (2015).
81 Gaudier, M., Schuwirth, B. S., Westcott, S. L., Wigley, D. B. . Structural basis of DNA replication origin recognition by an ORC protein. *Science* **317**, 1213-1216, doi:10.1126/science.1143664 (2007).
82 Sakakibara, N., Kelman, L. M. & Kelman, Z. How is the archaeal MCM helicase assembled at the origin? Possible mechanisms. *Biochemical Society Transactions* **37**, 7-11, doi:10.1042/BST0370007 (2009).
83 Makarova, K. S., Koonin, E. V. & Kelman, Z. The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. *Biology direct* **7**, 1-10, doi:10.1186/1745-6150-7-7 (2012).
84 Tosun, Ç. *Role of Orc9 in DNA Damage Response in Haloferax volcanii* Master of Research thesis, University of Nottingham, (2019).
85 Rawls, K. S. *Characterization of glycerol metabolism and related metabolic pathways in the haloarchaeaon Haloferax volcanii*. (University of Florida, 2010).

86 Kogoma, T. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiology and Molecular Biology Reviews* **61**, 212-238, doi:10.1128/mmbr.61.2.212-238.1997 (1997).
87 Masai, H. & Arai, K. Mechanisms of primer RNA synthesis and D-loop/R-loop-dependent DNA replication in Escherichia coli. *Biochimie* **78**, 1109-1117, doi:10.1016/S0300-9084(97)86737-5 (1996).
88 Neale, M. J. & Keeney, S. Clarifying the mechanics of DNA strand exchange in meiotic recombination. *Nature* **442**, 153-158, doi:10.1038/nature04885 (2006).
89 Michel, B. *et al.* Rescue of arrested replication forks by homologous recombination. *Proceedings of the National Academy of Sciences* **98**, 8181-8188, doi:10.1073/pnas.111008798 (2001).
90 Venkitaraman, A. R. Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell* **108**, 171-182, doi:10.1016/S0092-8674(02)00615-3 (2002).
91 Haldenby, S., White, M. F. & Allers, T. RecA family proteins in archaea: RadA and its cousins. *Biochemical Society Transactions* **37**, 102-107, doi:10.1042/BST0370102 (2009).
92 Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D. & Rehrauer, W. M. Biochemistry of homologous recombination in Escherichia coli. *Microbiological reviews* **58**, 401-465, doi:10.1128/mr.58.3.401-465.1994 (1994).
93 Heyer, W. D., Ehmsen, K. T. & Liu, J. Regulation of homologous recombination in eukaryotes. *Annual review of genetics* **44**, 113-139, doi:10.1146/annurev-genet-051710-150955 (2010).
94 Clark, A. J. & Margulies, A. D. Isolation and characterization of recombination-deficient mutants of Escherichia coli K12. *Proceedings of the National Academy of Sciences of the United States of America* **53**, 451, doi:10.1073/pnas.53.2.451 (1965).
95 Shinohara, A., Ogawa, H. & Ogawa, T. Rad51 protein involved in repair and recombination in S. cerevisiae is a RecA-like protein. *Cell* **69**, 457-470, doi:10.1016/0092-8674(92)90447-K (1992).
96 Sehorn, M. G., Sigurdsson, S., Bussen, W., Unger, V. M. & Sung, P. Human meiotic recombinase Dmc1 promotes ATP-dependent homologous DNA strand exchange. *Nature* **429**, 433-437, doi:10.1038/nature02563 (2004).
97 Ogawa, T., Yu, X., Shinohara, A. & Egelman, E. H. Similarity of the yeast RAD51 filament to the bacterial RecA filament. *Science* **259**, 1896-1899, doi:10.1126/science.8456314 (1993).
98 Aboussekhra, A., Chanet, R., Adjiri, A. & Fabre, F. Semidominant suppressors of Srs2 helicase mutations of Saccharomyces cerevisiae map in the RAD51 gene, whose sequence predicts. *Molecular and cellular biology* **12**, 3224-3234, doi:10.1128/mcb.12.7.3224-3234.1992 (1992).
99 Yang, S., Yu, X., Seitz, E. M., Kowalczykowski, S. C. & Egelman, E. H. Archaeal RadA protein binds DNA as both helical filaments and octameric rings. *Journal of molecular biology* **314**, 1077-1085, doi:10.1006/jmbi.2001.5213.
100 Sandler, S. J., Satin, L. H., Samra, H. S. & Clark, A. J. recA-like genes from three archaean species with putative protein products similar to Rad51 and Dmcl proteins of the yeast Saccharomyces cerevisiae. *Nucleic acids research* **24**, 2125-2132 (1996).
101 Story, R., Weber, I. & Steitz, T. The structure of the E. coli recA protein monomer and polymer. *Nature* **355**, 318–325, doi:10.1038/355318a0 (1992).

102 Shin, D. S. *et al.* Full-length archaeal Rad51 structure and mutants: mechanisms for RAD51 assembly and control by BRCA2. *The EMBO journal* **22**, 4566-4576, doi:10.1093/emboj/cdg429 (2003).
103 Conway, A. B. *et al.* Crystal structure of a Rad51 filament. *Nature structural & molecular biology* **11**, 791-796, doi:10.1038/nsmb795 (2004).
104 Kurumizaka, H. *et al.* A possible role of the C-terminal domain of the RecA protein. A gateway model for double-stranded DNA binding. *Journal of Biological Chemistry* **271**, 33515-33524, doi:10.1074/jbc.271.52.33515 (1996).
105 Galkin, V. E. *et al.* The Rad51/RadA N-terminal domain activates nucleoprotein filament ATPase activity. *Structure* **14**, 983-992, doi:10.1016/j.str.2006.04.001 (2006).
106 Lee, C. D. & Wang, T. F. The N-terminal domain of Escherichia coli RecA have multiple functions in promoting homologous recombination. *Journal of Biomedical Science* **16**, 37, doi:10.1186/1423-0127-16-37 (2009).
107 Komori, K. *et al.* Domain analysis of an archaeal RadA protein for the strand exchange activity. *Journal of Biological Chemistry* **275**, 33791-33797, doi:10.1074/jbc.M004556200 (2000).
108 Wardell, K. *Regulation of Homologous Recombination in the Archaeon Haloferax volcanii.* Doctor of Philosophy thesis, University of Nottingham, (2013).
109 Joo, C. *et al.* Real-time observation of RecA filament dynamics with single monomer resolution. *Cell* **126**, 515-527, doi:10.1016/j.cell.2006.06.042 (2006).
110 Bell, J. C., Plank, J. L., Dombrowski, C. C. & Kowalczykowski, S. C. Direct imaging of RecA nucleation and growth on single molecules of SSB-coated ssDNA. *Nature* **491**, 274-278, doi:10.1038/nature11598 (2012).
111 Nishinaka, T., Ito, Y., Yokoyama, S. & Shibata, T. An extended DNA structure through deoxyribose-base stacking induced by RecA protein. *Proceedings of the National Academy of Sciences* **94**, 6623-6628, doi:10.1073/pnas.94.13.6623 (1997).
112 Nishinaka, T., Doi, Y., Hara, R. & Yashima, E. Elastic behavior of RecA-DNA helical filaments. *Journal of molecular biology* **370**, 837-845, doi:10.1016/j.jmb.2007.05.044 (2007).
113 Flory, J., Tsang, S. S. & Muniyappa, K. Isolation and visualization of active presynaptic filaments of recA protein and single-stranded DNA. *Proceedings of the National Academy of Sciences* **81**, 7026-7030, doi:10.1073/pnas.81.22.7026 (1984).
114 Chen, Z., Yang, H. & Pavletich, N. P. Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* **453**, 489-484, doi:10.1038/nature06971 (2008).
115 van der Heijden, T. *et al.* Real-time assembly and disassembly of human RAD51 filaments on individual DNA molecules. *Nucleic acids research* **35**, 5646-5657, doi:10.1093/nar/gkm629 (2007).
116 Mine, J. *et al.* Real-time measurements of the nucleation, growth and dissociation of single Rad51-DNA nucleoprotein filaments. *Nucleic acids research* **35**, 7171-7187, doi:10.1093/nar/gkm752 (2007).
117 Hilario, J., Amitani, I., Baskin, R. J. & Kowalczykowski, S. C. Direct imaging of human Rad51 nucleoprotein dynamics on individual DNA molecules. *Proceedings of the National Academy of Sciences* **106**, 361-368, doi:10.1073/pnas.0811965106 (2009).

118 Cox, M. M. Regulation of bacterial RecA protein function. *Critical reviews in biochemistry and molecular biology* **42**, 41-63, doi:10.1080/10409230701260258 (2007).
119 Sung, P. & Robberson, D. L. DNA strand exchange mediated by a RAD51-ssDNA nucleoprotein filament with polarity opposite to that of RecA. *Cell* **82**, 453-461, doi:10.1016/0092-8674(95)90434-4 (1995).
120 Baumann, P. & West, S. C. The human Rad51 protein: polarity of strand transfer and stimulation by hRP-A. *The EMBO journal* **16**, 5198-5206, doi:10.1093/emboj/16.17.5198 (1997).
121 Ruigrok, R. W. *et al.* The inactive form of recA protein: the 'compact' structure. *The EMBO journal* **12**, doi:10.1002/j.1460-2075.1993.tb05626.x (1993).
122 Egelman, E. H. & Stasiak, A. Structure of helical RecA-DNA complexes: Complexes formed in the presence of ATP-gamma-S or ATP. *Journal of molecular biology* **191**, 677-697, doi:10.1016/0022-2836(86)90453-5 (1986).
123 Robertson, R. B. *et al.* Visualizing the disassembly of S. cerevisiae Rad51 nucleoprotein filaments. *Journal of molecular biology* **388**, 703-720, doi:10.1016/j.jmb.2009.03.049 (2009).
124 Thomas, M., White, R. L. & Davis, R. W. Hybridization of RNA to double-stranded DNA: Formation of R-loops. *Proceedings of the National Academy of Sciences* **73**, 2294-2298, doi:10.1073/pnas.73.7.2294 (1976).
125 Drolet, M. *et al.* Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. *Proceedings of the National Academy of Sciences* **92**, 3526-3530, doi:10.1073/pnas.92.8.3526 (1995).
126 Costantino, L. & Koshland, D. The Yin and Yang of R-loop biology. *Current opinion in cell biology* **34**, 39-45, doi:10.1016/j.ceb.2015.04.008 (2015).
127 Garcia-Rubio, M. *et al.* Yra1-bound RNA-DNA hybrids cause orientation-independent transcription-replication collisions and telomere instability. *Genes & development* **32**, 965-977, doi:10.1101/gad.311274.117 (2018).
128 Santos-Pereira, J. M. & Aguilera, A. R loops: new modulators of genome dynamics and function. *Nature Reviews Genetics* **16**, 583-597, doi:10.1038/nrg3961 (2015).
129 Skourti-Stathaki, K. & Proudfoot, N. J. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes & development* **28**, 1384-1396, doi:10.1101/gad.242990.114 (2014).
130 Boguslawski, S. J. *et al.* Characterization of monoclonal antibody to DNA· RNA and its application to immunodetection of hybrids. *Journal of immunological methods* **89**, 123-130, doi:10.1016/0022-1759(86)90040-2 (1986).
131 Hu, Z., Zhang, A., Storz, G., Gottesman, S. & Leppla, S. H. An antibody-based microarray assay for small RNA detection. *Nucleic Acids Res* **34**, e52, doi:10.1093/nar/gkl142 (2006).
132 Ginno, P. A., Lim, Y. W., Lott, P. L., Korf, I. & Chedin, F. GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res* **23**, 1590-1600, doi:10.1101/gr.158436.113 (2013).
133 Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chedin, F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell* **45**, 814-825, doi:10.1016/j.molcel.2012.01.017 (2012).

134 Chan, Y. A. *et al.* Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. *PLoS Genet* **10**, e1004288, doi:10.1371/journal.pgen.1004288 (2014).
135 Sanz, L. A. *et al.* Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Mol Cell* **63**, 167-178, doi:10.1016/j.molcel.2016.05.032 (2016).
136 Xu, W. *et al.* The R-loop is a common chromatin feature of the Arabidopsis genome. *Nat Plants* **3**, 704-714, doi:10.1038/s41477-017-0004-x (2017).
137 El Hage, A., Webb, S., Kerr, A. & Tollervey, D. Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genet* **10**, e1004716, doi:10.1371/journal.pgen.1004716 (2014).
138 Wahba, L., Costantino, L., Tan, F. J., Zimmer, A. & Koshland, D. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes & development* **30**, 1327-1338, doi:10.1101/gad.280834 (2016).
139 Westover, K. D., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* **119**, 481-489, doi:10.1016/j.cell.2004.10.016 (2004).
140 Liu, L. F. & Wang, J. C. Supercoiling of the DNA template during transcription. *Proceedings of the National Academy of Sciences* **84**, 7024-7027, doi:10.1073/pnas.84.20.7024 (1987).
141 Roy, D., Yu, K. & Lieber, M. R. Mechanism of R-loop formation at immunoglobulin class switch sequences. *Molecular and cellular biology* **28**, 50-60, doi:10.1128/MCB.01251-07 (2008).
142 Hegazy, Y. A., Fernando, C. M. & Tran, E. J. The balancing act of R-loop biology: The good, the bad, and the ugly. *Journal of Biological Chemistry* **295**, 905-913, doi:10.1074/jbc.REV119.011353 (2020).
143 Cheng, B., Zhu, C. X., Ji, C., Ahumada, A. & Tse-Dinh, Y. C. Direct interaction between Escherichia coli RNA polymerase and the zinc ribbon domains of DNA topoisomerase I. *J Biol Chem* **278**, 30705-30710, doi:10.1074/jbc.M303403200 (2003).
144 Drolet, M., Bi, X. & Liu, L. F. Hypernegative supercoiling of the DNA template during transcription elongation in vitro. *Journal of Biological Chemistry* **269**, 2068-2074, doi:10.1016/s0021-9258(17)42136-3 (1994).
145 Masse, E. & Drolet, M. Escherichia coli DNA topoisomerase I inhibits R-loop formation by relaxing transcription-induced negative supercoiling. *Journal of Biological Chemistry* **274**, 16659-16664, doi:10.1074/jbc.274.23.16659 (1999).
146 Phoenix, P., Raymond, M. A., Masse, E. & Drolet, M. Roles of DNA topoisomerases in the regulation of R-loop formation in vitro. *J Biol Chem* **272**, 1473-1479, doi:10.1074/jbc.272.3.1473 (1997).
147 Dominguez-Sanchez, M. S., Barroso, S., Gomez-Gonzalez, B., Luna, R. & Aguilera, A. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet* **7**, e1002386, doi:10.1371/journal.pgen.1002386 (2011).
148 Gomez-Gonzalez, B. *et al.* Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J* **30**, 3106-3119, doi:10.1038/emboj.2011.206 (2011).
149 Huertas, P. & Aguilera, A. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Molecular cell* **12**, 711-721, doi:10.1016/j.molcel.2003.08.010 (2003).

150 Wahba, L., Amon, J. D., Koshland, D. & Vuica-Ross, M. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol Cell* **44**, 978-988, doi:10.1016/j.molcel.2011.10.017 (2011).
151 Roy, D., Zhang, Z., Lu, Z., Hsieh, C. L. & Lieber, M. R. Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. *Mol Cell Biol* **30**, 146-159, doi:10.1128/MCB.00897-09 (2010).
152 Kasahara, M., Clikeman, J. A., Bates, D. B. & Kogoma, T. RecA protein-dependent R-loop formation in vitro. *Genes & development* **14**, 360-365, doi:10.1101/gad.14.3.360 (2000).
153 Zaitsev, E. N. & Kowalczykowski, S. C. A novel pairing process promoted by Escherichia coli RecA protein: inverse DNA and RNA strand exchange. *Genes & development* **14**, 740-749, doi:10.1101/gad.14.6.740 (2000).
154 Wahba, L., Gore, S. K. & Koshland, D. The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. *Elife* **2**, e00505, doi:10.7554/eLife.00505 (2013).
155 Hong, X., Cadwell, G. W. & Kogoma, T. Escherichia coli RecG and RecA proteins in R-loop formation. *The EMBO journal* **14**, 2385-2392, doi:10.1002/j.1460-2075.1995.tb07233.x (1995).
156 Cao, Y. & Kogoma, T. Requirement for the polymerization and 5'--> 3'exonuclease activities of DNA polymerase I in initiation of DNA replication at oriK sites in the absence of RecA in Escherichia coli rnhA mutants. *Journal of bacteriology* **175**, 7254-7259, doi:10.1128/jb.175.22.7254-7259.1993 (1993).
157 Luder, A. & Mosig, G. Two alternative mechanisms for initiation of DNA replication forks in bacteriophage T4: priming by RNA polymerase and by recombination. *Proceedings of the National Academy of Sciences* **79**, 1101-1105, doi:10.1073/pnas.79.4.110 (1982).
158 Carles-Kinch, K. & Kreuzer, K. N. RNA-DNA hybrid formation at a bacteriophage T4 replication origin. *Journal of molecular biology* **266**, 915-926, doi:10.1006/jmbi.1996.0844 (1997).
159 Kreuzer, K. N. & Brister, J. R. Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives. *Virology journal* **7**, 1-16, doi:10.1186/1743-422X-7-358 (2010).
160 Belanger, K. G. & Kreuzer, K. N. Bacteriophage T4 initiates bidirectional DNA replication through a two-step process. *Molecular cell* **2**, 693-701, doi:10.1016/S1097-2765(00)80167-7 (1998).
161 Kogoma, T. A novel Escherichia coli mutant capable of DNA replication in the absence of protein synthesis. *Journal of molecular biology* **121**, 55-69, doi:10.1016/0022-2836(78)90262-0 (1978).
162 Ogawa, T., Pickett, G. G., Kogoma, T. & Kornberg, A. 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. *Proceedings of the National Academy of Sciences* **81**, 1040-1044, doi:10.1073/pnas.81.4.1040 (1984).
163 Kaguni, J. M. & Kornberg, A. Topoisomerase I confers specificity in enzymatic replication of the *Escherichia coli* chromosomal origin. *Journal of Biological Chemistry* **259**, 8578-8583, doi:10.1016/s0021-9258(17)39769-7 (1984).

164 Saini, N. *et al.* Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature* **502**, 389-392, doi:10.1038/nature12584 (2013).
165 Donnianni, R. A. & Symington, L. S. Break-induced replication occurs by conservative DNA synthesis. *Proceedings of the National Academy of Sciences* **110**, 13475-13480, doi:10.1073/pnas.1309800110 (2013).
166 Malkova, A., Ivanov, E. L. & Haber, J. E. Double-strand break repair in the absence of RAD51 in yeast: a possible role for break-induced DNA replication. *Proceedings of the National Academy of Sciences* **93**, 7131-7136, doi:10.1073/pnas.93.14.7131 (1996).
167 Ira, G. & Haber, J. E. Characterization of RAD51-independent break-induced replication that acts preferentially with short homologous sequences. *Molecular and cellular biology* **22**, 6384-6392, doi:10.1128/MCB.22.18.6384-6392.2002 (2002).
168 Woods, W. G. & Dyall‐Smith, M. L. Construction and analysis of a recombination‐deficient (radA) mutant of Haloferax volcanii. *Molecular microbiology* **23**, 791-797, doi:10.1073/pnas.74.11.5088 (1997).
169 Moalic, Y. *et al.* Role of Ori in Thermococcus barophilus. *bioRxiv*, doi:10.1101/2021.04.27.441579 (2021).
170 Gehring, A. M. *et al.* Genome Replication in *Thermococcus kodakarensis* Independent of Cdc6 and an Origin of Replication. *Frontiers in microbiology* **8**, 2084, doi:10.3389/fmicb.2017.02084 (2017).
171 Ohbayashi, R. *et al.* Evolutionary Changes in DnaA-Dependent Chromosomal Replication in Cyanobacteria. *Frontiers in microbiology* **11**, 786, doi:10.3389/fmicb.2020.00786 (2020).
172 Yang, H. *et al.* Activation of a dormant replication origin is essential for Haloferax mediterranei lacking the primary origins. *Nature Communications* **6**, 8321, doi:10.1038/ncomms9321 (2015).
173 Wu, Z., Yang, H., Liu, J., Wang, L. & Xiang, H. Association between the dynamics of multiple replication origins and the evolution of multireplicon genome architecture in haloarchaea. *Genome biology and evolution* **6**, 2799-2810, doi:10.1093/gbe/evu219 (2014).
174 Sollier, J. & Cimprich, K. A. Breaking bad: R-loops and genome integrity. *Trends in cell biology* **25**, 514-522, doi:10.1016/j.tcb.2015.05.003 (2015).
175 García-Muse, T. & Aguilera, A. R Loops: From Physiological to Pathological Roles. *Cell* **179**, 604-618, doi:10.1016/j.cell.2019.08.055 (2019).
176 Crossley, M. P., Bocek, M. & Cimprich, K. A. R-Loops as Cellular Regulators and Genomic Threats. *Molecular cell* **73**, 398-411, doi:10.1016/j.molcel.2019.01.024 (2019).
177 Brambati, A., Zardoni, L., Nardini, E., Pellicioli, A. & Liberi, G. The dark side of RNA:DNA hybrids. *Mutation Research/Reviews in Mutation Research* **784**, 108300, doi:10.1016/j.mrrev.2020.108300 (2020).
178 Aguilera, A. & Garcia-Muse, T. R loops: from transcription byproducts to threats to genome stability. *Molecular cell* **46**, 115-124, doi:10.1016/j.molcel.2012.04.009 (2012).
179 Aguilera, A. The connection between transcription and genomic instability. *The EMBO journal* **21**, 195-201, doi:10.1093/emboj/21.3.195 (2002).

180 Herman, R. K. & Dworkin, N. B. Effect of gene induction on the rate of mutagenesis by ICR-191 in Escherichia coli. *Journal of bacteriology* **106**, 543-550, doi:10.1128/jb.106.2.543-550.1971 (1971).
181 Brock, R. Differential mutation of the beta-galactosidase gene of Escherichia coli. *Mutation research* **11**, 181-186 (1971).
182 Beletskii, A., Grigoriev, A., Joyce, S. & Bhagwat, A. S. Mutations induced by bacteriophage T7 RNA polymerase and their effects on the composition of the T7 genome. *Journal of molecular biology* **300**, 1057-1065, doi:10.1006/jmbi.2000.3944 (2000).
183 Datta, A. & Jinks-Robertson, S. Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science* **268**, 1616-1619, doi:10.1126/science.7777859 (1995).
184 Wright, B. E., Schmidt, K. H. & Minnick, M. F. Mechanisms by which transcription can regulate somatic hypermutation. *Genes and Immunity* **5**, 176-182, doi:10.1038/sj.gene.6364053 (2004).
185 Storb, U. *et al.* Somatic hypermutation of immunoglobulin genes is linked to transcription. *Somatic Diversification of Immune Responses*, 11-19, doi:10.1007/978-3-642-71984-4_2 (1998).
186 Freudenreich, C. H. R-loops: targets for nuclease cleavage and repeat instability. *Current genetics* **64**, 789-794, doi:10.1007/s00294-018-0806-z (2018).
187 Li, X. & Manley, J. L. Cotranscriptional processes and their influence on genome stability. *Genes & development* **20**, 1838-1847, doi:10.1101/gad.1438306 (2006).
188 Aguilera, A. & Gaillard, H. Transcription and recombination: when RNA meets DNA. *Cold Spring Harbor perspectives in biology* **6**, doi:10.1101/cshperspect.a016543 (2014).
189 Ikeda, H. & Matsumoto, T. Transcription promotes recA-independent recombination mediated by DNA-dependent RNA polymerase in Escherichia coli. *Proceedings of the National Academy of Sciences* **76**, 4571-4575, doi:doi.org/10.1073/pnas.76.9.4571 (1979).
190 Ikeda, H. & Kobayashi, I. Involvement of DNA-dependent RNA polymerase in a recA-independent pathway of genetic recombination in Escheria coli. *Proceedings of the National Academy of Sciences* **74**, 3932-3936, doi:10.1073/pnas.74.9.3932 (1977).
191 Voelkel-Meiman, K., Keil, R. L. & Roeder, G. S. Recombination-stimulating sequences in yeast ribosomal DNA correspond to sequences regulating transcription by RNA polymerase I. *Cell* **48**, 1071-1079, doi:10.1016/0092-8674(87)90714-8 (1987).
192 Stewart, S. E. & Roeder, G. S. Transcription by RNA polymerase I stimulates mitotic recombination in Saccharomyces cerevisiae. *Molecular and cellular biology* **9**, 3464-3472, doi:10.1128/mcb.9.8.3464-3472.1989 (1989).
193 Dul, J. L. & Drexler, H. Transcription stimulates recombination II. Generalized transduction of Escherichia coli by phages T1 and T4. *Virology* **162**, 471-477, doi:10.1016/0042-6822(88)90489-8 (1988).
194 Thomas, B. J. & Rothstein, R. Elevated recombination rates in transcriptionally active DNA. *Cell* **56**, 619-630, doi:10.1016/0092-8674(89)90584-9 (1989).
195 Nickoloff, J. A. & Reynolds, R. J. Transcription stimulates homologous recombination in mammalian cells. *Molecular and Cellular Biology* **10**, 4837-4845, doi:10.1128/mcb.10.9.4837-4845.1990 (1990).

196 Vilette, D., Uzest, M., Ehrlich, S. D. & Michel, B. DNA transcription and repressor binding affect deletion formation in Escherichia coli plasmids. *The EMBO Journal* **11**, 3629-3634, doi:10.1002/j.1460-2075.1992.tb05447.x (1992).
197 Gaillard, H., Herrera-Moyano, E. & Aguilera, A. Transcription-associated genome instability. *Chemical reviews* **113**, 8638-8661, doi:10.1021/cr400017y (2013).
198 Postow, L. *et al.* Positive torsional strain causes the formation of a four-way junction at replication forks. *Journal of Biological Chemistry* **276**, 2790-2796, doi:10.1074/jbc.M006736200 (2001).
199 Olavarrieta, L., Hernández, P., Krimer, D. B. & Schvartzman, J. B. DNA Knotting Caused by Head-on Collision of Transcription and Replication. *Journal of Molecular Biology* **322**, 1-6, doi:10.1016/s0022-2836(02)00740-4 (2002).
200 Garcia-Rubio, M. L. & Aguilera, A. Topological constraints impair RNA polymerase II transcription and causes instability of plasmid-borne convergent genes. *Nucleic acids research* **40**, 1050-1064, doi:10.1093/nar/gkr840 (2012).
201 Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M. E. & Nudler, E. Linking RNA polymerase backtracking to genome instability in E. coli. *Cell* **146**, 533-543, doi:10.1016/j.cell.2011.07.034 (2011).
202 Santos-Rosa, H. & Aguilera, A. Isolation and genetic analysis of extragenic suppressors of the hyper-deletion phenotype of the Saccharomyces cerevisiae hpr1 delta mutation. *Genetics* **139**, 57-66, doi:10.1093/genetics/139.1.57 (1995).
203 Piruat, J. I. & Aguilera, A. A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. *The EMBO journal* **17**, 4859-4872, doi:10.1093/emboj/17.16.4859 (1998).
204 Piruat, J. I. & Aguilera, A. Mutations in the yeast SRB2 general transcription factor suppress hpr1-induced recombination and show defects in DNA repair. *Genetics* **143**, 1533-1542, doi:10.1093/genetics/143.4.1533 (1996).
205 Fan, H. Y., Cheng, K. K. & Klein, H. L. Mutations in the RNA polymerase II transcription machinery suppress the hyperrecombination mutant hpr1 Δ of Saccharomyces cerevisiae. *Genetics* **142**, 749-759, doi:10.1093/genetics/142.3.749 (1996).
206 Lin, Y., Dent, S. Y., Wilson, J. H., Wells, R. D. & Napierala, M. R loops stimulate genetic instability of CTG.CAG repeats. *Proceedings of the National Academy of Sciences* **107**, 692-697, doi:10.1073/pnas.0909740107 (2010).
207 Grabczyk, E., Mancuso, M. & Sammarco, M. C. A persistent RNA.DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP in vitro. *Nucleic acids research* **35**, 5351-5359, doi:10.1093/nar/gkm589 (2007).
208 Marriott, H. *Genome Architecture and DNA Replication in Haloferax volcanii* Doctor of Philosophy thesis, University of Nottingham, (2017).
209 Cline, S. W., Schalkwyk, L. C. & Doolittle, W. F. Transformation of the archaebacterium Halobacterium volcanii with genomic DNA. *Journal of bacteriology* **171**, 4987-4991, doi:10.1139/m89-022 (1989).
210 Holmes, M. L., Nuttall, S. D. & Dyall-Smith, M. L. Construction and use of halobacterial shuttle vectors and further studies on Haloferax DNA gyrase. *Journal of Bacteriology* **173**, 3807-3813, doi:10.1128/jb.173.12.3807-3813.1991 (1991).
211 Howley, P. M., Israel, M. A., Law, M. F. & Martin, M. A. A rapid method for detecting and mapping homology between heterologous DNAs. Evaluation of polyomavirus

genomes. *Journal of Biological Chemistry* **254**, 4876-4883, doi:10.1016/s0021-9258(17)30093-5 (1979).
212 Beverley, S. M. Estimation of circular DNA size using γ-irradiation and pulsed-field gel electrophoresis. *Analytical biochemistry* **177**, 110-114, doi:10.1016/0003-2697(89)90023-7 (1989).
213 Bitan-Banin, G., Ortenberg, R. & Mevarech, M. Development of a gene knockout system for the halophilic archaeon *Haloferax volcanii* by use of the *pyrE* gene. *Journal of bacteriology* **185**, 772-778, doi:10.1128/JB.185.3.772-778.2003 (2003).
214 Szybalski, W. & Bryson, V. Genetic studies on microbial cross resistance to toxic agents I., Cross resistance of Escherichia coli to fifteen antibiotics. *Journal of bacteriology* **64**, 489-499, doi:10.1128/jb.64.4.489-499.1952 (1952).
215 McCready, S. The repair of ultraviolet light-induced DNA damage in the halophilic archaebacteria, Halobacterium cutirubrum, Halobacterium halobium and Haloferax volcanii. *Mutation Research/DNA Repair* **364**, 25-32, doi:10.1016/0921-8777(96)00018-3 (1996).
216 Wilbanks, E. G. *et al.* A workflow for genome-wide mapping of archaeal transcription factors with ChIP-seq. *Nucleic acids research* **40**, e74, doi:10.1093/nar/gks063 (2012).
217 Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306-2309, doi:10.1126/science.290.5500.2306 (2000).
218 Komori, K. *et al.* Both RadA and RadB are involved in homologous recombination in Pyrococcus furiosus. *Journal of Biological Chemistry* **275**, 33782-33790, doi:10.1074/jbc.M004557200 (2000).
219 Sung, P. Yeast Rad55 and Rad57 proteins form a heterodimer that functions with replication protein A to promote DNA strand exchange by Rad51 recombinase. *Genes & development* **11**, 1111-1121, doi:10.1101/gad.11.9.1111 (1997).
220 Sugiyama, T. & Kowalczykowski, S. C. Rad52 protein associates with replication protein A (RPA)-single-stranded DNA to accelerate Rad51-mediated displacement of RPA and presynaptic complex formation. *Journal of Biological Chemistry* **277**, 31663-31672, doi:10.1074/jbc.M203494200 (2002).
221 Liu, J., Doty, T., Gibson, B. & Heyer, W. D. Human BRCA2 protein promotes RAD51 filament formation on RPA-covered single-stranded DNA. *Nature structural & molecular biology* **17**, 1260-1262, doi:10.1038/nsmb.1904 (2010).
222 Wardell, K. *et al.* RadB acts in homologous recombination in the archaeon Haloferax volcanii, consistent with a role as recombination mediator. *DNA Repair* **55**, 7-16, doi:10.1016/j.dnarep.2017.04.005 (2017).
223 Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854-858, doi:10.1038/nature07730 (2009).
224 Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497-1502, doi:10.1126/science.1141319 (2007).
225 Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature methods* **4**, 651-657, doi:10.1038/nmeth1068 (2007).
226 Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837, doi:10.1016/j.cell.2007.05.009 (2007).

227 Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560, doi:10.1038/nature06008 (2007).
228 Narlikar, L. & Jothi, R. in *Next Generation Microarray Bioinformatics* 305-322 (Springer, 2012).
229 Liao, Y. *et al.* CdrS Is a Global Transcriptional Regulator Influencing Cell Division in Haloferax volcanii. *mBio* **12**, e0141621, doi:10.1128/mBio.01416-21 (2021).
230 Mondragon, P. *et al.* TrmB family transcription factor as a thiol-based regulator of oxidative stress response. *bioRxiv*, doi:10.1101/2022.03.05.483106 (2022).
231 Sakrikar, S. & Schmid, A. K. An archaeal histone-like protein regulates gene expression in response to salt stress. *Nucleic acids research* **49**, 12732-12743, doi:10.1093/nar/gkab1175 (2021).
232 Wang, K. *Transcription regulation and growth phase transition in hyperthermoacidophilic archaea*, Department of Molecular Biosciences, The Wenner-Gren Institute, Stockholm, (2018).
233 Cockram, C. A., Filatenkova, M., Danos, V., El Karoui, M. & Leach, D. R. Quantitative genomic analysis of RecA protein binding during DNA double-strand break repair reveals RecBCD action in vivo. *Proc Natl Acad Sci U S A* **112**, E4735-4742, doi:10.1073/pnas.1424269112 (2015).
234 Smith, G. R. How RecBCD enzyme and Chi promote DNA break repair and recombination: a molecular biologist's view. *Microbiol Mol Biol Rev* **76**, 217-228, doi:10.1128/MMBR.05026-11 (2012).
235 Kang, K. *et al.* Epigenomic Analysis of RAD51 ChIP-seq Data Reveals cis-regulatory Elements Associated with Autophagy in Cancer Cell Lines. *Cancers (Basel)* **13**, doi:10.3390/cancers13112547 (2021).
236 Hopp, T. P. *et al.* A short polypeptide marker sequence useful for recombinant protein identification and purification. *Bio/technology* **6**, 1204-1210, doi:10.1038/nbt1088-1204 (1988).
237 Einhauer, A. & Jungbauer, A. Affinity of the monoclonal antibody M1 directed against the FLAG peptide. *Journal of chromatography* **921**, 25-30, doi:10.1016/S0021-9673(01)00831-7 (2001).
238 Einhauer, A. & Jungbauer, A. The FLAG™ peptide, a versatile fusion tag for the purification of recombinant proteins. *Journal of biochemical and biophysical methods* **49**, 455-465, doi:10.1016/S0165-022X(01)00213-5 (2001).
239 Zhang, H. Y., Wei, J. W., Qian, W. & Deng, C. Y. Analysis of HrpG regulons and HrpG-interacting proteins by ChIP-seq and affinity proteomics in Xanthomonas campestris. *Molecular plant pathology* **21**, 388-400, doi:10.1111/mpp.12903 (2020).
240 Shin, S.-I. *et al.* Z-DNA-forming sites identified by ChIP-Seq are associated with actively transcribed regions in the human genome. *DNA Research* **23**, 477-486, doi:10.1093/dnares/dsw031 (2016).
241 Perkins, T. T. *et al.* ChIP-seq and transcriptome analysis of the OmpR regulon of Salmonella enterica serovars Typhi and Typhimurium reveals accessory genes implicated in host colonization. *Molecular microbiology* **87**, 526-538, doi:10.1111/mmi.12111 (2013).
242 Xiong, X. *et al.* A Scalable Epitope Tagging Approach for High Throughput ChIP-Seq Analysis. *ACS synthetic biology* **6**, 1034-1042, doi:10.1021/acssynbio.6b00358 (2017).

243 Savic, D. *et al.* CETCh-seq: CRISPR epitope tagging ChIP-seq of DNA-binding proteins. *Genome research* **25**, 1581-1589, doi:10.1101/gr.193540.115 (2015).
244 Jones, N. J. *The Role of Homologous Recombination in Haloferax volcanii* Doctor of Philosophy thesis, University of Nottingham, (2019).
245 Sancar, G. B. Enzymatic photoreactivation: 50 years and counting. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **451**, 25-37, doi:10.1016/S0027-5107(00)00038-5 (2000).
246 Sinha, R. P. & Hader, D. P. UV-induced DNA damage and repair: a review. *Photochemical & Photobiological Sciences* **1**, 225-236, doi:10.1039/b201230h (2002).
247 Unscear, S. Effects of Ionizing Radiation. *United Nations, New York*, 453-487 (2000).
248 Cadet, J. *et al.* Hydroxyl radicals and DNA base damage. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **424**, 9-21, doi:10.1016/S0027-5107(99)00004-4 (1999).
249 Wallace, S. S. Enzymatic processing of radiation-induced free radical damage in DNA. *Radiation research* **150**, S60-S79, doi:10.2307/3579809 (1998).
250 Patel, H. *et al.* nf-core/chipseq: nf-core/chipseq v1.2.2 - Rusty Mole. doi:10.5281/ZENODO.4711243 (2021).
251 Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048, doi:10.1093/bioinformatics/btw354 (2016).
252 Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**, 1813-1831, doi:10.1101/gr.136184.111 (2012).
253 Love, M., Anders, S. & Huber, M. Differential gene expression analysis based on the negative binomial distribution. *Genome Biol* **15**, 550 (2014).
254 Liu, T. (Github, 2015).
255 Trieselmann, B. A. & Charlebois, R. L. Transcriptionally active regions in the genome of the archaebacterium Haloferax volcanii. *Journal of bacteriology* **174**, 30-34, doi:10.1128/jb.174.1.30-34.1992 (1992).
256 Eichler, J. & Maupin-Furlow, J. Post-translation modification in Archaea: lessons from Haloferax volcanii and other haloarchaea. *FEMS microbiology reviews* **37**, 583-606, doi:10.1111/1574-6976.12012 (2013).
257 Ludt, K. & Soppa, J. Influence of Origin Recognition Complex Proteins on the Copy Numbers of Three Chromosomes in *Haloferax volcanii*. *Journal of bacteriology* **200**, doi:10.1128/JB.00161-18 (2018).
258 Haque, R. U., Paradisi, F. & Allers, T. *Haloferax volcanii* as immobilised whole cell biocatalyst: new applications for halophilic systems. *Applied microbiology and biotechnology* **103**, 3807-3817, doi:10.1007/s00253-019-09725-y (2019).
259 Large, A. *et al.* Characterization of a tightly controlled promoter of the halophilic archaeon *Haloferax volcanii* and its use in the analysis of the essential cct1 gene. *Molecular microbiology* **66**, 1092-1106, doi:10.1111/j.1365-2958.2007.05980.x (2007).
260 Torrey, T. A. & Kogoma, T. Suppressor mutations (*rin*) that specifically suppress the recA+ dependence of stable DNA replication in *Escherichia coli* K-12. *Molecular and General Genetics MGG* **187**, 225-230, doi:10.1007/BF00331121 (1982).

261 Kogoma, T., Barnard, K. G. & Hong, X. RecA, Tus protein and constitutive stable DNA replication in *Escherichia coli rnhA* mutants. *Molecular and General Genetics MGG* **244**, 557-562, doi:10.1007/BF00583907 (1994).
262 Stuckey, R., Garcia-Rodriguez, N., Aguilera, A. & Wellinger, R. E. Role for RNA:DNA hybrids in origin-independent replication priming in a eukaryotic system. *Proceedings of the National Academy of Sciences* **112**, 5779-5784, doi:10.1073/pnas.1501769112 (2015).
263 Filee, J., Forterre, P., Sen-Lin, T. & Laurent, J. Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins. *Journal of Molecular Evolution* **54**, 763-773, doi:10.1007/s00239-001-0078-x (2002).
264 Ito, J. & Braithwaite, D. K. Compilation and alignment of DNA polymerase sequences. *Nucleic acids research* **19**, 4045, doi:10.1093/nar/19.15.4045 (1991).
265 Nakamura, T. M. & Cech, T. R. Reversing time: origin of telomerase. *Cell* **92**, 587-590, doi:10.1016/S0092-8674(00)81123-X (1998).
266 Allen, J. M. *et al.* Roles of DNA polymerase I in leading and lagging-strand replication defined by a high-resolution mutation footprint of ColE1 plasmid replication. *Nucleic acids research* **39**, 7020-7033, doi:10.1093/nar/gkr157 (2011).
267 Hübscher, U., Maga, G. & Spadari, S. Eukaryotic DNA polymerases. *Annual review of biochemistry* **71**, 133-163, doi:10.1146/annurev.biochem.71.090501.150041 (2002).
268 Raia, P., Delarue, M. & Sauguet, L. An updated structural classification of replicative DNA polymerases. *Biochemical Society Transactions* **47**, 239-249, doi:10.1042/BST20180579 (2019).
269 Kazlauskas, D., Krupovic, M., Guglielmini, J., Forterre, P. & Venclovas, Č. Diversity and evolution of B-family DNA polymerases. *Nucleic acids research* **48**, 10142-10156, doi:10.1093/nar/gkaa760 (2020).
270 Makarova, K. S., Krupovic, M. & Koonin, E. V. Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replication machinery. *Frontiers in microbiology* **5**, 354, doi:10.3389/fmicb.2014.00354 (2014).
271 Ishino, Y., Komori, K., Cann, I. K. & Koga, Y. A novel DNA polymerase family found in Archaea. *Journal of bacteriology* **180**, 2232-2236, doi:10.1128/JB.180.8.2232-2236.1998 (1998).
272 MacNeill, S. A. The haloarchaeal chromosome replication machinery. *Biochemical Society Transactions* **37**, 108-113, doi:10.1042/BST0370108 (2009).
273 Sauguet, L., Raia, P., Henneke, G. & Delarue, M. Shared active site architecture between archaeal PolD and multi-subunit RNA polymerases revealed by X-ray crystallography. *Nature communications* **7**, 12227, doi:10.1038/ncomms12227 (2016).
274 Cubonova, L. *et al.* Archaeal DNA polymerase D but not DNA polymerase B is required for genome replication in *Thermococcus kodakarensis*. *Journal of bacteriology* **195**, 2322-2328, doi:10.1128/JB.02037-12 (2013).
275 Technologies, O. N. *Advantages of nanopore sequencing*, <https://nanoporetech.com/how-it-works/advantages-nanopore-sequencing> (
276 Tyson, J. R. *et al.* MinION-based long-read sequencing and assembly extends the Caenorhabditis elegans reference genome. *Genome research* **28**, 266-274, doi:10.1101/gr.221184.117 (2018).
277 Madden, T. The BLAST sequence analysis tool. *The NCBI handbook* (2003).

278 Belton, J. M. *et al.* Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268-276, doi:10.1016/j.ymeth.2012.05.001 (2012).
279 Lioy, V. S. *et al.* Dynamics of the compartmentalized Streptomyces chromosome during metabolic differentiation. *Nature Communications* **12**, 5221, doi:10.1038/s41467-021-25462-1 (2021).
280 Le, T. B. & Laub, M. T. Transcription rate and transcript length drive formation of chromosomal interaction domain boundaries. *The EMBO journal* **35**, 1582-1595, doi:10.15252/embj.201593561 (2016).

Chapter 7- Appendices

DNA sequencing methodologies written and carried out by the Deep Seq facility:

Oxford Nanopore MinION sequencing

Whole-genome sequencing libraries were prepared from nine *Haloferax volcanii* genomic DNA samples. DNA samples were diluted 1/10 with nuclease-free water and needle sheared 30X with a 26G needle (BD; 300300). Sheared DNA was cleaned up using 0.5X AMPure XP beads (Beckman Coulter; A63882) and eluted in nuclease-free water. DNA concentrations were quantified using the Qubit 4 Fluorometer (Thermo Fisher Scientific) and the Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific; Q32853). Sequencing libraries were prepared using the Native Barcoding genomic DNA protocol (Oxford Nanopore Technologies; Version: NBE_9065_v109_revS_14Aug2019). This protocol used the Ligation Sequencing Kit (Oxford Nanopore Technologies; SQK-LSK109) and the Native Barcoding Expansion 1-12 kit (Oxford Nanopore Technologies; EXP-NBD104). Barcoded samples were pooled in equimolar amounts for the final sequencing-adapter ligation. All purification steps were performed using AMPure XP beads (Beckman Coulter; A63882). The final barcoded library pool was loaded onto a MinION flow cell (Oxford Nanopore Technologies; FLO-MIN106 R9.4.1) and run on the GridION X5 mk1.

ChIP-Seq

ChIP-Seq libraries were prepared using the NEBNext Ultra II DNA library Prep Kit for Illumina (NEB; E7645) and NEBNext Multiplex Oligos for Illumina; 96 Unique Dual Index Primer Pairs (NEB; E6440). For each condition, libraries were prepared from two input sample and three ChIP samples. Sample concentrations were measured using the Qubit Fluorometer and the Qubit dsDNA HS Assay Kit (ThermoFisher Scientific; Q32854). Sample fragment-length distributions were assessed using the Agilent 4200 TapeStation and the High Sensitivity D5000 ScreenTape assay (Agilent; 5067-5592, 5067-5593) and again using the Agilent 2100 Bioanalyzer and the Agilent High Sensitivity DNA Kit (Agilent; 5067-4626).

Due to the variability in fragment-size distributions, a double-sided SPRI size-selection step was performed during the clean-up step following adapter ligation. The first bead addition volume was 15 ul and the second was 25 ul. PCR enrichment of the adaptor-ligated libraries was performed using 6 cycles of amplification.

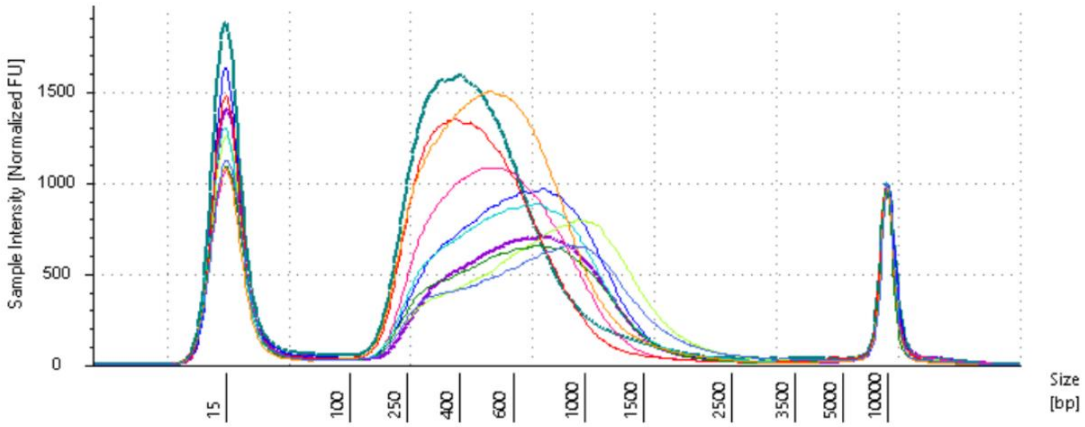
Libraries were quantified using the Qubit Fluorometer and the Qubit dsDNA HS Assay Kit (ThermoFisher Scientific; Q32854) and library fragment-size distributions were assessed using the Agilent 4200 TapeStation and the Agilent High Sensitivity D5000 ScreenTape assay (Agilent; 5067-5592, 5067-5593). Libraries were pooled in equimolar amounts, based on TapeStation region analysis, and final library quantification performed using the KAPA Library Quantification Kit for Illumina

(Roche; KK4824). The library pool was sequenced on the Illlumina NextSeq 500 on a NextSeq 500 High Output 150 cycle kits (Illumina; 20024907), to generate over 20-30 million pairs of 75-bp paired-end reads per sample.

Table 7.1: Notation of samples as detailed in Chapter 3 and in this chapter (Appendices). Sequencing of samples LOD1-WCE and HOD3-WCE was not necessary as three IP samples and two WCE samples were used per condition (LOD and HOD). These samples showed the largest fragment sizes out of the WCE samples.

Notation in Chapter 3	Deep Seq sample number	Barcode
LOD1 IP-X (LOD_IP_REP1)	ds1132_1	A11
LOD2 IP-X (LOD_IP_REP2)	ds1132_2	B11
LOD3 IP-X (LOD_IP_REP3)	ds1132_3	C11
HOD1 IP-X (HOD_IP_REP1)	ds1132_4	D11
HOD2 IP-X (HOD_IP_REP2)	ds1132_5	E11
HOD3 IP-X (HOD_IP_REP3)	ds1132_6	F11
LOD1 WCE-X	ds1132_7	
LOD2 WCE-X (LOD_WCE_REP2)	ds1132_8	H11
LOD3 WCE-X	ds1132_9	A12
HOD1 WCE-X (HOD_WCE_REP1)	ds1132_10	B12
HOD2 WCE-X	ds1132_11	C12
HOD3 WCE-X	ds1132_12	

High Sensitivity D5000 ScreenTape information for ChIP-Seq samples (carried out by the Deep Seq facility)

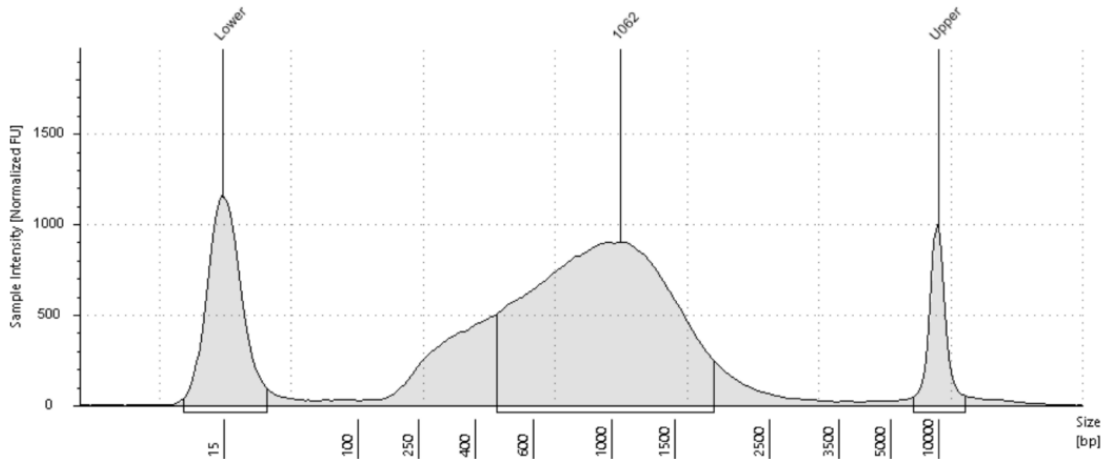


Samples

File	Indicator	Well ID	Comment
File 1		A2	ds1132_1
File 1		B2	ds1132_2
File 1		C2	ds1132_3
File 1		D2	ds1132_4
File 1		E2	ds1132_5
File 1		F2	ds1132_6
File 1		H2	ds1132_8
File 1		B1	ds1132_9
File 1		C1	ds1132_10
File 1		D1	ds1132_11

Figure 7.1: Final size distribution of fragments in the ChIP-Seq samples prepared in Chapter 3. Samples were size selected prior to library creation. Samples ds1132_7 (LOD1 WCE-X) and ds1132_12 (HOD3 WCE-X) were not included in this figure as their sequencing was not necessary and their fragment lengths were the longest of the WCE samples for the LOD and HOD conditions respectively (seen in figures below). Two standards (15bp and 10,000bp) were used for calibration. The majority of fragments can be seen falling between 100 and 1500bp.

G2: ds1132_7

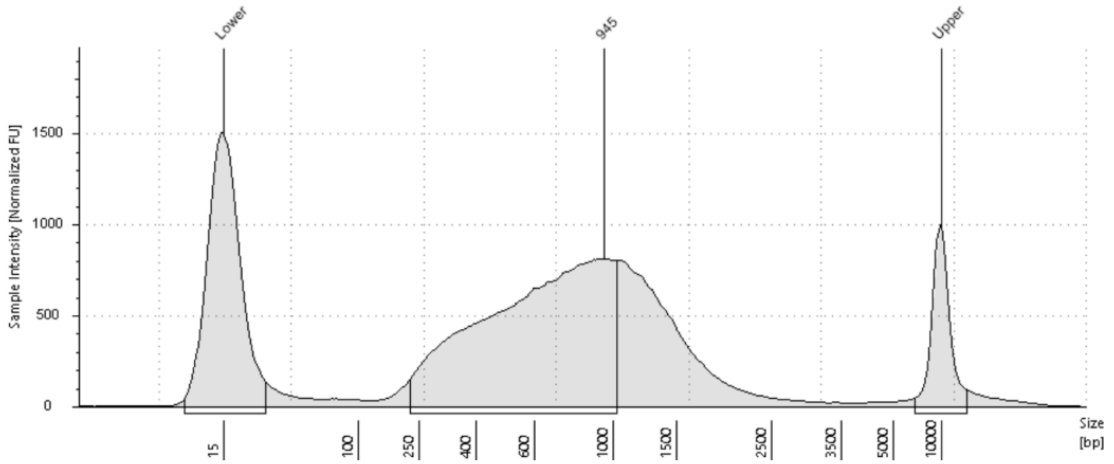


Sample Table

Well	Conc. [pg/µl]	Sample Description	Alert	Observations
G2	1730	ds1132_7		

Figure 7.2: Fragment sizes in sample ds1132_7 (LOD1 WCE-X). The figure shows three peaks: the length of fragments in this sample at 1062bp, and two standard lengths used for calibration at 15bp and 10,000bp. This sample was not sequenced due to fragment length.

E1: ds1132_12



Sample Table

Well	Conc. [pg/µl]	Sample Description	Alert	Observations
E1	1270	ds1132_12		

Figure 7.3: Fragment sizes in sample ds1132_12 (HOD3 WCE-X). The figure shows three peaks: the length of fragments in this sample at 945bp, and two standard lengths used for calibration at 15bp and 10,000bp. This sample was not sequenced due to fragment length.

Chromosome	Start	End	Peaks	Samples	HOD_IP_REP1.pval	HOD_IP_REP2.pval	HOD_IP_REP3.pval	LOD_IP_REP1.pval	LOD_IP_REP2.pval	LOD_IP_REP3.pval	Annotation	Distance to TSS	Nearest PromoterID	Nearest gene	Gene name	Description
Main chromosome	541794	567045	8	4	NA	20.4938	4.68562;4.00772;11.244;3	5.59648	NA	7.16036	TTS (HVO_A0619)	461	HVO_A0618	HVO_A0618		ferredoxin-NAD+ reductase
Main chromosome	641615	646225	4	4	NA	38.0442	13.331	4.48588	NA	6.60607	promoter-TSS (HVO_A0524)	-763	HVO_A0524	HVO_A0524	3-hydroxyacyl-CoA	3-hydroxyacyl-CoA dehydrogenase
Main chromosome	898407	915128	7	4	NA	17.4065;12.5885	5.37078	4.72091;3.57163	NA	5.27401	TTS (HVO_0301)	742	HVO_0302	HVO_0302		hypothetical protein (ABC-type multidrug transport system, permease component)
Main chromosome	1299695	1337026	8	4	3.61193	40.219	16.6471;3.85494;7.86105	NA	NA	6.77394;4.49177	promoter-TSS (HVO_0761)	-276	HVO_0761	HVO_0761		hypothetical protein (Uncharacterized conserved protein)
Main chromosome	1679351	1683190	4	4	NA	22.6726	11.8996	4.64046	NA	4.29122	promoter-TSS (HVO_1143)	436	HVO_1145	HVO_1145	rps3aR	30S ribosomal protein S3a.eR
Main chromosome	2660736	2687823	9	4	NA	11.9013;26.9585	13.909	4.12497;4.5317;1	NA	7.57215	TTS (HVO_2167)	574	HVO_2168	HVO_2168	moxR3	methanol dehydrogenase regulatory protein
Main chromosome	2975710	2992724	4	4	NA	19.4962	9.53598	3.81337	NA	4.36235	promoter-TSS (HVO_2479)	308	HVO_2481	HVO_2481		phosphohydrolase
Main chromosome	3052114	3072434	4	4	3.39782	25.4962	12.2666	NA	NA	5.02916	TTS (HVO_2576)	288	HVO_2575	HVO_2575		translation elongation factor aEF-1 subunit alpha-like protein
Main chromosome	296722	302453	3	3	NA	9.59964	3.3382	4.25471	NA	NA	TTS (HVO_A0230)	454	HVO_A0231	HVO_A0231		hypothetical protein (rolling circle replication protein Rep)
Main chromosome	339411	344800	3	3	NA	10.6344	3.87908	4.05088	NA	NA	promoter-TSS (HVO_A0193)	9	HVO_A0194	HVO_A0194		glycosyltransferase
Main chromosome	583243	587522	3	3	NA	32.4685	11.5687	5.56949	NA	NA	promoter-TSS (HVO_A0582)	565	HVO_A0581	HVO_A0581		MmgE/PrpD family protein
Main chromosome	590548	610620	5	3	NA	21.1647	5.52447;7.14684;1.735	4.03598	NA	NA	TTS (HVO_A0565)	689	HVO_A0566	HVO_A0566		3-isopropylmalate dehydratase small subunit
Main chromosome	647462	656539	3	3	NA	10.7894	4.53735	4.91441	NA	NA	promoter-TSS (HVO_A0514)	343	HVO_A0515	HVO_A0515	livG4	branched-chain amino acid ABC transporter ATP-binding protein
Main chromosome	658179	659264	3	3	3.57427	19.0111	4.57156	NA	NA	NA	promoter-TSS (HVO_A0505)	30	HVO_A0505	HVO_A0505	fadA4	enoyl-CoA hydratase
Main chromosome	948902	952087	3	3	NA	5.62301	4.05688	4.85255	NA	NA	promoter-TSS (HVO_0349)	-445	HVO_0349	HVO_0349	rpoA1	DNA-directed RNA polymerase subunit A
Main chromosome	1010129	1015153	3	3	3.01779	22.805	7.23825	NA	NA	NA	promoter-TSS (HVO_0421)	-411	HVO_0421	HVO_0421	pinA	P-loop ATPase of the PilT family
Main chromosome	1155048	1164863	4	3	NA	42.3705	15.5355;4.81436	NA	NA	5.13758	promoter-TSS (HVO_0592)	-103	HVO_0592	HVO_0592		adenosylcobinamide amidohydrolase
Main chromosome	1476536	1490670	4	3	NA	32.3244	13.5084	4.41054;5.9185	NA	NA	promoter-TSS (HVO_0933)	685	HVO_0935	HVO_0935	fdhA	formate dehydrogenase subunit alpha
Main chromosome	1575614	1578834	3	3	NA	10.3618	11.0584	8.37978	NA	NA	promoter-TSS (HVO_1032)	-714	HVO_1032	HVO_1032		methyltransferase
Main chromosome	1655128	1678130	3	3	NA	15.2781	8.73725	3.89036	NA	NA	promoter-TSS (HVO_1130)	97	HVO_1130	HVO_1130		hypothetical protein (Uncharacterized conserved protein)
Main chromosome	1747425	1775312	5	3	NA	18.1562	5.5949;8.77618;6.62205	NA	NA	4.64182	promoter-TSS (HVO_1234)	-138	HVO_1234	HVO_1234	phr3	cryptochrome/photolyase-related protein Phr3
Main chromosome	1825409	1873972	5	3	NA	25.0172;31.6796	10.1572;11.6353	3.82138	NA	NA	promoter-TSS (HVO_1334)	208	HVO_1333	HVO_1333	lhr2	ATP-dependent DNA helicase
Main chromosome	1886789	1926763	6	3	NA	25.8749;27.7229	9.11556;15.2222	3.851	NA	NA	TTS (HVO_1398)	329	HVO_1396	HVO_1396		P-hydroxybenzoate hydroxylase
Main chromosome	1942970	1961149	3	3	NA	15.2078	6.44233	NA	NA	4.88165	promoter-TSS (HVO_1444)	330	HVO_1445	HVO_1445		hypothetical protein (Uncharacterized conserved protein)
Main chromosome	1962765	1970768	3	3	NA	22.4307	9.92551	4.90362	NA	NA	promoter-TSS (HVO_1458)	-175	HVO_1459	HVO_1459		trehalose utilization protein
Main chromosome	1972516	1993496	3	3	NA	30.5143	6.34801	NA	NA	3.92047	promoter-TSS (HVO_1476)	461	HVO_1477	HVO_1477	comA	endonuclease
Main chromosome	2053976	2064071	4	3	NA	14.4862	2.55315;2.71886	4.32445	NA	NA	promoter-TSS (HVO_1553)	678	HVO_1554	HVO_1554	traB	TraB family protein
Main chromosome	2270502	2277606	3	3	NA	17.7088	2.95178	3.96524	NA	NA	promoter-TSS (HVO_1773)	184	HVO_1772	HVO_1772		hypothetical protein (Uncharacterized conserved protein)
Main chromosome	2284925	2286879	3	3	7.07014	21.849	2.51317	NA	NA	NA	TTS (HVO_1783)	899	HVO_1783	HVO_1783		hypothetical protein (Uncharacterized conserved protein)
Main chromosome	2406741	2439793	5	3	NA	29.1471	6.97124;6.86923;6.17079	6.37751	NA	NA	promoter-TSS (HVO_1939)	-523	HVO_1939	HVO_1939	mutLa	DNA mismatch repair protein MutL
Main chromosome	2502424	2509295	4	3	NA	10.7859	8.59917;5.13507	NA	NA	5.59596	promoter-TSS (HVO_2024)	341	HVO_2025	HVO_2025		HTR-like protein
Main chromosome	2621199	2623116	3	3	NA	7.53467	5.61218	5.25452	NA	NA	promoter-TSS (HVO_2121)	263	HVO_2122	HVO_2122	dppF4	putative dipeptides/oligopeptides ABC transporter ATP-binding protein
Main chromosome	2707688	2716320	3	3	NA	21.627	10.0537	4.07885	NA	NA	promoter-TSS (HVO_2211)	169	HVO_2210	HVO_2210		hypothetical protein (Lrp/AsnC family C-terminal domain)
Main chromosome	2718179	2731073	3	3	NA	17.7551	9.08219	5.06189	NA	NA	TTS (HVO_2223)	421	HVO_2225	HVO_2225		succinylglutamate desuccinylase
Main chromosome	2901055	2907787	3	3	NA	7.70568	3.96545	5.4242	NA	NA	TTS (HVO_2399)	444	HVO_2400	HVO_2400		hypothetical protein (Zn finger protein)
Main chromosome	2957777	2967133	3	3	NA	18.7372	13.1215	4.82872	NA	NA	promoter-TSS (HVO_2454)	-259	HVO_2455	HVO_2455	trpF	N-(5phosphoribosyl)anthranilate isomerase
Main chromosome	2994305	3007534	4	3	NA	7.0744;17.9549	7.24876	5.65463	NA	NA	TTS (HVO_2499)	526	HVO_2500	HVO_2500	cat2	cationic amino acid transporter
Main chromosome	3025095	3036990	4	3	NA	16.4342;4.65535	7.52158	4.45439	NA	NA	promoter-TSS (HVO_2528)	-1	HVO_2529	HVO_2529		short-chain family oxidoreductase
Main chromosome	3038902	3050264	4	3	NA	30.9432;4.01351	9.64164	3.89462	NA	NA	promoter-TSS (HVO_2543)	275	HVO_2545	HVO_2545	rpl18	50S ribosomal protein L18
Main chromosome	3184572	3199699	4	3	NA	11.2975;10.2974	6.01638	5.13562	NA	NA	promoter-TSS (HVO_2709)	200	HVO_2711	HVO_2711		hypothetical protein (Predicted membrane protein, DoxD family)
Main chromosome	3212348	3224243	5	3	NA	35.8956;6.82268	10.3458;2.29486	NA	NA	3.8774	promoter-TSS (HVO_2736)	-107	HVO_2736	HVO_2736	tmcA	hypothetical protein (Predicted P-loop ATPase fused to an acetyltransferase)
Main chromosome	3241391	3254038	3	3	NA	12.7003	4.25499	5.90502	NA	NA	promoter-TSS (HVO_2770)	438	HVO_2771	HVO_2771		hypothetical protein (Uncharacterized conserved protein, YjeF-related)
Main chromosome	3346641	3351499	3	3	3.96022	16.4578	3.91534	NA	NA	NA	promoter-TSS (HVO_2878)	720	HVO_2877	HVO_2877		putative sugar transporter
pHV3	49459	54759	2	2	NA	15.8983	3.67039	NA	NA	NA	promoter-TSS (HVO_B0044)	-106	HVO_B0044	HVO_B0044		
pHV3	56757	80818	2	2	NA	18.5334	6.8092	NA	NA	NA	promoter-TSS (HVO_B0056)	245	HVO_B0057	HVO_B0057		
pHV3	146238	159121	3	2	NA	27.5371	2.70236;2.89584	NA	NA	NA	promoter-TSS (HVO_B0136)	532	HVO_B0137	HVO_B0137		
pHV3	193268	197805	2	2	NA	19.1772	2.03112	NA	NA	NA	promoter-TSS (HVO_B0169)	-218	HVO_B0169	HVO_B0169		
pHV3	310155	319635	3	2	NA	20.0494	4.49981;3.31444	NA	NA	NA	TTS (HVO_B0264)	273	HVO_B0265	HVO_B0265		
pHV3	348915	362011	2	2	NA	19.4781	5.48044	NA	NA	NA	TTS (HVO_B0308)	696	HVO_B0309	HVO_B0309		
pHV3	386792	395258	2	2	NA	15.5097	3.89817	NA	NA	NA	promoter-TSS (HVO_B0335)	-258	HVO_B0336	HVO_B0336		
pHV3	422693	428384	2	2	NA	21.5163	3.93393	NA	NA	NA	promoter-TSS (HVO_B0369)	72	HVO_B0369	HVO_B0369		
pHV3	433297	436065	2	2	NA	18.1404	2.23109	NA	NA	NA	TTS (HVO_B0380)	352	HVO_B0379	HVO_B0379		
Main chromosome	5013	23862	4	2	NA	10.0366	2.8683;3.48618;5.19933	NA	NA	NA	promoter-TSS (HVO_0015)	433	HVO_0014	HVO_0014		
Main chromosome	35375	37803	2	2	NA	7.67969	2.6944	NA	NA	NA	TTS (HVO_0037)	563	HVO_0038	HVO_0038		
Main chromosome	39627	48537	3	2	NA	14.7754	2.78273;4.49744	NA	NA	NA	promoter-TSS (HVO_0043)	252	HVO_0044	HVO_0044		
Main chromosome	68112	70943	2	2	NA	21.2677	4.87157	NA	NA	NA	promoter-TSS (HVO_0066)	-79	HVO_0068	HVO_0068		
Main chromosome	74928	97101	4	2	NA	17.6309	5.69748;4.78595;3.07518	NA	NA	NA	promoter-TSS (HVO_0084)	-149	HVO_0085	HVO_0085		
Main chromosome	107846	111780	2	2	NA	9.47507	2.91993	NA	NA	NA	promoter-TSS (HVO_0112)	259	HVO_0113	HVO_0113		
Main chromosome	113337	115768	2	2	NA	18.049	3.77696	NA	NA	NA	TTS (HVO_0120)	244	HVO_0121	HVO_0121		
Main chromosome	130440	138036	2	2	NA	11.678	4.18911	NA	NA	NA	promoter-TSS (HVO_0143)	-210	HVO_0143	HVO_0143		
Main chromosome	139567	144373	2	2	NA	19.8696	4.25831	NA	NA	NA	promoter-TSS (HVO_0154)	242	HVO_0153	HVO_0153		
Main chromosome	145982	157778	3	2	NA	15.0958	4.25108;1.81036	NA	NA	NA	promoter-TSS (HVO_0166)	221	HVO_0167	HVO_0167		
Main chromosome	159272	163444	2	2	NA	15.0227	3.2676	NA	NA	NA	promoter-TSS (HVO_0179)	291	HVO_0176	HVO_0176		
Main chromosome	170369	173671	2	2	NA	25.0049	6.82345	NA	NA	NA	promoter-TSS (HVO_0190)	3	HVO_0191	HVO_0191		
Main chromosome	175804	185182	2	2	NA	16.7966	7.66189	NA	NA	NA	promoter-TSS (HVO_0200)	11	HVO_0200	HVO_0200		
Main chromosome	187139	190934	2	2	NA	27.0776	5.97336	NA	NA	NA	promoter-TSS (HVO_0208)	-186	HVO_0208	HVO_0208		
Main chromosome	194973	199043	2	2	NA	33.1001	8.49557	NA	NA	NA	promoter-TSS (HVO_0215)	259	HVO_0216	HVO_0216		
Main chromosome	201682	215814	4	2	NA	27.774	11.3219;3.35749;2.38295	NA	NA	NA	promoter-TSS (HVO_0229)	14	HVO_0230	HVO_0230		
Main chromosome	219003	225149	3	2	NA	25.8376	5.89623;5.67416	NA	NA	NA	promoter-TSS (HVO_0244)	377	HVO_0245	HVO_0245		
Main chromosome	227070	231133	2	2	NA	23.3202	3.89609	NA	NA	NA	promoter-TSS (HVO_0254)	-68	HVO_0255	HVO_0255		
Main chromosome	258460	262251	2	2	NA	17.033	3.0036	NA	NA	NA	promoter-TSS (HVO_A0270)	-611	HVO_A0270	HVO_A0270		
Main chromosome	308196	314302	2	2	NA	41.3085	10.1898	NA	NA	NA	promoter-TSS (HVO_A0219)	-540	HVO_A0219	HVO_A0219		
Main chromosome	334773	336134	2	2	NA	9.83361	2.28141	NA	NA	NA	TTS (HVO_A0198)	492	HVO_A0197	HVO_A0197		
Main chromosome	349280	364144	3	2	NA	21.9985	6.2035;6.08605	NA	NA	NA	TTS (HVO_A0176)	1562	HVO_A0176	HVO_A0176		
Main chromosome	372623	378755	3	2	NA	15.4445	5.79736;6.79319	NA	NA	NA	promoter-TSS (HVO_A0159)	194	HVO_A0158	HVO_A0158		

Main chromosome	383014	384280	2	2	NA	5.8668	5.74352	NA	NA	NA	promoter-TSS (HVO_A0152)	-46	HVO_A0153	HVO_A0153
Main chromosome	570909	574525	2	2	NA	22.2856	4.2124	NA	NA	NA	TTS (HVO_A0595)	693	HVO_A0594	HVO_A0594
Main chromosome	577367	581604	2	2	NA	17.141	5.38692	NA	NA	NA	promoter-TSS (HVO_A0585)	-631	HVO_A0587	HVO_A0587
Main chromosome	613292	615688	2	2	NA	18.4025	2.34557	NA	NA	NA	TTS (HVO_A0554)	1197	HVO_A0553	HVO_A0553
Main chromosome	617603	623715	2	2	NA	16.5227	6.8471	NA	NA	NA	promoter-TSS (HVO_A0548)	213	HVO_A0549	HVO_A0549
Main chromosome	629507	635321	2	2	NA	41.2099	11.1077	NA	NA	NA	TTS (HVO_A0535)	898	HVO_A0536	HVO_A0536
Main chromosome	637208	639690	2	2	NA	17.2126	4.2062	NA	NA	NA	promoter-TSS (HVO_A0530)	-686	HVO_A0530	HVO_A0530
Main chromosome	662715	666317	2	2	NA	22.7446	6.27773	NA	NA	NA	TTS (HVO_A0497)	177	HVO_A0498	HVO_A0498
Main chromosome	671287	677762	2	2	NA	23.1355	8.13913	NA	NA	NA	TTS (HVO_A0487)	286	HVO_A0488	HVO_A0488
Main chromosome	681266	682551	2	2	NA	23.2888	4.40322	NA	NA	NA	TTS (HVO_A0482)	809	HVO_A0481	HVO_A0481
Main chromosome	691539	694157	2	2	NA	10.9837	4.28448	NA	NA	NA	promoter-TSS (HVO_A0471)	-97	HVO_A0471	HVO_A0471
Main chromosome	812342	814291	2	2	NA	7.84102	1.78087	NA	NA	NA	exon (HVO_A0346, exon 1 of 1	178	HVO_A0346	HVO_A0346
Main chromosome	817290	818225	2	2	NA	12.5915	2.49527	NA	NA	NA	promoter-TSS (HVO_A0340)	680	HVO_A0341	HVO_A0341
Main chromosome	821965	826149	2	2	NA	5.29935	5.15788	NA	NA	NA	TTS (HVO_A0335)	1220	HVO_A0335	HVO_A0335
Main chromosome	831250	832372	2	2	NA	11.7761	2.21723	NA	NA	NA	promoter-TSS (HVO_A0326)	758	HVO_A0328	HVO_A0328
Main chromosome	849452	853972	2	2	NA	13.97	5.00655	NA	NA	NA	TTS (HVO_A0304)	969	HVO_A0305	HVO_A0305
Main chromosome	868731	869715	2	2	NA	6.58828	3.0024	NA	NA	NA	exon (HVO_A0292, exon 1 of 1	948	HVO_A0292	HVO_A0292
Main chromosome	918447	925748	2	2	NA	21.9561	4.99317	NA	NA	NA	promoter-TSS (HVO_0318)	-372	HVO_0318	HVO_0318
Main chromosome	927507	934121	2	2	NA	16.664	7.35802	NA	NA	NA	promoter-TSS (HVO_0325)	172	HVO_0324	HVO_0324
Main chromosome	942950	946243	2	2	NA	32.0903	9.26504	NA	NA	NA	promoter-TSS (HVO_0340)	28	HVO_0340	HVO_0340
Main chromosome	955037	958107	2	2	NA	8.7651	2.42685	NA	NA	NA	promoter-TSS (HVO_0353)	-31	HVO_0353	HVO_0353
Main chromosome	961067	962526	2	2	NA	17.7029	6.26481	NA	NA	NA	promoter-TSS (HVO_0358)	47	HVO_0358	HVO_0358
Main chromosome	976172	977051	2	2	NA	4.22411	5.18572	NA	NA	NA	TTS (HVO_0382)	610	HVO_0385	HVO_0385
Main chromosome	980884	983683	2	2	NA	27.6999	6.51994	NA	NA	NA	promoter-TSS (HVO_0391)	40	HVO_0392	HVO_0392
Main chromosome	984022	992573	2	2	NA	18.6708	4.94211	NA	NA	NA	TTS (HVO_0395)	810	HVO_0396	HVO_0396
Main chromosome	994770	998607	2	2	NA	28.0288	6.76008	NA	NA	NA	TTS (HVO_0405)	1581	HVO_0406	HVO_0406
Main chromosome	1000802	1004692	2	2	NA	34.5821	8.92273	NA	NA	NA	promoter-TSS (HVO_0411)	166	HVO_0412	HVO_0412
Main chromosome	1016346	1029825	3	2	NA	27.5954	9.29051;3.85782	NA	NA	NA	promoter-TSS (HVO_0433)	4	HVO_0433	HVO_0433
Main chromosome	1036636	1041128	2	2	NA	4.7913	1.98482	NA	NA	NA	promoter-TSS (HVO_0453)	-66	HVO_0453	HVO_0453
Main chromosome	1043313	1046476	2	2	NA	8.91515	5.14544	NA	NA	NA	promoter-TSS (HVO_0459)	-339	HVO_0460	HVO_0460
Main chromosome	1049250	1050738	2	2	NA	28.648	2.68726	NA	NA	NA	TTS (HVO_0466)	820	HVO_0465	HVO_0465
Main chromosome	1052291	1064776	4	2	NA	19.8099	4.82326;5.7222;4.46609	NA	NA	NA	promoter-TSS (HVO_0472)	288	HVO_0474	HVO_0474
Main chromosome	1066785	1071279	2	2	NA	10.0443	2.79426	NA	NA	NA	promoter-TSS (HVO_0484)	531	HVO_0485	HVO_0485
Main chromosome	1073228	1074162	2	2	NA	13.9886	5.18453	NA	NA	NA	TTS (HVO_0492)	193	HVO_0494	HVO_0494
Main chromosome	1081175	1082752	2	2	NA	10.4353	2.60493	NA	NA	NA	TTS (HVO_0506)	539	HVO_0508	HVO_0508
Main chromosome	1087360	1092802	2	2	NA	10.6042	3.19905	NA	NA	NA	promoter-TSS (HVO_0520)	-50	HVO_0520	HVO_0520
Main chromosome	1101708	1102578	2	2	NA	13.5239	3.97851	NA	NA	NA	promoter-TSS (HVO_0536)	309	HVO_0535	HVO_0535
Main chromosome	1104168	1126299	3	2	NA	20.4385	7.03123;6.97961	NA	NA	NA	promoter-TSS (HVO_0548)	-490	HVO_0548	HVO_0548
Main chromosome	1135613	1140797	2	2	NA	27.0892	7.70126	NA	NA	NA	promoter-TSS (HVO_0566)	94	HVO_0567	HVO_0567
Main chromosome	1143395	1150432	2	2	NA	14.8834	5.5014	NA	NA	NA	TTS (HVO_0573)	134	HVO_0574	HVO_0574
Main chromosome	1180804	1192062	2	2	NA	31.6895	11.5049	NA	NA	NA	promoter-TSS (HVO_0619)	-949	HVO_0619	HVO_0619
Main chromosome	1194617	1195819	2	2	NA	11.0862	2.74792	NA	NA	NA	promoter-TSS (HVO_0628)	-896	HVO_0628	HVO_0628
Main chromosome	1206395	1210228	2	2	NA	56.7696	14.9693	NA	NA	NA	promoter-TSS (HVO_0638)	371	HVO_0637	HVO_0637
Main chromosome	1212238	1215229	2	2	NA	22.9698	6.10257	NA	NA	NA	promoter-TSS (HVO_0646)	264	HVO_0644	HVO_0644
Main chromosome	1219946	1232277	4	2	NA	23.1691	4.4912;2.17074;11.2376	NA	NA	NA	promoter-TSS (HVO_0663)	120	HVO_0662	HVO_0662
Main chromosome	1234614	1239319	3	2	NA	17.4672	3.09104;6.8578	NA	NA	NA	promoter-TSS (HVO_0672)	99	HVO_0673	HVO_0673
Main chromosome	1247307	1252221	2	2	NA	7.24775	3.51834	NA	NA	NA	promoter-TSS (HVO_0684)	199	HVO_0685	HVO_0685
Main chromosome	1259089	1260028	2	2	NA	13.8632	3.58676	NA	NA	NA	promoter-TSS (HVO_0692)	489	HVO_0693	HVO_0693
Main chromosome	1261614	1274370	3	2	NA	17.8616	7.66708;2.8333	NA	NA	NA	promoter-TSS (HVO_0702)	-379	HVO_0705	HVO_0705
Main chromosome	1277035	1278539	2	2	NA	29.9024	6.60185	NA	NA	NA	promoter-TSS (HVO_0717)	649	HVO_0716	HVO_0716
Main chromosome	1281153	1286468	2	2	NA	18.9316	6.03549	NA	NA	NA	promoter-TSS (HVO_0723)	571	HVO_0724	HVO_0724
Main chromosome	1293776	1297741	2	2	NA	16.3473	2.53709	NA	NA	NA	TTS (HVO_0739)	313	HVO_0740	HVO_0740
Main chromosome	1339676	1343675	2	2	NA	27.0061	8.28612	NA	NA	NA	promoter-TSS (HVO_0786)	69	HVO_0786	HVO_0786
Main chromosome	1345317	1355686	3	2	NA	20.3413	4.38703;9.27503	NA	NA	NA	promoter-TSS (HVO_0794)	284	HVO_0795	HVO_0795
Main chromosome	1357291	1362565	2	2	NA	27.3001	8.11256	NA	NA	NA	promoter-TSS (HVO_0803)	-181	HVO_0805	HVO_0805
Main chromosome	1365020	1381140	2	2	NA	20.772	6.18252	NA	NA	NA	promoter-TSS (HVO_0817)	-1	HVO_0817	HVO_0817
Main chromosome	1383340	1391589	2	2	NA	23.7327	6.33498	NA	NA	NA	promoter-TSS (HVO_0833)	43	HVO_0833	HVO_0833
Main chromosome	1399834	1405535	2	2	NA	28.5173	11.2623	NA	NA	NA	exon (HVO_0854, exon 1 of 1)	1111	HVO_0854	HVO_0854
Main chromosome	1416957	1423929	2	2	NA	9.12571	4.81388	NA	NA	NA	promoter-TSS (HVO_0868)	-369	HVO_0868	HVO_0868
Main chromosome	1425649	1427320	2	2	NA	7.96113	3.92776	NA	NA	NA	promoter-TSS (HVO_0870)	-12	HVO_0871	HVO_0871
Main chromosome	1430615	1435801	3	2	NA	12.1604;7.64625	3.89978	NA	NA	NA	promoter-TSS (HVO_0879)	414	HVO_0880	HVO_0880
Main chromosome	1438884	1445301	2	2	NA	39.86	8.08431	NA	NA	NA	promoter-TSS (HVO_0890)	87	HVO_0890	HVO_0890
Main chromosome	1447531	1449334	2	2	NA	3.44211	3.70259	NA	NA	NA	TTS (HVO_0895)	886	HVO_0895	HVO_0895
Main chromosome	1457126	1459216	2	2	NA	7.29904	3.40824	NA	NA	NA	TTS (HVO_0903)	378	HVO_0904	HVO_0904
Main chromosome	1462941	1464795	2	2	NA	3.9214	1.95942	NA	NA	NA	TTS (HVO_0911)	394	HVO_0912	HVO_0912
Main chromosome	1493575	1495401	2	2	NA	10.0685	6.24501	NA	NA	NA	promoter-TSS (HVO_0950)	397	HVO_0948	HVO_0948
Main chromosome	1497390	1512780	4	2	NA	21.9911;26.1126	9.20025	NA	NA	NA	promoter-TSS (HVO_0962)	42	HVO_0962	HVO_0962
Main chromosome	1514861	1524089	3	2	NA	24.1548;18.6262	8.32463	NA	NA	NA	promoter-TSS (HVO_0972)	603	HVO_0973	HVO_0973
Main chromosome	1533919	1546375	2	2	NA	11.3656	5.57654	NA	NA	NA	promoter-TSS (HVO_0996)	475	HVO_0995	HVO_0995
Main chromosome	1547656	1558068	2	2	NA	31.7525	14.6126	NA	NA	NA	TTS (HVO_1007)	994	HVO_1008	HVO_1008
Main chromosome	1559669	1571139	2	2	NA	19.12	10.2001	NA	NA	NA	promoter-TSS (HVO_1019)	-692	HVO_1019	HVO_1019
Main chromosome	1580761	1581470	2	2	NA	6.21038	4.9675	NA	NA	NA	promoter-TSS (HVO_1034)	-105	HVO_1035	HVO_1035
Main chromosome	1583120	1586117	2	2	NA	26.1948	9.10198	NA	NA	NA	promoter-TSS (HVO_1039)	417	HVO_1040	HVO_1040
Main chromosome	1588799	1594412	2	2	NA	22.2252	11.8835	NA	NA	NA	promoter-TSS (HVO_1049)	335	HVO_1048	HVO_1048
Main chromosome	1595844	1599394	2	2	NA	7.87298	4.53087	NA	NA	NA	promoter-TSS (HVO_1056)	157	HVO_1055	HVO_1055
Main chromosome	1601866	1612298	2	2	NA	15.114	6.72771	NA	NA	NA	promoter-TSS (HVO_1063)	40	HVO_1063	HVO_1063

Main chromosome	1614583	1615492	2	2	NA	1.85082	2.34415	NA	NA	NA	promoter-TSS (HVO_1073)	64	HVO_1073	HVO_1073
Main chromosome	1618986	1619654	2	2	NA	3.86601	3.34245	NA	NA	NA	promoter-TSS (HVO_1077)	-589	HVO_1077	HVO_1077
Main chromosome	1622153	1626385	3	2	NA	4.44456	2.69078;3.8304	NA	NA	NA	promoter-TSS (HVO_1083)	-407	HVO_1083	HVO_1083
Main chromosome	1628542	1632201	2	2	NA	29.9554	17.9795	NA	NA	NA	exon (HVO_1088, exon 1 of 1)	-1027	HVO_1089	HVO_1089
Main chromosome	1633775	1644129	3	2	NA	23.9692;3.06725	12.9399	NA	NA	NA	promoter-TSS (HVO_1097)	344	HVO_1098	HVO_1098
Main chromosome	1647443	1653408	2	2	NA	14.2485	5.44096	NA	NA	NA	promoter-TSS (HVO_1110)	-367	HVO_1112	HVO_1112
Main chromosome	1687672	1691103	2	2	NA	15.0908	8.54398	NA	NA	NA	promoter-TSS (HVO_1154)	-422	HVO_1155	HVO_1155
Main chromosome	1700237	1709791	3	2	NA	10.9496;17.2962	9.81604	NA	NA	NA	promoter-TSS (HVO_1171)	114	HVO_1172	HVO_1172
Main chromosome	1714727	1719038	2	2	NA	2.39666	4.17958	NA	NA	NA	promoter-TSS (HVO_1190)	750	HVO_1189	HVO_1189
Main chromosome	1720674	1726922	2	2	NA	6.54139	4.56424	NA	NA	NA	TTS (HVO_1195)	473	HVO_1196	HVO_1196
Main chromosome	1729509	1730770	2	2	NA	12.195	6.08912	NA	NA	NA	TTS (HVO_1201)	659	HVO_1202	HVO_1202
Main chromosome	1738431	1739218	2	2	NA	9.68292	3.44918	NA	NA	NA	promoter-TSS (HVO_1214)	-586	HVO_1214	HVO_1214
Main chromosome	1777340	1794302	2	2	NA	25.5958	11.1179	NA	NA	NA	promoter-TSS (HVO_1262)	601	HVO_1261	HVO_1261
Main chromosome	1797262	1808221	2	2	NA	22.4863	8.90845	NA	NA	NA	TTS (HVO_1278)	259	HVO_1280	HVO_1280
Main chromosome	1809919	1823181	2	2	NA	25.6765	11.617	NA	NA	NA	promoter-TSS (HVO_1298)	-484	HVO_1298	HVO_1298
Main chromosome	1880580	1885542	2	2	NA	8.02365	3.9947	NA	NA	NA	TTS (HVO_1371)	101	HVO_1370	HVO_1370
Main chromosome	1935228	1937361	2	2	NA	9.3221	3.23777	NA	NA	NA	promoter-TSS (HVO_1426)	-280	HVO_1427	HVO_1427
Main chromosome	1999707	2015924	4	2	NA	17.2552	8.13244;3.18933;1.8076	NA	NA	NA	promoter-TSS (HVO_1503)	693	HVO_1504	HVO_1504
Main chromosome	2034297	2043344	3	2	NA	29.3056	5.5274;10.2546	NA	NA	NA	promoter-TSS (HVO_1536)	-86	HVO_1536	HVO_1536
Main chromosome	2046698	2051267	2	2	NA	8.00665	3.34718	NA	NA	NA	promoter-TSS (HVO_1544)	244	HVO_1545	HVO_1545
Main chromosome	2066437	2070004	2	2	NA	14.7074	2.14438	NA	NA	NA	promoter-TSS (HVO_1565)	234	HVO_1564	HVO_1564
Main chromosome	2084940	2089916	2	2	NA	15.059	2.28911	NA	NA	NA	promoter-TSS (HVO_1579)	-318	HVO_1579	HVO_1579
Main chromosome	2098363	2107239	2	2	NA	12.7166	4.21144	NA	NA	NA	TTS (HVO_1594)	941	HVO_1595	HVO_1595
Main chromosome	2129847	2132013	2	2	NA	35.9299	6.11561	NA	NA	NA	TTS (HVO_1628)	1019	HVO_1629	HVO_1629
Main chromosome	2138001	2140297	2	2	NA	27.9766	2.94415	NA	NA	NA	TTS (HVO_1638)	868	HVO_1639	HVO_1639
Main chromosome	2148035	2158599	2	2	NA	14.3045	2.12012	NA	NA	NA	promoter-TSS (HVO_1655)	544	HVO_1654	HVO_1654
Main chromosome	2164483	2167234	2	2	NA	22.4238	2.2381	NA	NA	NA	promoter-TSS (HVO_1669)	261	HVO_1670	HVO_1670
Main chromosome	2204289	2223693	2	2	NA	14.2758	2.02237	NA	NA	NA	exon (HVO_1711, exon 1 of 1)	1457	HVO_1711	HVO_1711
Main chromosome	2259748	2263894	2	2	NA	15.7362	1.8111	NA	NA	NA	promoter-TSS (HVO_1756)	-180	HVO_1758	HVO_1758
Main chromosome	2299187	2325614	2	2	NA	14.9528	2.84893	NA	NA	NA	TTS (HVO_1812)	494	HVO_1813	HVO_1813
Main chromosome	2343998	2354112	3	2	NA	24.1846	3.25874;2.9825	NA	NA	NA	TTS (HVO_1855)	244	HVO_1854	HVO_1854
Main chromosome	2369600	2373503	2	2	NA	15.6532	2.30635	NA	NA	NA	promoter-TSS (HVO_1879)	881	HVO_1880	HVO_1880
Main chromosome	2378131	2384896	2	2	NA	13.4152	3.66644	NA	NA	NA	promoter-TSS (HVO_1894)	767	HVO_1895	HVO_1895
Main chromosome	2386894	2388308	2	2	NA	18.0418	3.98999	NA	NA	NA	TTS (HVO_1904)	416	HVO_1904	HVO_1904
Main chromosome	2392017	2394380	2	2	NA	18.738	1.94827	NA	NA	NA	exon (HVO_1908, exon 1 of 1)	1387	HVO_1908	HVO_1908
Main chromosome	2443474	2446105	2	2	NA	6.67116	2.11582	NA	NA	NA	promoter-TSS (HVO_1959)	85	HVO_1959	HVO_1959
Main chromosome	2449988	2465266	3	2	NA	23.1649	10.6144;3.42546	NA	NA	NA	TTS (HVO_1975)	322	HVO_1976	HVO_1976
Main chromosome	2470088	2471496	2	2	NA	11.5867	4.43215	NA	NA	NA	promoter-TSS (HVO_1989)	-457	HVO_1989	HVO_1989
Main chromosome	2474786	2477640	2	2	NA	14.437	3.32708	NA	NA	NA	promoter-TSS (HVO_1994)	386	HVO_1995	HVO_1995
Main chromosome	2485569	2487198	2	2	NA	18.2925	5.37046	NA	NA	NA	promoter-TSS (HVO_2005)	866	HVO_2006	HVO_2006
Main chromosome	2493732	2500780	2	2	NA	25.1889	10.7384	NA	NA	NA	promoter-TSS (HVO_2015)	-407	HVO_2015	HVO_2015
Main chromosome	2513052	2515161	2	2	NA	14.0215	4.51242	NA	NA	NA	promoter-TSS (HVO_2034)	-283	HVO_2034	HVO_2034
Main chromosome	2551698	2553126	2	2	NA	7.83683	2.76691	NA	NA	NA	promoter-TSS (HVO_2068)	100	HVO_2068	HVO_2068
Main chromosome	2583312	2584627	2	2	NA	14.9079	1.90846	NA	NA	NA	exon (HVO_2087, exon 1 of 1)	1032	HVO_2087	HVO_2087
Main chromosome	2586134	2587434	2	2	NA	6.72226	1.9563	NA	NA	NA	exon (HVO_2088, exon 1 of 1)	866	HVO_2088	HVO_2088
Main chromosome	2591411	2600530	2	2	NA	12.9058	5.0736	NA	NA	NA	promoter-TSS (HVO_2095)	33	HVO_2095	HVO_2095
Main chromosome	2627776	2630144	2	2	NA	28.9513	10.1342	NA	NA	NA	TTS (HVO_2127)	102	HVO_2128	HVO_2128
Main chromosome	2632541	2638128	2	2	NA	17.3015	7.79421	NA	NA	NA	promoter-TSS (HVO_2133)	-26	HVO_2135	HVO_2135
Main chromosome	2640032	2643487	2	2	NA	26.0588	11.0416	NA	NA	NA	promoter-TSS (HVO_2142)	438	HVO_2144	HVO_2144
Main chromosome	2647025	2650768	2	2	NA	15.4023	3.32518	NA	NA	NA	promoter-TSS (HVO_2149)	522	HVO_2149A	HVO_2149A
Main chromosome	2652927	2658157	2	2	NA	18.3517	10.7571	NA	NA	NA	TTS (HVO_2155)	492	HVO_2156	HVO_2156
Main chromosome	2689343	2706172	2	2	NA	20.77	7.46063	NA	NA	NA	promoter-TSS (HVO_2194)	180	HVO_2193	HVO_2193
Main chromosome	2734746	2741943	3	2	NA	5.62656;3.87539	6.3218	NA	NA	NA	promoter-TSS (HVO_2242)	330	HVO_2244	HVO_2244
Main chromosome	2799014	2804184	2	2	NA	20.2826	9.60011	NA	NA	NA	promoter-TSS (HVO_2296)	-362	HVO_2296	HVO_2296
Main chromosome	2806570	2830892	4	2	NA	27.7722;15.8468	10.9368;2.2026	NA	NA	NA	TTS (HVO_2315)	841	HVO_2316	HVO_2316
Main chromosome	2834244	2835409	2	2	NA	8.5598	2.45922	NA	NA	NA	TTS (HVO_2333)	577	HVO_2334	HVO_2334
Main chromosome	2836917	2852744	5	2	NA	20.3403;9.86792	9.42925	NA	NA	NA	promoter-TSS (HVO_2344)	240	HVO_2343	HVO_2343
Main chromosome	2854058	2856480	2	2	NA	12.9916	5.99994	NA	NA	NA	promoter-TSS (HVO_2355)	-27	HVO_2355	HVO_2355
Main chromosome	2860159	2864581	2	2	NA	16.6364	8.28697	NA	NA	NA	promoter-TSS (HVO_2362)	-224	HVO_2362	HVO_2362
Main chromosome	2867823	2873493	2	2	NA	11.3671	8.90992	NA	NA	NA	promoter-TSS (HVO_2370)	237	HVO_2369	HVO_2369
Main chromosome	2877114	2878262	2	2	NA	13.353	5.84305	NA	NA	NA	promoter-TSS (HVO_2378)	-728	HVO_2378	HVO_2378
Main chromosome	2883407	2896625	3	2	NA	14.8128;10.7925	6.80486	NA	NA	NA	TTS (HVO_2385)	213	HVO_2386	HVO_2386
Main chromosome	2912117	2915370	2	2	NA	10.4524	6.8857	NA	NA	NA	TTS (HVO_2411)	555	HVO_2410	HVO_2410
Main chromosome	2918038	2920374	2	2	NA	11.7315	3.13402	NA	NA	NA	promoter-TSS (HVO_2414)	101	HVO_2415	HVO_2415
Main chromosome	2925971	2933092	2	2	NA	9.43531	4.23967	NA	NA	NA	promoter-TSS (HVO_2424)	-21	HVO_2425	HVO_2425
Main chromosome	2936045	2936585	2	2	NA	6.77716	4.05141	NA	NA	NA	promoter-TSS (HVO_2432)	513	HVO_2433	HVO_2433
Main chromosome	2938716	2939395	2	2	NA	9.56862	2.80219	NA	NA	NA	promoter-TSS (HVO_2436)	-658	HVO_2436	HVO_2436
Main chromosome	2941603	2943123	2	2	NA	17.5324	7.86481	NA	NA	NA	TTS (HVO_2436)	329	HVO_2437	HVO_2437
Main chromosome	2971891	2973171	2	2	NA	13.9547	6.21283	NA	NA	NA	promoter-TSS (HVO_2466)	-211	HVO_2468	HVO_2468
Main chromosome	3010188	3023037	3	2	NA	21.1829	9.97265;9.76766	NA	NA	NA	promoter-TSS (HVO_2515)	601	HVO_2516	HVO_2516
Main chromosome	3075494	3085482	3	2	NA	16.7828;11.7722	7.87101	NA	NA	NA	promoter-TSS (HVO_2596)	57	HVO_2596	HVO_2596
Main chromosome	3087044	3103859	2	2	NA	28.6576	10.3294	NA	NA	NA	promoter-TSS (HVO_2610)	-171	HVO_2612	HVO_2612
Main chromosome	3110875	3117062	2	2	NA	29.6211	13.8731	NA	NA	NA	promoter-TSS (HVO_2629)	273	HVO_2628	HVO_2628
Main chromosome	3118976	3137980	2	2	NA	16.138	7.3101	NA	NA	NA	promoter-TSS (HVO_2643)	-269	HVO_2645	HVO_2645
Main chromosome	3140934	3143304	2	2	NA	26.7094	6.02358	NA	NA	NA	promoter-TSS (HVO_2660)	482	HVO_2658	HVO_2658
Main chromosome	3146388	3167852	2	2	NA	21.0703	10.8546	NA	NA	NA	promoter-TSS (HVO_2672)	627	HVO_2671	HVO_2671

Main chromosome	3170104	3172371	2	2	NA		9.61505		5.57705	NA	NA	NA	TTS (HVO_2688)	355	HVO_2689	HVO_2689
Main chromosome	3178432	3182765	2	2	NA		14.9618		6.03301	NA	NA	NA	promoter-TSS (HVO_2699)	212	HVO_2698	HVO_2698
Main chromosome	3204252	3209197	2	2	NA		6.91766		3.13868	NA	NA	NA	TTS (HVO_2725)	358	HVO_2726	HVO_2726
Main chromosome	3226975	3239856	3	2	NA		27.8787;17.1758		8.01027	NA	NA	NA	promoter-TSS (HVO_2758)	-509	HVO_2758	HVO_2758
Main chromosome	3258910	3261483	2	2	NA		10.7569		3.44248	NA	NA	NA	promoter-TSS (HVO_2792)	-32	HVO_2792	HVO_2792
Main chromosome	3280657	3281511	2	2	NA		4.41339		4.41332	NA	NA	NA	exon (HVO_2815, exon 1 of 1)	1078	HVO_2815	HVO_2815
Main chromosome	3294722	3295919	2	2	NA		6.91945		2.30708	NA	NA	NA	exon (HVO_2827, exon 1 of 1)	1423	HVO_2827	HVO_2827
Main chromosome	3329695	3335176	2	2	NA		17.6194		3.96357	NA	NA	NA	TTS (HVO_2860)	928	HVO_2859	HVO_2859
Main chromosome	3336734	3340068	2	2	NA		6.72149		2.12293	NA	NA	NA	TTS (HVO_2863)	270	HVO_2865	HVO_2865
Main chromosome	3354838	3357993	2	2	NA		22.9245		4.43644	NA	NA	NA	promoter-TSS (HVO_2884)	220	HVO_2883	HVO_2883
Main chromosome	3363376	3373007	3	2	NA		14.22	2.17077;3.21386		NA	NA	NA	promoter-TSS (HVO_2892A)	439	HVO_2894	HVO_2894
Main chromosome	3374619	3386000	4	2	NA		16.298	3.2869;4.78016;2.38006		NA	NA	NA	promoter-TSS (HVO_2907)	-592	HVO_2909	HVO_2909
Main chromosome	3387048	3394220	2	2	NA		14.4937		7.58511	NA	NA	NA	promoter-TSS (HVO_2922)	498	HVO_2921	HVO_2921
Main chromosome	3399735	3400545	2	2	NA		19.8464		4.9649	NA	NA	NA	promoter-TSS (HVO_2932)	75	HVO_2934	HVO_2934
Main chromosome	3417497	3427039	3	2	NA		13.993	6.22412;4.1139		NA	NA	NA	TTS (HVO_2952)	114	HVO_2951	HVO_2951
Main chromosome	3431521	3434281	2	2	NA		15.1449		2.97102	NA	NA	NA	promoter-TSS (HVO_2961)	-72	HVO_2961	HVO_2961
Main chromosome	3447655	3455828	2	2	NA		11.9881		2.96358	NA	NA	NA	promoter-TSS (HVO_2983)	224	HVO_2984	HVO_2984
Main chromosome	3457773	3477881	4	2	NA		14.8066	4.57211;4.93399;2.388		NA	NA	NA	TTS (HVO_3001)	1454	HVO_3000	HVO_3000
pHV1	4602	5302	2	2	NA	NA		NA		NA		10.1341	8.81252 promoter-TSS (HVO_C0004)	312	HVO_C0005	HVO_C0005
pHV1	16466	21699	2	2		50.2114	NA		4.05607	NA	NA	NA	promoter-TSS (HVO_C0026)	175	HVO_C0028	HVO_C0028
pHV1	30228	68874	5	2		49.5311	NA		3.93873;3.66128;6.093;8.9	NA	NA	NA	promoter-TSS (HVO_C0051)	595	HVO_C0050	HVO_C0050
pHV1	71868	82493	2	2		27.6478	NA		5.05846	NA	NA	NA	promoter-TSS (HVO_C0075)	395	HVO_C0076	HVO_C0076
pHV3	1728	2582	1	1	NA		11.0254	NA		NA	NA	NA	TTS (HVO_B0001)	600	HVO_B0002	HVO_B0002
pHV3	5466	6721	1	1	NA		1.83684	NA		NA	NA	NA	TTS (HVO_B0006)	1070	HVO_B0005	HVO_B0005
pHV3	27730	28918	1	1	NA		4.02203	NA		NA	NA	NA	promoter-TSS (HVO_B0025)	-35	HVO_B0026	HVO_B0026
pHV3	31094	34506	1	1	NA		17.8855	NA		NA	NA	NA	TTS (HVO_B0028)	783	HVO_B0029	HVO_B0029
pHV3	40981	41610	1	1	NA		2.12605	NA		NA	NA	NA	TTS (HVO_B0037)	901	HVO_B0038	HVO_B0038
pHV3	43598	44282	1	1	NA		2.82657	NA		NA	NA	NA	promoter-TSS (HVO_B0038A)	737	HVO_B0039	HVO_B0039
pHV3	82796	87732	1	1	NA		12.1321	NA		NA	NA	NA	promoter-TSS (HVO_B0075)	-360	HVO_B0075	HVO_B0075
pHV3	94859	96095	1	1	NA		2.60297	NA		NA	NA	NA	promoter-TSS (HVO_B0084)	-14	HVO_B0084	HVO_B0084
pHV3	112148	113251	1	1	NA		7.65095	NA		NA	NA	NA	TTS (HVO_B0099)	241	HVO_B0100	HVO_B0100
pHV3	115375	116382	1	1	NA		3.02913	NA		NA	NA	NA	promoter-TSS (HVO_B0103)	-180	HVO_B0103	HVO_B0103
pHV3	128830	129735	1	1	NA		3.42061	NA		NA	NA	NA	promoter-TSS (HVO_B0116)	13	HVO_B0116	HVO_B0116
pHV3	140612	142576	1	1	NA		3.20868	NA		NA	NA	NA	TTS (HVO_B0125)	649	HVO_B0126	HVO_B0126
pHV3	188667	189150	1	1	NA		2.95887	NA		NA	NA	NA	promoter-TSS (HVO_B0163)	-96	HVO_B0163	HVO_B0163
pHV3	200850	204438	1	1	NA		6.52508	NA		NA	NA	NA	promoter-TSS (HVO_B0173)	-355	HVO_B0173	HVO_B0173
pHV3	206026	206960	1	1	NA		4.31843	NA		NA	NA	NA	promoter-TSS (HVO_B0176)	-629	HVO_B0176	HVO_B0176
pHV3	208942	211512	1	1	NA		15.4238	NA		NA	NA	NA	promoter-TSS (HVO_B0179)	168	HVO_B0180	HVO_B0180
pHV3	215030	216965	1	1	NA		8.09839	NA		NA	NA	NA	promoter-TSS (HVO_B0186)	-270	HVO_B0186	HVO_B0186
pHV3	220936	225243	1	1	NA		10.8289	NA		NA	NA	NA	TTS (HVO_B0191)	361	HVO_B0192	HVO_B0192
pHV3	237077	237566	1	1	NA		3.37892	NA		NA	NA	NA	promoter-TSS (HVO_B0204)	-771	HVO_B0204	HVO_B0204
pHV3	241245	245252	1	1	NA		9.09026	NA		NA	NA	NA	promoter-TSS (HVO_B0206)	290	HVO_B0208	HVO_B0208
pHV3	246919	252018	1	1	NA		11.6022	NA		NA	NA	NA	TTS (HVO_B0214)	1502	HVO_B0214	HVO_B0214
pHV3	255715	256217	1	1	NA		3.45422	NA		NA	NA	NA	promoter-TSS (HVO_B0217)	-602	HVO_B0217	HVO_B0217
pHV3	272871	273619	1	1	NA		5.73028	NA		NA	NA	NA	TTS (HVO_B0230)	595	HVO_B0231	HVO_B0231
pHV3	283186	284724	1	1	NA		21.0247	NA		NA	NA	NA	promoter-TSS (HVO_B0235)	-630	HVO_B0235	HVO_B0235
pHV3	289815	293302	1	1	NA		16.5568	NA		NA	NA	NA	promoter-TSS (HVO_B0243)	665	HVO_B0244	HVO_B0244
pHV3	303957	306568	1	1	NA		6.95626	NA		NA	NA	NA	promoter-TSS (HVO_B0256)	567	HVO_B0257	HVO_B0257
pHV3	321433	329241	1	1	NA		13.1787	NA		NA	NA	NA	promoter-TSS (HVO_B0272)	87	HVO_B0272	HVO_B0272
pHV3	332512	335798	1	1	NA		7.45729	NA		NA	NA	NA	promoter-TSS (HVO_B0284)	-362	HVO_B0284	HVO_B0284
pHV3	337855	340229	1	1	NA		4.25449	NA		NA	NA	NA	promoter-TSS (HVO_B0291)	610	HVO_B0290	HVO_B0290
pHV3	344556	345232	1	1	NA		4.345	NA		NA	NA	NA	promoter-TSS (HVO_B0298)	543	HVO_B0296	HVO_B0296
pHV3	364396	365452	1	1	NA		5.69117	NA		NA	NA	NA	promoter-TSS (HVO_B0316)	844	HVO_B0317	HVO_B0317
pHV3	375465	378135	1	1	NA		5.77814	NA		NA	NA	NA	TTS (HVO_B0326)	667	HVO_B0325	HVO_B0325
pHV3	397575	399878	1	1	NA		7.7313	NA		NA	NA	NA	promoter-TSS (HVO_B0344)	363	HVO_B0345	HVO_B0345
pHV3	401893	404550	1	1	NA		4.70334	NA		NA	NA	NA	TTS (HVO_B0348)	379	HVO_B0349	HVO_B0349
pHV3	409264	415859	1	1	NA		9.32509	NA		NA	NA	NA	promoter-TSS (HVO_B0355)	-386	HVO_B0356	HVO_B0356
pHV3	418847	420121	1	1	NA		2.0763	NA		NA	NA	NA	TTS (HVO_B0362)	660	HVO_B0363	HVO_B0363
Main chromosome	2665	3218	1	1	NA		2.33306	NA		NA	NA	NA	promoter-TSS (HVO_0002)	0	HVO_0003	HVO_0003
Main chromosome	27069	28000	1	1	NA		1.87617	NA		NA	NA	NA	promoter-TSS (HVO_0026)	561	HVO_0027	HVO_0027
Main chromosome	32415	32951	1	1	NA		5.64533	NA		NA	NA	NA	TTS (HVO_0032)	407	HVO_0033	HVO_0033
Main chromosome	50127	56104	1	1	NA		4.68795	NA		NA	NA	NA	promoter-TSS (HVO_0053)	-795	HVO_0053	HVO_0053
Main chromosome	64871	66170	1	1	NA		9.31407	NA		NA	NA	NA	exon (HVO_0065, exon 1 of 1)	1477	HVO_0065	HVO_0065
Main chromosome	98685	99850	1	1	NA		12.6504	NA		NA	NA	NA	promoter-TSS (HVO_0100)	-54	HVO_0100	HVO_0100
Main chromosome	103593	104666	1	1	NA		5.8929	NA		NA	NA	NA	promoter-TSS (HVO_0106)	-115	HVO_0106	HVO_0106
Main chromosome	117430	120655	1	1	NA		11.2837	NA		NA	NA	NA	TTS (HVO_0125)	423	HVO_0126	HVO_0126
Main chromosome	122400	123762	1	1	NA		2.57846	NA		NA	NA	NA	promoter-TSS (HVO_0132)	-300	HVO_0132	HVO_0132
Main chromosome	127334	127918	1	1	NA		6.3957	NA		NA	NA	NA	promoter-TSS (HVO_0136)	404	HVO_0137	HVO_0137
Main chromosome	166425	168502	1	1	NA		5.71036	NA		NA	NA	NA	promoter-TSS (HVO_0186)	318	HVO_0185	HVO_0185
Main chromosome	194119	194696	1	1	NA		2.58633	NA		NA	NA	NA	TTS (HVO_0214)	1022	HVO_0214	HVO_0214
Main chromosome	253372	254265	1	1	NA		3.49757	NA		NA	NA	NA	TTS (HVO_A0274)	569	HVO_A0275	HVO_A0275
Main chromosome	330941	332605	1	1	NA		4.00435	NA		NA	NA	NA	promoter-TSS (HVO_A0199)	761	HVO_A0200	HVO_A0200
Main chromosome	531623	532119	1	1		3.35504	NA		NA	NA	NA	NA	promoter-TSS (HVO_A0003)	-235	HVO_A0003	HVO_A0003
Main chromosome	537555	539620	1	1	NA		12.1048	NA		NA	NA	NA	promoter-TSS (HVO_A0635)	-105	HVO_A0635	HVO_A0635
Main chromosome	684869	685869	1	1	NA		9.37133	NA		NA	NA	NA	TTS (HVO_A0478)	157	HVO_A0479	HVO_A0479
Main chromosome	688513	688968	1	1	NA		4.81671	NA		NA	NA	NA	promoter-TSS (HVO_A0475)	-460	HVO_A0475	HVO_A0475

Main chromosome	701396	702544	1	1	NA		5.3742	NA		NA	NA	NA		promoter-TSS (HVO_A0461)	636	HVO_A0463	HVO_A0463
Main chromosome	755707	756106	1	1	NA		4.28571	NA		NA	NA	NA		promoter-TSS (HVO_A0409)	-418	HVO_A0409	HVO_A0409
Main chromosome	855898	856675	1	1	NA		13.2614	NA		NA	NA	NA		promoter-TSS (HVO_A0303)	448	HVO_A0302	HVO_A0302
Main chromosome	862469	863014	1	1	NA		5.51368	NA		NA	NA	NA		promoter-TSS (HVO_A0297)	224	HVO_A0296	HVO_A0296
Main chromosome	865135	865749	1	1	NA	NA			3.97991	NA	NA	NA		promoter-TSS (HVO_A0295)	-94	HVO_A0295	HVO_A0295
Main chromosome	888412	889876	1	1	NA		4.29705	NA		NA	NA	NA		TTS (HVO_0283)	646	HVO_0282	HVO_0282
Main chromosome	893441	894337	1	1	NA		12.2105	NA		NA	NA	NA		TTS (HVO_0285)	672	HVO_0286	HVO_0286
Main chromosome	936336	937286	1	1	NA		6.8982	NA		NA	NA	NA		promoter-TSS (HVO_0330)	-10	HVO_0331	HVO_0331
Main chromosome	938795	939518	1	1	NA	NA			2.89131	NA	NA	NA		promoter-TSS (HVO_0336)	683	HVO_0335	HVO_0335
Main chromosome	966806	967597	1	1	NA		13.8576	NA		NA	NA	NA		promoter-TSS (HVO_0363)	-183	HVO_0366	HVO_0366
Main chromosome	970422	971197	1	1	NA		8.42573	NA		NA	NA	NA		promoter-TSS (HVO_0377)	-197	HVO_0377	HVO_0377
Main chromosome	1033480	1034456	1	1	NA		3.06035	NA		NA	NA	NA		promoter-TSS (HVO_0450)	152	HVO_0449	HVO_0449
Main chromosome	1077111	1077584	1	1	NA		4.74045	NA		NA	NA	NA		promoter-TSS (HVO_0500)	-91	HVO_0500	HVO_0500
Main chromosome	1085596	1086157	1	1	NA		9.43967	NA		NA	NA	NA		promoter-TSS (HVO_0515)	301	HVO_0514	HVO_0514
Main chromosome	1129295	1130450	1	1	NA		3.69551	NA		NA	NA	NA		promoter-TSS (HVO_0558)	167	HVO_0559	HVO_0559
Main chromosome	1166687	1167739	1	1	NA		8.88315	NA		NA	NA	NA		TTS (HVO_0600)	419	HVO_0601	HVO_0601
Main chromosome	1199083	1202863	1	1	NA		8.41109	NA		NA	NA	NA		TTS (HVO_0631)	1151	HVO_0631	HVO_0631
Main chromosome	1216999	1218041	1	1	NA		10.329	NA		NA	NA	NA		TTS (HVO_0649)	738	HVO_0649	HVO_0649
Main chromosome	1241225	1242559	1	1	NA		14.8038	NA		NA	NA	NA		promoter-TSS (HVO_0677)	217	HVO_0679	HVO_0679
Main chromosome	1256148	1256860	1	1	NA		6.02801	NA		NA	NA	NA		promoter-TSS (HVO_0690)	-20	HVO_0690	HVO_0690
Main chromosome	1291338	1292287	1	1	NA		4.20409	NA		NA	NA	NA		promoter-TSS (HVO_0735)	516	HVO_0734	HVO_0734
Main chromosome	1396046	1396602	1	1	NA		4.788	NA		NA	NA	NA		promoter-TSS (HVO_0844)	96	HVO_0845	HVO_0845
Main chromosome	1412527	1413209	1	1	NA	NA			3.25212	NA	NA	NA		promoter-TSS (HVO_0862)	647	HVO_0861	HVO_0861
Main chromosome	1454402	1454909	1	1	NA		6.7785	NA		NA	NA	NA		promoter-TSS (HVO_0901)	-133	HVO_0901	HVO_0901
Main chromosome	1526785	1527634	1	1	NA	NA			2.45462	NA	NA	NA		promoter-TSS (HVO_0981)	-44	HVO_0981	HVO_0981
Main chromosome	1573148	1574498	1	1	NA	NA			3.51995	NA	NA	NA		promoter-TSS (HVO_1028)	802	HVO_1030	HVO_1030
Main chromosome	1732810	1734294	1	1	NA	NA			2.49949	NA	NA	NA		promoter-TSS (HVO_1206)	-109	HVO_1206	HVO_1206
Main chromosome	1741173	1741929	1	1	NA	NA			2.79486	NA	NA	NA		TTS (HVO_1216)	506	HVO_1217	HVO_1217
Main chromosome	1876783	1878305	1	1	NA		7.06068	NA		NA	NA	NA		promoter-TSS (HVO_1363)	76	HVO_1364	HVO_1364
Main chromosome	1994704	1996494	1	1	NA		2.23754	NA		NA	NA	NA		promoter-TSS (HVO_1489)	198	HVO_1490	HVO_1490
Main chromosome	2071703	2072451	1	1	NA		11.1126	NA		NA	NA	NA		promoter-TSS (HVO_1569)	384	HVO_1568	HVO_1568
Main chromosome	2074327	2075386	1	1	NA		6.91087	NA		NA	NA	NA		promoter-TSS (HVO_1570)	-896	HVO_1570	HVO_1570
Main chromosome	2081127	2082840	1	1	NA		10.8817	NA		NA	NA	NA		TTS (HVO_1573)	129	HVO_1574	HVO_1574
Main chromosome	2093859	2096319	1	1	NA		7.93238	NA		NA	NA	NA		promoter-TSS (HVO_1587)	-251	HVO_1588	HVO_1588
Main chromosome	2109905	2112376	1	1	NA		6.02352	NA		NA	NA	NA		TTS (HVO_1604)	278	HVO_1605	HVO_1605
Main chromosome	2113875	2117395	1	1	NA		7.73444	NA		NA	NA	NA		promoter-TSS (HVO_1609)	-88	HVO_1609	HVO_1609
Main chromosome	2119423	2120096	1	1	NA		7.92136	NA		NA	NA	NA		promoter-TSS (HVO_1615)	119	HVO_1616	HVO_1616
Main chromosome	2122770	2127969	1	1	NA		16.3279	NA		NA	NA	NA		TTS (HVO_1622)	864	HVO_1624	HVO_1624
Main chromosome	2134152	2136292	1	1	NA		5.16175	NA		NA	NA	NA		promoter-TSS (HVO_1635)	167	HVO_1634	HVO_1634
Main chromosome	2161822	2162598	1	1	NA		3.5825	NA		NA	NA	NA		promoter-TSS (HVO_1665)	-13	HVO_1665	HVO_1665
Main chromosome	2172579	2175586	1	1	NA		8.28753	NA		NA	NA	NA		promoter-TSS (HVO_1681)	-35	HVO_1682	HVO_1682
Main chromosome	2177875	2180334	1	1	NA		13.0239	NA		NA	NA	NA		TTS (HVO_1686)	112	HVO_1685	HVO_1685
Main chromosome	2190991	2191720	1	1	NA		4.54917	NA		NA	NA	NA		TTS (HVO_1696)	145	HVO_1697	HVO_1697
Main chromosome	2194140	2202404	1	1	NA		11.3223	NA		NA	NA	NA		promoter-TSS (HVO_1701)	176	HVO_1702	HVO_1702
Main chromosome	2225938	2227656	1	1	NA		2.71267	NA		NA	NA	NA		exon (HVO_1723, exon 1 of 1)	1034	HVO_1723	HVO_1723
Main chromosome	2230646	2231377	1	1	NA	NA			NA	NA	NA		4.53282	TTS (HVO_1726)	1075	HVO_1726	HVO_1726
Main chromosome	2256132	2258108	1	1	NA		16.9389	NA		NA	NA	NA		exon (HVO_1751, exon 1 of 1)	-1209	HVO_1750	HVO_1750
Main chromosome	2265813	2268422	1	1	NA		17.9303	NA		NA	NA	NA		promoter-TSS (HVO_1762)	595	HVO_1763	HVO_1763
Main chromosome	2279457	2280832	1	1	NA		4.54853	NA		NA	NA	NA		TTS (HVO_1778)	1973	HVO_1779	HVO_1779
Main chromosome	2290177	2291247	1	1	NA		17.4657	NA		NA	NA	NA		TTS (HVO_1789)	479	HVO_1790	HVO_1790
Main chromosome	2293663	2297491	1	1	NA		7.28313	NA		NA	NA	NA		promoter-TSS (HVO_1794)	4	HVO_1796	HVO_1796
Main chromosome	2341416	2341783	1	1		5.0106	NA		NA	NA	NA	NA		TTS (HVO_1845)	1699	HVO_1845	HVO_1845
Main chromosome	2356634	2362410	1	1	NA		8.38765	NA		NA	NA	NA		TTS (HVO_1868)	962	HVO_1868	HVO_1868
Main chromosome	2375296	2376435	1	1	NA		14.0419	NA		NA	NA	NA		TTS (HVO_1885)	236	HVO_1886	HVO_1886
Main chromosome	2396157	2404810	1	1	NA		5.67571	NA		NA	NA	NA		promoter-TSS (HVO_1916)	-134	HVO_1916	HVO_1916
Main chromosome	2467686	2468110	1	1	NA		3.03668	NA		NA	NA	NA		promoter-TSS (HVO_1986)	-270	HVO_1986	HVO_1986
Main chromosome	2479754	2480231	1	1	NA		4.52721	NA		NA	NA	NA		TTS (HVO_1998)	1584	HVO_1999	HVO_1999
Main chromosome	2482612	2483541	1	1	NA		8.94652	NA		NA	NA	NA		promoter-TSS (HVO_2002)	-590	HVO_2002	HVO_2002
Main chromosome	2489043	2489712	1	1	NA		6.88689	NA		NA	NA	NA		promoter-TSS (HVO_2007)	353	HVO_2008	HVO_2008
Main chromosome	2573787	2575477	1	1	NA		5.98511	NA		NA	NA	NA		exon (HVO_2081, exon 1 of 1)	1361	HVO_2081	HVO_2081
Main chromosome	2603239	2604213	1	1	NA		2.4211	NA		NA	NA	NA		promoter-TSS (HVO_2104)	65	HVO_2105	HVO_2105
Main chromosome	2880126	2881405	1	1	NA	NA			3.74269	NA	NA	NA		TTS (HVO_2379)	1312	HVO_2379	HVO_2379
Main chromosome	2909472	2910614	1	1	NA		6.44586	NA		NA	NA	NA		promoter-TSS (HVO_2405)	204	HVO_2406	HVO_2406
Main chromosome	2947769	2949020	1	1	NA	NA			3.7232	NA	NA	NA		promoter-TSS (HVO_2444)	-150	HVO_2444	HVO_2444
Main chromosome	2949656	2950504	1	1	NA		4.5722	NA		NA	NA	NA		promoter-TSS (HVO_2446)	97	HVO_2446	HVO_2446
Main chromosome	3256967	3257355	1	1	NA		5.30427	NA		NA	NA	NA		promoter-TSS (HVO_2787)	233	HVO_2788	HVO_2788
Main chromosome	3265960	3266351	1	1	NA		3.24241	NA		NA	NA	NA		TTS (HVO_2798)	584	HVO_2799	HVO_2799
Main chromosome	3273384	3274287	1	1	NA		2.36847	NA		NA	NA	NA		exon (HVO_2808, exon 1 of 1)	1002	HVO_2808	HVO_2808
Main chromosome	3278286	3278950	1	1	NA		4.02859	NA		NA	NA	NA		promoter-TSS (HVO_2812)	-665	HVO_2814	HVO_2814
Main chromosome	3292311	3293198	1	1	NA		3.27127	NA		NA	NA	NA		promoter-TSS (HVO_2826)	88	HVO_2826	HVO_2826
Main chromosome	3314139	3317585	1	1	NA		8.20526	NA		NA	NA	NA		TTS (HVO_2842)	505	HVO_2844	HVO_2844
Main chromosome	3327364	3327734	1	1	NA		1.7424	NA		NA	NA	NA		TTS (HVO_2851)	415	HVO_2852	HVO_2852
Main chromosome	3396461	3397694	1	1	NA		6.49008	NA		NA	NA	NA		promoter-TSS (HVO_2930)	-35	HVO_2930	HVO_2930
Main chromosome	3408704	3410759	1	1	NA		2.84408	NA		NA	NA	NA		TTS (HVO_2943)	143	HVO_2944	HVO_2944
Main chromosome	3414309	3415427	1	1	NA		2.02536	NA		NA	NA	NA		promoter-TSS (HVO_2946)	22	HVO_2946	HVO_2946
Main chromosome	3436505	3441341	1	1	NA		7.68173	NA		NA	NA	NA		TTS (HVO_2969)	1355	HVO_2968	HVO_2968

Main chromosome	3443180	3444896	1	1	NA	2.46325	NA	NA	NA	NA	TTS (HVO_2974)	545	HVO_2975	HVO_2975
Main chromosome	3481487	3481900	1	1	NA	NA	NA	NA	NA	4.44053	promoter-TSS (HVO_3012)	180	HVO_3014	HVO_3014
pHV1	149	3261	1	1	65.3528	NA	NA	NA	NA	NA	promoter-TSS (HVO_C0002)	84	HVO_C0002	HVO_C0002
pHV1	6037	7373	1	1	13.654	NA	NA	NA	NA	NA	promoter-TSS (HVO_C0006)	184	HVO_C0009	HVO_C0009
pHV1	9308	9814	1	1	11.4521	NA	NA	NA	NA	NA	TTS (HVO_C0011)	265	HVO_C0012	HVO_C0012
pHV1	11111	15230	1	1	41.8612	NA	NA	NA	NA	NA	promoter-TSS (HVO_C0015)	-368	HVO_C0018	HVO_C0018
pHV1	22586	28442	1	1	NA	NA	NA	NA	6.38085	NA	promoter-TSS (HVO_C0036)	-888	HVO_C0036	HVO_C0036
pHV1	70137	70450	1	1	22.7451	NA	NA	NA	NA	NA	promoter-TSS (HVO_C0071)	-119	HVO_C0071	HVO_C0071
pHV1	83192	84192	1	1	NA	NA	NA	NA	4.92289	NA	promoter-TSS (HVO_C0084)	-28	HVO_C0085	HVO_C0085
pHV1	84285	84891	1	1	24.9222	NA	NA	NA	NA	NA	promoter-TSS (HVO_C0085)	-63	HVO_C0088	HVO_C0088

Table 7.2: Intervals of enrichment. Where “Main chromosome” text is red, the interval is found upon pHV4 (integrated onto the main chromosome in all laboratory strains). Start and End in the table headings denote the start and end of the interval. NA in p-value columns demonstrates that this sample is not enriched for RadA binding in this interval. Gene names and descriptions are shown for intervals identified as being enriched for RadA binding in three samples or four samples.

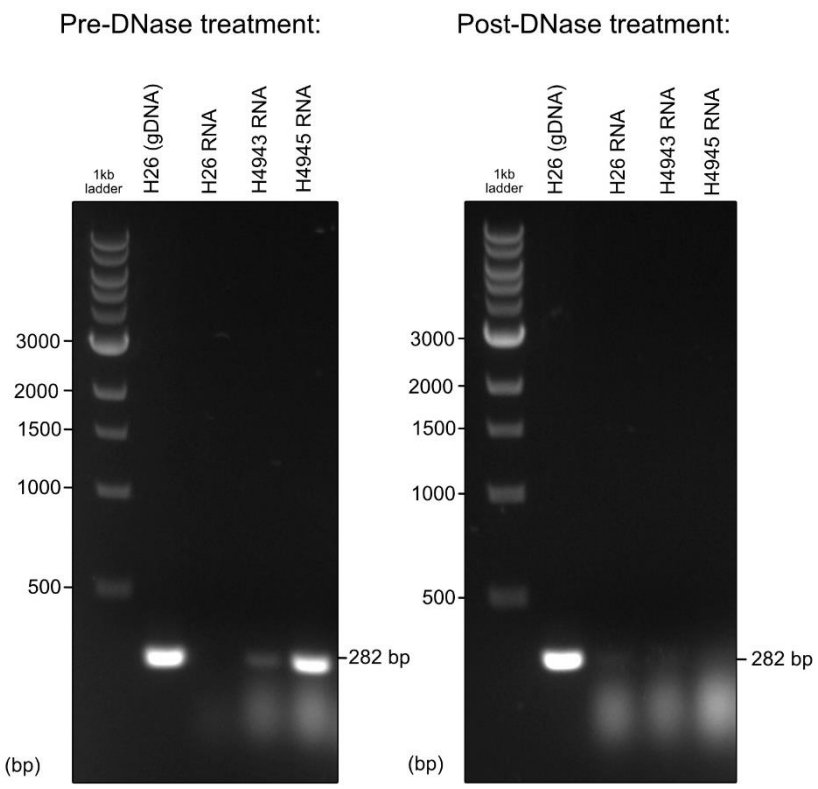


Figure 7.4: PCR amplification of RNA extracted from strains H26, H4943 and H4945, pre- and post-DNAse treatment. Oligonucleotides o331 and o332 were used to amplify a 282bp region of the radA gene. H26 genomic DNA was used as a positive control. There is a reduction in amplification post-DNase treatment of the RNA samples.

Pustell DNA matrices created using contig sequences (assembled using Flye in MacVector) and the H26 reference genome[49]:

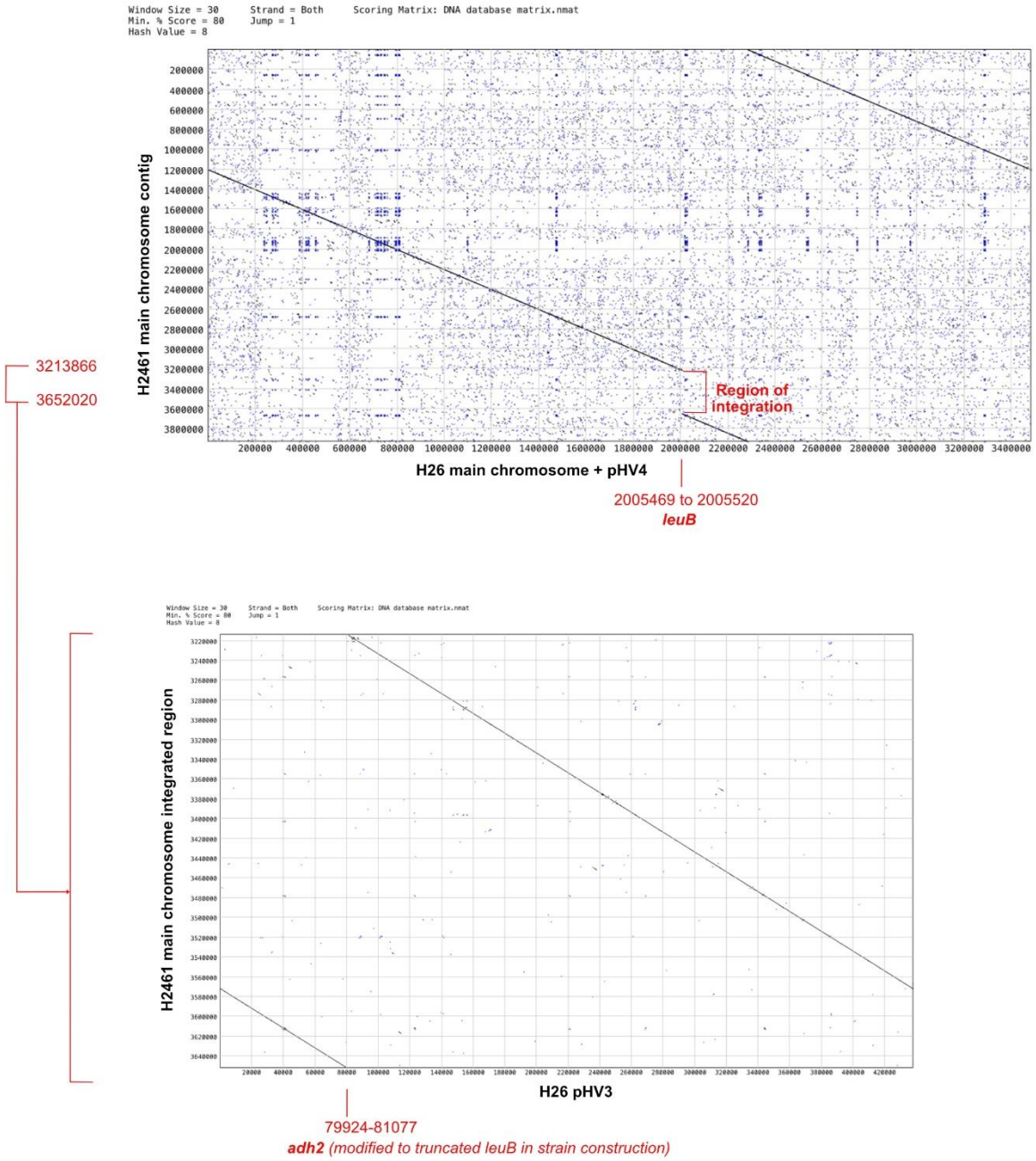


Figure 7.5: Pustell DNA matrices of H2461 Oxford Nanopore MinION sequencing data identifying the loci of pHV3 integration.

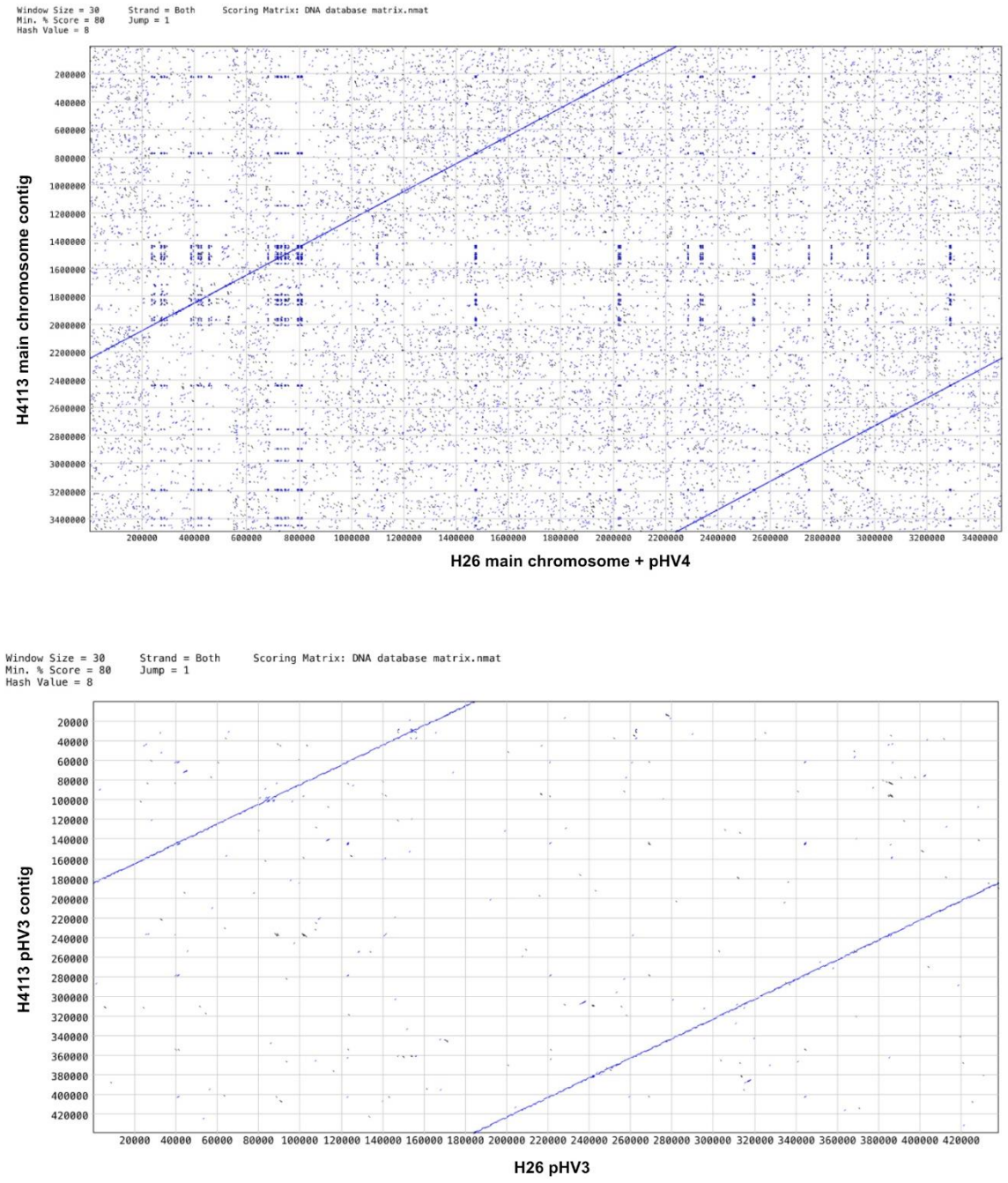


Figure 7.6: Pustell DNA matrices of H4113 Oxford Nanopore MinION sequencing data. Two individual contigs were mapped against the corresponding component of the H26 reference genome showing no discontinuities. pHV3 remains as an episome.

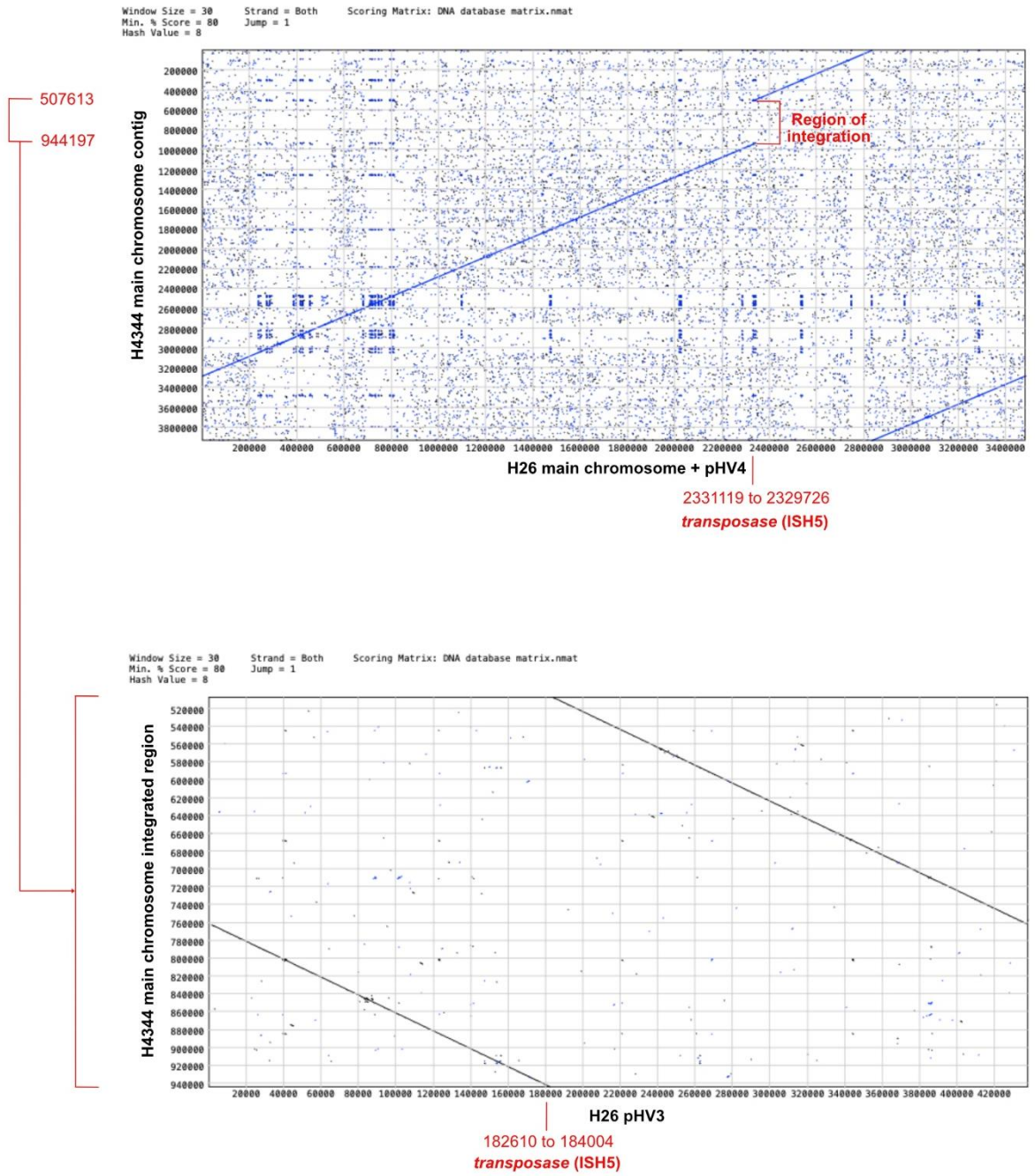


Figure 7.7: Pustell DNA matrices of H4344 Oxford Nanopore MinION sequencing data identifying the loci of pHV3 integration.

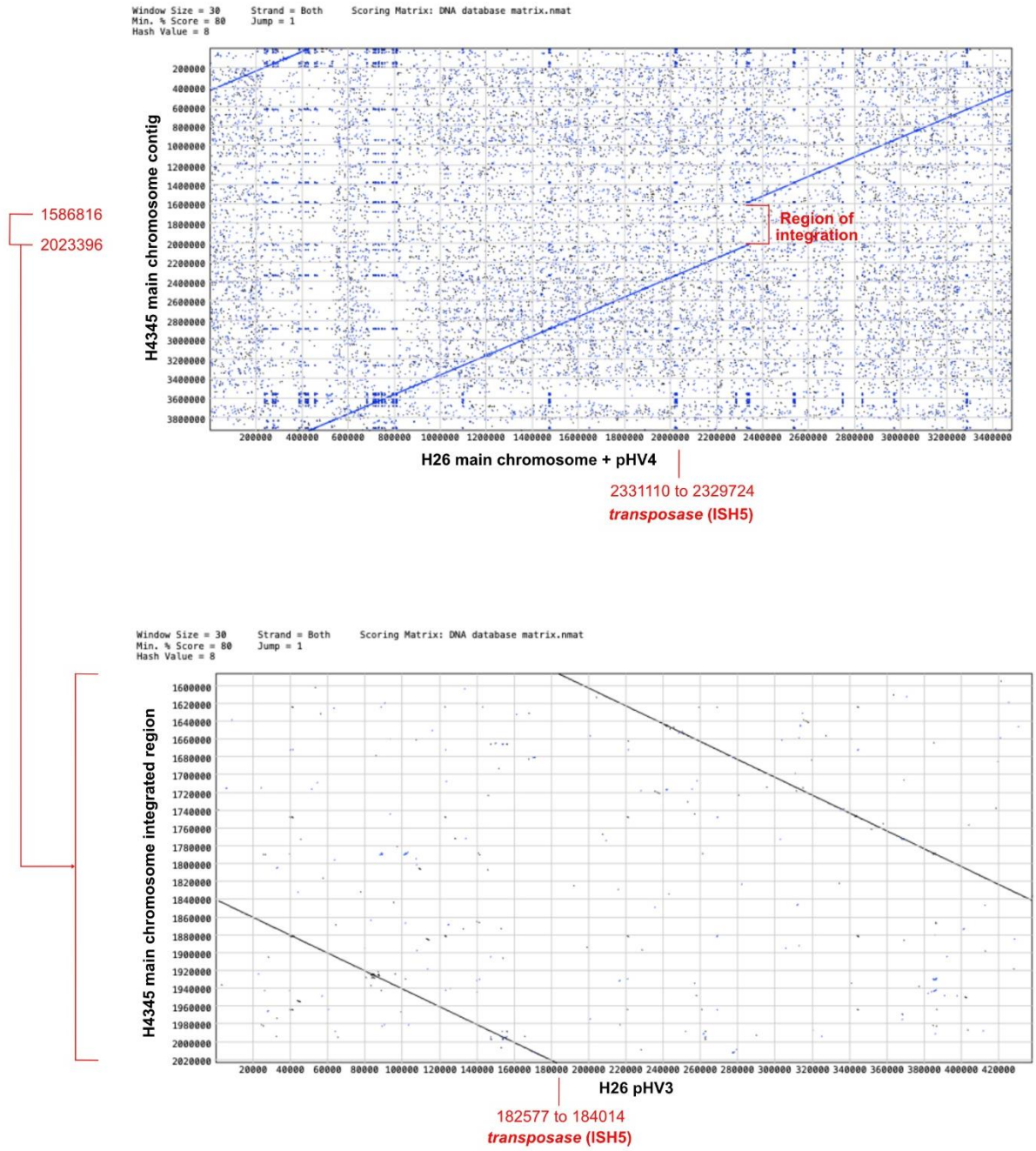


Figure 7.8: Pustell DNA matrices of H4345 Oxford Nanopore MinION sequencing data identifying the loci of pHV3 integration.

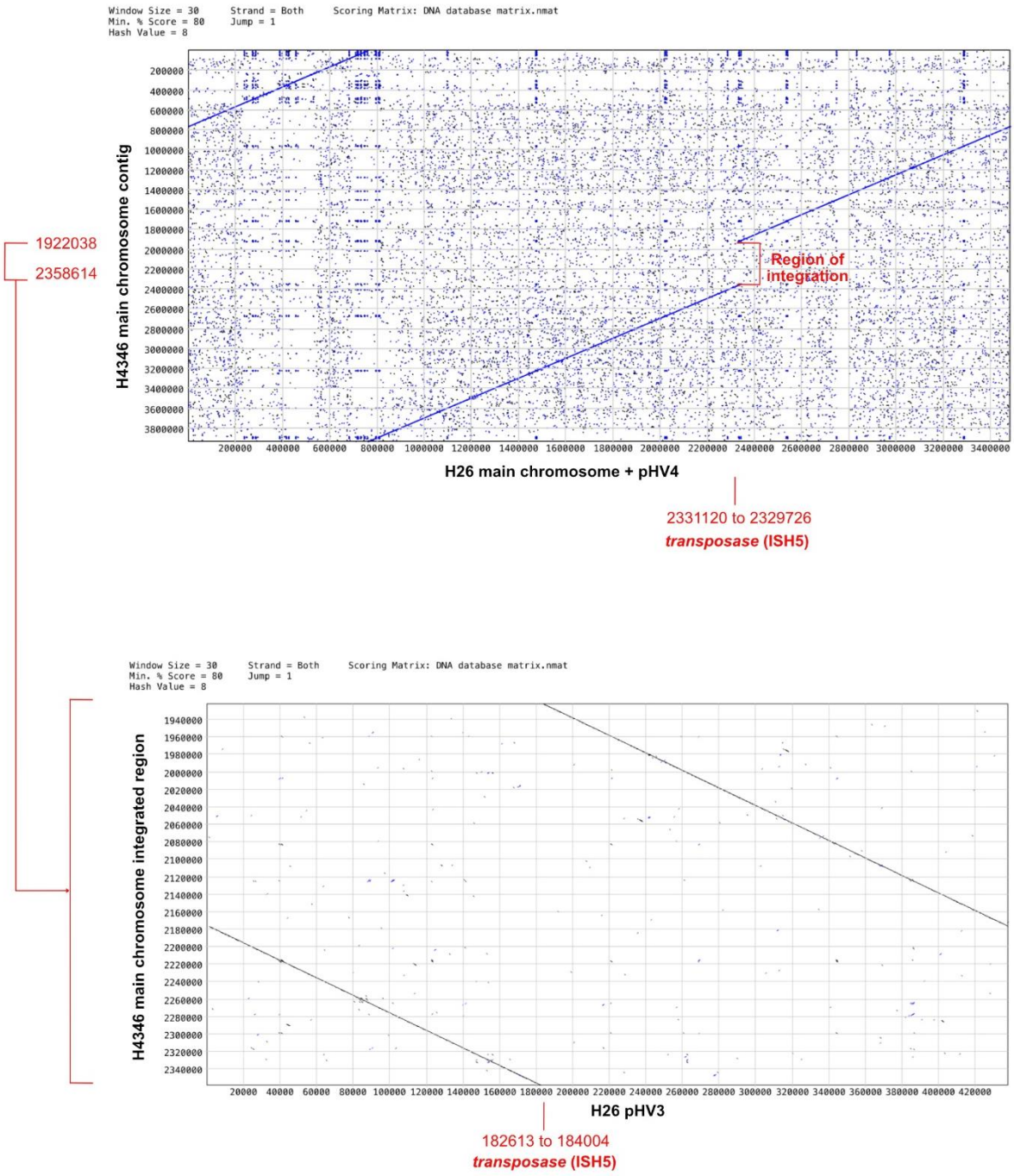


Figure 7.9: Pustell DNA matrices of H4346 Oxford Nanopore MinION sequencing data identifying the loci of pHV3 integration.

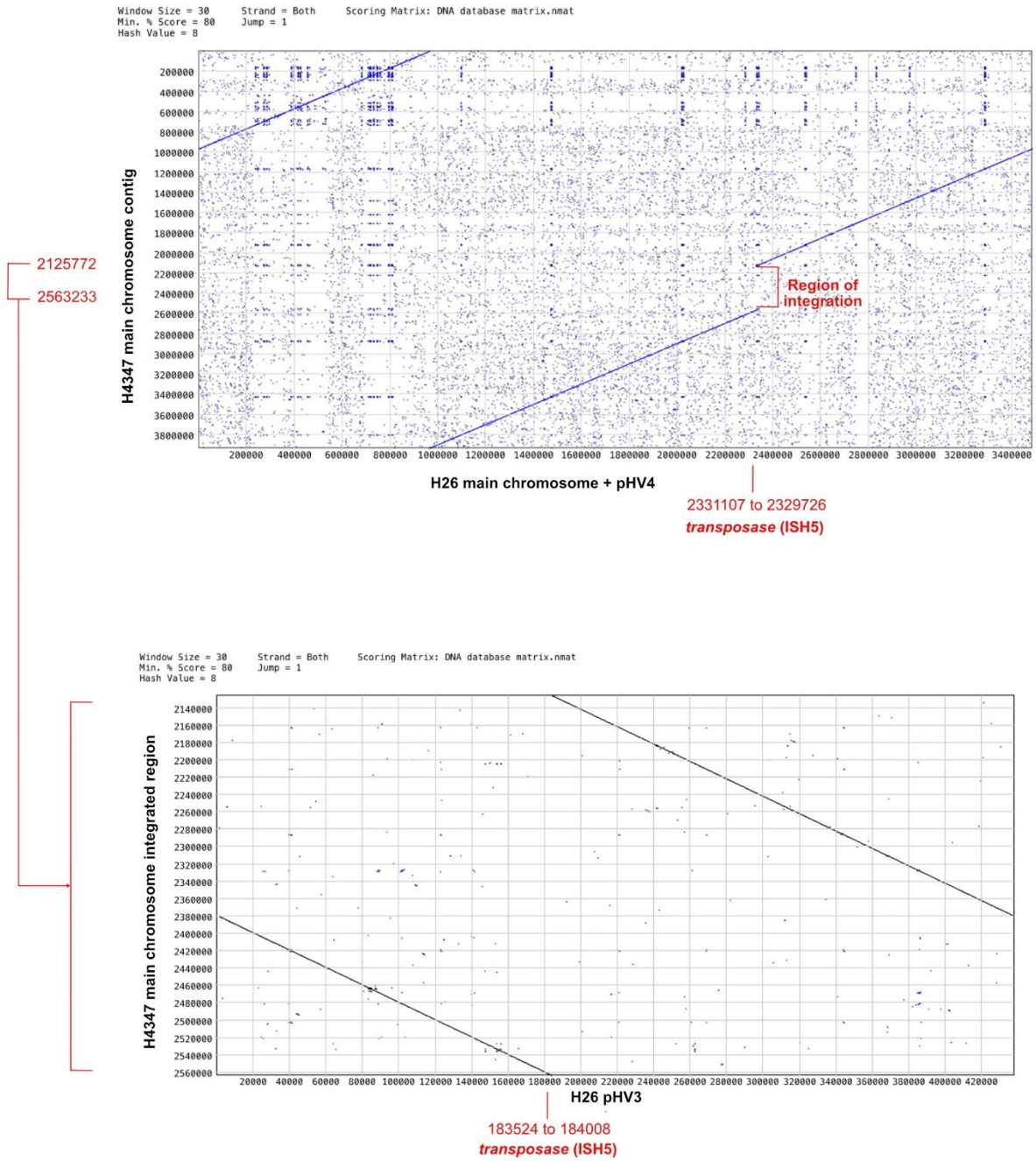
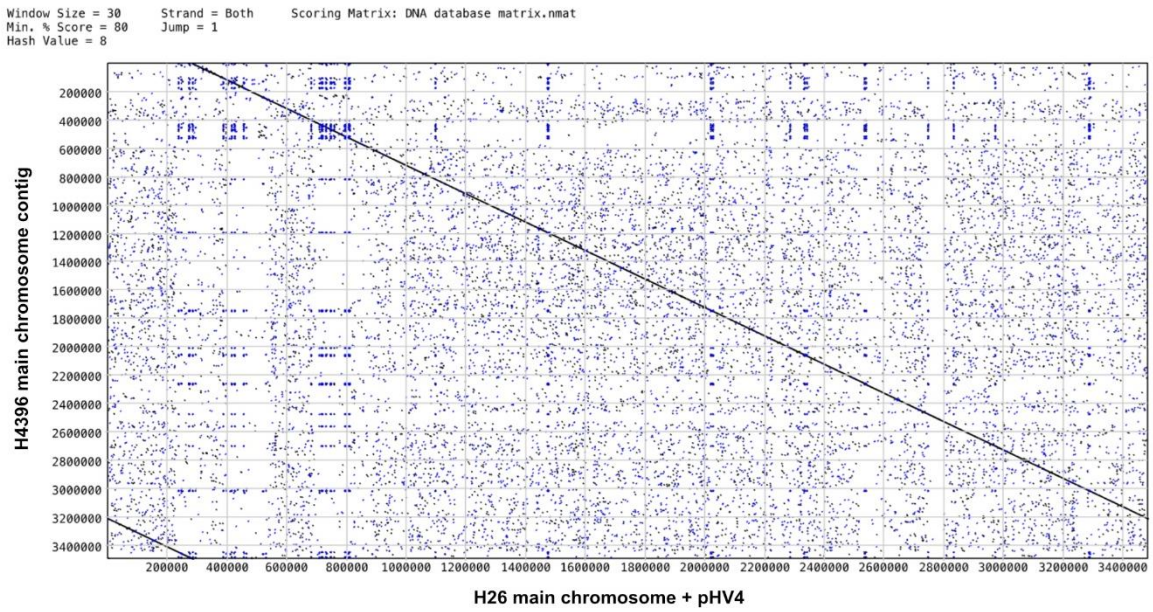


Figure 7.10: Pustell DNA matrices of H4347 Oxford Nanopore MinION sequencing data identifying the loci of pHV3 integration.



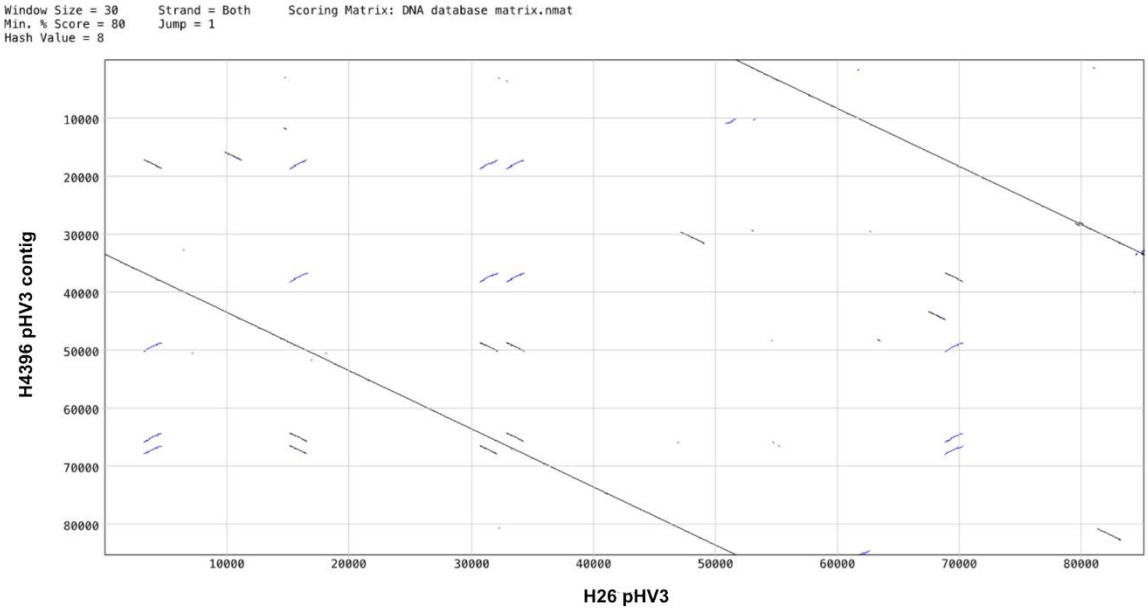
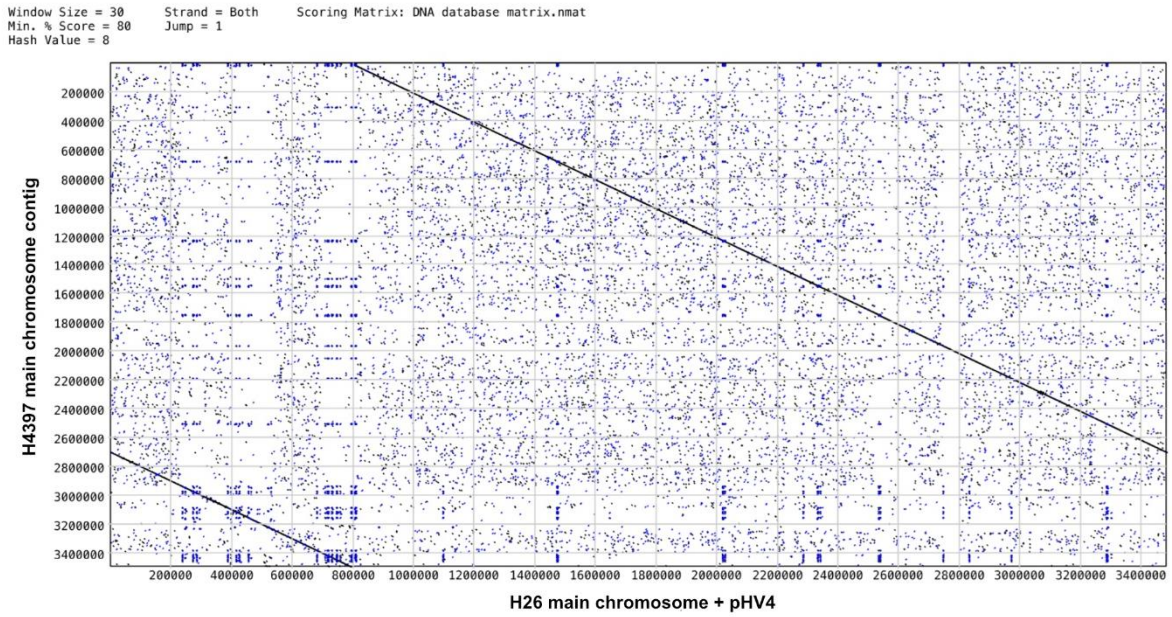


Figure 7.11: Pustell DNA matrices of H4396 Oxford Nanopore MinION sequencing data. Two individual contigs were mapped against the corresponding component of the H26 reference genome showing no discontinuities. pHV3 remains as an episome.



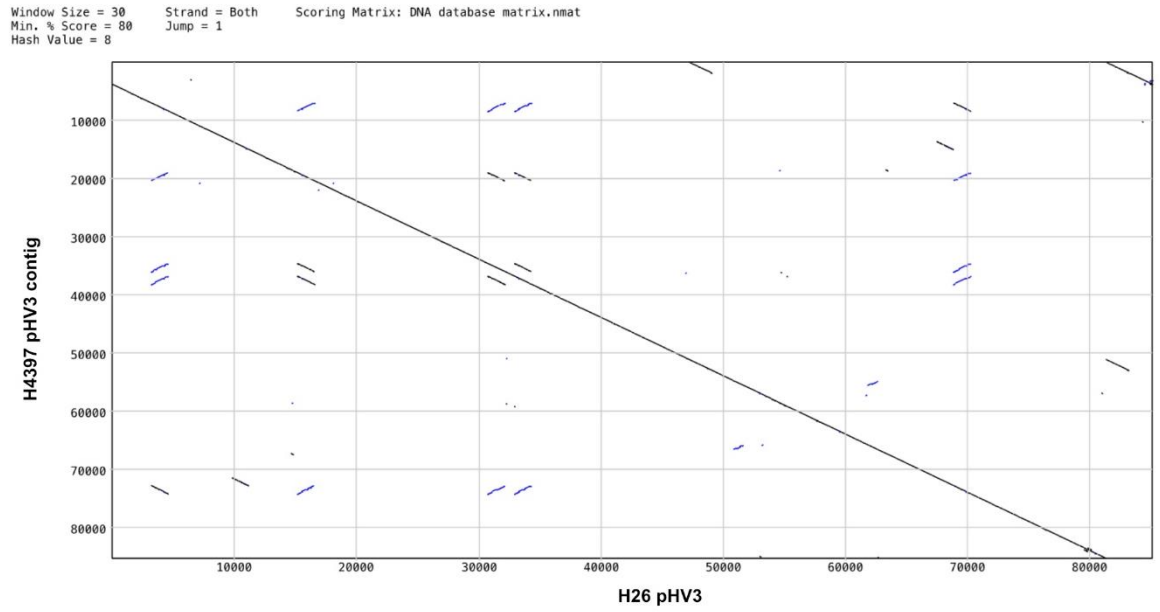


Figure 7.12: Pustell DNA matrices of H4397 Oxford Nanopore MinION sequencing data. Two individual contigs were mapped against the corresponding component of the H26 reference genome showing no discontinuities. pHV3 remains as an episome.

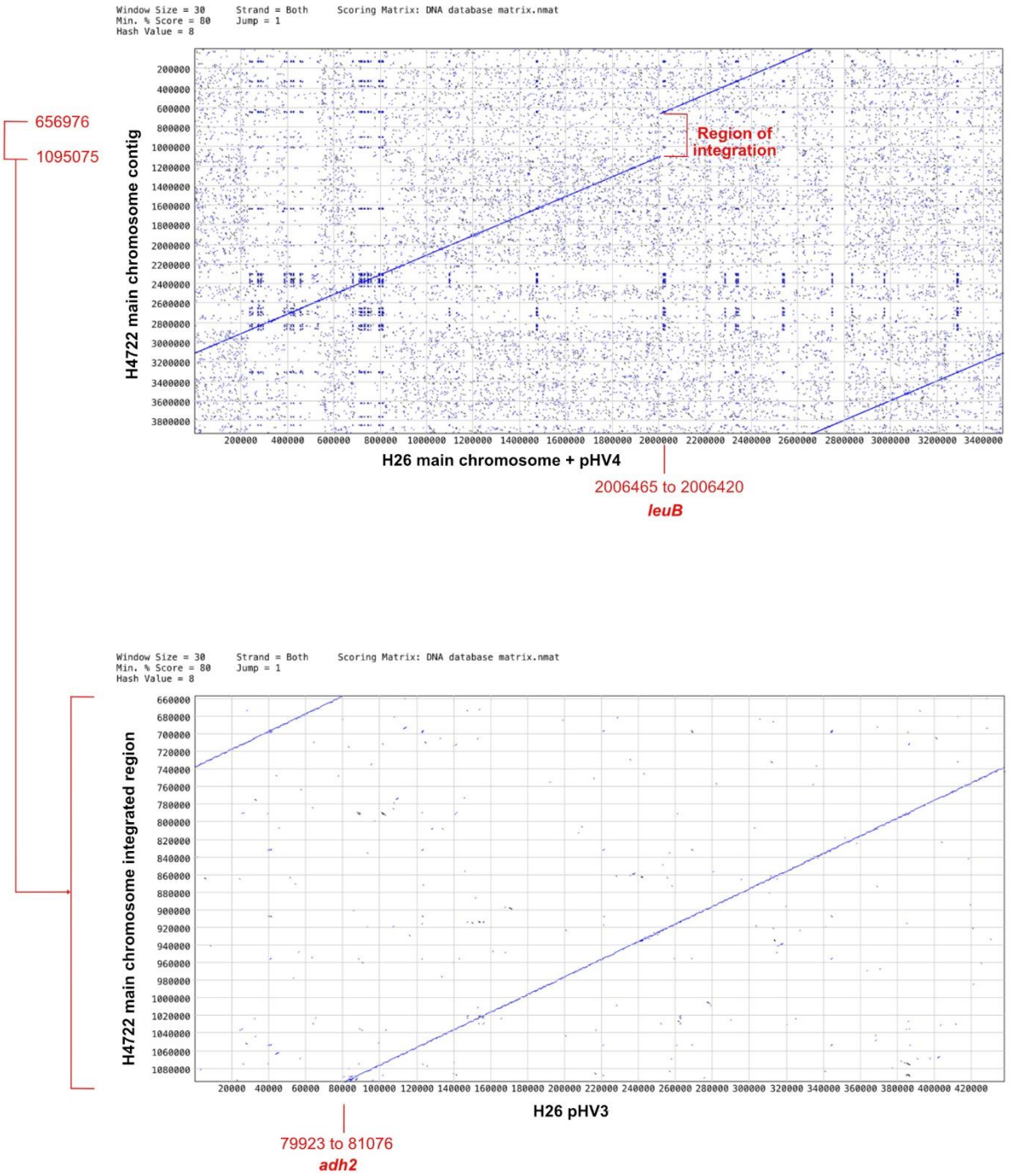
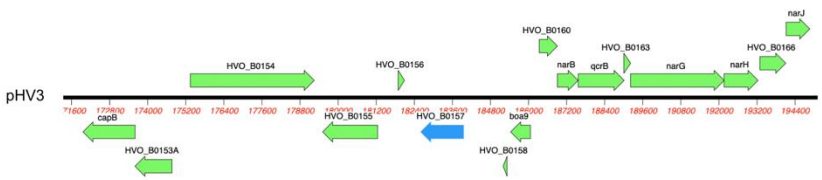
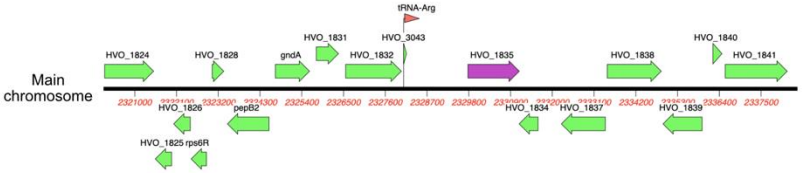


Figure 7.13: Pustell DNA matrices of H4722 Oxford Nanopore MinION sequencing data identifying the loci of pHV3 integration.

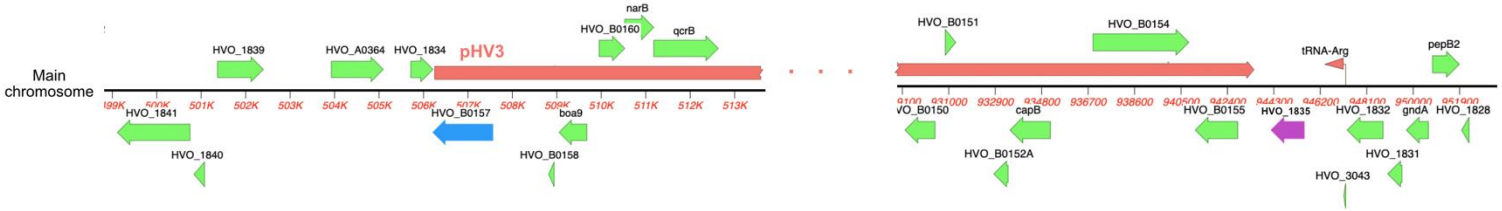
Wild type HVO_B0157 region:



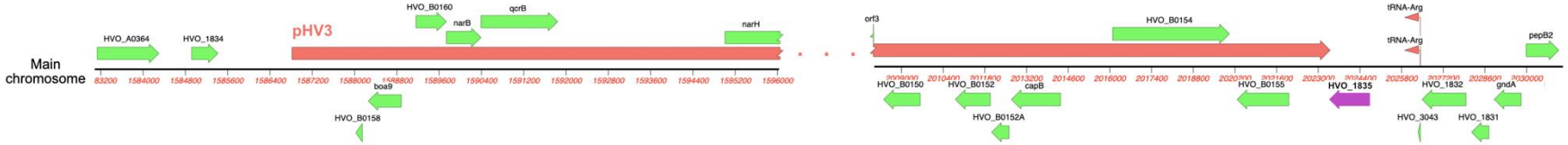
Wild type HVO_0275 region:



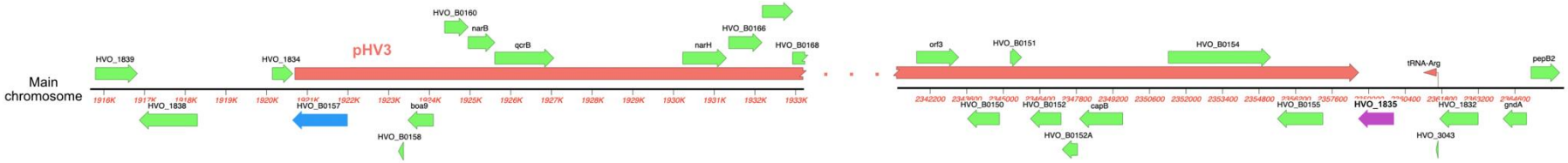
H4344:



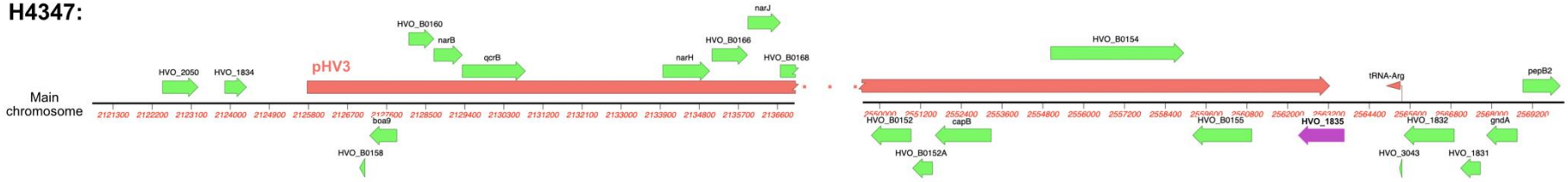
H4345:



H4346:



H4347:



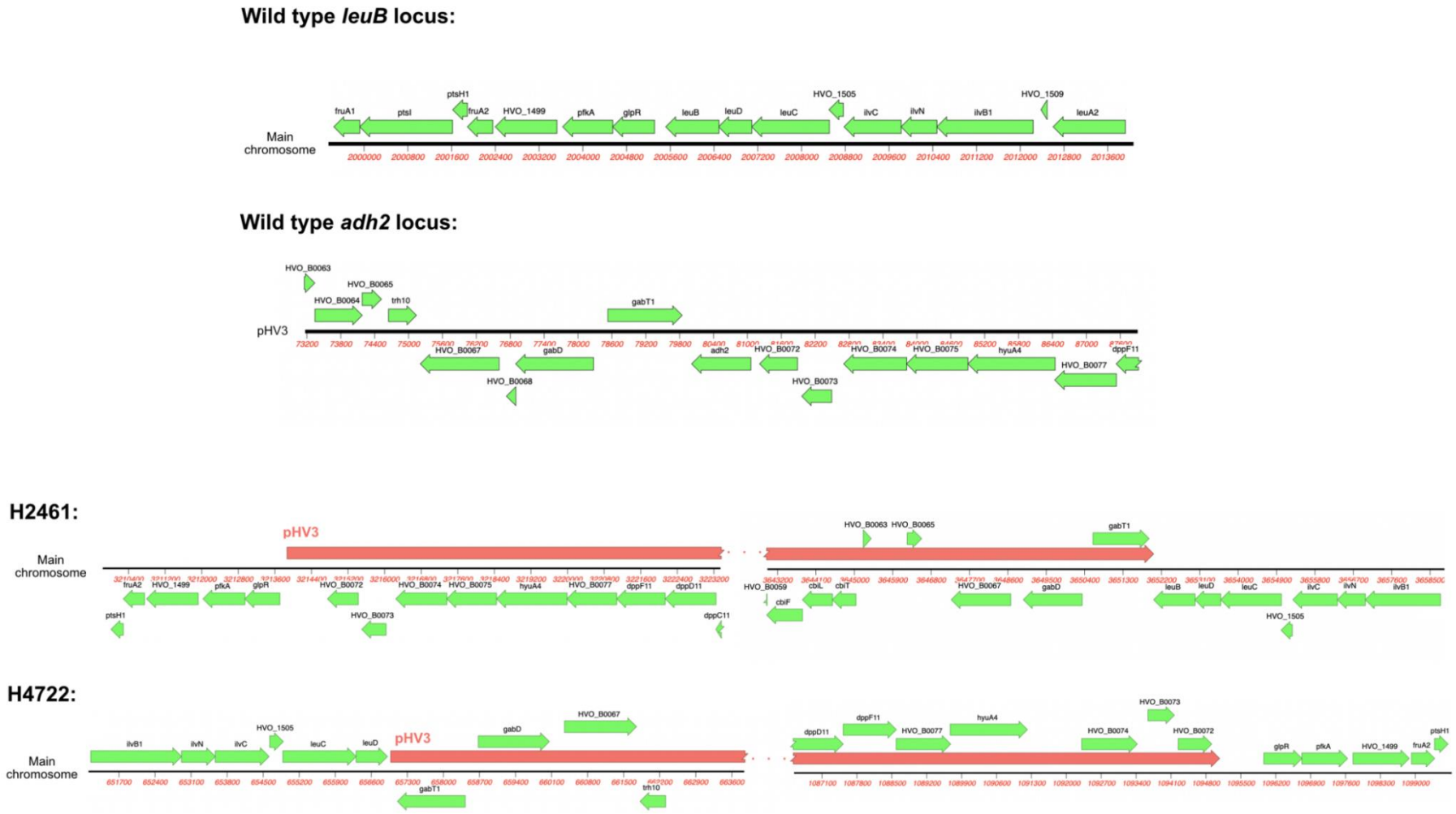
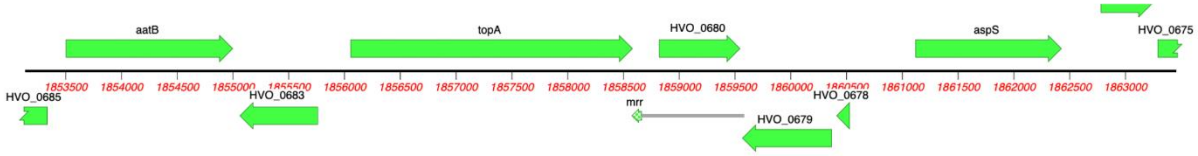
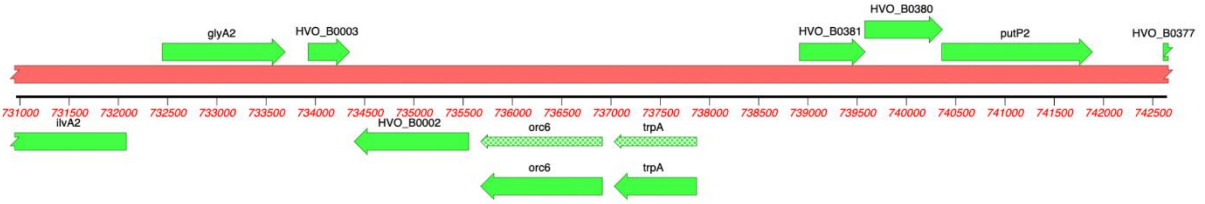


Figure 7.14: Regions of the genome of each of the strains shown at which transitions between pHV3 and the main chromosome occur. Wild-type HVO_1835, HVO_B0157, *leuB* and *adh2* regions are also shown.

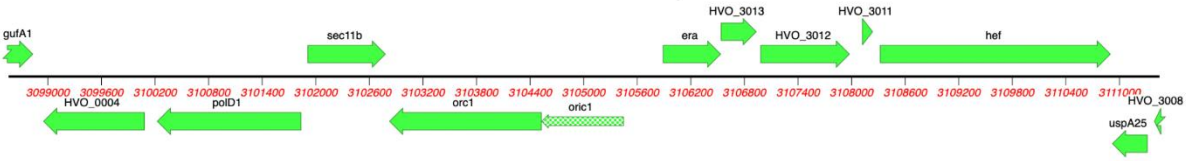
***Δmrr*:**



***orc6, ΔoripHV3::trpA+*:**



***oriC1*:**



***polB1*:**

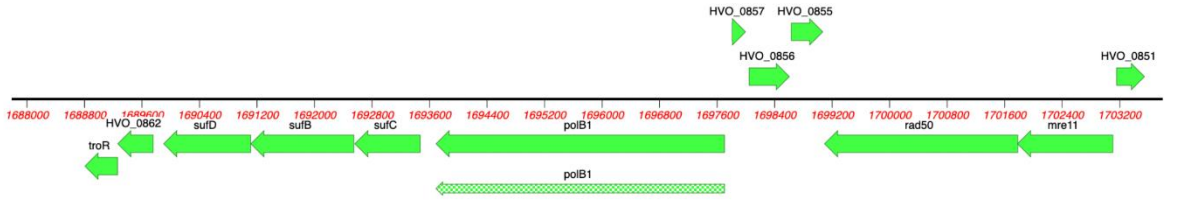


Figure 7.15: Genes aligned to the H4722 genome. Green checkerboard-filled arrows indicate sequences that have been aligned to the H4722 genome. Non-checkerboard green arrows are auto-annotated genes (detailed in Chapter 4). *mrr* has been deleted and only a small region of the gene aligns to the *mrr* region. *orc6* is present but *trpA* is found where *oripHV3* is found natively. *oriC1* and *polB1* are both present.