Origins "R" Us:
Investigating the Role of R-loops in Origin-Independent DNA Replication

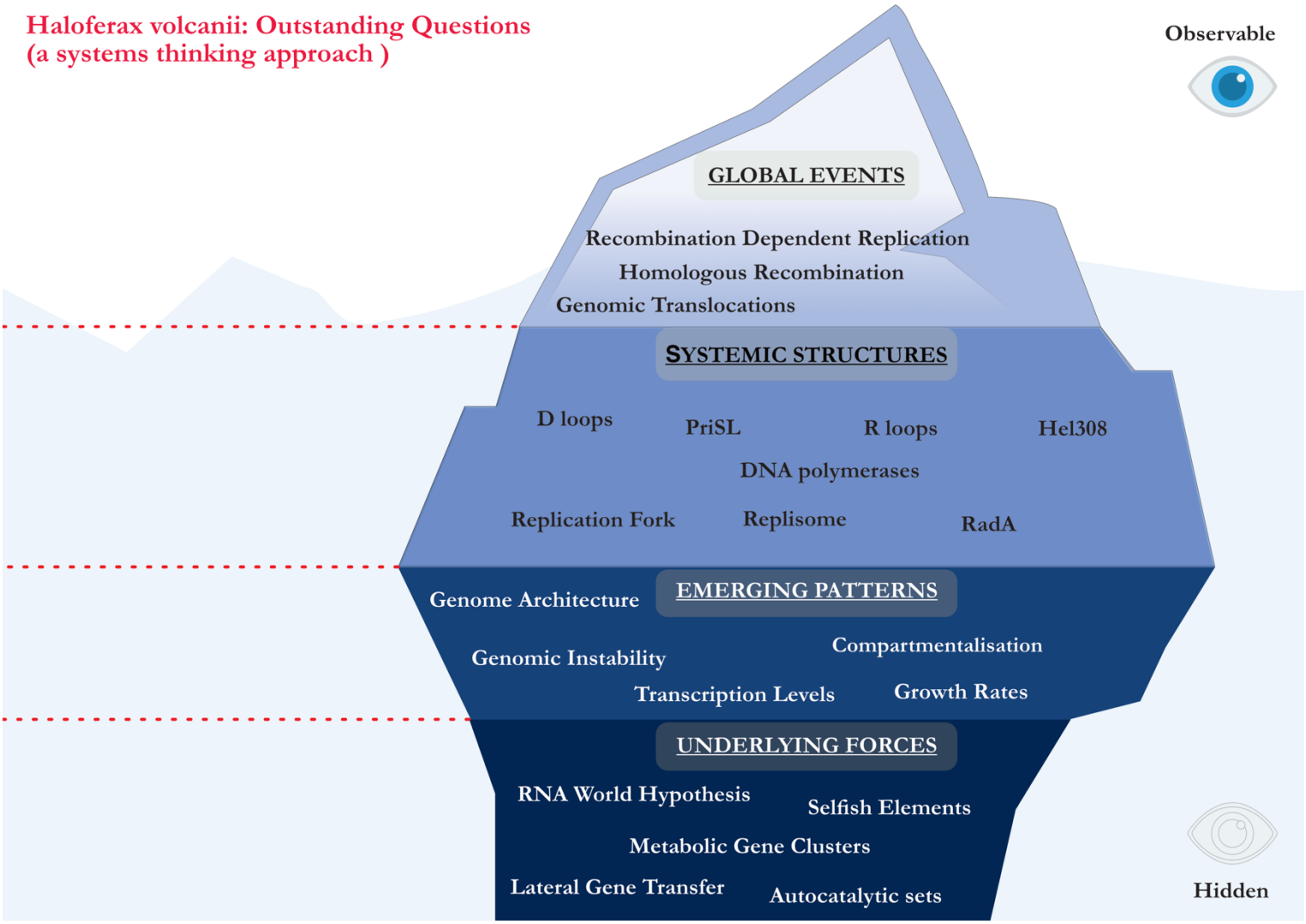
Anastasia Serdyuk, BSc (Hons)

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

2024

University of Nottingham
UK | CHINA | MALAYSIA

Graphical Abstract[1]



[1]The infographics in this review were created using Affinity Designer Software version 2.0.3 – by the author, unless stated otherwise where relevant citations for the reproductions are indicated.

Abstract

The ability of a molecule to self-replicate has been implicated to be the driving force behind the evolution of cellular life from the primordial RNA world. Thus the physicochemical properties of genomic replication are conserved in all three domains of life; Eukaryotes, Bacteria, and Archaea. One of these fundamental properties is the requirement of a terminal hydroxyl group for *de novo* DNA synthesis. The canonical DNA replication mechanism involves initiation from specific chromosomal sequences – origins of replication. However, an alternative mechanism – recombination dependent replication – has been observed in every domain; the cells are able to replicate without an origin while utilising the 3' end of a recombination intermediate (directly from R-loops, or indirectly from D-loops) to initiate synthesis from any location on the chromosome. Our understanding of the steps and enzymology of the full replisome assembly from recombination intermediates remains fragmentary. This is due to the small number of culturable model organisms that can replicate in an origin-independent manner. One of these organisms is the halophilic archaeon *Haloferax volcanii,* which growths faster in the absence of origins, but is easy to culture and is amenable to genetic manipulation. *H.volcanii* possesses a unique genome architecture: a main chromosome, and 3 mini-chromosomes; each containing multiple origins which can all be deleted except for the origin on pHV3. Moreover, attempted deletions of the pHV3 origin have resulted in genomic rearrangements, where pHV3 is integrated onto the main chromosome. The reasons for that are unknown, except for the low transcription levels on pHV3 detected in previous transcriptomic analyses.

To investigate the correlation between the levels of transcription, and the ability to delete the origin on pHV3, we have generated strains with increased transcription levels on pHV3 by employing three parallel research lines: through (1) engineering a tryptophan-inducible promoter for regulatory expression of the *adh2* gene on the pHV3 mini- chromosome, and (2) generating deletion constructs of *rnhA* and *rnhC* genes which degrade R-loops. Another way is (3) to stabilise the D-loop intermediate structure by attempting to delete the origin by transforming Hel308 deletion (ΔHel308) and point mutation (Hel308-**D**145**N**, Hel308-**F**316**A**, Hel308-**R**743**A**) strains with plasmids carrying Δ*ori-pHV3*, Δ*orc6* (initiator protein gene), and Δ*ori-pHV3* with simultaneous deletion of *orc6*.

Our results – for the first time – confirm the possibility of origin deletion on pHV3 in wild-type and Hel308-**R**743**A,** strains, orc6 initiator deletion (Δorc6) in Hel308-**D**145**N,** Hel308-**F**316**A,** and Hel308-**R**743**A** point mutation strains, as well as combined pHV3 origin with orc6 deletions in ΔHel308 and Hel308-**R**743**A** strains. This supports the initial hypothesis that the deletion of Hel308 may be involved in the stabilisation of D-loop structures, and be a key factor in the replisome assembly process.

However, it remains an open question whether this origin-deletion has occurred with pHV3 maintaining its episomal state. For that, further analysis using pulsed-field gel electrophoresis would confirm if pHV3 had integrated onto the main chromosome.

Overall, our work provides the foundational basis to warrant further study in the correlation between the increased levels of available recombination intermediates – through the stabilisation of D-loops/R-loops, or upregulation of gene expression, and hence transcription – and the deletion of origins in the model archaeon *Haloferax volcanii*.

Key Words: DNA replication; DNA Recombination; Recombination-Dependent Replication; D-loops; R-loops; Hel308; Archaea; Halophiles; *Haloferax volcanii;* Genome Architecture; Origin Deletion; Origins of Life; Evolution

Table of Contents

List of Abbreviations

Amp	Ampicillin
ARS	Autonomously Replicating Sequence
ATP	Adenosine Triphosphate
BIR	Break Induced Replication
DSB	Double Strand Break
Cas	Casamino acids
Cdc6	Cell Division Cycle 6
Cdc7	Cell Division Cycle 7
CDK	Cyclin Dependent Kinase
Cdt1	Chromatin licensing and DNA replication factor 1
ChIP	Chromatin Immunoprecipitation
CMG	Cdc45-MCM-GINS complex
cPCR	colony Polymerase Chain Reaction
DNA	Deoxyribonucleic Acid
ssDNA	single-stranded Deoxyribonucleic Acid
D-Loop````	Displacement Loop
DSB	Double Strand Break
DS	Downstream
DUE	DNA Unwinding Element
DRIP	DNA:RNA immunoprecipitation
GINS complex	**G**-**I**nchi-**N**i-**S**an [Japanese for 5-1-2-3 = Sld5, Psf1, Psf2, and Psf3 subunits]
HGT	Horizontal Gene Tranfer
LUCA	Last Universal Common Ancestor
MCM	Minichromosome Maintenance Complex
MFA	Marker Frequency Analysis
ORB	Origin Recognition Box
ORC	Origin Recognition Complex
OriC	Origin of Chromosomal Replication
PCR	Polymerase Chain Reaction
Pif-1	Petite Integration Frequency 1
Pre-RC	Pre- Replicative Complex

RCR	Rolling Circle Replication
RDR	Recombination Dependent Replication
RIP	Replication Initiation Point (mapping)
R-Loop	RNA Displacement Loop
RPA	Replication Protein A
cSDR	constitutive Stable DNA Replication
iSDR	inducible Stable DNA Replication
S-phase	Synthesis phase
US	Upstream
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galacto pyranoside

Chapter 1 DNA Replication: the When, Where & How

1.1 Setting the Scene

The core genetic information processing pathways and associated machinery – which promote cellular life and its propagation – are conserved across the three domains of life: Eukarya, Bacteria, and Archaea. It is now widely accepted that the genomic content of every organism is contained as DNA within the chromosomes of its cells. And before these cells can divide, the entire genome must undergo accurate and timely duplication in order to be distributed equally, and remaining identical to the original copy, into the daughter cells. Faithful duplication requires that the genome is replicated only once per cell division cycle. Errors in DNA replication present a threat to not only the viability of an individual cell (i.e., chromosome rearrangements and breakage leading to cell apoptosis), but also to the entire organism – where DNA lesions impede replication progression in the form of stalled or blocked replication forks. Eventually, accumulation of sustained DNA damage leads to genomic instability (i.e., global replication stress) – a hallmark of cancer aetiology[1]. Thus proper genome maintenance is dependent on the cooperation of several tightly linked processes termed the 'Three Rs': Replication, Recombination, and Repair.

Much scientific effort has been directed at understanding how the assembly of replication machinery is coordinated in space and time, and what safety mechanisms are activated should any errors arise. Given the prevalence of genomic instability in human disease, it is unsurprising that DNA replication constitutes one of the most active research areas in today's field of molecular biology. Therefore, this section aims to provide a historical perspective on the period of uncertainty (i.e., the 'Replication Problem'), and the subsequent burst of research (i.e., the 'Molecular Revolution'), that has led us to the latest guiding paradigm – a theoretical framework – for DNA replication mechanisms.

1.2 The DNA Replication Problem

The breakthrough marking the beginning of molecular biology was when the 'transforming principle' from pneumococcal bacteria was discovered to be made of DNA, as opposed to of protein[2,3]. Avery and colleagues provided the initial evidence of the genetic material's chemical composition. The implication of DNA playing a role in the transmission of genetic information – regarded as an axiom today – presented a theoretical challenge against the existing protein-centred hypothesis[4,5]. Despite the isolated substance being resistant to trypsin, chymotrypsin, and ribonuclease agents, Mirsky argued that there was a possibility of protein impurities in Avery's samples, thus igniting a long debate on the true nature of the transforming principle. Because there was no counter evidence, the ability of DNA to transform all organisms was regarded as a working hypothesis for almost a decade.

The main obstacle to the acceptance of Avery's work was that the DNA polymer was thought of as 'too simple': how can a single molecular entity consisting of a 4-base nucleotide sequence permit such diversity in genes across all kingdoms of life? As a response, Chargaff proposed that the amount of adenine was equal to thymine, and cytosine to guanine[6] – with the respective ratios differing across species – thereby overthrowing Levene's tetranucleotide hypothesis[7]. Hershey-Chase's experiments supporting DNA's genetic role[8], were readily accepted despite the 25% of protein contamination[9], and provided an essential clue. The same year, genes previously seen as a hypothetical abstraction were 'rediscovered' as concrete, structural entities. Watson and Crick[10] had known that the key to understanding the biological processes of heredity was contained in the tertiary structure of DNA, so they applied the available crystallographic[11,12], and biochemical data[13] to assign a right-handed double helix model[14] to the enigmatic molecule. Assuming 'form follows function', they proposed that the complementary nucleotide bases between coiled helices were held by hydrogen bonds (eventually termed Watson-Crick interactions). Another important consequence of the model was that it implied a self-duplicating mechanism, whereby one strand of the helix acts as a template to direct the synthesis of the new strand through phosphodiester bond formation between the sugars and the nitrogenous bases, arranged in an antiparallel orientation. Their suggestion raised contention among other leaders in the field who noticed a problem with their model; namely, the mechanism of helical unwinding through hydrogen bond breaking before synthesis[15]. Delbrück drew attention to the plectonemic coiling of the helix, and argued against the semi-conservative replication by suggesting an alternative dispersive mode (*see* figure 1). The replication debate was settled when Meselson and

Stahl provided evidence of semi-conservative replication[16]. Watson concluded: "Nor does the need to untwist the DNA molecule to separate the two intertwined strands represent a real problem"(for review *see*[17]).

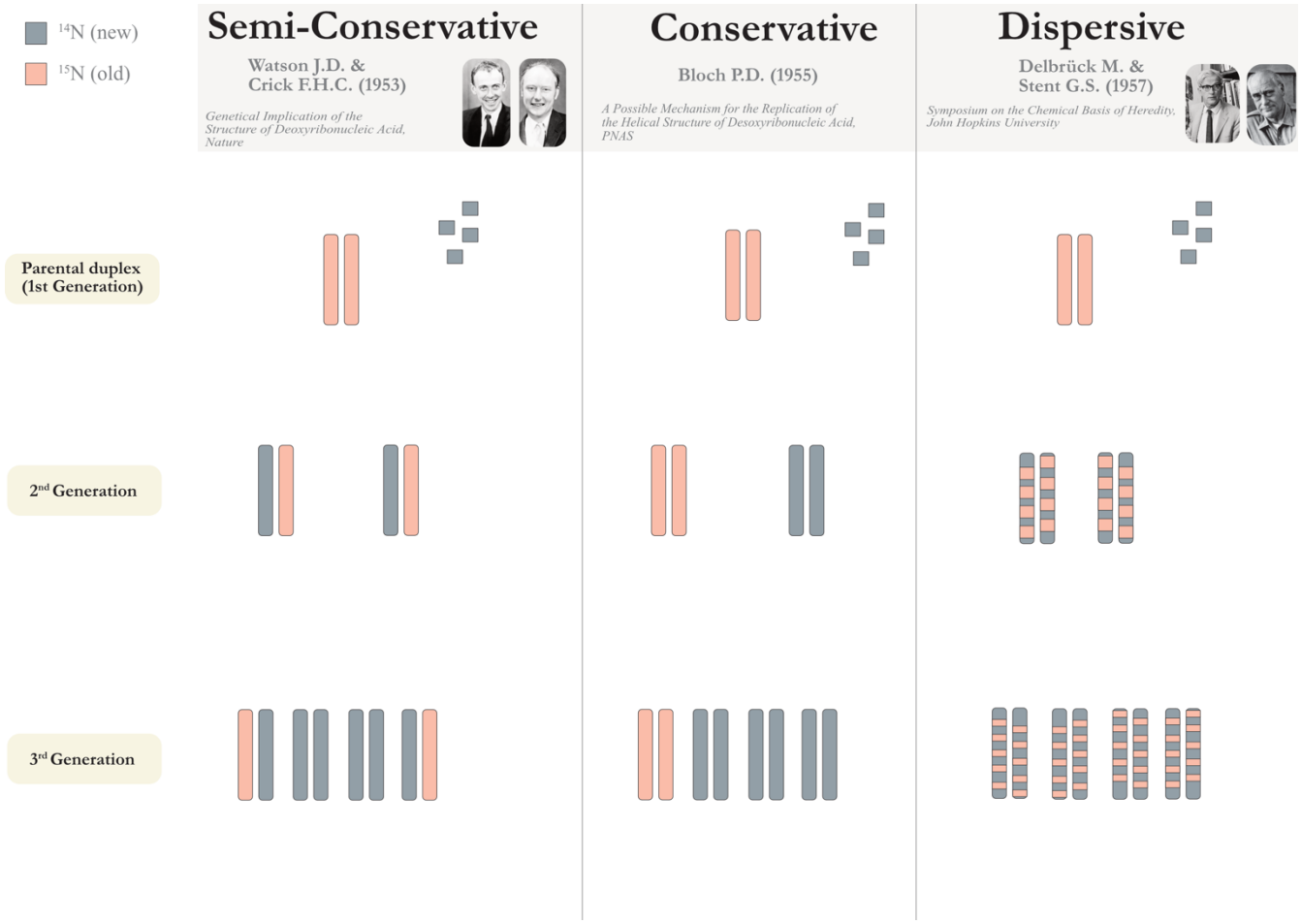


Figure 1. **Proposed Mechanisms of DNA Replication: Semi-Conservative, Conservative, and Dispersive.**

The schematic represents the expected outcomes according to each mode of replication represented by the pioneering groups during the molecular revolution.
In **semi-conservative** replication, the two parental strands separate, with each strand acting as a template to direct the synthesis through complementary base pairing, with the resulting daughter duplex consisting of the newly synthesised, and conserved parental strands. In **conservative** replication both parental strands are conserved, and in **dispersive** mode the double helix remains unwound, while segments break and re-join through crossing over thus the newly synthesised DNA appears 'dispersed' in the daughter strands.
Meselson and Stahl demonstrated **semi-conservative replication** by taking an alternative approach to radioactive labelling (in contrast to the Phage group's use of bacteriophage, rendering inconclusive results) – and growing *E.coli* cells in $^{14}NH_4Cl/^{15}NH_4Cl$ mediums containing N15 ('heavy') and N14 ('light') nitrogen isotopes, to measure the gradient densities every generation. This elegant experiment was conducted using a combination of Avery's DNA isolation, and Hershey-Chase's isotope labelling techniques.

1.2.1 The DNA Polymerase Puzzle

The key evidence for DNA's genetic role, and its semi-conservative mode of replication came from Kornberg's lab, where the chemical process of DNA synthesis was reconstituted *in vitro,* followed by the purification of the "catalytic extracts" which contained the enzyme required for phosphodiester bond formation, and chain elongation – discovered as DNA polymerase I – in *E.coli*[18,19]. Previous analyses have shown that the precursor to strand formation must be an activated nucleoside 5'- phosphate[20]. Analogously to glucose-1-phosphate being activated to uridine-diphosphate glucose in glycogen synthesis[21], Kornberg's group generated four 32-P labelled nucleotide bases – dATP,dCTP,dTTP, and dGTP – to serve as starting units for the synthetic DNA strand extension reaction. From that, they have formed their initial hypotheses regarding the enzymatic mechanisms, and the chemical composition of the replication products:

(1) Is the synthesised DNA strand identical to its template?

Does DNA synthesis proceed in a template directed manner like Watson-Crick's model would suggest, and is the newly synthesised DNA therefore a faithful copy of its template? The 'nearest neighbour' technique of 32-P labelled nucleotides revealed that the frequency of nucleotide pairs, and the complementary base ratios between the 'starting', and the synthesised strand remained identical – serving as corroboratory evidence for the antiparallel orientation outlined in the double helix model. The authors were surprised to find that all four nucleotide bases, as well as DNA polymerase and Mg^{2+}, were required; if the template substrate served as a simple primer, why were all four nucleotides a necessity? This has prompted further questioning:

(2) Does replication proceed in a template directed manner as predicted by Watson and Crick, catalysed by DNA polymerase?

When "DNA primers" containing differing ratios (i.e., 0.5 to 1.9) of nucleotide base pairs were used – the synthesised product maintained the initial nucleotide pair ratios, and was independent of the concentrations of the individual bases thus indicating template-directed replication.

These conclusions have laid the foundation for DNA replication research using bacterial models that occupied scientists for the next 70 years and counting. During the Nobel Prize acceptance lecture, Kornberg compared DNA to a "tape recording" in that:

"exact copies can be made from it so that this information can be used again and elsewhere in time and space."

But how are these copies during DNA polymerase directed synthesis made "exact"?

Considering that the nucleotide pool contains an unequal proportion of the four bases, what are regulatory mechanisms that ensure accurate nucleotide selectivity? At the base-pairing selection step, (1) the correct nucleotide must be selected for the polymerisation reaction through correct geometric pairing with the polymerase, and (2) the preceding nucleotide in the primer terminus is "proofread" for accurate base pairing before the addition of the second nucleotide (*see* figure 2). This highlights another universal hallmark of DNA (i.e., replicative) polymerases – that is, the inability of *de novo* synthesis. Replicative polymerases are incapable of performing the initial phosphodiester bond formation between two dNTPs – in contrast to RNA polymerases – thus they must add nucleotides to a pre-existing RNA primer site at the template, synthesised by a specialised RNA polymerase called primase, and extend it from 3'OH end of the single-stranded DNA template strand (n.b., implications of the 3' prime end requirement form a recurrent theme throughout this review, and is made relevant in various systems).

While preparing their second manuscript, Kornberg's group faced a problem: they were unable to remove deoxyribonuclease activity from the polymerase. It was later found that the reason was the presence of 3'- 5' exonuclease domain[22] that carries proofreading, as well as mispaired nucleotide excision (i.e. editing) mechanisms. When the second replicative enzyme – called DNA polymerase III – was isolated from *E.coli*, it took on the role as the primary enzyme responsible for the elongation of the majority of the bacterial chromosome. Since then, sequence conservation[23] and biochemical[24] studies have led to the classification of DNA polymerases from all three domains of life into six families: A, B, C, D, X and Y – with the first four polymerases responsible for high fidelity DNA replication, while X and Y are more specialised forms of lesion bypass, and translesion synthesis polymerases involved in DNA Repair[25]. PolA, PolB, and PolC are homologous to PolI, PolII, and PolIII families in *E.coli*[26] respectively, with family B most commonly found in Eukaryotes, families A and C in Bacteria, and families B and D in Archaea. Despite differences in function between polymerases, the archetypal DNA-dependent polymerase (figure 2**B**) is composed of a core polymerisation catalytic site, which itself is composed of fingers, palm, and thumb subdomains, as well as a separate 3'-5' exonuclease domain that proceeds in an opposing direction to DNA synthesis. In case of a mismatched base pair, the catalytic step is slowed down, and the nascent strand terminus is 'shuttled' from the polymerisation to the exonuclease active site of the DNA polymerase (figure 2**C**) for the excision of the incorrect nucleotide through bond hydrolysis. Such structural distribution of enzymatic function is exemplified by the crystal structure of the multidomain *E.coli* Pol I Klenow fragment[27] which retains 3'-5' exonuclease (proofreading) , and 5'-3' polymerisation activities,

thereby contributing to replication fidelity through intrinsic proofreading and strand displacement synthesis[28] abilities. It is worthwhile to note that the polymerase is a molecular motor capable of translocation along the template strand which proceeds chiefly in terms of chemical thermodynamics. In other words, DNA polymerase acts as a "channel" for the copying of genetic information, by the "reading" of each nucleotide on the template strand, and "writing in" of the complementary nucleotide through a nucleotidyl transfer reaction, where the paired nucleotides are stabilised by hydrogen bonds and base stacking interactions. This ability to convert "information" through a physical reaction or "work" has led some authors to propose that the polymerase functions analogously to a Maxwell's demon[29,30]. The "memory" of an organism's genetic information is embedded within the DNA polymer's structure, where DNA replication is the reversible process of "retrieving" and "storing" of this information – with **information processing** and **assimilation** being the defining features of a **complex system** or a living organism. The RNA-first scenario proposes that while modern genetic apparatus requires to be encoded by proteins, in the early pre-DNA environments (i.e., the RNA world), the ancient RNA-dependent RNA polymerase or the ribozyme harboured the ability to self-replicate, and thus played a major role in the RNA to DNA transition. The DNA molecule – due to it inherent stability – has replaced RNA as the main genetic material; however, the imposed directionality of 5' triphosphate addition in DNA synthesis is itself an artefact of the RNA world metabolism. The structural similarities of polymerase families A and B, as well as viral RNA polymerases all suggest a common origin[31].

Following this reasoning, the polymerase is the earliest form of a self-reproducing system that has evolved from prebiotic conditions; whose ability to harbour both "information" and "function" has been the driving factor of evolution itself. Thus it can be assumed that the basic physicochemical forces underpinning DNA replication are both conserved and fundamental in all living systems. Various kinetic studies (for review *see* [32] and citations therein) using DNA polymerases have therefore led to a minimal model[33] of the polymerisation process; its mechanics are outlined in figure 2**A**

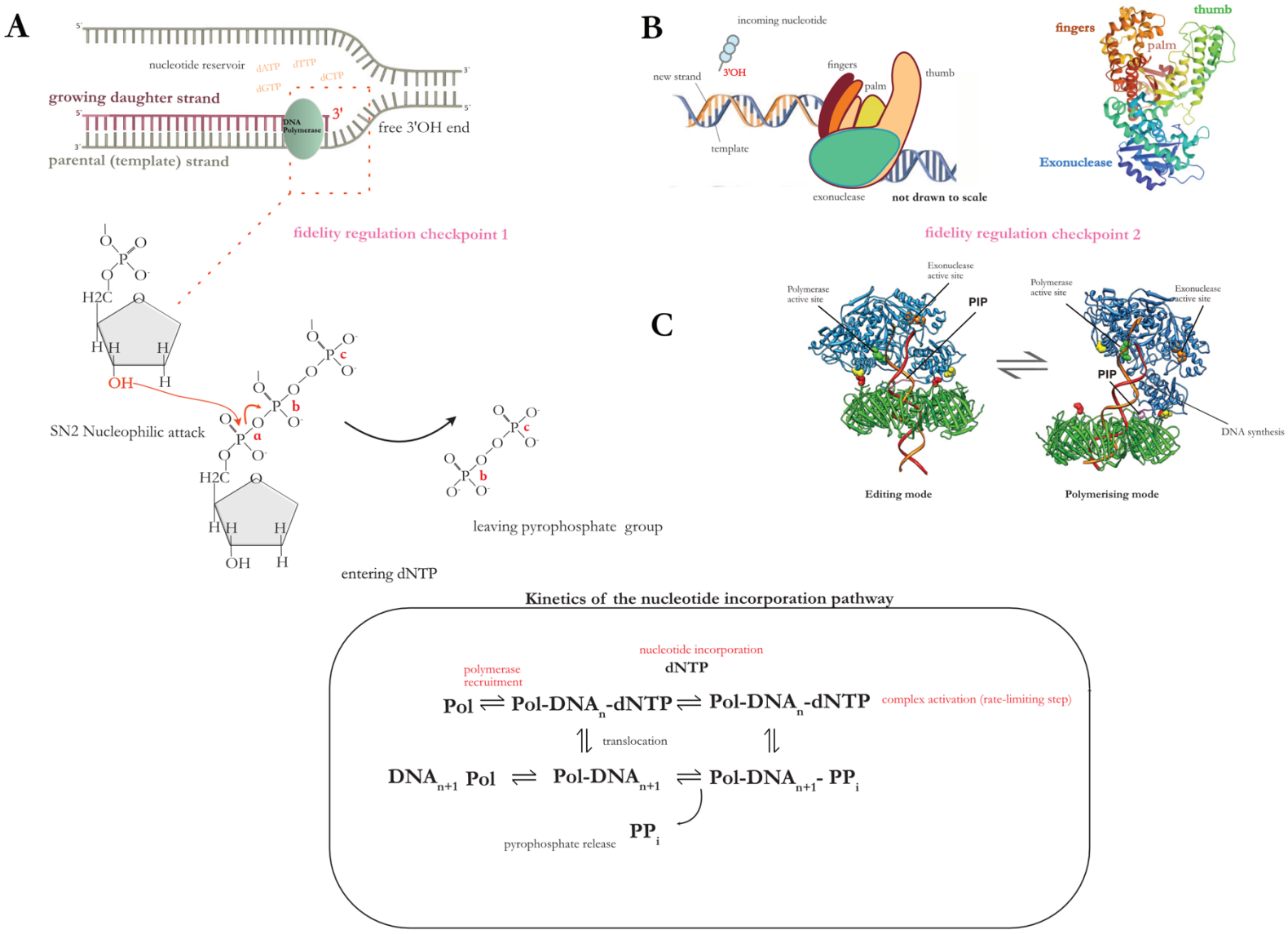


Figure 2.

(A)Universal mechanism of nucleotide incorporation during the polymerisation step in DNA replication. Among all studied replicative polymerases, phosphodiester bond formation occurs via a conserved stepwise mechanism.

(B) The side chains of the 'fingers' domain (refer to diagram 2**B**; schematic of the polymerase multidomain organisation (left); crystal structure of *E.coli* Pol I Klenow fragment (right), adapted from [28] (PBD ID:1KFD)) bind the incoming dNTP, and position it in the conserved palm domain (i.e., the catalytic unit). The active site of the palm contains two essential aspartic acid residues which coordinate the two divalent ions necessary for the nucleotidyl transfer reaction: the activated 3'OH on the nascent strand terminus performs a nucleophilic attack on the α-phosphate of the dNTP thus resulting in phophodiester bond formation through a condensation reaction. The inorganic pyrophosphate group (PP_i) bond is hydrolysed, and the free energy change ensures forward translocation of the polymerase along the template.

(C) The dNTP substrate can only undergo activation in its 5' position, which is what imposes the strict unidirectionality of DNA replication. What is the reason behind this universal requirement, if the 5'OH is just as capable of a nucleophilic attack? The answer lies in the proofreading function of the polymerase; the addition of one nucleotide per synthesis step ensures fidelity, and polymerase repurposing for multiple enzymatic reactions without dissociating from the DNA is bioenergetically convenient. The thumb domain assists in the switching of the polymerase between polymerisation to editing modes (figure 2**C**; tertiary structure of the *P.furiosus* PCNA-PolB-DNA complex switching from pol and exo states, taken from[34]).

Polymerase selectivity is one of the major contributors to overall fidelity of replication, with the proofreading function increasing accuracy of copying by 10^2-10^3 fold at nucleotide level[34]. At a more global scale, however, the order of replication events must be regulated both temporally and spatially. The genome must be replicated during the synthesis (S-phase) stage before cell division, and at the same time must occur *only once* per cell division cycle to avoid over-replication. Aberrant replication initiation events can lead to chromosome copy number alterations (i.e., aneuploidy or polyploidy), and promote genomic instability through the accumulation of mutations. Thus the formation of the replication bubble must occur at a specific locus of the chromosome – dictated by the location of the replication origin – and proceed in a timely manner in accordance with the cell division cycle, as well as transcription and DNA repair events[35]. It is therefore unsurprising that the main regulatory step through which this is imposed is replication initiation.

1.2.2 Replicon Model: leading paradigm for the study of DNA Replication

It is helpful to think of initiation of any biological event as a result of the direct or combined action of regulatory elements; on specific substrates, as well as the negative or positive effects these elements elicit upon binding. Early models of gene expression control were centred around its repression – for example, the (lac) operon model of bacterial gene regulation, as proposed by Jacob and Monod[36], states that gene expression is controlled by a regulatory circuit formed through specific interaction between a trans-acting repressor factor and a cis-acting operator. The authors reached a – what may currently seem rather short-sighted – conclusion that these genetic control mechanisms operate solely through inhibition, and that the removal of these repressive effects is the main event that activates protein synthesis. With the lack of integrative approaches, progress in bacterial cell biology research had come to an impasse; there was a fundamental gap in knowledge on the integrative action of molecular mechanisms within the cell. Jacob and colleagues had expressed this growing sentiment at the 1969 Cold Spring Harbor symposium[37]:

"we still know very little about the general system which integrates cellular controls, the regulation of DNA replication, the formation of bacterial membrane, and the process of cellular division with its equipartition of the DNA copies"

Following the discovery of extra-chromosomal, self-replicating genetic elements – called episomes: a term now used interchangeably with plasmids[38] – Jacob *et al*[39] proposed a simple replicon model for replication initiation in *E.coli* circular chromosome. In their model, an

individual unit of replication – the **replicon** – is defined by the specific chromosomal sequence called a replicator (i.e., replication origin or ori ; 'operator of replication')[40], from which replication is initiated upon the interaction with the trans-acting, diffusible initiator protein (whose own structural gene is found frequently in proximity to the native origin) in a sequence-specific manner. This in turn triggers the recruitment of a helix unwinding element called helicase that acts as a stable platform for the assembly of replication machinery – collectively referred to as the **replisome** – in a concerted manner forming a replication bubble from the single strands for elongation to occur (*see* figure 2**A**,**B**). A defining feature of the replicon unit is that it encodes specific determinants (that is, the replicator and the initiator) which allow it to process control signals allowing it to autonomously replicate as one whole. This is in contrast to other autonomously replicating sequences such as episomal plasmids. The replicon hypothesis provides an explanation for the phenomenon in *E.coli* called plasmid incompatibility which arises due to the competition of several plasmids for the same initiation factors, thus preventing stable inheritance[41]. An observant reader may notice that the replicon is a reworking of the earlier lac operon model, combined with the idea of a diffusible factor interacting with the membrane during bacterial conjugation[42] – the operon repressor is analogous to the initiator, and the operator to the replicator, with one critical distinction being that the initiator acts as an activator in a positive interaction with the origin. However, due to the nature of replication being inherently autocatalytic, regulation cannot be complete without the reciprocal actions of both activation and repression mechanisms that occur during distinct stages of the cell cycle. If the rate of replication is determined by the frequency of initiation events, what are the distinct factors that regulate origin firing in space and time?

1.2.3 The Divided Genome: Nature's Riddle

The following section discusses the limitations of the single replicon model – the findings that stimulated its subsequent reworkings, and a revision of the commonly accepted terms such as replicon unit, origin of replication *etc.,*

Diversity in Replication Factors

The replicon model was shown to be highly adaptable to most bacterial systems, with limitations arising when extended to more complex genomes, such as the ones of higher eukaryotes. Due to genetic simplicity (i.e., a circular chromosome with a single bidirectional origin) and ease of culture, *E.coli* served as the leading model for the identification of ARS elements through the cloning of candidate replicator fragments into a marked plasmid vector, selected for their ability to self-replicate, and remain as a separate unit within the host cell. Using this simple ARS assay, the *E.coli* replicator OriC was identified[43] thus making the replicon model a guiding paradigm for the replication regulation and origin prediction[44] in bacterial systems. Supporting evidence for the replicon model came from the isolation of the *E.coli* initiator – a 473 amino acid protein called DnaA[41,42] was shown to bind with high affinity to DNA containing the sequence for OriC, in an ATP-dependent manner[45]. The DnaA initiator protein, which binds to the specific 9-aa consensus sequence called DnaA box (clustered within the 250-aa OriC region; the DnaA gene itself is usually found adjacent to the origin) which controls the replication of the entire chromosome was found to be highly conserved among bacterial species[46]. In eukaryotes, such as budding (*s.cerevisiae*) and fission (*s.pombe*) yeast, the ARS technique developed through bacterial genetics led to the isolation[47] and sequence analysis[48] of yeast ARS elements – 100bp long, with a characteristic AT-rich consensus sequence (5'-[A/T]T̲T̲TAT[A/G]T̲T̲T[A/T]-3') – that serve as putative replicators. From that, the eukaryotic initiator multiprotein complex ORC was purified from budding yeast in 1992[49].

One may think of a replicator as a specific initiation site or control point for an individual event of bidirectional replication; the single origin model in bacteria served as a useful starting point for the identification of several initiator proteins under set physiological growth conditions. Under stressful growing conditions (e.g., arrested protein synthesis) the paradigm is flipped: both OriC and DnaA are shown to be dispensable during SDR in thymine-starved *E.coli* cells (*see* section 2.3). While the hetero-hexameric ORC initiator is conserved in eukaryotes, with orthologues found from yeast to humans[50], the cis-acting replicators or clusters of origins are highly diverse among different species. In the majority of bacteria, the dual DnaA- OriC interaction occurs in a sequence specific manner to replicate the single circular chromosome. This is in contrast to eukaryotic replication systems, which typically possess many linear chromosomes that are larger in size, on which there are clusters of origins – where one round of replication may initiate from hundreds to thousands of origins, as depicted in early autoradiography studies[51]. The way the replicon model falls short is that it fails to address spatial

and temporal regulation of initiation which occurs in eukaryotes such as fission yeast with an excess of activation-capable origins, and more fluid control mechanisms [52].

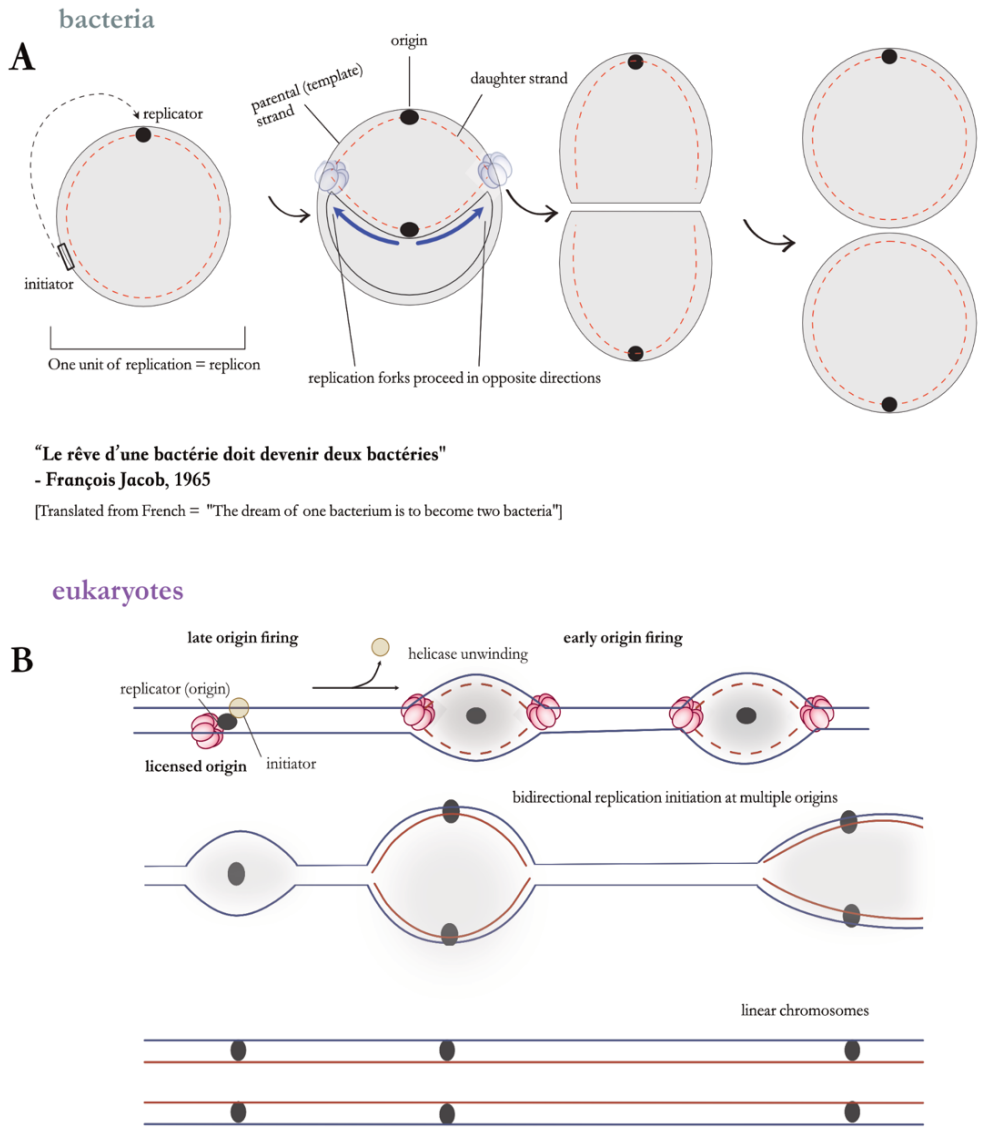


Figure 3.

(A) Early model of the replicon hypothesis in bacterial systems.
(B) Adaptation of the replicon model to eukaryotic genomes. The earlier model was reworked to accommodate the multiple origin organisation in eukaryotes, from studies in ARS elements in budding yeast. In eukaryotes, origins are fired asynchronously during S-phase. For an origin to be 'activated', it must first be licensed through the recruitment of various replication factors. Differential timing of origin activation – marked by early and late replication initiation events – prevents over-replication or aberrant re-replication events. Thus a single set of initiation factors activates hundreds- to-thousands of replication origins on a single eukaryotic linear chromosome.

Many Origins, One Chromosome: Time to Revisit the Single Replicon Model?

Table 1. **Glossary of revised terms used in this review.** For more detailed descriptions see[53,54]

Glossary †

Chromosome
= the largest or the primary replicon, containing the majority of core and essential genes

Mega-Plasmid
= characterised by lack of core genes, thus being disposable for cell viability in normal environmental conditions

Mini-Chromosome
= synonymous with mega-plasmid

Plasmid
= like mega-plasmids, lack essential genes but are smaller in size (suggested 350Kb cut-off, or 10% of genome)

Extra-chromosomal element
= a genetic element which is distinct from the main chromosome, and is capable of replicating independently, outside of the host cell (an example would include a plasmid).

Chromid
= an intermediate genetic element that falls between a plasmid and a chromosome, with a single replicon. Main distinction is the presence of at least one essential gene that renders the chromid non-dispensable. Additional regulatory elements that synchronises its replication with the cell cycle may be present.

Replicon
= in bacterial systems; a DNA molecule with a single origin of replication. In Archaea with multiple origins on a circular chromosome, use of the canonical term should be avoided.

Replisome
= a multi-protein complex of replication machinery which is assembled at the replication fork to form a replication complex at the start of DNA replication

Replication Fork
= a Y-shaped structure that forms as the DNA helicase unwinds the DNA double helix , and forms a region for the assembly of the replisome for replication to proceed bidirectionally from the replication bubble

† list of common terms encountered in literature that are subject to misinterpretation

The replicon model was constructed on the dogma that bacterial domain members can be defined by the possession of a single, circular chromosome which encodes a conserved set of essential genes[55] (*see* glossary) – however the paradigm was overturned when alphaproteobacteria containing a secondary replicon carrying essential genes were discovered[56], and the expansion to other members of the bacterial domain stimulated a revision of these historically used terms.

Moreover, it has been shown that 10% of bacterial genomes differ from *E.coli* in that they contain numerous replicons which can be both circular or linear[54].

Advances in genome sequencing during the 1970s have led to the identification of the third domain of life – the archaea[57] – which provided a novel platform for comparative molecular biology. Search similarity techniques to previously known origins in other domains for *bona fide* origins in archaea have not given results; the nature of the archaeal origin, or if replication was initiated through origins at all, remained unknown long after the archaeal genomes were first sequenced[58]. Since archaea bear a morphological resemblance to bacteria in terms of their chromosomal structure, it was initially proposed that they contain a single replication origin. Indeed, using codon (GGTC) skew analysis, Myllykalio and coworkers[59] have identified the first replication origin (i.e, OriC) in the hyperthermophile *Pyrococcus abbysi,* corroborated with experimental evidence from 2-dimensional gel[60] and RIP mapping[61]. The first archaea with multiple origins to be mapped using gel analysis came from the *Sulfolobus* genus[62] – stimulating a major shift in thinking at the time. Through the use of MFA techniques[63] , it was also shown that bidirectional replication occurs from each of the three origins. These three origins were also found to be involved in complex cross-interaction with the adjacently encoded initiator proteins, Orc 1-1 and Orc1-3 [64], as well as a WhiP (winged-helix initiation protein)[65]. But what makes the *Solfolobus* genome especially intriguing is that the genomic region adjacent to OriC3 appears to be 'captured' from a virus or an extra-chromosomal element of viral origin[65] (*see* glossary for **extra-chromosomal** element). The staggering sequence diversity of the *Sulfolobus* origins (OriC1-3) also hints at independent derivation through horizontal gene transfer.[62]

Similarly to *Sulfolobus*, other archaeal genomes[65–67] were also found to be composed of multiple replicons, with each replicon containing more than one replication origin. What is the exact definition of a replicon or a single replication control point given the divided genomic architecture, and the cross-interaction between multiple replicator-initiator systems? DiCenzo and Finan[53] suggest that classical terms such as 'replicon', should be used with caution – if not at all discarded – when describing genomes that fall outside the canonical *E.coli* model. It was also assumed *a priori* that genome replication cannot be initiated without replication origins. However some archaeal species, such as *Haloferax volcanii*, demonstrate that replication without origins occurs faster, and without any phenotypic deficits[68]. Thus we arrive at another critical juncture; what is then the initial evolutionary purpose for replication origins? Are origins of replication ancestral genetic elements or were they recently obtained through horizontal gene transfer? If so, at what point in evolutionary history have origins been captured? And perhaps more importantly,

could the extra-chromosomal elements capture postulate be extended to explain the evolution of multiple initiation sites and the linearisation of the chromosome in eukaryotes?
Given the above lines of evidence, the reader might then arrive at the conclusion that the organisation of multireplicon genomes is far from stochastic – that their maintenance must hold some functional or evolutionary purpose[53]. In fact, genome rearrangements such as insertion-deletion events from mobile genetic elements[69], and origin transfer[67] between species was the driving force that shaped genomic organisation in the *Haloarchaea* class of archaea. What existing studies have failed to resolve is the reasoning behind the 'hidden cost' of the multipartite genome– that is, increased complexity. What are the genetic events that led to the expansion into multiple replicons, and do they confer any advantage to the cell?

Taking the conjecture that the modern eukaryotic cell evolved from a lineage of archaea containing multiple origins – the study of archaeal replication origins can therefore provide an understanding of the complex mechanisms in eukaryotes, and potentially give insight into some of the selection pressures present at the primordial times of the LUCA. For this task, the ideal model would be an archaeon that is easy to culture within laboratory conditions, and one which would be amenable to genetic manipulation (*see* chapter **4**).

Therefore the next two chapters aim to familiarise with the events starting from origin-recognition, leading up in stages to full replisome assembly – with a special focus on the archaeal domain – before continuing into some exceptional cases of replication (e.g., recombination-dependent replication; or RDR), and their implications.

1.3 Where Do We Start? DNA Replication Initiation across the Three Domains of Life

The initiatory steps leading up to replisome formation can be broadly classified into four distinct stages: (I) origin recognition, (II) pre-RC assembly, (III) replicative helicase activation and DNA unwinding, and (IV) loading of replicative DNA polymerases along with other enzymes which support the replisome (V) to ensure high processivity. (*see* figure **5** and table **2**) Stages of replication has been separated for comparative analysis between the three domains of life. Each model organism therein was purposefully chosen to demonstrate the evolutionary transitions in genomic organisation.

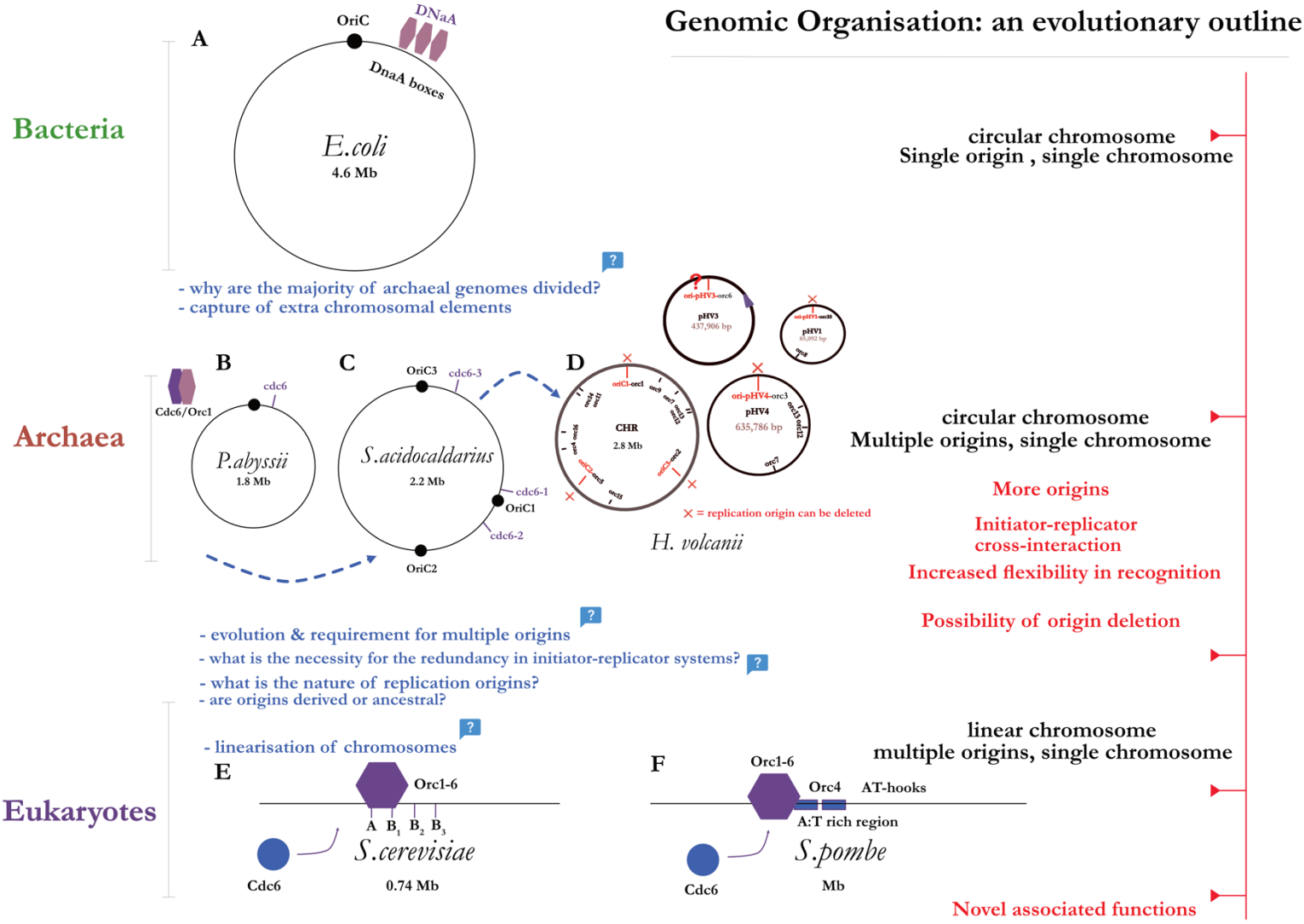


Figure 4. Evolutionary timeline of replicator diversification across the 3 domains.

(adapted from Schwob)

1.3.1 Bacteria

Before the DNA polymerase can associate and extend the DNA strand, the double helix must first be unwound. This requires the assembly of a higher-order nucleoprotein complex (i.e., pre-RC) which will then recruit the helicase. As previously discussed, the typical bacterial origin, OriC, appears once per chromosome for most bacteria, and its sequence is encoded adjacent to the initiator protein DNaA; hinting at a possible coordination between the levels of initiator and initiation rate (i.e., origin firing). The classic mechanism describes a single monomer of DnaA

binding to the consensus sequence – consequently named the DnaA box – to induce a 'bend' in the interaction site, and thus facilitate DNA melting. However as with most biological systems, the molecular reality is much more complex. Per single OriC, there have been a total of 12 characterised DnaA boxes to this date[70] – all with varying degrees of conservation to the original consensus. Each DnaA protein monomer binds to the respective DNA box; R1, R2, R4 (high affinity sites) or I, τ, and C (low affinity sites which lie in between of the R-sites)[35,71]. DnaA-ATP and DNaA-ADP can both bind to the sites, depending on the affinity, with the domain III containing ATP-activating and DNA-binding functions[72]. This brings us to the hallmark feature of the protein: its multimodular structure is what confers it with multifunctionality and the ability to coordinate entire replisome assembly. DnaA belongs to the AAA+ superfamily of ATPases (that is, ATPases associated with various cellular activities), and thus has an evolutionary relationship with the eukaryotic (Orc1) and archaeal (Orc1/Cdc6) initiator proteins[73] which bear structural similarities[74].

The collective binding induces topological stress (i.e., superhelical torsion) on the dsDNA, which in turn unwinds the adjacent AT-rich region – termed the DUE (DNA unwinding element)[75]. This creates a stable open complex structure or 'bubble' (i.e., the pre-RC) to which the helicase loading protein DnaC, through its interaction with domain I of DnaA, binds two hexamers of DnaB helicase and loads them unto the ssDNA region, at opposite orientations. The helicase then recruits DnaG primase which itself binds to the DnaB-DnaC complex thus leading to ATP-ADP hydrolysis stimulating helicase activation. The DnaB pair of helicases unwind with the directionality of 5' to 3' – hence the helicases work in opposing directions, and establish a bidirectional replication fork to which replication machinery can be loaded (for the latest overview of bacterial initiation, *see*[76]).

1.3.2 Eukaryotes

Contrary to their bacterial counterpart, progress to fully characterise the eukaryotic origins and initiation process was lagging. This points to an obvious difference: the size of the genome. Take a simple model eukaryote – *S. cerevisiae* or budding yeast – and compare it to the bacterial model of *E.coli*: the genomes are 12.2 Mb v. 4.6 Mb, respectively. Hence the first point of contrast in Eukaryotic initiation is increased spatiotemporal control (*see* figure 5**B)** to ensure accurate replication of a larger genome. There are multiple origins on a single chromosome – with increased flexibility of initiator interaction as origins become less defined, and have less sequence

conservation (with notable exception being *S. pombe*, which compared to *S.cerevisiae,* lacks distinct sequences, apart from rich AT regions[77]). Thus we witness another emerging trend: with the increasing number of origins, there is an overall decrease in their specificity (*see* figure 4). Schwob[78] posits an intriguing explanation: accumulation of recombination intermediates at replication origins in fission yeast drives genomic instability, which in turn may have promoted replicator diversification and redundancy as a counteractive mechanism[79]. In other words, replicator flexibility may have evolved to balance out the ill-effects of genomic instability – proposing an evolutionary relation between replication, recombination, and genomic stability in Eukaryotes. Once per mitotic cell cycle, the genome must be replicated with utmost precision due to the selective pressure of genomic instability and cell death as a result of over or under replication. This is reflected in the tight control mechanisms which couple the process of initiation to the stages of the cell cycle which are centred around preventing re-replication.

The major difficulty that came in characterising eukaryotic origins is that there are multiple origins on a single chromosome that lack discernible sequence motifs, and that the origins in higher eukaryotes are largely defined through complex chromatin interactions (i.e., a subset of origins, termed a 'cluster', can be activated according to the developmental phase[80]). This led to the development of the two-state model of initiation (refer to figure 5B) which corresponds to the levels of CDK activity[81]: the origins are 'licensed' and the pre-RC established during the G1 phase of low CDK and increased DDK (DBF4-dependent kinase or Cdc7) levels, and then subsequently activated during S-phase. Analogously to previously defined bacterial systems, the ORC – a six-subunit AAA+ ATPase – binds to the ARS sequence in an ATP-dependent manner. However unlike DnaA, ORC-ATP binding cannot directly unwind the DNA region[82]. Upon ORC binding, Cdc6[83] – a factor displaying sequence homology to the ORC subunit Orc1, suggesting common ancestry – is recruited to form a ring-shaped structure. Concomitantly, the Cdt1 initiator protein[84,85] acts as a chaperone to recruit the MCM2-7 helicase; together, this forms the intermediate ORC-Cdc6-Cdt1-MCM2-7 of the pre-RC, where the dsDNA can feed into the pore of the resulting MCM double hexamer[86]. Like DnaB, the MCM molecule must also be activated through ATP hydrolysis reaction – thus many MCM hexamers are loaded following ATP hydrolysis by Cdc6 and ORC, and Cdt1 release in an iterative fashion[87] (for an excellent review on MCM loading, *see*[88]). We hence arrive at another checkpoint control point: the activation of the MCM2-7 helicase (the 'core') depends on the additional proteins Cdc45 and GINS, and together they form the CMG complex which acts as a replicative helicase[89]. The transition from the G to S-phase of the cell cycle is guarded by the increase in the Cdc7 and

CDKs. Cdc7 directly phosphorylates the N-terminus of the MCM2-7 alongside a tripartite complex consisting of Sld2-Sid3-Dbp11 factors (SDS complex)[90], which mediates CMG formation and activates the helicase. The above stepwise model is what constitutes the 'origin firing' step; the duplex is unwound, and the replicative polymerase ε (pol-ε) alongside other replisome components are loaded.

1.3.3 Archaea

Archaeal chromosomes are circular and small akin to bacterial, yet they usually come with multiple replication origins on most chromosomes, and share homology with eukaryotic replication factors; the archaeal domain hence represents a unique fusion of bacterial and eukaryotic features. In fact, the first archaeal gene encoding an initiator protein – a distinct sequence located downstream of the replication origin (*OriC)*[59,60]–was first identified in the hyperthemophile *Pyrococcus* genome. Due to it sequence homology to regions of the eukaryotic Orc1 and Cdc6 – it was subsequently named Cdc/Orc1 (n.b., for clarity, the archaeal initiator will be referred to as simply 'Orc' for the rest of this review). This led to some authors to speculate that the eukaryotic Cdc6 and archaeal Orc1 have diversified from a gene duplication leading back to a common ancestor[91]. Interestingly, the same study found that the promoter region for the DNA polymerase subunit genes (i.e., DP1 and DP2) overlapped with the *Pyrococcus* OriC sequence, providing a first hint at the replication initiation control through transcription[92]. The same year, a mutational analysis and sequence alignment study proposed a structure of archaeal cdc6 ortholog, and its functional implications in pre-RC assembly[93]. The crystal structure of *Pyrococcus* cdc6 protein reveals its multidomain organisation; with domains I and II having an AAA+ ATPase module, and domain III being composed of a winged-helix (WH) fold. Soon after the initial discovery, the postulated mechanism of origin-binding was confirmed *in vitro*[94]. The purified Orc protein was shown to bind to the origin-recognition sequences termed the ORBs (a conserved 13-base repeat), or mini-ORB elements in *Sulfolobus* [62] flanking the AT-rich DUE element within the origin region[61].

The inverted position of the ORBs on either side of the DUE is what precisely determines the polarity of Orc binding. The Orc initiators bend the DNA through their N-terminal AAA+ domain; an extra layer of complexity is added through varying binding affinities between Orc proteins determined by its WH domain. Initially, *in vitro* studies[95] in P.*furiosus* led authors to prematurely conclude that Orc binds in an ATP-independent manner, with the resulting

structural distortion[96] of the binding site leading to the unwinding of duplex. Intuitively, one would presume that the binding mechanism is analogous to that of DnaA within the bacterial domain. And while the initiator-origin recognition motif interaction is conserved, duplex unwinding upon Orc binding, the helicase recruitment mechanisms, and higher order complex assembly remain a contested topic. This is partly due to the differing methods used to study initiator-origin binding mechanisms. Biochemical studies[35,95,97,98] support strand unwinding upon Orc binding, leading to higher order assembly, while early structural analyses present an obvious conflict. Some authors support DNA unwinding following strand distortion due to the topological stress induced by AAA+ domain binding[96], while others assert that the base-pairing is maintained even after strand distortion[96,99]. This discrepancy persists within other species of archaea: biochemical analysis in *M.thermoautotrophicus*[100] and *A.pernix*[100] support higher order complex assembly, while *Sulfolobus* appear to be in contradiction (reviewed in[101], p. 60). Nevertheless, it became apparent that Orc binding and the subsequent topological changes serve as an important step in initiation; yet again, we see that the archaeal initiator mirrors the eukaryotic ORC in its main role of helicase recruitment rather than the direct origin melting of DnaA.

Contrary to early *Pyrococcus* studies[95], Orc needs to be ATP-bound for its activation; however, *in vitro* studies in the same species have shown that the loading of the helicase itself occurs via an ATP-independent mechanism[102]. Soon after MCM2-7 emerged as a candidate for the eukaryotic helicase, a number of MCM homologues were identified in archaea, with each species containing at least one homologue (for review, *see* [103]). Although the biochemical properties of the archaeal MCM were known – that is, 3' to 5' DNA translocation capabilities, ssDNA and dsDNA binding, and ATPase activities[104] – the mechanism of MCM loading by Orc remained to be elucidated. Work from Bell lab – consistent with earlier ChiP studies[60,94] – has shown that the homohexameric open-ring MCM directly binds to the ATP-bound Orc protein *in vitro*[105,106]. Here, ATP binding and MCM release following ATP hydrolysis serves as a regulatory switch to confer MCM loading to a particular temporal window: a primitive version of spatiotemporal control observed in eukaryotes. Recent atomic force microscopy techniques provided further experimental verification that MCM from can interact with DNA in a variety of conformations under physiological conditions[107]. An important distinction from the eukaryotic MCM2-7 which is only active when part of the CMG complex, is that the archaeal MCM displays intrinsic helicase activity in some species[108]. In others, paradoxically, MCM requires the binding of cdc6 homologues to be activated[109].

The rest of the replisome is then loaded: GAN or GINS-associated nuclease (i.e., Cdc45 or RecJ), and GINS factors which modulate the helicase activity, as well as the PCNA, RFC, primase (PriSLX), RPA, and the polymerases (B/D)[110].

Taking the above evidence together, it becomes apparent that the distribution of functions of replication proteins is highly diverse among archaeal species, as studies reveal a complex interactome leading up to full replisome assembly – understanding of which still remains fragmentary.

Table 2. Overview of replication machinery found across the 3 domains of life.

Replisome Assembly Step	Eukaryotes	Archaea	Bacteria
STAGE I Origin Recognition	ORC (Orc 1,2,3,4,5,6)	Orc/Cdc6[a,b]	DnaA
STAGE II Pre-RC formation	Cdc6/Cdt1	Orc/Cdc6[a,b] WhiP[b]	DnaA
STAGE III DNA duplex melting	MCM (Mcm2,3,4,5,6,7) six subunits -heterohexamer	MCM[a,b] 6 subunits -homohexamer	(DnaA)n
STAGE IV Helicase Loading	cdc6 cdt1	Orc/Cdc6[a,b]	DnaC helicase loader [DnaI - helicase loader in gram+ bacteria] DnaB helicase [DnaC - helicase in gram+ bacteria]
STAGE V Polymerase - replisome assembly	CMG complex PCNA clamp RFC clamp loader PriSL primase	GINS[a,b] PCNA clamp[a,b] RFC clamp loader[a,b] PriSL[a,b]/PriX[b]primase	β clamp DnaG primase
	B-Family polymerases (ε , δ)	Polymerase B[a,b] Polymerase D[a]	C-Family Polymerase III
Recommended Literature	Costa & Diffley (2022)[86]	Greci & Bell (2020)[89] Pérez-Arnaiz et al (2020)[108] (focus on *H.volcanii*)	Wegrzyn & Konieczny (2024)[75]

[a] euryarchaeota
[b] crenarchaeota (n.b., old nomenclature, now referred to as thermoproteota)

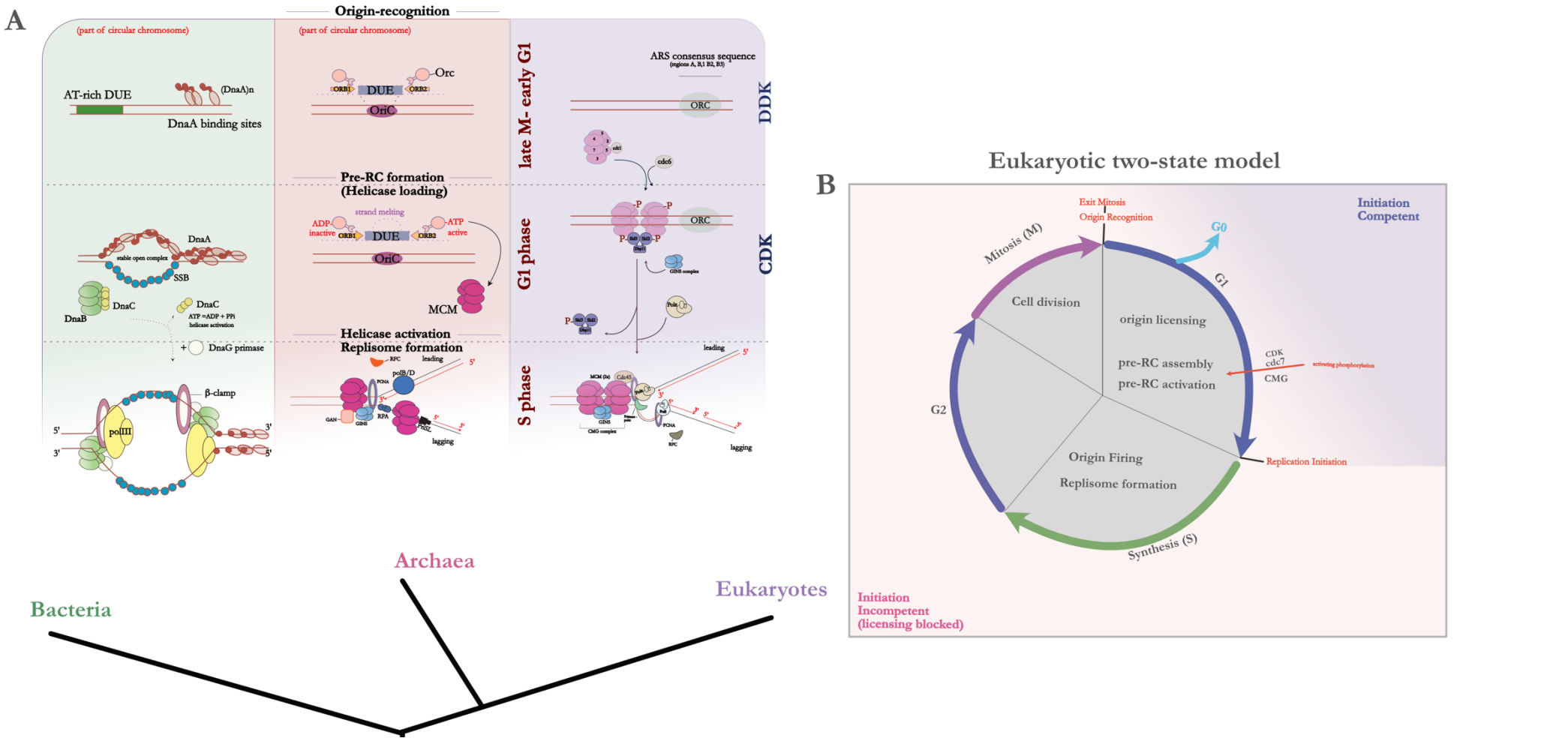


Figure 5.

(A) Replication initiation mechanism and associated replication factors – from origin recognition to full replisome assembly – across the 3 domains of life.
(B) A schematic diagram representing the temporal control of DNA replication stages in Eukaryotes. Origin licensing through phosphorylation by various CDKs serves as a major control point for the transition between the G1 to S stage of the cell cycle; hence the two-state model provides a temporal window in which origins are 'initiation competent'.

Chapter 2 DNA Replication & Recombination: A Dynamic Interplay

2.1 Adding a Level of Complexity: The Asymmetry of DNA Replication

When the first photographic micrograph of the *E.coli* replication[111] was presented – just before the 1963 Cold Spring Harbor Symposium – Monod raised a critical question: How is simultaneous bidirectional replication achieved, given that the DNA Polymerase I can only add nucleotides to the hydroxyl end of the strand (i.e., from 5' to 3')? The answer arrived five years later, confirming the asymmetric nature of DNA replication: one strand is replicated continuously (i.e., the leading stand – from 5' to 3'), while the lagging strand is replicated in the opposite direction, and in segments called Okazaki fragments; that is, discontinuously[112]. In bacteria, it was observed that there are more guanine nucleotides compared to cytosines within the leading strand; these strand-specific biases (a technique termed GC skew analysis) can thus be exploited to not only distinguish the leading from lagging strands, but also locate putative origins of replication and termination sites in archaea[113]. The lagging strand differs vastly by its enzymology: RNA primase is required to synthesise the 3' end primers, SSB protein to protect the exposed ssDNA, RNaseH to remove the RNA primases (DnaG in bacteria, PriSL in eukaryotes, and its homologue PriXLS in archaea), and finally; ligase to seal the synthesised fragments together. It becomes apparent that this universal requirement for a terminal 3'OH group for DNA polymerase-mediated extension is observed in all living forms across all three domains (as well as non-living viruses). Taken together, one could simplify strand extension to three fundamental requirements: (1) a terminal hydroxyl group provided by a primer or a recombination intermediate (2) DNA polymerase, and (3) interactions with additional factors to help load the replisome.

It is hence tempting to speculate that the LUCA relied on using a template-independent RNA polymerase (particularly due to its innate ability to bond nucleotides within its active site) due to the pressures of using RNA as a sole genetic material during the transition from the RNA world[114], with DNA polymerase being a later invention. Intriguingly, comparative genomics has revealed that the main components of the replisome do not share homology between bacteria

and archaea/eukaryotes, with a notable exception of sliding clamps; the primordial cell relied on a separate set of enzymes to replicate its RNA genome[115].

2.2 Recombination Dependent Replication

The viral origin hypothesis – enunciated by Forterre – states that horizontal gene transfer from mobile genetic elements and viruses has contributed to the evolution of the vast array of replication machinery in archaea and eukaryotes[116]. In fact, simple replicators – such as T4 and ΦX174 bacteriophages – possess a unique ability to bypass the primer requirement through rolling circle replication (i.e, mode of RDR), where a simple nick generated by RCR endonuclease is sufficient thus representing the simplest strategies of replication initiation[117,118]. It is worthwhile to investigate the mechanisms of the differing methods employed to overcome the primer requirement, as the 'clues' provided may enable to better characterise the ancestral features of replication initiation.

2.2.1 Clue № 1. Lessons from Viral Models

Viruses served as invaluable models of the replisome – for example, Alberts proposed the 'trombone' model to explain the coordination of the leading and lagging strands. Studies into the cell cycle of T4 bacteriophage were the first to propose a connecting link between replication and recombination, and initiated a research line into recombination processes, which were regarded as a rudimentary 'cut-and-paste' mechanism[119]. As early as in 1980, Mosig (for review of author's work and citations therein, *see* [120]) suggested that the replication of the bacteriophage occurs through homologous recombination. In the early stages, the replication initiated from fixed origins; however, at a later stage of the process the the 3 prime end of the lagging strand cannot be extended. This results in the recruitment of the DNA strand exchange protein called UvsX to the 3' ssDNA – thereby resulting in the formation of the D-loop through strand invasion. A D-loop can therefore be defined an intermediate structure that is formed during processes involving homologous recombination, whereby a single strand invades the dsDNA molecule in a strand exchange event. The homologous region is used as a template for the DNA synthesis and subsequent ligation events to repair a break in the DNA molecule. A similar mechanism is employed as part of the natural life cycle of bacteriophage T4. Although in the early stages of the cycle, mostly origins are used – some origins utilise the 3' ends of R-loops (i.e., a three-stranded nucleic acid structure, which involves a RNA-DNA hybrid from a transcript, displacing a DNA strand – commonly occurring due to aborted transcription) to directly prime

replication[120,121]. This tells us about the important distinction when it comes to priming replication through recombination intermediates: R loops possess an advantage over D-loops due to their ability to serve as direct primers. The 3'ssDNA ends are generated (either as a natural part of the replication process or through end processing via the 5 to 3' exonuclease activity of T4-encoded RNaseHs). The necessity of these DNA breaks for RDR initiation was confirmed using *in vivo* models of artificially created DBSs. Then, UvsX protein promoted strand exchange (n.b., UvsX have also been noted to be involved in branch migration and complementary DNA reannealing) to form the D-loop (*see* figure **6**). Several authors have questioned the necessity of this two-way mode of replication, as a similar mechanism has been utilised in bacteria. Is there any functional advantage, if *de novo* replication in T4 bacteriophage requires not only a D-loop formed, but also terminal redundancy supplemented by homologous sequences from a second copy of the genome. Syeda[122] reasons that even though RDR restrains genomic structure and ploidy – compared to canonical origin-dependent replication – RDR constitutes an *ad hoc* mechanism to overcome replicative blocks and ensure replication restart. This presents origins of replication as strict control points that have been favoured through evolution to replace a potentially dysregulated RDR initiation mode of replication.

The UvsX protein in bacteriophages also displays some sequence similarity to the bacterial recA belongs to the RecA/Rad51/RadA superfamily of recombinases, found within the bacterial, eukaryotic, and archaeal domains, respectively. Bacterial RecA, Rad51family members, and archaeal RadA are all homologous to each other. And although it is tempting to speculate that the viral recombinase follows the same pattern, due to some reports of weak homology of UvsX to RecA[119], structural analyses reveal that RecA has evolved through convergent evolution; UvsX and RecA/Rad51/RadA are orthologous[123].

The pressing problem in RDR initiation research line is the missing gap between the initial D-loop formation, and the molecular mechanisms leading up to full replisome assembly; in all 3 domains. However in origin-independent replication *E.coli*, DNA footprinting assays have revealed that PriA is able to not only recognise the D-loop structure, but can also recruit the φX174-like primasome, and lead to the formation of the replication fork[124,125]. PriA belongs to the 3'−5' DExH helicases of the Superfamily 2 class**.** It becomes apparent that interactions between the helicase and the recombination intermediate may serve as a potential clue to the full elucidation of the replisome assembly mechanism; however, research into the interactions that occur between the recombination intermediate, and the proteins which assist in the assembly of

the replisome has been lacking. This is partly due to the difficulty of deleting origins in eukaryotic models and sustaining viable originless cells. Archaea encode homologues to number of eukaryotic replication proteins, but in addition have a very flexible genome that allows for genetic manipulation, and a platform to investigate origin-independent mechanisms; implications of which can be extended to other life forms.

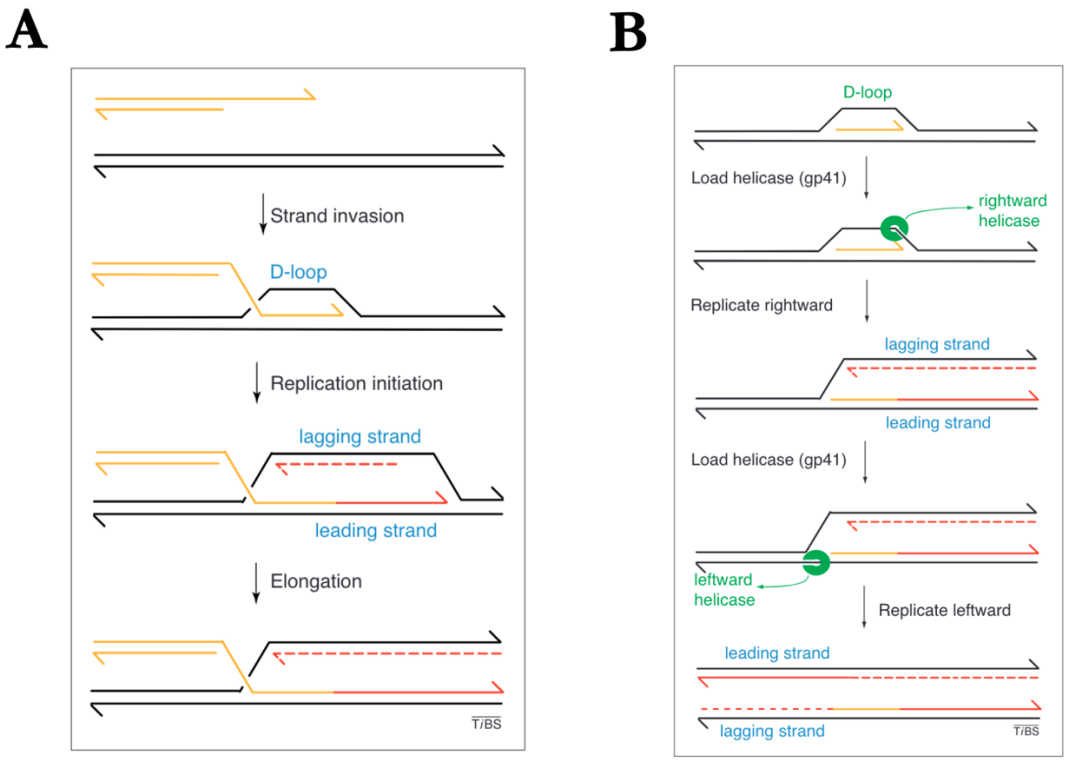


Figure 6. Schematic diagram outlining the steps in the RDR process that occurs during the T4 bacteriophage lifecycle. **(A)** Model of D-loop formation through the strand invasion mechanism, where the 3'end of the DNA strand from the previous replication cycle is used to prime and initiate the next round of replication. The (yellow) invading strand primes continuous replication in red on the leading strand, and the discontinuous line denotes lagging strand synthesis. The model was later revised to accentuate the functional involvement of the helicases (**B)** in RDR. (modified from Kreuzer[119])

2.2.2 Clue № 2. Break-induced DNA Replication in Eukaryotes

A form of RDR exists within the eukaryotic domain – and is termed BIR. The first evidence. came from studies in *s.cerevisiae* when observing the telomere maintenance mechanisms within cells which lack telomerase[126] Anand[127] emphasizes the lack of progress in understanding the conversion of D-loop structures into replisomes in BIR of budding yeast, as no homologues of the bacterial PriA have been discovered in eukaryotes. A potential lead is that a subunit of pol-δ – a PolB-like polymerase – has been shown to be essential in all BIR events. In higher eukaryotes, such as humans, HelQ helicase interacts with pol-δ to inhibit DNA synthesis, and in turn, promotes DNA repair pathways such as synthesis-dependent strand annealing[128]. One of the experimental methods employed was to induce artificial chromosomal DSBs using site-specific endonucleases, thereby stimulating strand invasion and initiation through BIR[129] The simple model involves a 3'end resection of the DBS, exposing a DNA strand which invades a homologous DNA molecule sequence to form a D-loop. The 3'end acts as a primer in BIR to initiate synthesis through a migrating bubble thus resulting in conservative inheritance. All pre-RC components of canonical origin-dependent replication were shown to act during BIR, in addition to recombination proteins such as Rad51, Rad52, Rad54, Rad55, and Rad57 which catalyse D loop formation[130]. As with other helicases, it is still unknown through which interactions MCM is recruited to the D-loop.

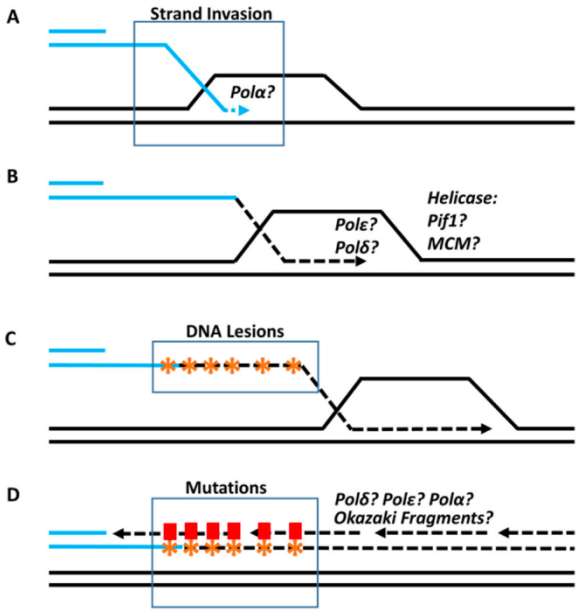


Figure 7.

A distinguishing feature of Rad51-dependent BIR, is that it occurs via a bubble migration mechanism. Polα is implicated in the formation of the D-loop, and the replication factors that have been speculated to be involved are indicated. Figure taken from [128]

2.2.3 Clue № 3 Origin-Independent Replication Initiation in Bacteria and Archaea

Kogoma and Lark[131] provided the first experimental evidence of an origin-independent replication process occurring in bacteria, expanding on their earlier paper which characterised *E.coli* replication which continued through several rounds despite thymine deficiency. Replication initiation through tightly controlled actions of DnaA and OriC is the preferred pathway due to it being highly efficient. However, the 'cost' of such mechanism is repeated protein synthesis of all the replication components with every cycle – an energy consuming process. In SOS-induced cells, and low-nutrient environments – an alternative pathway termed iSDR is therefore activated. Interestingly, *E.coli* with deletions for RNase HI (i.e, *rnhA*) have displayed another subcategory of SDR: that is, constitutive SDR or cSDR. As Δ*rnhA E.coli* were able to grow without DnaA and OriC, it was postulated that the increase in R-loop formation due to the deletion of RNase HI, can promote replication. This avoids the use of an initiator protein which is sequence-specific, as replication can initiate at different sites across the genome. In both cSDR and iSDR, RecA and PriA have been shown to be essential[132]. Historically, replication initiation from R loops has not received much traction, owing to the lack of an experimental way to track R-loop formations *in vivo*. With the advent of DRIP (i.e, S9.6 antibody which binds to DNA:RNA hybrids) , and DRIP-seq (high throughput sequencing) techniques, it became possible to characterise these structures. More importantly, the harmful biological implications of excessive R loop accumulation have been implicated in human diseases like cancer, which stimulated a revival in research of the correlation between R-loops and genomic instability[133].

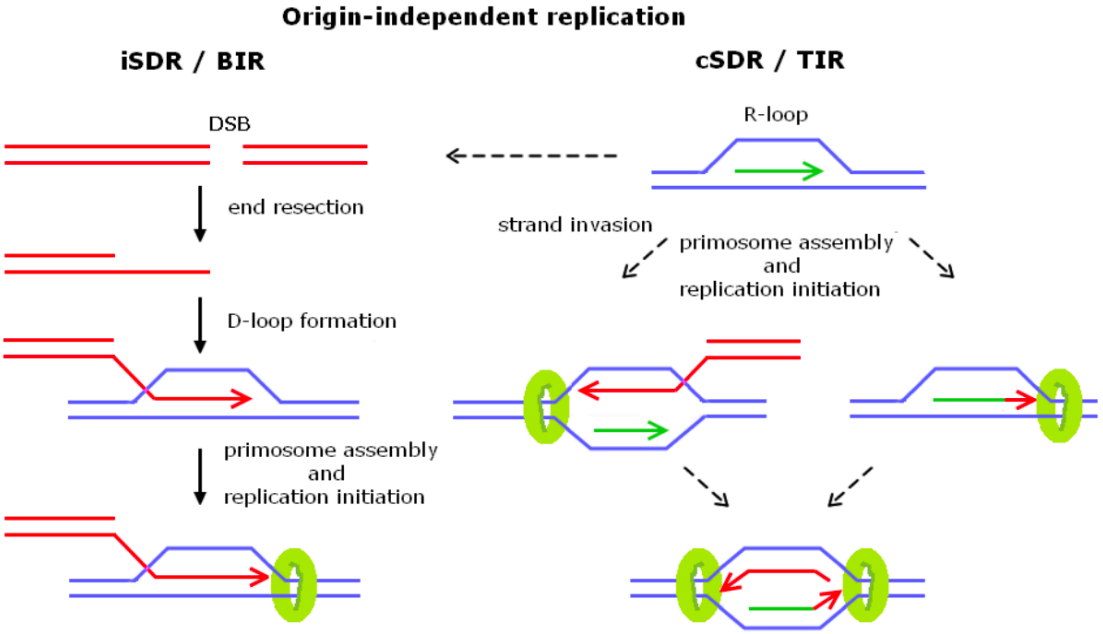


Figure 8. The two pathways of stable DNA replication in *E.coli* which occurs independent of DnaA and OriC. iSDR differes from cSDR in that it occurs through D-loop formation, as opposed to R-loops.

Figure taken from [130]

There is a balancing act between efficiency of DnaA-dependent replication and the energy saving advantage of SDR, which allows the bacterium to survive in adverse conditions, but occurs with low sequence specificity, and hence is inefficient for proper survival and growth in the normal environment [134] It was therefore believed that origin-independent replication was only needed for ensuring survival of the cell in harsh environments, at the expense of replication accuracy.

This paradigm was overturned when a paper in 2013 issue of *Nature* reported that *Haloferax volcanii* – a halophilic species found within the archaeal domain – is able to not only survive but also display a 7.5% faster growth phenotype having all of its origins deleted, compared to the wild-type strains[135]. There were several intriguing features. Firstly, replication profiles of genome copy numbers along the length of the chromosome revealed that this type of replication does not initiate from a fixed sequence; but rather, in a stochastic manner with initiation points dispersed all over the genome. Another observation was that when the RadA recombinase gene was put under a tryptophan inducible promoter to regulate its levels, originless cells displayed an absolute requirement for this protein. Basing on this body of evidence, alongside the previous known cases of similar type of replication mechanism in *E.coli*[136], the authors suggested that a RDR mechanism must be involved, where RadA catalyses D-loop formation. The next line of questioning involved the replication machinery that is assembled during RDR; with MCM being a major player in the recruitment to the D-loop structure. The indispensability of RadA in

archaeal RDR was only confirmed this year[137], where it was shown to fluctuate according to the growth stage.

This phenomenon does not extend to archaeal cells deleted for individual origins which display a growth disadvantage; given the known cases of sexual mating involving HGT in *H.volcanii*[138], the authors made a suggestion that origins behave akin to selfish genetic elements, which prioritise the maintenance of their own ploidy. This could explain the discrepancy between the deletions of individual origins which have no growth advantage, however the picture was only beginning to emerge. This discovery stimulated the birth of a new subfield – the study of the necessity and the nature of replication origins within the archaeal domain. In the more phylogenetically distant thermophilic archaeon – *Thermococcus kodakarensis* – the single origin can also be deleted and have no deleterious consequences on the phenotype[139]. Similarly, results from MFA technique were consistent with the hypothesis of dispersed sites of replication initiation during RDR. The picture becomes less clear when in a closely related species to *H.volcanii* – *H.mediterranei* – genuine origin deletion cannot be achieved as a dormant origin becomes activated[140]. What is then so special about the replication origins in *H.volcanii*? Finding an answer to this question may reveal new insights on the fundamental characteristics of replication origins that were previously unknown, due to being 'concealed' during normal replication processes.

Chapter 3 Archaea - The Third Domain of Life

"The transition from a paradigm in crisis to a new one from which a new tradition of normal science can emerge is far from a cumulative process, one achieved by an articulation or extension of the old paradigm. Rather it is a reconstruction of the field from new fundamentals, a reconstruction that changes some of the field's most elementary theoretical generalizations as well as many of its paradigm methods and applications."

- Thomas Kuhn, *Structure of Scientific Revolutions* (1962)[141]

3.1 Shifting the Paradigm

3.1.1 The Eukaryote-Prokaryote Dichotomy : a model in crisis

In the 1960s, it became increasingly evident that the bacteriological classification using classical taxonomical methods was coming to a phylogenetic impasse. Thus the ready acceptance of the term 'prokaryote' served as a comfortable – albeit simplified – way to eliminate all confusion as to what constitutes a bacterium; the cells were either nucleated (i.e., a eukaryote) or they were not (i.e., a prokaryote)[142]. Stanier and Niel[142] expanded on this concept to provide a structural basis for the distinction of bacterial cells from other life forms in that (1) bacteria lack internal compartmentalisation through inner membranes, (2) have distinct cellular apparatus, (3) divide through binary fission, and (4) possess a peptidoglycan cell wall. Hence this dichotomy which dominated the field throughout most of the 20th century, was not originally implied to reflect phylogenetic truth; but rather, the deepest structural discontinuity found within the tree of life[143].

Brock's isolation of thermophilic microorganisms in Yellowstone springs (i.e., deep water vents with waters reaching the temperatures of 90 degrees Celsius)[144] drew attention to extremophiles and their unique properties. Extremophiles comprise a large group of microorganisms which have adapted to survive in harsh environmental conditions, such as halophiles in high salinity, acidophiles and alkaliphiles that can thrive in the extremes of the pH value (from below 0 to above 10.5)[145] Improvements in the culturing methods in the early-mid 1970s led to the isolation

of methanogens (i.e., anaerobic archaea capable of producing methane gas as a by-product)[146], halophiles (with optimal growth recorded at 20% of NaCl)[147], as well as thermoacidophiles[148]. Notice how the early papers refer to these microorganisms as 'archaeabacteria'; it was regarded as common knowledge that the tree of life consists of a bipartite division hence archaea were classified as a second group of bacteria, separated from the 'true' or eubacteria.

3.1.2 Carl Woese's Discovery

Woese believed that the studies of protein synthesis at the time lacked an evolutionary underpinning, which was the reason for their lack of progress. His background in biophysics endowed him with a unique perspective: in a letter to Crick, he expressed how he intended to study of the conservation of proteins, and their variation amongst different domains of life[149]. Woese saw the potential in the current technology of Sanger sequencing[150], and utilised it to sequence the small subunit of 16S rRNA which appeared to be evolved from a common ancestor. From that, Woese and his postdoc Fox concluded that bacteria and archaebacteria constitute separate domains on the tree of life[146] (*see* figure 11) – and have redrawn the evolutionary tree to show tripartite division between Eukarya, Bacteria, and Archaea (i.e., the -bacteria suffix has been removed to highlight archaea's evolutionary distinction). Woese was highly criticised for the reductionist approach of attempting to rewrite the entire tree of life using a single molecule. To his defence, Zillig proposed the homology[151–153] between the RNAP molecules within the 3 domains, thus strengthening the proposal, leading to the establishment a new tripartite tree model[57,154].

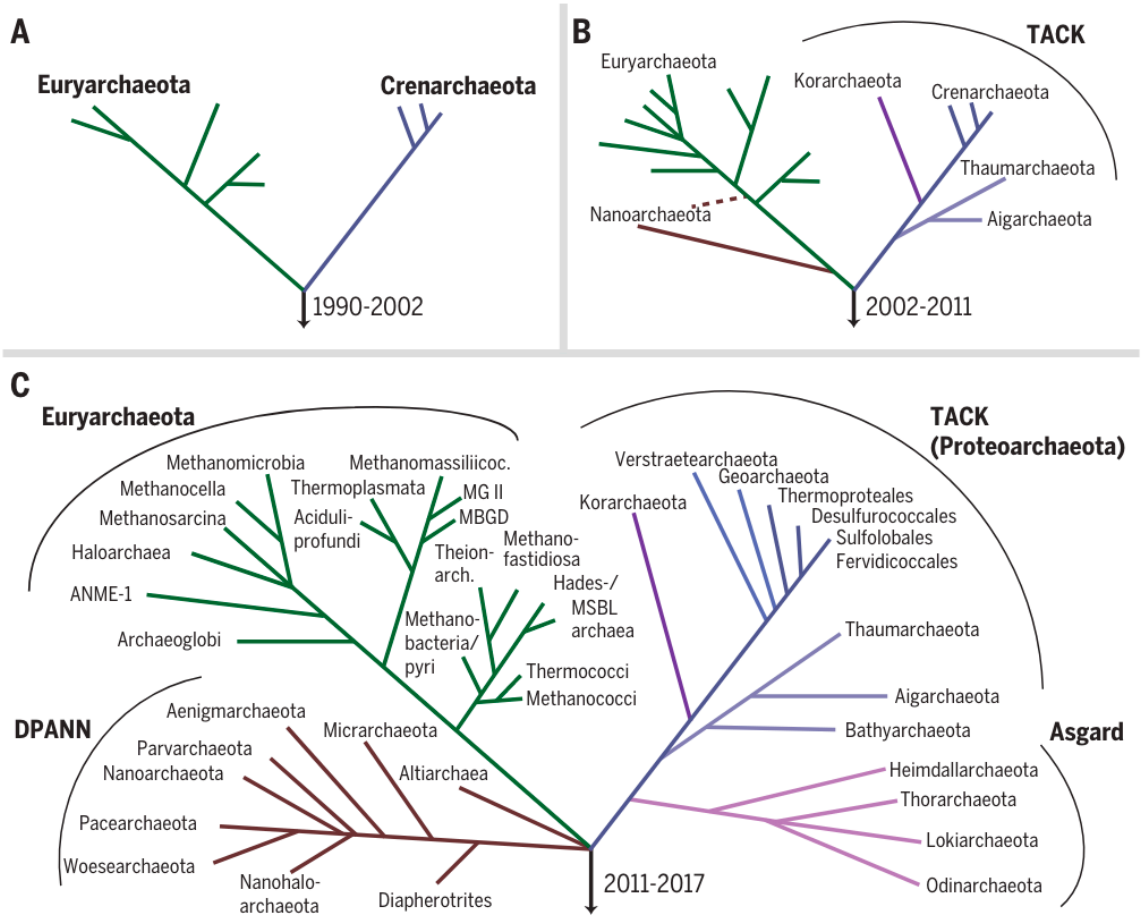


Figure 9. The changing view of the tree of life; (A) the archaeal domain was separated into 2 main superphyla Euryarchaeota and Crenarchaeota. Since then, many new superphyla were added (B,C) due to improved culturing methods and phylogenetic analyses.

Figure taken from [155]

Table 3. Comparison of the major cellular features between the three domains. (Taken from [156])

Trait	Bacteria	Archaea	Eukarya
Carbon linkage of lipids	Ester	Ether	Ester
Phosphate backbone of lipids	Glycerol-3-phosphate	Glycerol-1-phosphate	Glycerol-3-phosphate
Metabolism	Bacterial	Bacterial-like	Eukaryotic
Core transcription apparatus	Bacterial	Eukaryotic-like	Eukaryotic
Translation elongation factors	Bacterial	Eukaryotic-like	Eukaryotic
Nucleus	No	No	Yes
Organelles	No	No	Yes
Methanogenesis	No	Yes	No
Pathogens	Yes	No	Yes

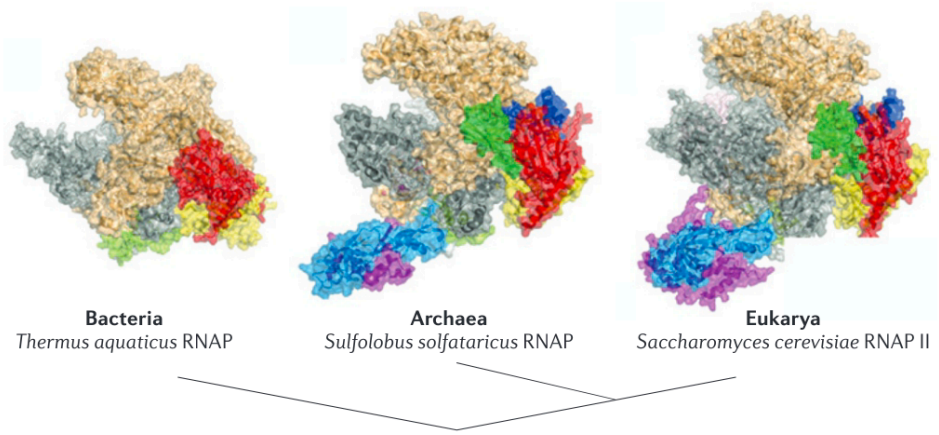


Figure 10.The homologous subunits shown in the RNAP structures, suggesting shared ancestry.

Taken from [151]

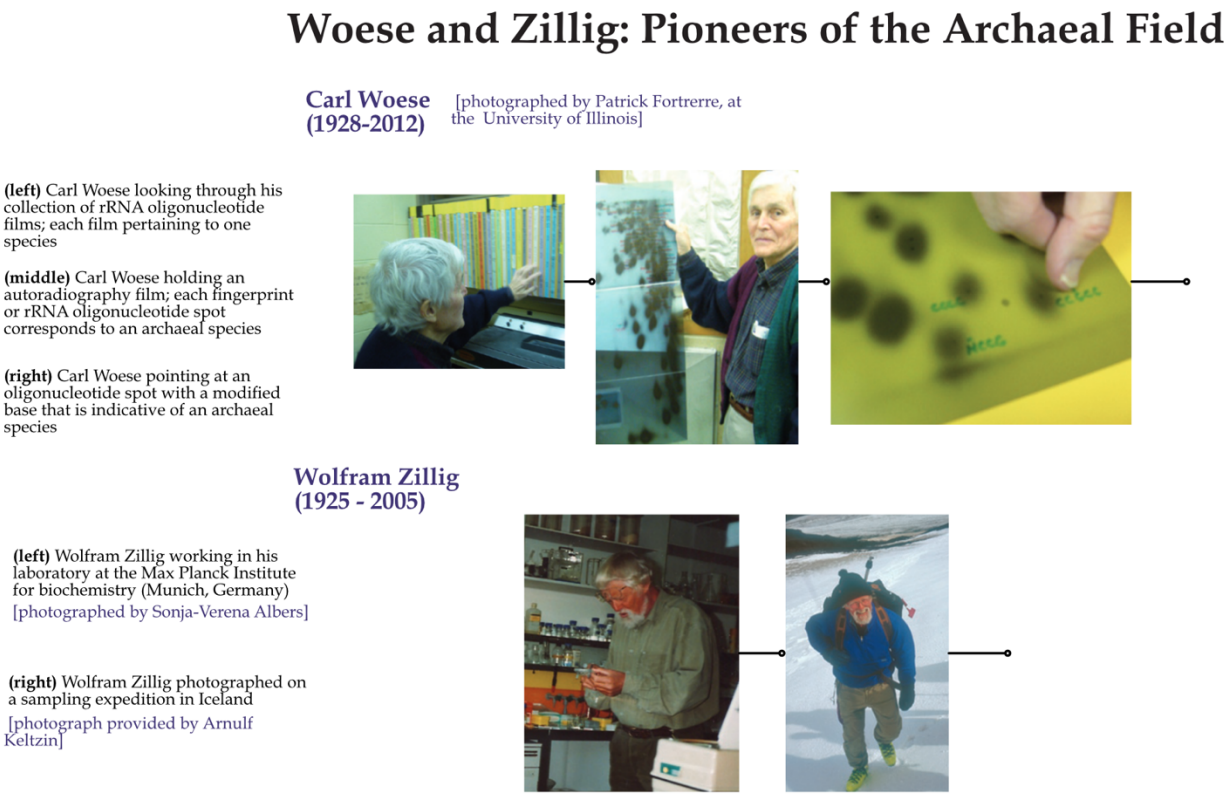


Figure 11. A historical perspective on the development of new approaches to study archaea by the two leading scientists of the field.

(modified from Albers *et al.*, 2013)

Impact & General Project Aims

3.2 BBSRC's Research Theme: Understanding the Rules of Life

The implications of our work fit within the BBSRC research framework which prioritises the study of the fundamental processes, such as replication and recombination, within a variety of biological systems. Elucidating the core principles of biological processes can enable to solve problems which fall within biotechnology, healthcare, and ecological sectors.

3.3 On the Impact of Research Methodology

According to Woese[157], the above discussion encapsulates the 'two great problems' that have paved the way towards our current epistemological position : (1) the 'rediscovery' of the gene, and the constituents of a cell (i.e., molecular reductionism; chapters **1-2**) –and the (2) more complex, integrative method of questioning regarding evolution, and the structural organisation of a biological entity (chapter **3)**

The harsh reductionist approach, which dominated the first half of 20th century biology, supported the belief that complex biological processes can be explained solely through the analysis of their molecular components. However, it became evident that the same methodological approach cannot be extended to the emergence of such processes[158]. Thus the explosive advancement in basic and applied molecular biology alongside the advent of high-throughput genomic sequencing techniques (i.e., with the rise of '-omics' technology during the genomic revolution), has brought along a Kuhnian shift to the adaptation of more integrative approaches such as systems biology[159–161] Ultimately, the recognition of Archaea as the third form of life has provided a novel holistic framework for subsequent studies in molecular genetics[162].

In this project, we have chosen to undertake a combination of methods. Molecular reductionism is essential to understand how an isolated component functions within a biological system (e.g., deletion of RNase H genes leading to an increase in the frequency of R-loops, thereby promoting recombination dependent replication, and enabling the deletion the replication origin). However, it is the use of genomic techniques, such as next-generation sequencing[163],

which will enable us to fully characterise the dynamic interactions between these components and the system as a whole (e.g., use of DRIP-seq to map the genome-wide distribution of R-loops, and the correlation between the frequency of R-loops and genomic rearrangements). The complementary use of comparative genomics will therefore provide novel insights into the interplay between DNA replication, transcription, and genomic stability.

3.4 Overall Project Objectives: An Infographic

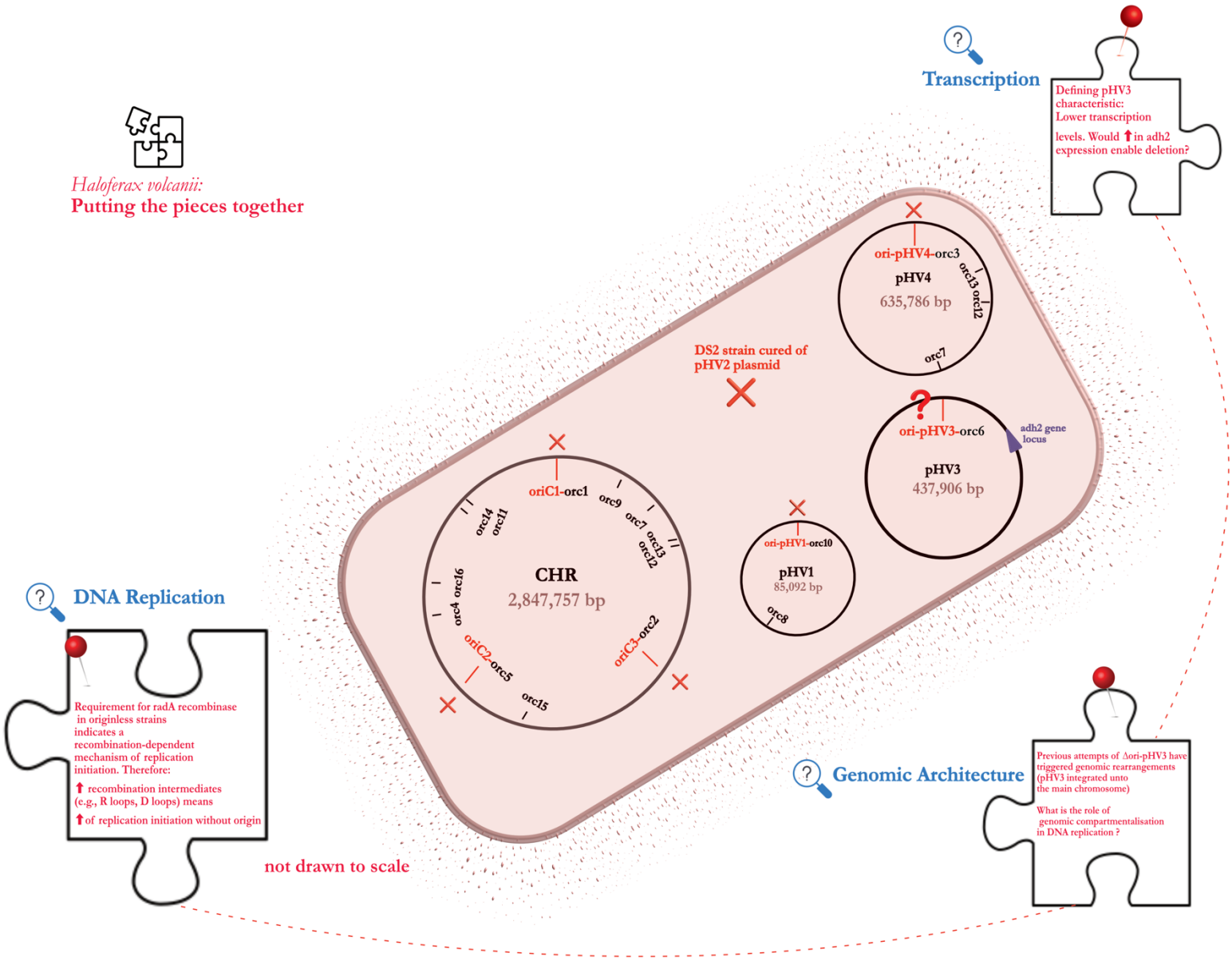


Figure 12.
A summary diagram of the methodological framework used – delineating the project into three parallel strands or approaches. Crosses denote possibility of generating origin-deleted mutant on each chromosome. The native origin is marked in red, with the origin-associated and non-native orc initiator proteins localised on the chromosome in black.

Haloferax volcanii belongs to a group of extremophiles adapted to high salt environments (i.e., halophiles) within the third domain of life – Archaea. They are relatively easy to cultivate, non-pathogenic to humans, and their wide array of previously characterised genetic tools and in-house mutant strains establishes an ideal genetic system to investigate the rules of life.

The genome of *H.volcanii* is composed of a single, circular chromosome (2.9Mb in size), as well as 3 smaller mini-chromosomes (i.e., pHV1, pHV3, pHV4), and a plasmid (pHV2). So far, attempts to delete the origin on

the mini-chromosome pHV3 (provided that pHV3 remains separate from the main chromosome) have not been successful. This poses several questions: (1) What are exact genetic features of pHV3 which make it distinct from the other mini-chromosomes, and (2) would the regulation of such features on pHV3 result in successful origin deletion?

Transcriptomic analysis performed by Trieselmann and Charlebois in 1992[164] has provided initial evidence of low transcriptional levels on pHV3. This has also been confirmed through recent RNA-seq analysis (personal commun., Darya Ausannikava, University of Nottingham).

Therefore, 3 parallel research avenues (*see* figure 1) will be investigated throughout the duration of this project using a range of genetic and biochemical techniques:

(1) Generation of a mutant strain where the *adh2* gene is under the control of a **strong, inducible promoter** (i.e., tryptophan inducible). This would enable us to **increase the transcription levels on pHV3** in a controlled manner.

(2) Increase the probability of **D-loop** formation and stability, and in turn, the probability of recombination-dependent replication occurring through the generation of Hel308 (a native helicase which degrades D-loops) knock-out strains. Simultaneously, strains with single amino acid substitutions within the Hel308 binding domain would achieve the same desired result of D-loop stabilisation.

(3) In a similar fashion to (2), increase the probability of **R-loop** formation and stability, through the generation of strains deleted for the *rnh* genes which encode a family of R-loop degrading ribonucleases (i.e., RNAase H)

Through the combination of the above methods, we are aiming to delete the origin of replication on the pHV3 mini-chromosome – confirmed through colony hybridisation, and pulsed-gel electrophoresis (*see* figure 2). Once origin-deleted strains have been successfully established, and characterised (flow cytometry and growth assays) the longer-term aims of this project would involve sequencing and bioinformatic analysis (CHIP-seq, DRIP-seq) to investigate the correlation between the frequency of R-loops, as well as provide a basis for further investigation of the relationship between genomic rearrangements (i.e., frequency of R loops) and genetic architecture of *H.volcanii* through Hi-C (chromosome conformation capture) techniques.

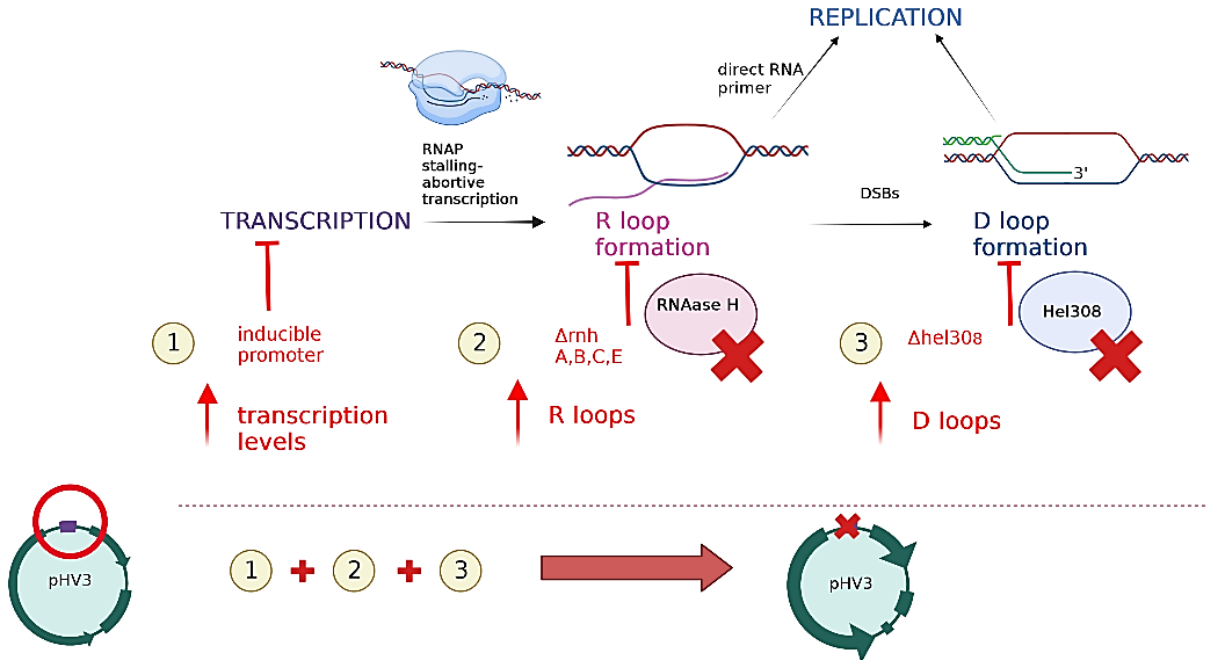


Figure 13: Schematic representation of the three parallel lines of investigation in this project.

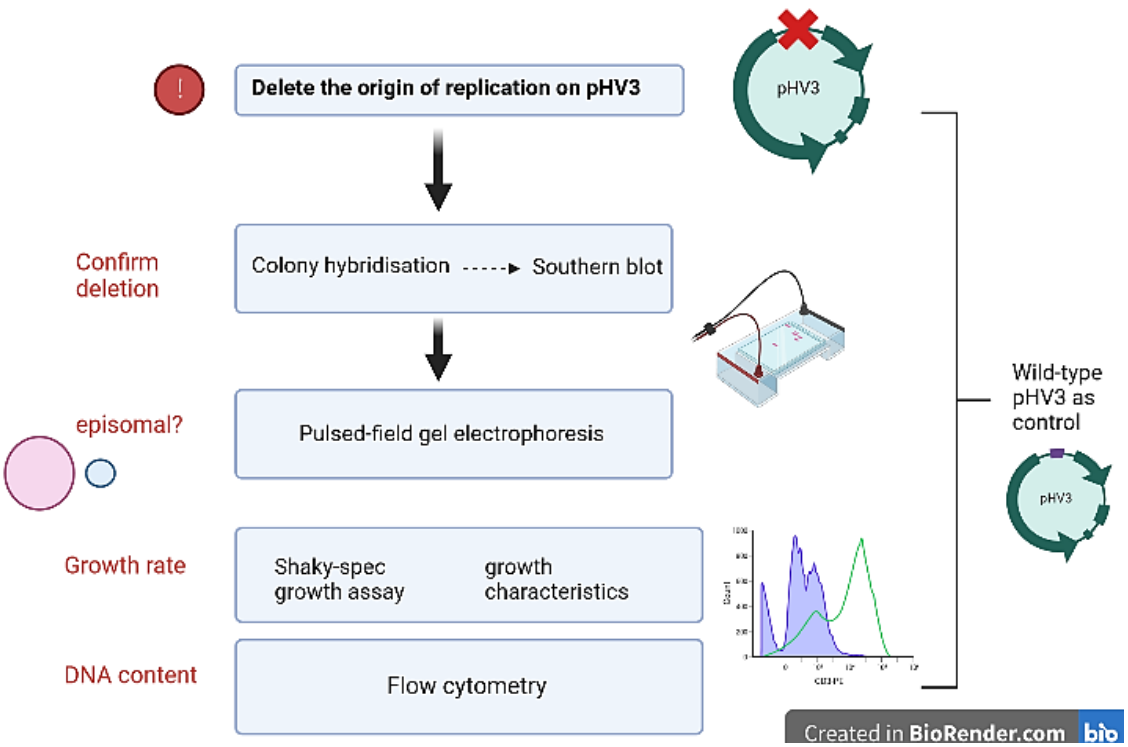


Figure 14. Schematic flow chart outlining the main experimental techniques following the confirmation of pHV3 origin deletion

Chapter 4 Materials and Methods

4.1 Materials

4.1.1 *Escherichia coli* strains

Table 4. *Escherichia coli* strains used in this study.

Strain	Genotype	Description	Description
XL1Blue MRF'	*endA1, gyrA96 (Na1R), lac [F' proAB lacIqZ⊿M15 Tn10 (TetR)], ⊿(mcrA)183, ⊿(mcrCBhsdSMR-mrr)173, recA1, relA1, supE44, thi-1*	Vector allowing for standard cloning procedures, and Standard cloning strain enabling blue/white selection in conjunction with pBluescriptII SK+ plasmid derivatives. Tetracycline resistant. Restriction endonuclease and recombination deficient, dam+. From Stratagene.	Vector used for standard cloning procedures with pBluescriptII SK+ plasmid derivatives, followed by blue-white selection methods using x-gal. Dam+, tetracycline resistant, recombination and restriction endonuclease resistant. Obtained from StrateGene® .

N2338 (GM121)	*F–, ara-14, dam-3, dcm-6, fhuA31, galK2, galT22, hsdR3, lacY1, leu-6, thi-1, thr-1, tsx-78*	dam- mutant dcm- mutant	Used for unmethylated DNA preparation in Haloferax volcanii transformations (dam- requirement). (Allers *et al.*, 2004[165])

4.1.2 *Haloferax volcanii* strains

Table 1.1. Haloferax volcanii laboratory strain numbers and their relative properties outlined in this study.

Starting Strain[a]	Pop-in Genotype	Pop-in Phenotype	Daughter Strain	Description	Source
Strains used for the deletion of the *rnhA* gene (Chapter 5.3)					
H5598 (ΔrnhB, ΔrnhB, ΔrnhE)	Δ*pyrE2* Δ*trpA* Δ*rnhB::trpA+* Δ*rnhE* *rnhA+::[ΔrnhA pyrE2+]*	-	**H5617** **H5618**	**Pop-out** of p2831 to give ΔrnhA. Construct allows TrpA+ selection for **ΔrnhB** and Ura-	This study (Chapter 5)
H5599 (ΔrnhB, ΔrnhB, ΔrnhE)	Δ*pyrE2* Δ*trpA* Δ*rnhB::trpA+* Δ*rnhE* *rnhA+::[ΔrnhA pyrE2+]*	-	Construct for the deletion of **ΔrnhB, ΔrnhB,** and **ΔrnhE** Could not delete *rnhA* – checked by colony hybridisation		
H5601	Δ*pyrE2* Δ*trpA* Δ*oriC1, ΔoriC2, ΔoriC3* Δ*ori-pHV4-2*	-	**H5619** **H5620**		

	Δ*rnhB::trpA+* Δ*rnhE* *rnhA+::[ΔrnhA pyrE2+]*			**Pop-out** of p2831 to give Δ*rnhA*. Can select for uracil auxotrophs (*pyrE2* counterselection) on Hv-Cas plates	
H5602	Δ*pyrE2* Δ*trpA* Δ*rnhE* *rnhA+::[ΔrnhA pyrE2+]*	Trp –	**H5621** **H5622**		
H5603	Δ*pyrE2* Δ*trpA* Δ*rnhE* *rnhA+::[ΔrnhA pyrE2+]*	Trp –	**H5623** **H5624**	**Pop-out** of p2831 to give Δ*rnhA*. Can select for uracil auxotrophs (*pyrE2* counterselection) on Hv-Cas plates	

Strains used for the deletion of the *rnhC* gene (Chapter 5.4)

H5598 **(ΔrnhB,** **ΔrnhB,** **ΔrnhE)**	Δ*pyrE2* Δ*trpA* Δ*rnhB::trpA+* Δ*rnhE* *rnhA+::[ΔrnhA pyrE2+]*	-	**H5617** **H5618**	**Pop-out** of p2831 to give ΔrnhA. Construct allows TrpA+ selection for Δ**rnhB** and Ura-	This study (Chapter 5)
H5599 **(ΔrnhB,** **ΔrnhB,** **ΔrnhE)**	Δ*pyrE2* Δ*trpA* Δ*rnhB::trpA+* Δ*rnhE* *rnhA+::[ΔrnhA pyrE2+]*	-	Construct for the deletion of **ΔrnhB,** Δ**rnhB,** and Δ**rnhE** Could not delete *rnhA* – checked by colony hybridisation		
H5601	Δ*pyrE2* Δ*trpA* Δ*oriC1, ΔoriC2, ΔoriC3* Δ*ori-pHV4-2* Δ*rnhB::trpA+* Δ*rnhE* *rnhA+::[ΔrnhA pyrE2+]*	-	**H5619** **H5620**	**Pop-out** of p2831 to give Δ*rnhA*. Can select for uracil auxotrophs (*pyrE2* counterselection) on Hv-Cas plates	

H5602	Δ*pyrE2* Δ*trpA* Δ*rnhE* *rnhA+::[ΔrnhA pyrE2+]*	Trp –	**H5621** **H5622**	
H5603	Δ*pyrE2* Δ*trpA* Δ*rnhE* *rnhA+::[ΔrnhA pyrE2+]*	Trp –	**H5623** **H5624**	**Pop-out** of p2831 to give Δ*rnhA*. Can select for uracil auxotrophs (*pyrE2* counterselection) on Hv-Cas plates

a = parental strains (i.e., pop-ins) contructed Andy Cubbon (2022)

Inducible gene replacement *(p.tnaA::adh2)* and *adh2* deletion strains (Chapter 6)

H98 (allers *et al,* 2004)	Δ*pyrE2* Δ*hdrB*	Ura - Thy -	**H5631** **H5632**	**Pop-in** of p2860 *(Δadh2, pyrE2+)* at the *adh2* locus of H98 to give *adh2* deletion (2 clones). (confirmed by colony hybridisation: see Chapter 6)	**Parental strain (H98)** allers *et al,* 2004 Constructed by TA **Daughter strains:** This study (Chapter 6)
			H5636	**Pop-in** of p2827 *(p.tnaA::adh2+ pyrE2+)* at the *adh2* locus. (upstream pop-in confirmed by PCR: *see* Chapter 5.3)	

H98	Δ*pyrE2* Δ*hdrB*		**H5637**	**Pop-in** of p2827 *(p.tnaA::adh2+ pyrE2+)* at the *adh2* locus. (downstream pop-in confirmed by PCR: *see* Chapter 5.3)
H5631 **H5632**	Δ*pyrE2* Δ*hdrB* Δ*adh2*		**H5653** **H5654**	**Pop-out** of p2860 *(Δadh2, pyrE2+)* at the *adh2* locus, to give Δ*adh2*. Deletion confirmed by Colony Hybridisation

Hel308 knock-out and point mutation strains (Chapter 7)

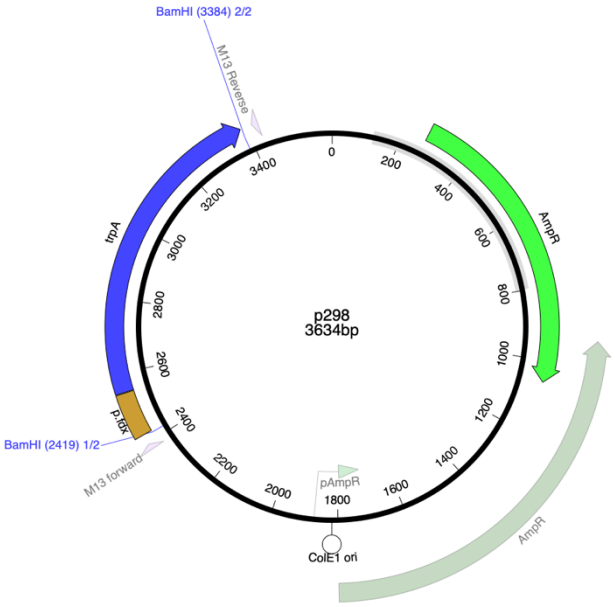
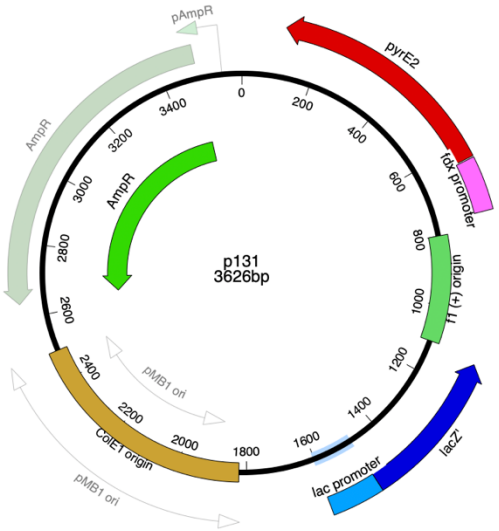
H4361 **(ΔHel308)**	Δ*pyrE2* *bgaHa-Bb* *leuB-Ag1* Δ*trpA* Δ*hel308*	Ura- Leu- Trp- X-gal red Growth: >>5 days (slow)	**H5638**	**Pop-in** of **p1801** *(ΔoripHV3::trpA+ pyrE2+)* to delete ori-pHV3	**Parental strains** (i.e., pop-ins) constructed by AC (2022) **Daughter strains** (i.e., pop-ins): this study. (Chapter 7)
			H5639	**Pop-in** of **p1802** *(Δorc6::trpA+ pyrE2+)* to delete orc6	
			H5640	**Pop-in** of **p1803** (Δori-pHV3 Δorc6::trpA+ pyrE2+) for simultaneous deletion of ori-pHV3 and orc6	

H2400 (D145N)	Δ*pyrE2* *bgaHa-Bb* *leuB-Ag1* Δ*trpA* *hel308-D145N*	Ura⁻ Leu⁻ Trp⁻ X-gal red Growth: 5 days	**H5641**	**Pop-in** of **p1801** *(ΔoripHV3::trpA+ pyrE2+)* to delete ori-pHV3
			H5642	**Pop-in** of **p1802** *(Δorc6::trpA+ pyrE2+)* to delete orc6
			H5643	**Pop-in** of **p1803** (Δori-pHV3 Δorc6::trpA+ pyrE2+) for simultaneous deletion of ori-pHV3 and orc6
H2397 (F316A)	Δ*pyrE2* *bgaHa-Bb* *leuB-Ag1* Δ*trpA* *Hel308-F316A*	Ura⁻ Leu⁻ Trp⁻ X-gal red Growth: 5 days	**H5644**	**Pop-in** of **p1801** *(ΔoripHV3::trpA+ pyrE2+)* to delete ori-pHV3
			H5645	**Pop-in** of **p1802** *(Δorc6::trpA+ pyrE2+)* to delete orc6
			H5646	**Pop-in** of **p1803** (Δori-pHV3 Δorc6::trpA+ pyrE2+) for simultaneous deletion of ori-pHV3 and orc6
H2398 (R743A)	Δ*pyrE2* *bgaHa-Bb* *leuB-Ag1* Δ*trpA* *hel308-R743A*	Ura⁻ Leu⁻ Trp⁻ X-gal red Growth: 5 days	**H5647**	**Pop-in** of **p1801** *(ΔoripHV3::trpA+ pyrE2+)* to delete ori-pHV3

			H5648	**Pop-in** of **p1802** *(Δorc6::trpA+ pyrE2+)* to delete orc6	
			H5649	**Pop-in** of **p1803** (Δori-pHV3 Δorc6::trpA+ pyrE2+) for simultaneous deletion of ori-pHV3 and orc6	
H164 (Wild type)	Δ*pyrE2* *bgaHa-Bb* *leuB-Ag1* Δ*trpA*	Ura⁻ Leu⁻ Trp⁻ X-gal red Growth: <5 days	**H5650**	**Pop-in** of **p1801** *(ΔoripHV3::trpA+ pyrE2+)* to delete ori-pHV3	
			H5651	**Pop-in** of **p1803** (Δori-pHV3 Δorc6::trpA+ pyrE2+) for simultaneous deletion ori-pHV3 and orc6	
			H5652	**Pop-in** of **p1803** (Δori-pHV3 Δorc6::trpA+ pyrE2+) for simultaneous deletion of ori-pHV3 and orc6	
Control strains for the deletion of Δori-pHV3, Δorc6, or Δori-pHV3 and Δorc6 – without Hel308 mutation background					
H4088	Δ*pyrE2* *leuB+::[pHV3]* Δ*trpA* Δ*oripHV3::trpA+*	Ura⁻	**-**	Pop-out of p1801 to give Δori-pHV3::trpA+ [confirmed through colony hybridisation and southern blot] **Pulsed-field gel:** pHV3 integrated onto the main chromosome	Hannah Marriott

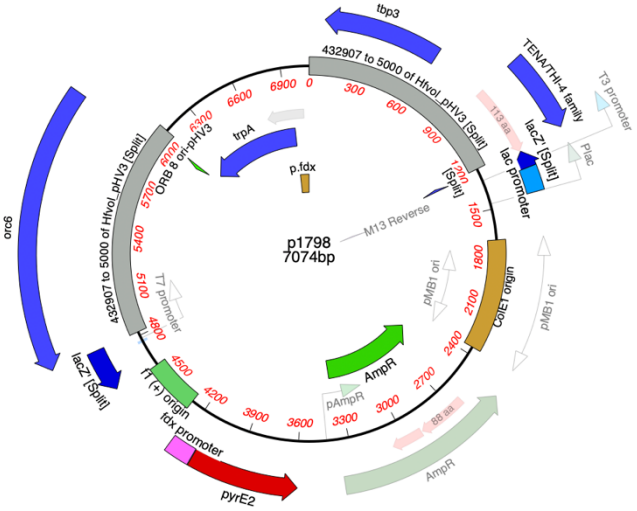
H4089	Δ*pyrE2* *leu*B+::*[pHV3]* Δ*trpA* Δ*oripHV3::trpA*+ Δ*orc6::trpA*+	Ura-	-	Pop out of p1802 to give Δorc6::trpA+ [confirmed through colony hybridisation and southern blot] **Pulsed-field gel:** pHV3 integrated onto the main chromosome	
H4090		Ura-		Pop out of p1803 to leave Δori-pHV3 Δorc6::trpA+ [confirmed through colony hybridisation and southern blot] **Pulsed-field gel:** pHV3 integrated onto the main chromosome	

4.1.3 Plasmids

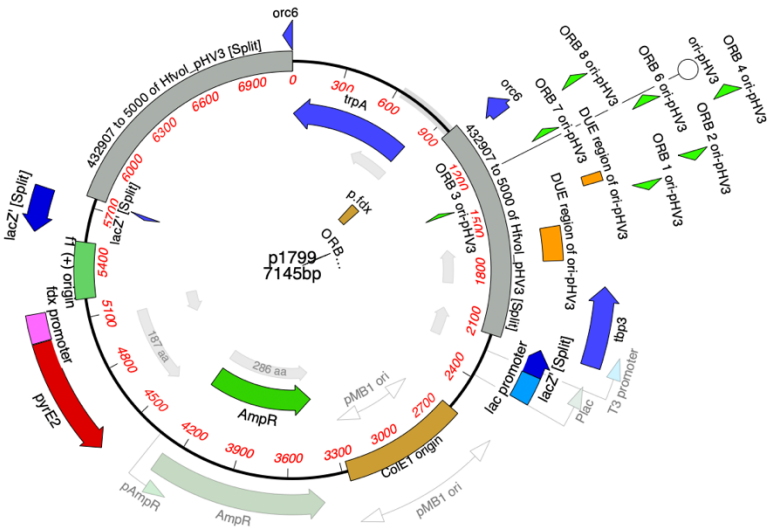
Strain	Construct	Use	Source	Plasmid Map
		Starter/vector plasmids for deletion constructs		
p298	Hf volcanii trpA ORF (PciI-SphI PCR fragment amplified from p49) in pIL11 between NcoI and SphI sites. (i.e., under ferredoxin (fdx) promoter of H. salinarium). XL1-blue host.	Used for making deletions in ΔtrpA background. To replace the use of p106, which has trpA flanked by one XbaI site, unlike p131 which has trpA flanked by two BamHI sites.	Lestini et al (2010)[166] Constructed by TA (2004)	
p131	pBluescript II SK+ with 0.7 kb BamHI/XbaI (both blunt-ended) p.fdx::pyrE2 fragment from pGB70 inserted at PsiI site. XL1-blue host.	Used for making deletions in ΔpyrE2 background. Derivative of pBluescript II, allowing for additional blue/white screening.	Allers (2004)[165] Constructed by TA	

For the deletion of Δori-pHV3, orc6 & Δori-pHV3/orc6 in Hel308 deletion and point mutation background strains

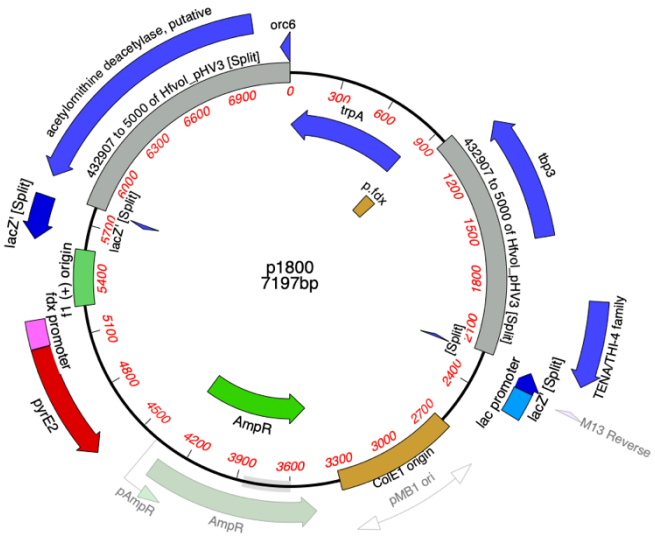
p1801	**Deletion construct: Δori-pHV3** p1520 with insertion of 965bp trpA+ BamHI fragment from p298 inserted at BglII site. Unmethylated; from Dam- strain. (dam+ version: p1798)	For the deletion of ori-pHV3, allowing for TrpA and uracil selection. TrpA marker linked to the origin deletion construct.	LM



p1802	**Deletion construct: Δorc6** p1521 with insertion of 965bp trpA+ BamHI fragment from p298 inserted at BglII site. (dam+ version: p1799)	For the deletion of *orc6* gene on the pHV3, allowing for TrpA and uracil selection. TrpA marker linked to the orc6 deletion construct.	LM

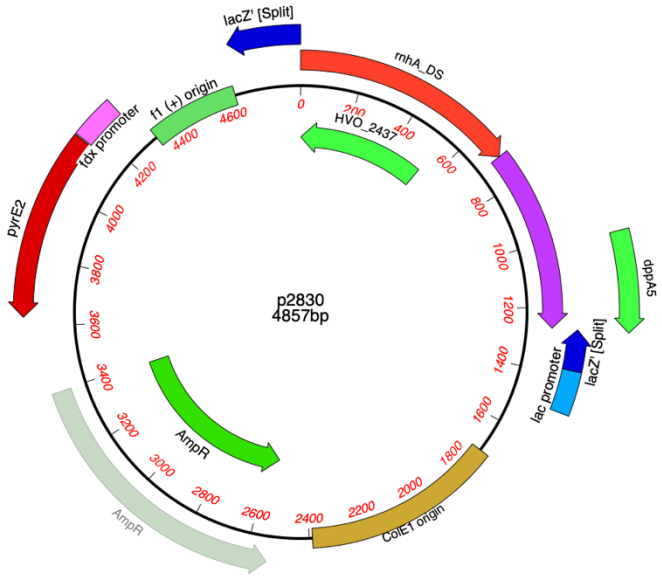


p1803	**Deletion construct: Δori-pHV3, Δorc6** p1522 with insertion of 965bp trpA+ BamHI fragment from p298 inserted at BglII site. (dam+ version: p1800)	For the simultaneous deletion of *orc6* gene, and the ori-pHV3, allowing for TrpA and uracil selection. TrpA marker linked to the origin-orc6 deletion construct.	LM



Generation of ΔrnhA deletion constructs

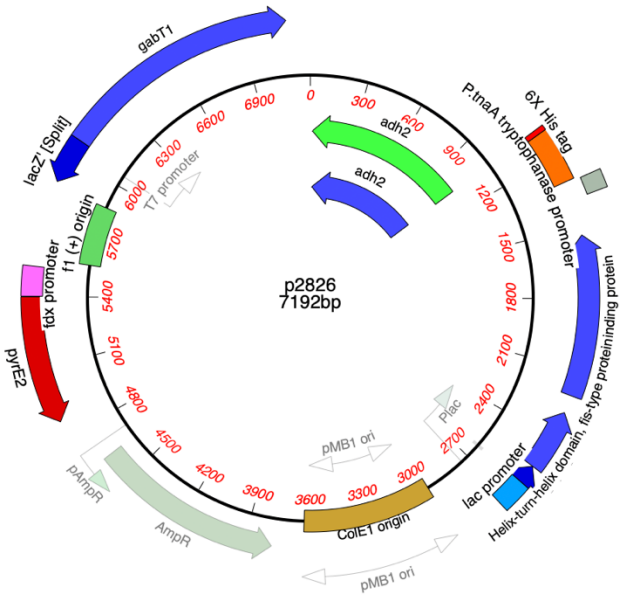
p2831	**Deletion construct: ΔrnhA** p131 with insertion of US and DS regions of rnhA between XhoI and EcoRI sites. Internal BamHI sites. Unmethylated DNA for Dam-	For the deletion of ΔrnhA	AC



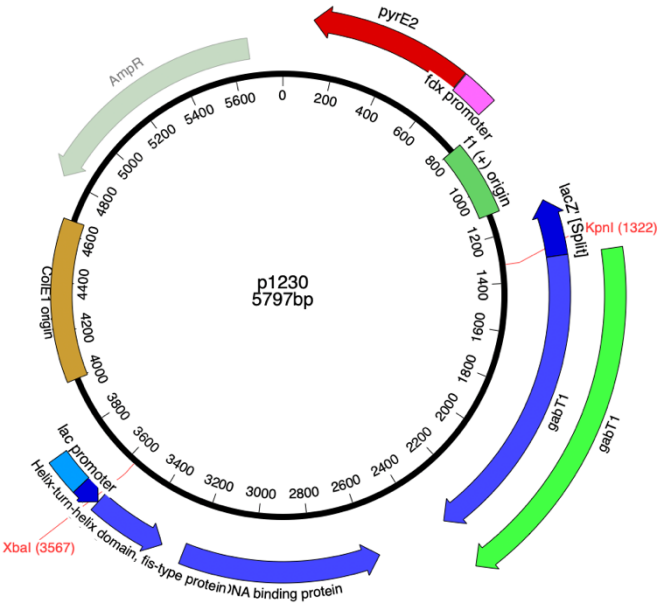
	strain transformations. (dam+ version: p2830)		

Generation of *adh2* under inducible tryptophanase promoter strains

p2827	**Inducible gene promoter construct p.tnaA::adh2** p1230 with insertion of 1394bp BamHI fragment from p2825, inserted at BamHI site in same orientation pyrE2. (dam+ version: p2826)	For the generation of strains containing *p.tnaA::adh2*	LM



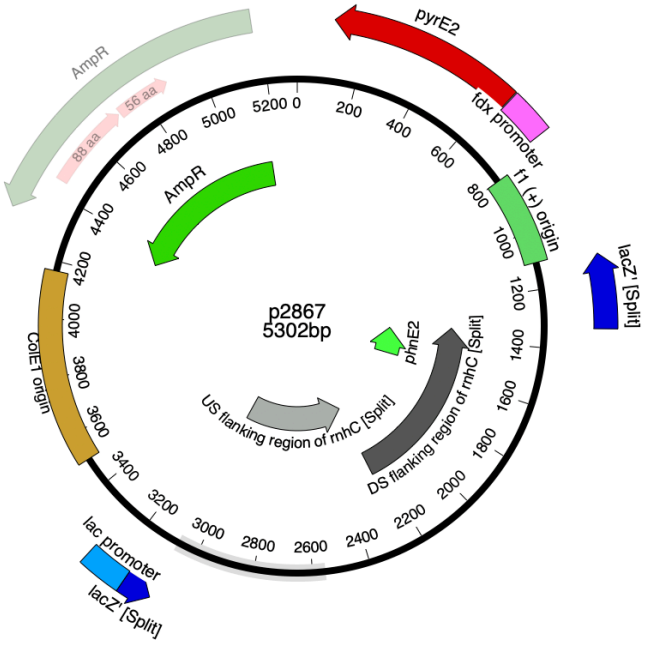
p2860	**Deletion construct: *Δadh2*** p131 with insertion of 2255 bp Δadh2 PCR construct, made from WT genomic DNA using internal primers with BamHI sites and external primers with	For the deletion of adh2 gene. Generated plasmids as contingency for colony hybridisation screening for the integration of p.tnaA::adh2 as an alternative to cPCR.	LM



	KpnI/XbaI sites, inserted at KpnI/XbaI site. (dam+ version: p1230)		

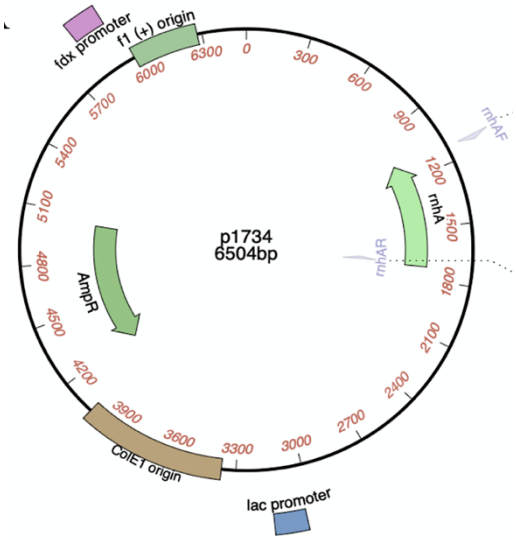
Generation of ΔrnhC deletion constructs

p2867	**Deletion construct: ΔrnhC** p131 with insertion of 1739 bp XbaI - XhoI fragment with rnhC deletion construct, amplified from p1736, 816bp US fragment and 923 bp DS fragment ligated at BamHI site, inserted at XhoI and XbaI sites.	For generation of ΔrnhC gene deletion construct strains.	This Study (Chapter 6)



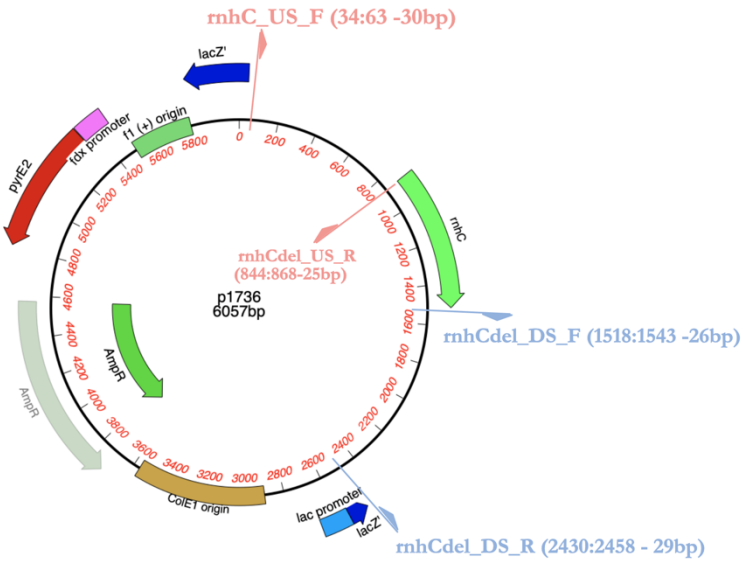
Genomic clones for colony hybridisation probe construction

p1734	**Genomic clone of *rnhA*** The fragment of 2891 bp was amplified from the genomic DNA, cut with XhoI and inserted into p131 vector.	*rnhA* probe construction to screen for ΔrnhA strains through colony hybridisation	Genomic Clone made by DA.



Genomic clones for generating gene deletion constructs through PCR amplification

p1736	**Genomic clone of *rnhC.*** Genomic clone containing rnhC gene: the fragment of 2462 bp was amplified from the genomic DNA, cut with XbaI and inserted into p131 vector.	.Construction of Δ*rnhC* strains through amplification and subsequent ligation of flanking regions of *rnhC* gene in p1736.	Genomic Clone made by DA.



TA = Thorsten Allers; LM = Laura Mitchell; AC = Andy Cubbon ,DA = Dasha Ausiannikava

4.1.4 Oligonucleotides

Name	Sequence (5' to 3')	Use
Oligonucleotides for gene deletion construction through PCR amplification of flanking regions		
rnhCdel_US_F (forward-sense)	CACTAGTTCTAGACGATAACCCCCGCGGCG	Amplification of **US** flanking region of p1736 *rnhC* genomic clone for the deletion of *rnhC*
rnhCdel_US_R (reverse-antisense)	GGTCACGGaTCCGGGGCCTCCGTCT	
rnhCdel_DS_F (forward-sense)	GGCGACGgatCCGCGAACTCCGCGGA	Amplification of **DS** flanking region of p1736 *rnhC* genomic clone for the deletion of *rnhC*
rnhCdel_DS_R (reverse-antisense)	GACGGTcTCGAGCATCTTCCCGACGATTC	
Oligonucleotides for screening the orientation of inducible gene construct (p.tnaA::adh2)		
Pair 1:		For diagnostic PCR (cPCR) of the p.tnaA::adh2 construct orientation : **"pop-ins"**
- dgabT1XhF (forward-sense)	GTCGAGTACGGAGCTCCCAGGCGCGTT	
- PtnaAFint (reverse-antisense)	GCCTGCCGATTACTTCACATTCGC	
Pair 2:		
p.tnaAintR (forward-sense)	GCGAATGTGAAGTAATCGGCAGGCG	
adh2USR (reverse-antisense)	TCGAACCCCACCAATACAGCCGACC	

(*PtnaAFint and p.tnaAintR have the ability to bind to the p.tnaA promoter region)		
adh2extR (forward-sense)	GGAACTGATTACCTCCTCGCTGGG	For diagnostic PCR (cPCR) of the p.tnaA::adh2 integration **"pop-outs"**
adh2intR (reverse-antisense)	CTCGTGACCGCGGTAGAACCAGAGG	

4.1.5 Media

Table 5. Growth media composition and selectable markers used in this study

***Haloferax volcanii* Growth Media**	
Sterilisation: Autoclave at 121˚C, for 15 minutes. **Storage:** Media stored in the dark, at room temperature. Agar plates stored in plastic bags (sealed to prevent desiccation), in the dark, and at room temperature. Before each use, the plates are dried upside down in an incubator for a minimum of 20 minutes. These conditions apply to all listed media/plates unless stated otherwise.	
18% Salt Water	30% SW; addition of 3mM CaCl2 after autoclaving, and once the solution is cool.
30% Salt Water	4 M NaCl, 148 mM MgCl2.6H2O, 122 mM MgSO4.7H2O, 94 mM KCl, 20 mM Tris.HCl at pH7.5.
Hv-YPC	**10 x YPC:** 5% yeast extract (Difco), 1% peptone (Oxoid), 1% casamino acids, 17.6 mM KOH. Not autoclaved, used immediately. **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 KPO4 Buffer (pH 7.0), 444 nM biotin, 2.5 µM thiamine. 18% SW, 1 x Ca, 30 mM Tris.HCl pH 7.0 autoclaved. Ensure that the solution has been cooled before the addition of other components. **Agar plates:** 1.6% agar (Bacto), 18% SW, 1 x YPC, 3mM CaCl2. Microwaved without 10 x YPC to dissolve agar. 10 x YPC added, then autoclaved. CaCl2 added prior to pouring, once cooled.
Hv-Ca	**Salts:** 362 mM CaCl2, 8.3%v/v trace elements (traditional), 615 µg/ml thiamine, 77 µg/ml biotin.

	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 KPO4 Buffer (pH 7.0), 444 nM biotin, 2.5 µM thiamine. 18% SW, 1 x Ca, 30 mM Tris.HCl pH 7.0 autoclaved. Ensure that the solution has been cooled before the addition of other components. **Agar plates:** 1.6% Agar (Bacto), 18% SW, 1 x Ca, 0.84% v/v of Hv-Ca Salts, 0.002%v/v of KPO4 Buffer (pH 7.0). Microwaved before addition of 10x Ca, Hv-Ca Salts, and KPO4 Buffer to dissolve agar. 10 x Ca added, autoclaved, then Hv-Ca Salts and KPO4 Buffer added prior to pouring.

***Haloferax volcanii* Media Supplements for Auxotrophic Selection**	
+Trp (Tryptophan)	(50 µg/ml final concentration). For ⊿*trpA* mutants.
+Thy (Thymidine)	(50 µg /ml final concentration) For ⊿*hdrB* mutants.
+5-FOA **(5-Fluoroorotic acid)**	For the selection of Δ*pyrE2*/ Ura- mutants 50 µg/ml (+ 10 µg/ml uracil final concentration)

Escherichia Coli **Growth Media**	
Sterilisation: Autoclave at 121˚C, for 15 minutes. **Storage:** Media stored in the dark, at room temperature. Agar plates stored in plastic bags (sealed to prevent desiccation), in the dark, and at room temperature. Before each use, the plates are dried upside down in an incubator for a minimum of 20 minutes. These conditions apply to all listed media/plates unless stated otherwise.	
LB **(Lysogeny Broth)**	**Broth:** 1% tryptone (Bacto), 0.5% yeast extract (Difco), 170 mM NaCl, 2nM NaOH, pH to 7.0. **Agar plates:** 300ml LB broth, 1.5% agar
Escherichia Coli **Growth Supplements**	
Amp (Ampicillin)	Added at final concentration of 50 μg/ml.
X-gal	Added at final concentration of 40 μg/ml.

4.1.6 Reagent Solutions and Buffers

H. *volcanii* transformation Reagents	
Buffered Spheroplasting Solution	1 M NaCl, 27 mM KCl, 50 mM Tris.HCl pH 8.5, 15% sucrose Filter sterilised through a 0.2 µM filter (Minisart Sartorius)
Unbuffered Spheroplasting Solution	1 M NaCl, 27 mM KCl, 15% sucrose, pH 7.5. Filter sterilised through a 0.2 µM filter (Minisart Sartorius)
Spheroplast Dilution Solution	23% SW, 15% sucrose, 37.5 mM CaCl2 Filter sterilised through a 0.2 µM filter (Minisart Sartorius)
60% Polyethylene Glycol 600 (PEG 600)	Per single transformation reaction: 150µl of PEG 600 100µl of unbuffered spheroplasting solution
DNA samples for transformation	5 µl 0.5 M EDTA, pH 8.0, 15 µl unbuffered spheroplasting solution, 10 µl DNA (~1-2 µg).
Regeneration Solution	18% SW, 1×YPC, 15% sucrose, 30 mM CaCl2 Filter sterilised through a 0.2 µM filter (Minisart Sartorius)
Plating Solution (or Transformant dilution solution)	18% SW, 15% sucrose, 30 mM CaCl2 Filter sterilised through a 0.2 µM filter (Minisart Sartorius)
E.*coli* transformation	
SOC broth	2% tryptone (Bacto), 0.5% yeast extract (Difco), 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM MgSO4, 20 mM glucose.
Colony lift	
2x SSPE	To make 20x SSPE: 3 M NaCl, 230 mM NaH2PO4, 32 mM EDTA, pH 7.4. Dilute with factor 1:10 to make 2x SSPE.
Denaturing Solution	1.5 M NaCl, 0.5 M NaOH

10% SDS	sodium dodecyl sulfate solution (kept at 30 °C)
Neutralising Solution	1.5 M NaCl, 0.5 M Tris.HCl, 1 mM EDTA.
Colony Hybridisation	
100 x Denhardt's Solution	2% Ficoll 400, 2 % polyvinyl pyrrolidone (PVP) 360, 2% BSA (bovine serum albumin, Fraction V).
20 x SSPE	3 M NaCl, 230 mM NaH2PO4, 32 mM EDTA, pH 7.4.
Low-stringency wash solution	2 x SSPE, 0.5% SDS
High-stringency wash solution	0.2% SSPE, 0.5% SDS
Pre-hybridisation Solution (Pre-Hyb)	6 x SSPE, 1% SDS, 5 x Denhardt's solution, 200 μg/ml salmon sperm DNA (Roche, boiled for 5 minutes prior to addition)
Hybridisation Solution (Hyb)	6 x SSPE, 1% SDS, 5% dextran sulphate
Agarose Gel Electrophoresis	
TBE (Tris/Borate/EDTA)	89 mM Tris.HCl, 89 mM boric acid, 2 mM EDTA.
TAE (Tris/Acetic acid/EDTA)	40 mM Tris.HCl, 20mM acetic acid, 1 mM EDTA.
Gel Loading Dye, Purple (6X)	2.5% Ficoll®-400 10 mM EDTA 3.3 mM Tris-HCl 0.08% SDS 0.02% Dye 1 8.0E-4% Dye 2 pH 8@25°C

	(According to New England Biolabs Supplier instructions)
SYBR™ Safe (Invitrogen)	(Z)-4-((3-Methylbenzo[d]thiazol-2(3H)-ylidene)methyl)-1-propylquinolin-1-ium 4-methylbenzenesulfonate

4.1.7 Other Reagents

Consumables

Round-botton tubes (Sarstedt)
1.5ml tubes (Eppendorf)

Chemicals

All chemical reagents were obtained from Sigma, unless stated otherwise.

Other Solutions

Sodium Acetate: 3 M NaAc pH 5.2 (filter sterilised)
TE buffer : 10 mM Tris.HCl pH 8.0, 1 mM EDTA

Oligonucleotides

All oligonucleotides were obtained from Eurofins Genomics, Germany.

Restriction Enzymes

All enzymes were obtained from NewEngland Biolabs (NEB).

Primer Design and Cloning Tools

All primer design and analysis was performed using MacVector with Assembler MacOS software version 18.5.1.

DNA Sequencing

The DNA sequences for the deletion constructs was obtained from the Deep Seq sequencing facility at the University of Nottingham.

4.2 Methods

4.2.1 *Haloferax volcanii:* General microbiology

Culturing and storage conditions

Haloferax volcanii cultures were grown on solid agar plates in a static incubator (LEEC) at 45℃ in sealed plastic bags to avoid desiccation, with growth rate of average 5-10 days.
Overnight cultures of 1-5ml volume were grown for at least 12h in a static incubator at 45℃ with 8rpm rotation. Cultures on solid agar plates could be stored in sealed plastic bags at room temperature – provided they are used within 7-10 days. For longer-term storage, a 20% v/v glycerol mixture is made (80% glycerol, 6% SW). The mixture is distributed into 2ml round-bottom tubes, which are then snap-frozen on dry ice, and stored at -80℃.

4.2.2 *Haloferax volcanii:* General manipulation

Gene Replacement/ Gene Deletion through PEG600 Transformation

The following methodology has been developed on the PEG-mediated transformation protocol originally outlined in 1987[167] and with practical steps described later, in 1989[168].

A 10ml (per 3 transformation reactions) overnight culture of 1-4 inoculated colonies is set up in a rotator at 45℃, 8rpm rotation speed. Using a spectrophotometer, the growth state of the culture is ensured to be at 0.6-0.8 (A_{650} absorption). The cells in the culture are pelleted in round-bottom tubes (14ml), at 3300 x g for a total of 8 minutes, at 25°C. The supernatant is removed, and the pellet is resuspended in 2ml of buffered spheroplasting solution. 2ml of the resulting mix is transferred into 2ml round-bottom tubes, to be pelleted again in a swing-bucket rotator at 3300 x g, for another 8 minutes. The resulting pellet is gently resuspended in 600µl of buffered spheroplasting solution, and 200 µl per transformation is distributed into fresh 2ml round-bottom tubes. For the spheroplast formation step, 0.5M of EDTA (pH of 8.0) is added on the side of the tube, and shaken horizontally so as to not to lyse the cells. The cells are left standing at room temperature for 10 minutes to allow the formation of the spheroplasts. Then, either the transforming DNA mixture or water control are added to the side of the tube, and left to incubate for 5 minutes. Then, 60% PEG600 is added to the side, and the tube is incubated at room temperature for 30 minutes.

Finally, a spheroplast dilution solution is added and incubated for 2 minutes at most, before pelleting the cells at 3300 x g, for 8 minutes. For the regeneration step, sterile conditions are ensured with work carried out under a bunsen burner. A ml of regeneration solution is added, and the whole pellet is transferred to a 4ml sterile tube, and left undisturbed for 1.5-2 hours to regenerate at 45 °C. At this stage, the necessary media supplements are added. Then, the tubes are placed overnight in the 8rpm rotator at 45°C.

The next day, 2ml of the cells are pelleted at the same conditions described above, the supernatant removed. The supernatant is removed, and the resulting pellet is resuspended in 1ml of plating solution, using filter tips for added sterility. Finally, 100µl of each dilution (10^{-1}, 10^{-2} for selective, and 10^{-6} for non-selective) is plated on Hv-Ca or Hv-YPC plates.

A "pop-in/pop-out" method developed in[169] is used as a gene knock-out/ gene-replacement system.

To select for colonies with integrated plasmids – the cells are plated on minimal media (Hv-Ca, with the required additional growth supplements), which selects for uracil autotrophy supplied by the pyrE2 gene on the plasmid construct. After around 5-10 days, the colonies are restreaked on a fresh agar plate containing the same media.

For the plasmid excision or "pop-out" stage, 5ml of the culture is incubated in the 8rpm rotator, at 45°C overnight, until at least 0.6 at A650 absorbance. The relief for uracil selection is achieved through diluting the culture 1/500 twice over two overnights in fresh 5ml of Hv-YPC media, with necessary growth additives such as thymine. This allows for the natural excision of the plasmid containing the pyrE2 gene, which is then selected for by plating the cells on Hv-YPC + 5-FOA.

4.2.3 Nucleic acid Quantification

The concentration and purity of the DNA samples was determined by calculating the absorbance ratios of 260/230nm and 260/280nm using the spectrophotometer NanoDrop 2000 (ThermoFisher Scientific).

4.2.4 Nucleic acid Manipulation

Generating gene deletion constructs through PCR amplification, restriction digest, and ligation

Following the amplification of the PCR fragments from the genomic clone containing the gene of interest, with upstream and downstream sequences in separate reactions – the sequences were incubated for 2 hours at 30° C with the appropriate internal cut site restriction enzyme (e.g., *BamH1*). The p131 vector is incubated with the external cut site restriction enzymes, overight to ensure complete digestion. Using T4 ligase, the US and DS fragments are ligated together, and incubated in a 10°C water bath, overnight. The same external cut sites of the ligated fragment, and the empty vector are cut by incubating with the relevant restriction enzymes. Finally, the fragment

is ligated into the empty p131 vector, and transformed into *E.coli* cells through electroporation. Refer to Table **6** for a graphical description of the cloning process steps.

4.2.5 Genotype Screening

Colony PCR (diagnostic)

This method is used when there is a substantial number of *Haloferax volcanii* colonies (i.e., more than 10), which require to be tested for the orientation of the integrated construct ("pop-in"), or for the presence of the desired fragment ("pop-out"). The individual colonies are picked with a sterile tip, and boiled in 100µl of water at 100 ˚C , for 10 minutes.

PCR Components	25µl Total Reaction	50µl Total Reaction
5× Q5 Reaction Buffer	5µl	10µl
10mM dNTPs	5µl	10µl
10µM Forward Primer	0.5µl	1 µl
10µM Reverse Primer	0.5µl	1 µl
Template DNA (boiled at 100˚ C, for 10 min)	1 µl	2 µl
Q5 Hot Start High-Fidelity DNA Polymerase	0.25 µl	0.25 µl
5× Q5 High GC enhancer	5 µl	10 µl
Nuclease-Free Water	7.75 µl	15.75 µl
PCR Conditions (Q5® Hot Start High-Fidelity DNA Polymerase)	**Step**	**Time and Temperature**
	Initial Denaturation	**98˚ C, 10 sec**
	Denaturation	**98˚ C, 30 sec**
	Extension	**Tm˚ C (35 Cycles), 20 sec per cycle**
	Final Extension	**72˚C ,30 sec/kb**
	Hold (Cool)	**10 ˚C , 10 min+**

Tm = The annealing temperature for each set of primers was calculated by applying the following equation[170]:

Equation 1. Calculation for the annealing temperature of the primer. The average temperature is taken between a pair of two primers.

$$85.5 + (16.6 \times log10[Na^+] + \big(0.41 \times (\%GC\ content)\big) - \big(100 - (\%\ homology)\big) - \left(\frac{600}{length}\right)$$

4.2.6 Genotype Screening: Colony hybridisation

Colony Lift

A number of colonies for testing is patched on Hv-YPC plates using autoclaved wooden toothpicks, and grown at 45℃ for 5 days at most. Cut circles of GE Healthcare Amersham Hybond – N+ positively charged membrane, are gently pressed on the plates with patched colonies for transfer for a minute. The transferred patched cells on the nylon membrane (filter) are sequentially lysed by transferring onto a Whatmanpaper soaked in 10% SDS for 10-15 minutes until a glossy appearance. Then, the filer are transferred onto the paper soaked in denaturing solution for 5 minutes until a blurry appearance. Finally, the filters are held on paper soaked in neutralising solution (and repeated twice with fresh solution), before being submerged for 30 seconds in 2×SSPE. The filters are then either air-dried or placed in an incubator for 5 minutes, before being crosslinked to the membrane with 120 mJ/cm2 UV.

Colony Hybridisation

The filters containing transferred patched cells from the colony lift were incubated with the prehybridization solution at 65℃ in a rotator overnight. The DNA mix is boiled at 100℃ for 5 minutes, and kept on ice before the addition of Hiprime random priming mix (Roche) and ^{32}P-dCTP and incubated for further 15-20 minutes, and purified on a BioRad P-30 column. The radiolabelled probe (50 ng of DNA and 0.74 MBq of [α-32P] dCTP) added to Salmon sperm DNA, is then added through a syringe into the hybridisation tube (n.b., the pre-hyb solution is removed, and replaced with hyb solution). After overnight incubation with rotation, the filters are washed with low stringency wash solution for the first 10, and 30 minutes. Then the next two

washes in 30 minute intervals, the filters are washed with high stringency wash solution. Before imaging, the filters are air-dried, wrapped in cling film, and exposed exposed to a phosphorimager screen (Fujifilm BAS Cassette 2325) overnight (a minimum of 24 hours). The image is scanned using a GE Healthcare Typhoon.

4.2.7 DNA Extraction and Purification

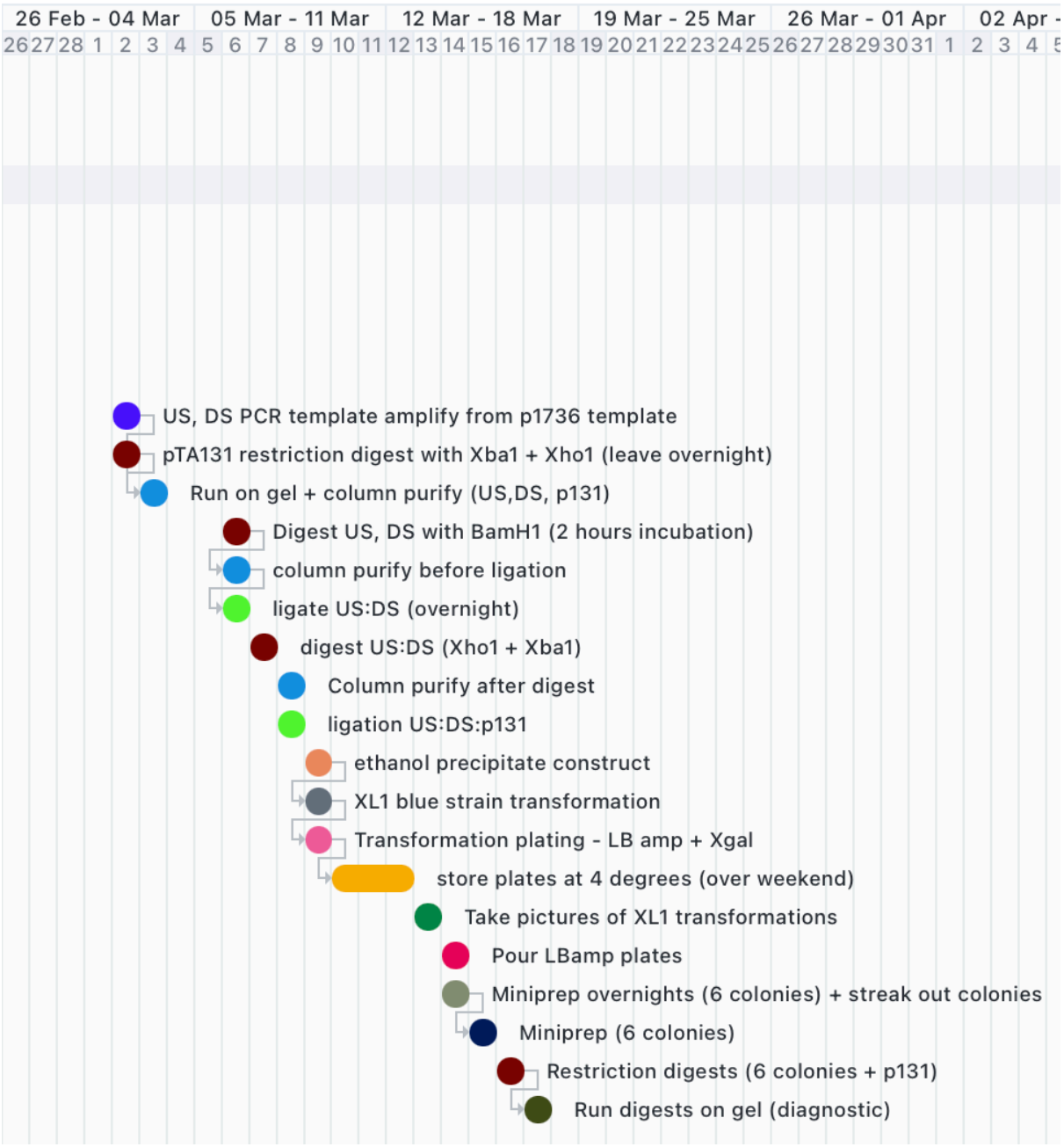
Isolation of high-copy plasmid DNA from E.coli (Maxiprep/Miniprep)

The steps are carried out according to the NucleoSpin® Plasmid protocol (section 5) in the Plasmid DNA purification Macherey-Nagel kit, with the provided reagents.

Extraction of DNA from agarose gels and PCR Clean-up

The steps are carried out according to the NucleoSpin® Gel and PCR Clean-up (DNA extraction from agarose gels; section 5.2) in the Plasmid DNA purification Macherey-Nagel kit, with the provided reagents.

Table 6. Timeline of the cloning process. The length of the bar indicates the number of days of incubation required.



Chapter 5 Construction of *rnh* strain deletions in an originless background

5.1 Generation of ΔrnhA strains through the pop-in/pop-out gene replacement approach

Haloferax strains deleted for the origins on the main chromosome (OriC1,2,3) were transformed to generate *rnhA* deletions, as described in the methods section.

To test for the gene deletion, the probe for the *rnhA* gene is constructed using PCR, and used in colony hybridisation; the lack of probe binding confirms the deletion of the *rnhA* gene.

5.1.1 Construction of the probe for the *rnhA* gene using a restriction digest

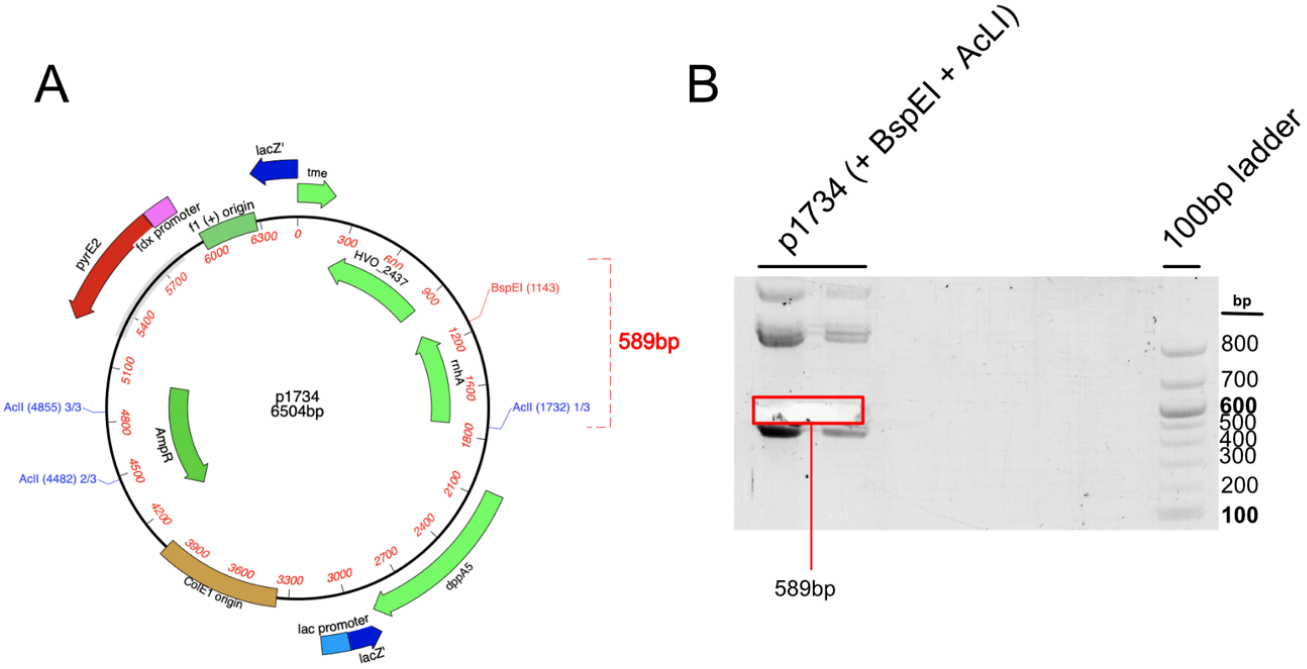


Figure 15. Schematic of the plasmid containing the (A,B) *rnhA* gene, with the flanking restriction sites used to obtain the 589bp fragment. (C) shows the *rnhA* gene fragment excision to be purified and used as a probe for filter hybridisation (*see* fig 5.1.2) of *rnhA* deletion mutants.

To obtain the *rnhA* gene fragment, a sequential digest using *ACLI* (cutting at 1732bp) and *BspE1* (cutting at 1143bp) restriction enzymes of the (fig 5.1.1 **A**) p1734 *rnhA* genomic clone was used.

The gel image (fig 5.1.1 **B**) indicates the excised restriction digest product of 589bp, which corresponds to the size of the gene fragment between the *ACLI* and *BspE1* restriction sites. The excised gel fragment was later purified using the methods described in the nucleic acid purification section.

5.1.2 Confirmation of Δ*rnhA* strains through colony hybridisation

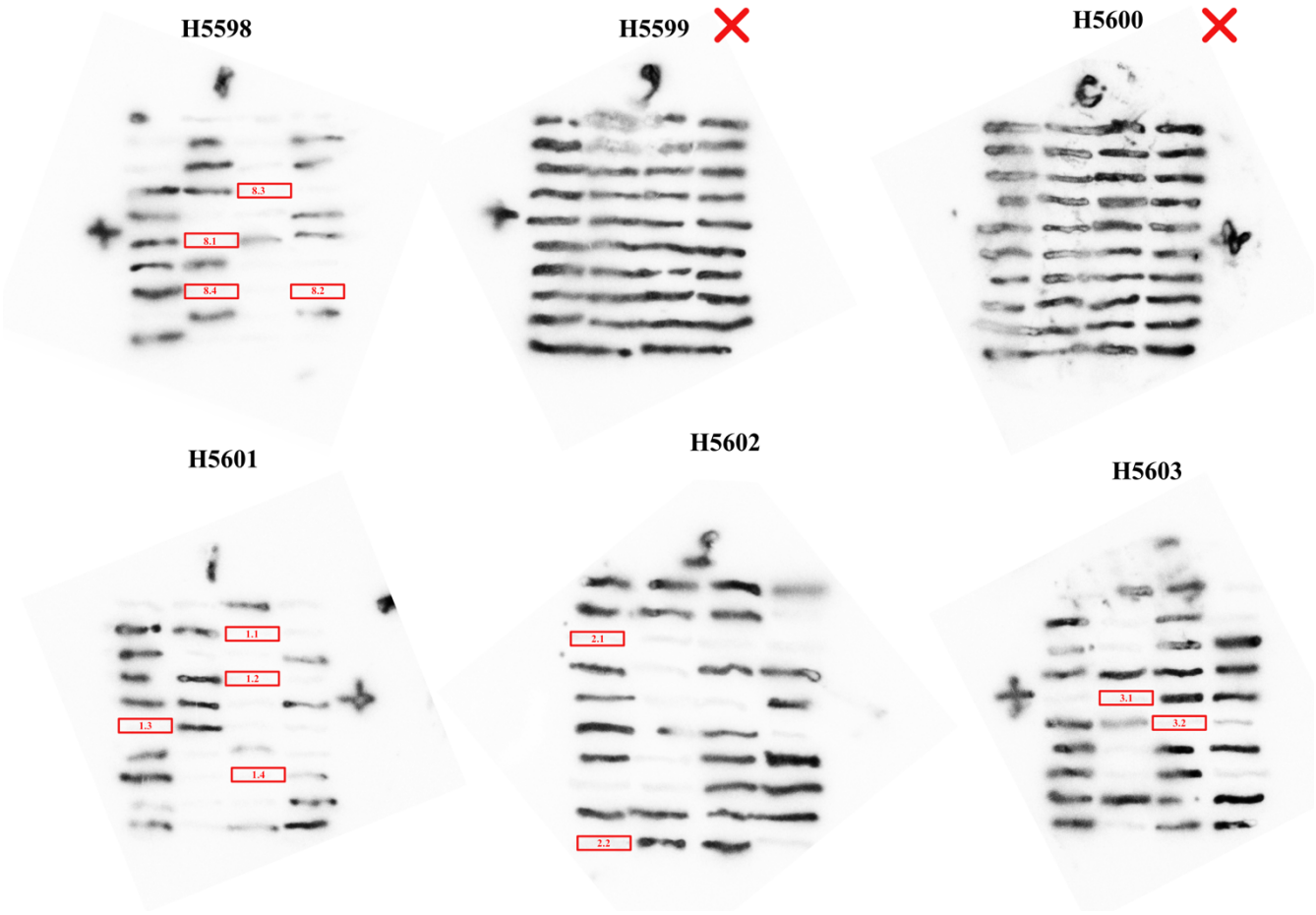


Figure 16. Colony hybridisation to confirm the deletion of the *rnhA* gene. The black patches indicate probe hybridisation (*rnhA* gene present), while the white patches show gene deletion. The red boxes around the patches show selected clones.

5.2 Generation of gene deletion (Δ*rnhC*) construct through PCR

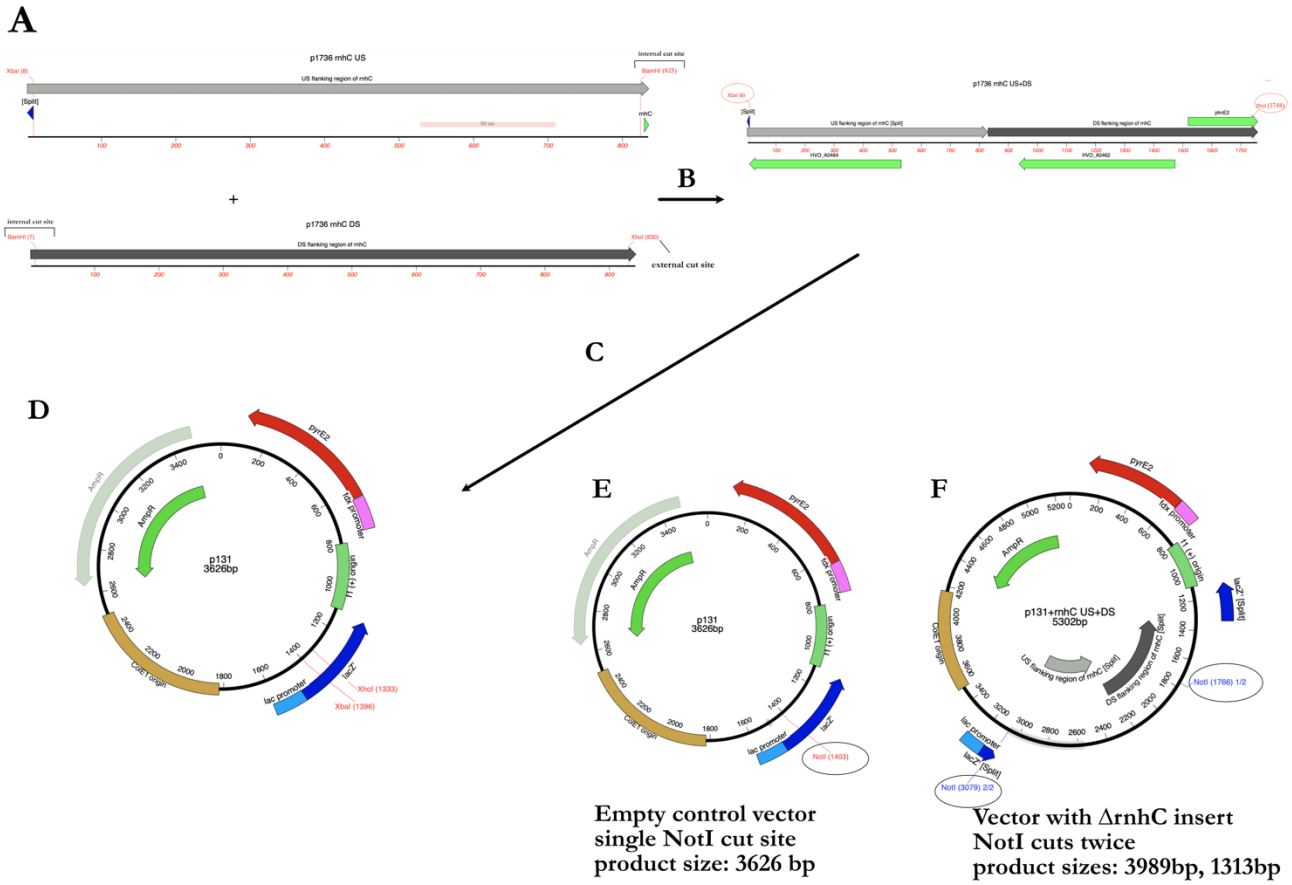


Figure 17. Schematic diagram of the entire cloning workflow.

5.2.1 Amplification of US and DS flanking regions of *rnhC* from the p1736 clone

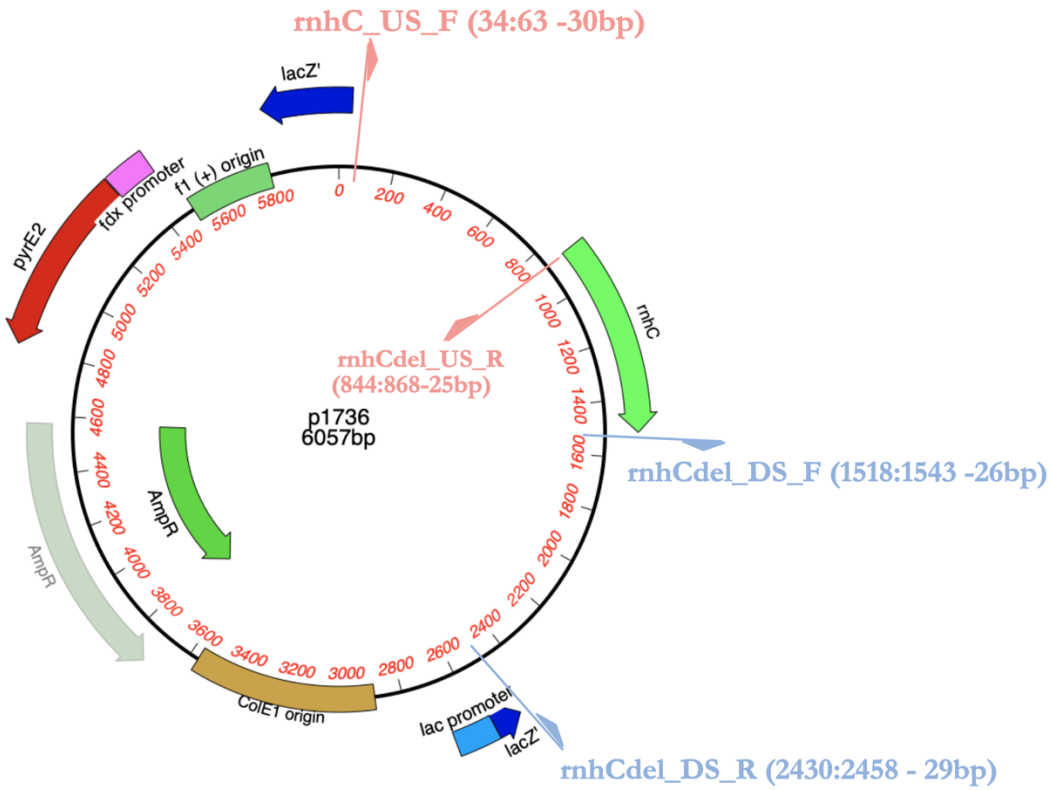


Figure 18. Plasmid map of genomic clone of rnhC – p1736 – and the relevant primer sites used to amplify the DS and US regions of the rnhC gene.

For a detailed schematic on the cloning process, refer to figure **17**.

A deletion construct for the *rnhC* gene was generated through the PCR amplification of US and DS fragments flanking the *rnhC* genomic clone, p1736. (figure **18**) The fragments were run on a gel (figure **19**) to ensure the correct sizes prior to ligation. The fragments are ligated together, and inserted into the empty p131 vector, followed by blue-white selection on LB- Amp + Xgal plates. The *rnhC* fragment is inserted into the Multiple Cloning Site, thus disrupting the lac-z operon, and the β-galactosidase pathway, rendering a white-coloured colony (figure **19**).

Finally, the DNA is extracted (miniprep) from six selected colonies on the X-gal plates, and the colonies are tested for the *rnhC* deletion fragment through a diagnostic digest. *NotI* cuts twice where there is an insert to give fragment sizes of 3989bp and 1313bp (figure **21**)

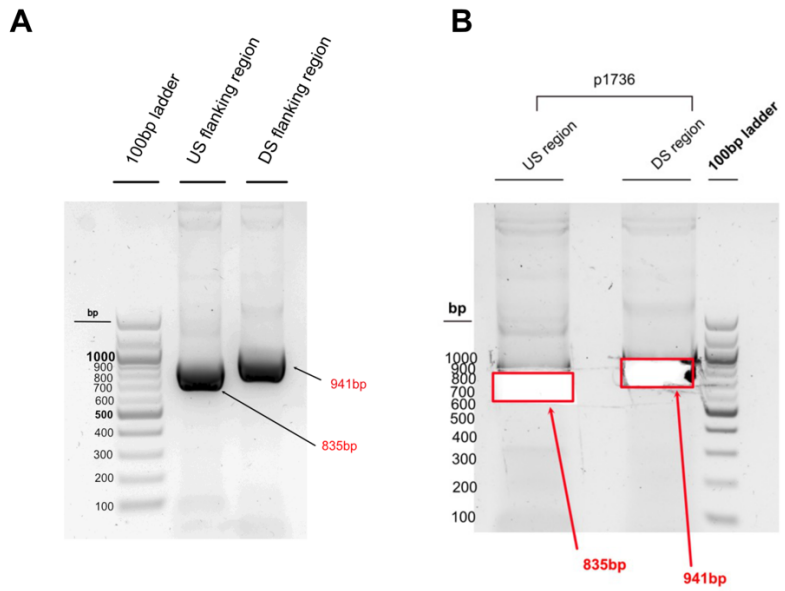


Figure 19. Diagnostic PCR for the flanking upstream (US; 835bp) and downstream (DS; 941bp) amplified regions for the deletion of the rnhC gene. (A) The black bands indicate the size of DNA fragments present on the gel (835bp and 941bp). In the biological repeat on the right **(B),** the red boxes show the positioning of the excised fragments from the gel, for the DNA to be purified using the agarose extraction method outlined in Ch. **4.2.7**)

5.2.2 Transformation of ΔrnhC insert into XL1 vector for blue-white screening on LB + Amp with X-gal plates

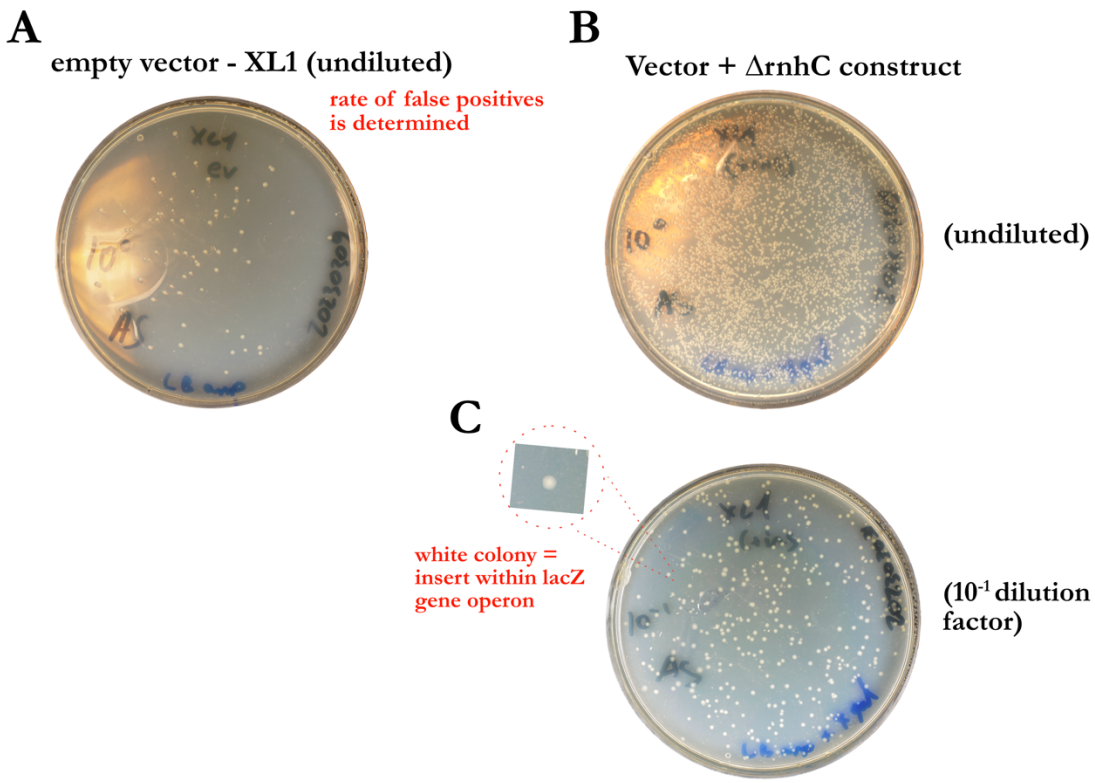


Figure 20. Blue-white selection on LB + Amp plates. The white colonies indicate the presence of the inserted fragment in the MCS.

5.2.3 Confirmation of *rnhC* deletion construct through a diagnostic digest

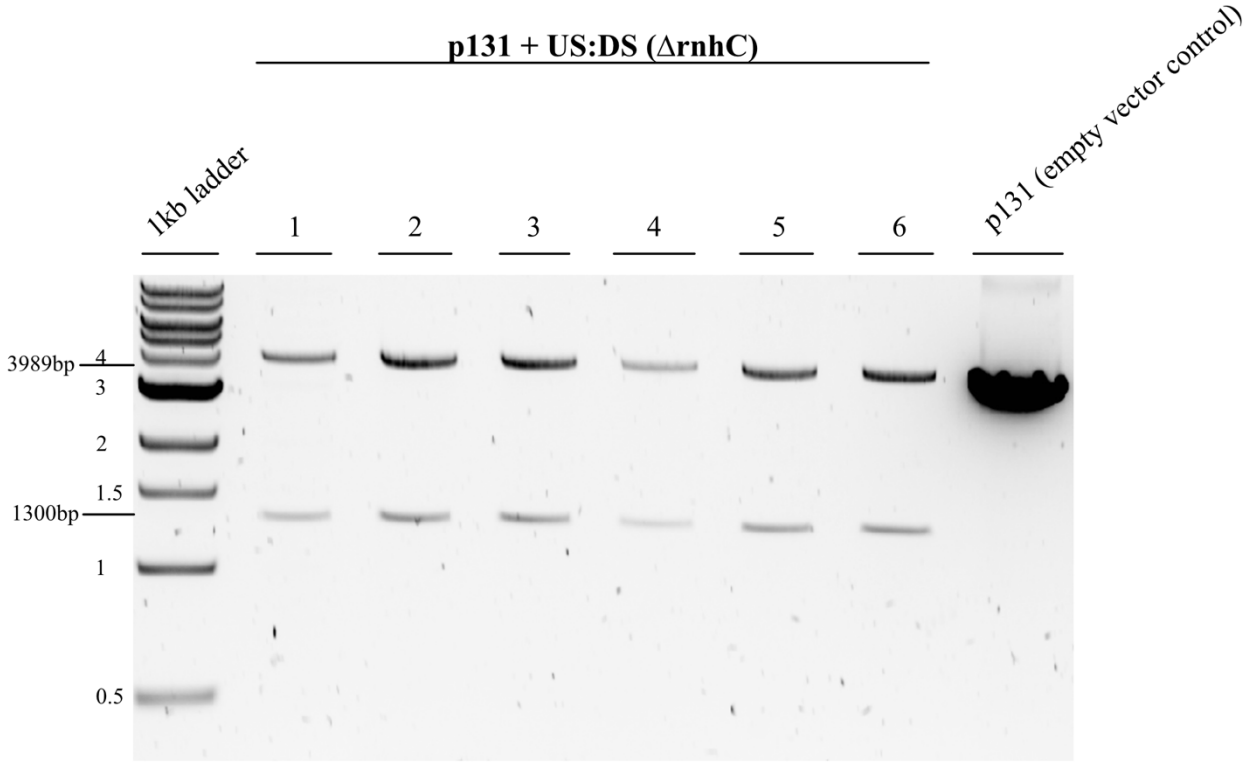


Figure 21. ΔrnhC genomic constructs made through ligating the flanking US and DS regions of the p1736 genomic clone *rnhC,* and ligated into the p131 vector. Two bands indicate the presence of the gene deletion construct, due to NotI cutting at two separate sites. All six clones contained the gene deletion.

5.3 Future work

A genome library of deletion constructs will be generated for the construction of the *rnhA,B,C,E* quadruple deletion mutants, transformed into *Haloferax volcanii* cells. From there, the deletion of the origin on pHV3 is going to be attempted. The episomal state of the pHV3 mini-chromosome is to be confirmed using pulsed-field gel electrophoresis. Should the deletion be successful, the correlation between the levels of R-loops, and genomic rearrangements will be characterised using DRIP-seq technology.

Chapter 6 Increasing the transcription levels on pHV3: inducible tryptophanase promoter constructs

6.1 Generating constructs of *adh2* gene under tryptophan-inducible promoter (p.tnaA::adh2)

The *Haloferax* starter strain H98, which is deficient for thymine, is transformed with a plasmid containing the p.tnaA::adh2 insertion (figure **22**). The orientation of the integrated plasmid is checked using PCR amplification (refer to figures **24-25**), with descriptions of used primers and expected bans fragments under the relevant figures.

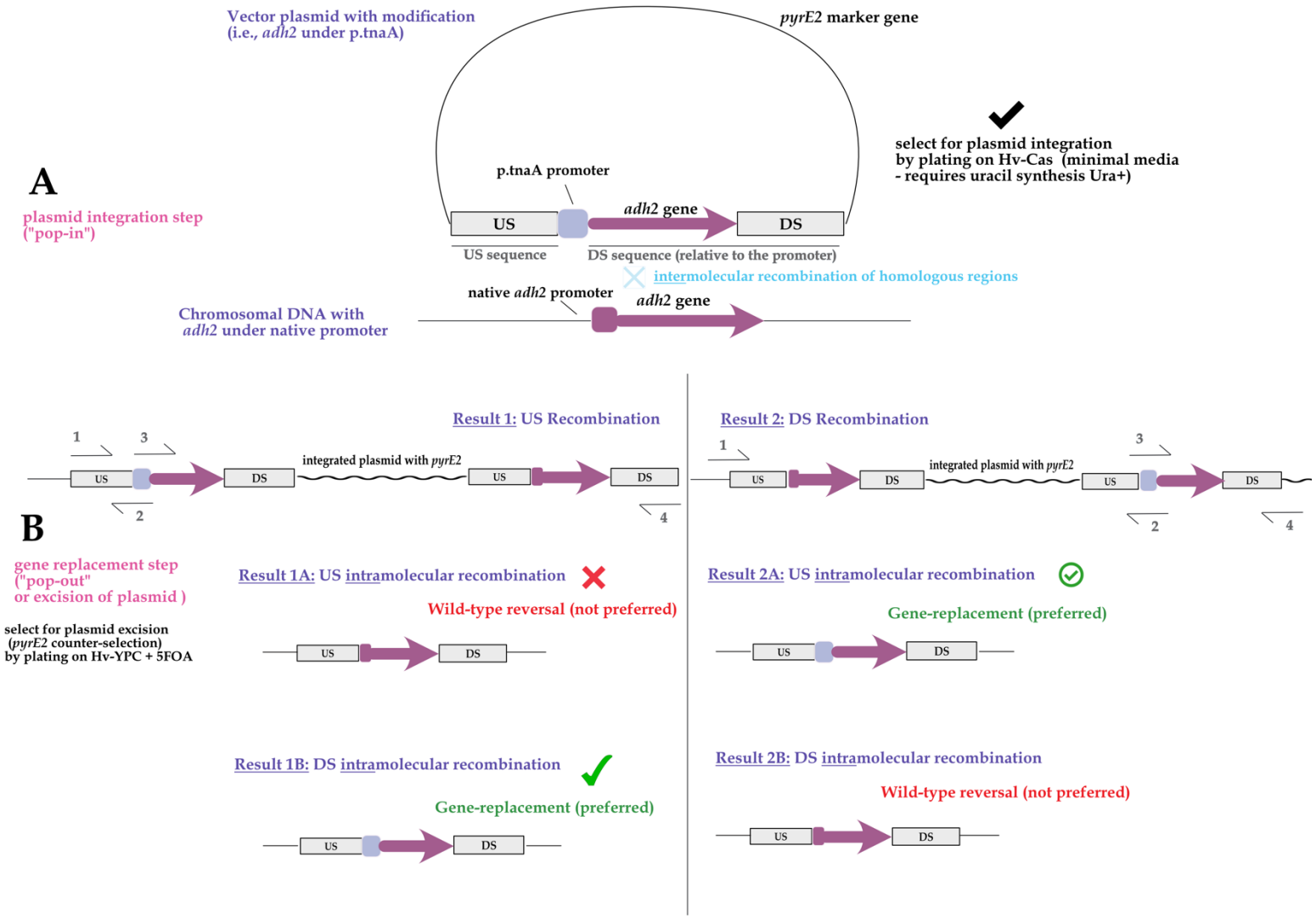


Figure 22. Schematic diagram of the gene-replacement method, and the possible results of the gene replacement. To test for the correct orientation of the fragment, only fragment which is amplified between the primer sites (shown as half-arrows) 3 and 4 is amplified. The DS pop-in fragment would be 3374bp, while the US pop-in 1738 bp.

6.2 Screening for the orientation of the ptnaA::adh2 inserted construct using colony PCR – "pop-in"

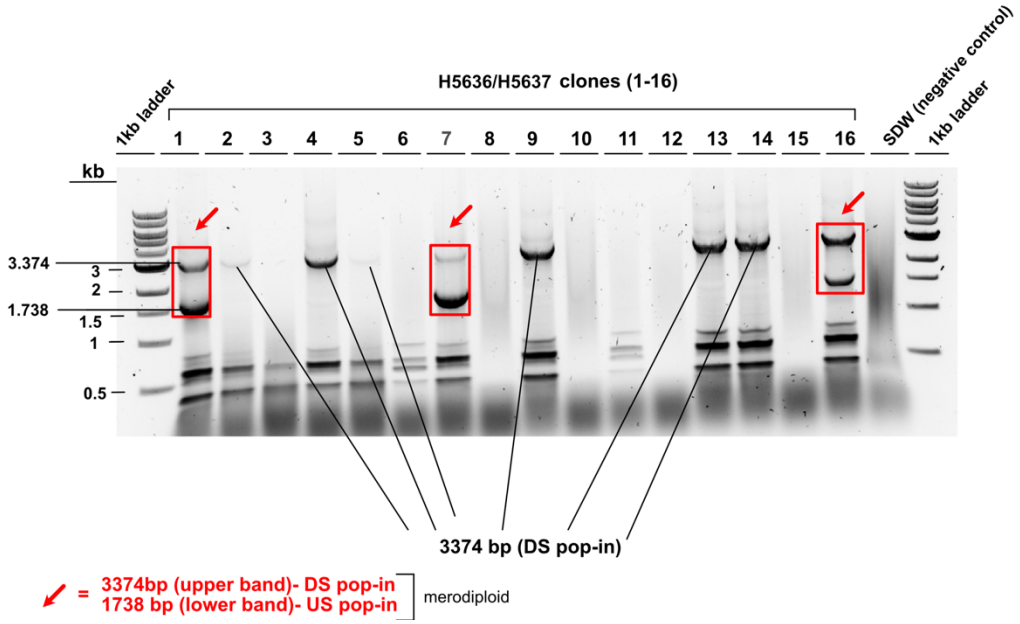


Figure 23. The double bands marked by red arrows on the gel indicate the presence of two fragments and thus are merodiploid. For a DS pop-in, clone 9 is chosen. For US pop-in, clone 7 is chosen.

To test for the correct orientation of the fragment, the primer pairs dgabT1XhF and p.tnaAFint, along with p.tnaAintR and adh2USR are used to amplify the fragment. The primers p.tnaAFint and p.tnaAintR can only bind to the p.tnaA promoter region. The correct sized fragment of 3374 bp, indicating a DS pop-in, is shown in figure **23.**

6.3 Screening for the <u>downstream</u> ptnaA::adh2 integrated construct using colony PCR

To check for the precense of the p.tnaA::adh2 fragment, PCR using a primer pair of adh2extR and adh2intR, would generate fragments of 677bp for a region containing the p.tnaA, and 358bp for a wild-type *adh2* gene region.

H5637 (clone 9 - DS pop-in)

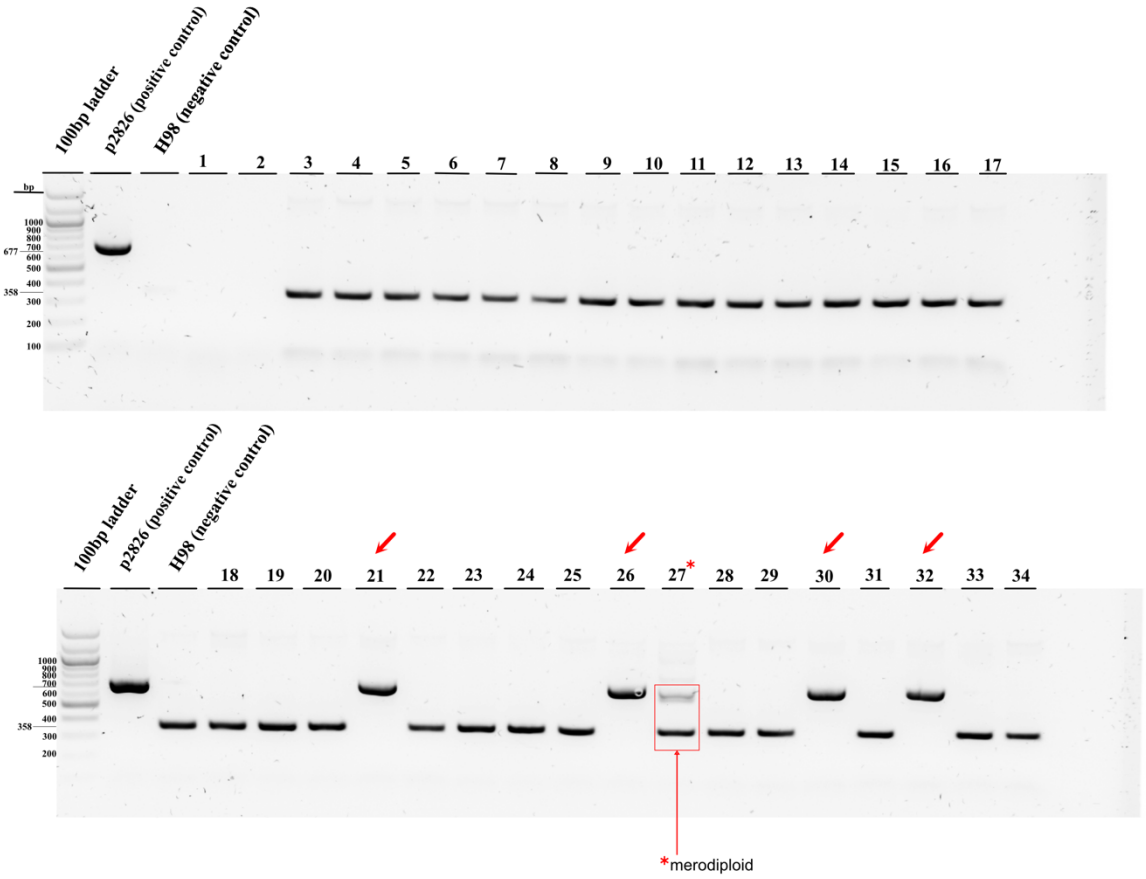


Figure 24. Arrows indicate that the larger sized fragment of 677bp which contains the p.tnaA::adh2 insert.

6.4 Screening for the upstream ptnaA::adh2 integrated construct using colony PCR

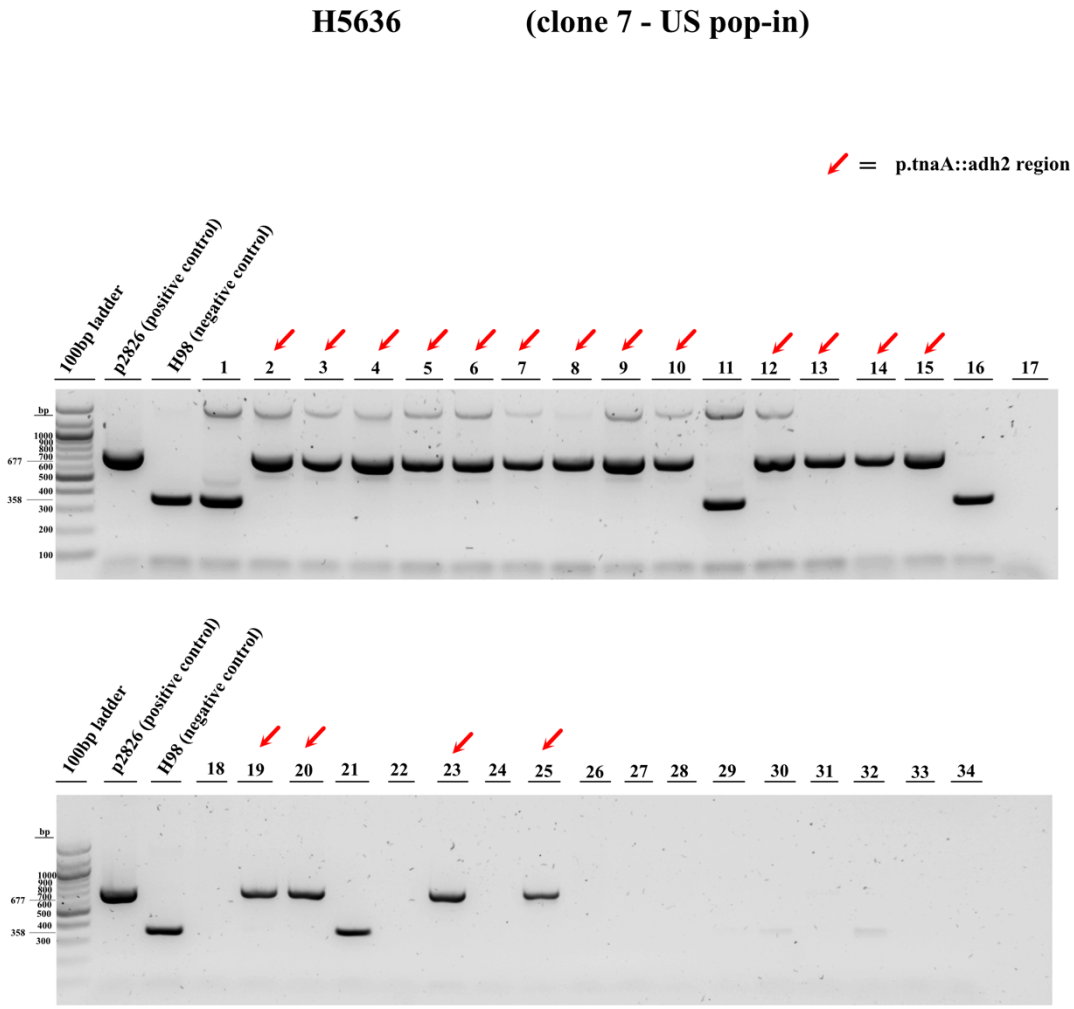


Figure 25. Clones containing the construct (not merodiploid) are indicated with red arrows.

6.5 Future work

Additional methods of upregulating gene transcription may be employed:

(1) Xylose[171,172]

(2) Iron siderophore cluster [173]

The results of this study have developed a system for controllable gene expression, which can be up or downregulated, and employed to test the effect of regulating the Orc initiator or RadA recombinase on the ability to delete the origin on the pHV3 chromosome.

Chapter 7 Hel308 mutation strains

7.1 Aims

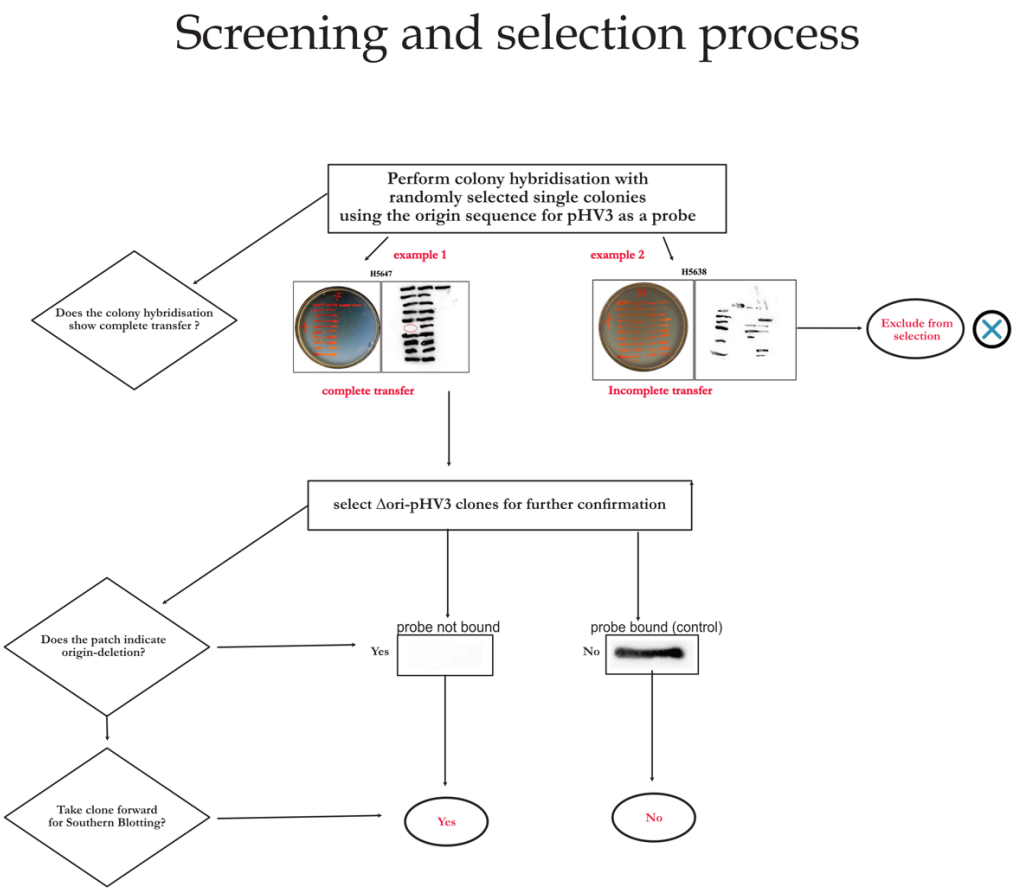


Figure 26. Methodological framework for clone selection.

7.2 Hel308 deletion and point mutation allows for the deletion of the Δori-pHV3

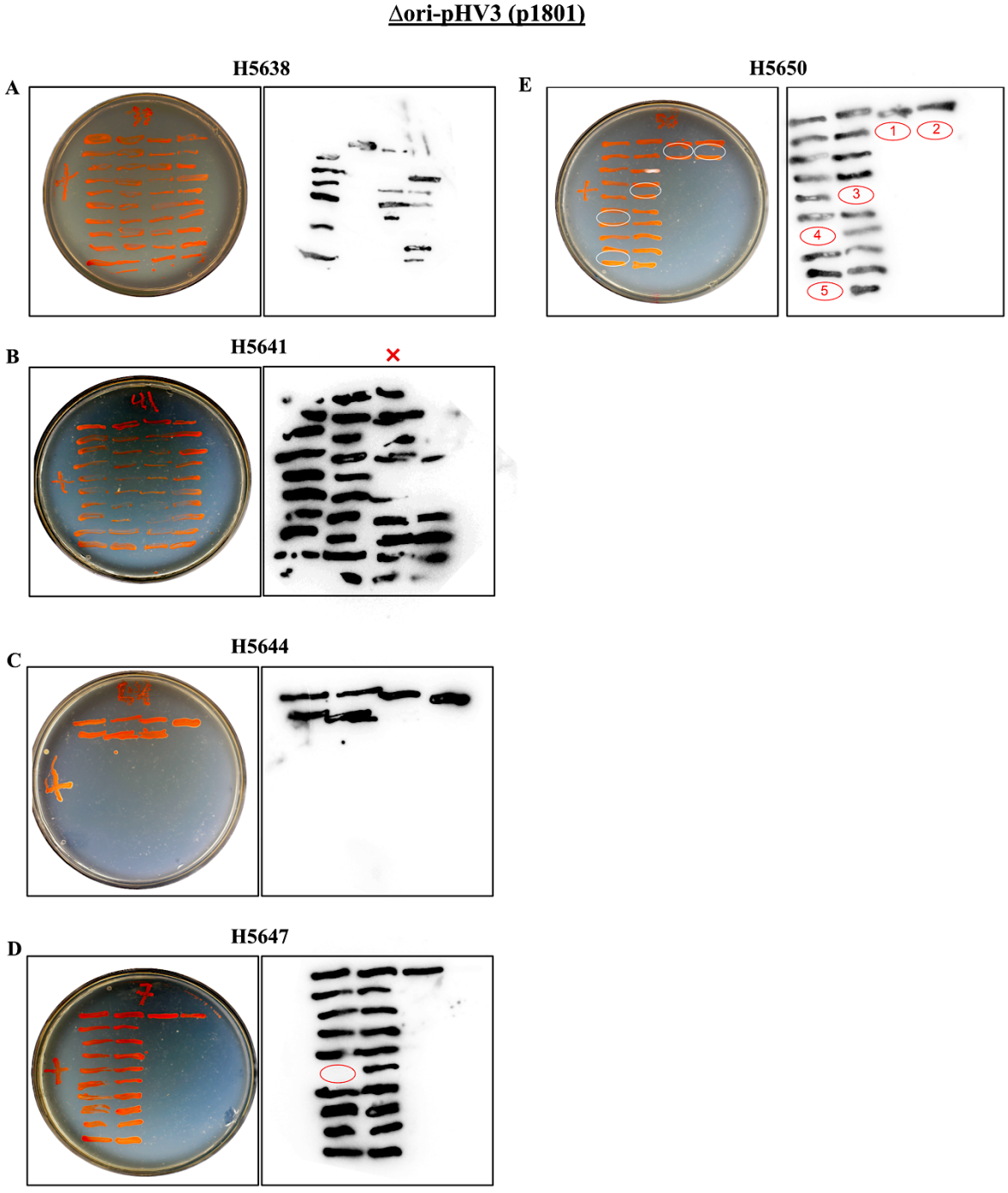


Figure 27. Colony hybridisation as a confirmation for Δori-pHV3 in Hel308 deletion and point mutation strains, using the ori-pHV3 sequence as a control probe.

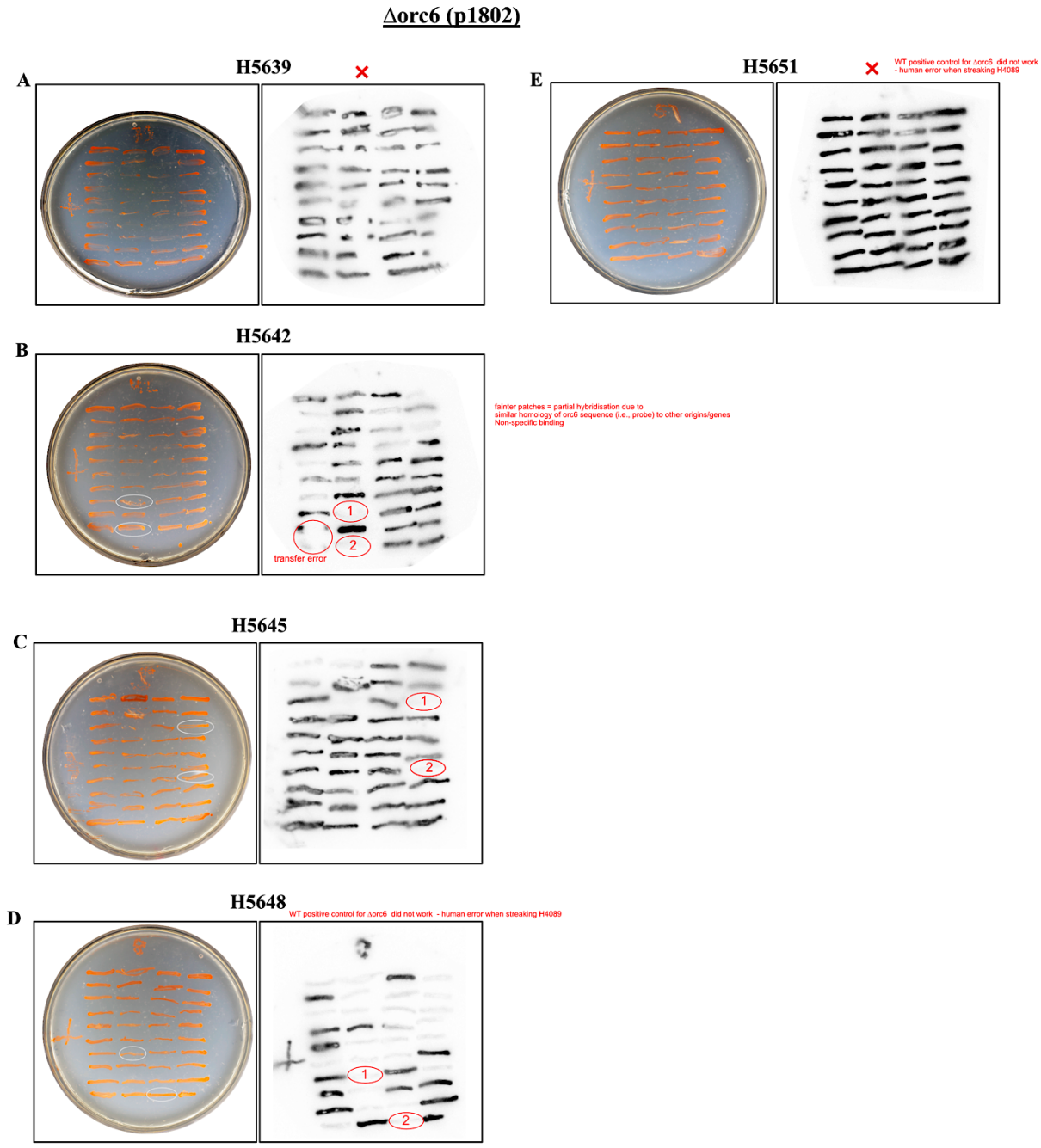


Figure 28. Colony hybridisation as a confirmation for Δorc6 in Hel308 deletion and point mutation strains, using the ori-pHV3 sequence as a control probe. In panels D and E, the left-hand plus sign (streaked parent strains with the orc6 gene sequence absent) serves as a positive control for the deletion of the *orc6*, as indicated by the lack of probe binding on the adjacent hybridisation filters.

Δori-pHV3, Δorc6 (p1803)

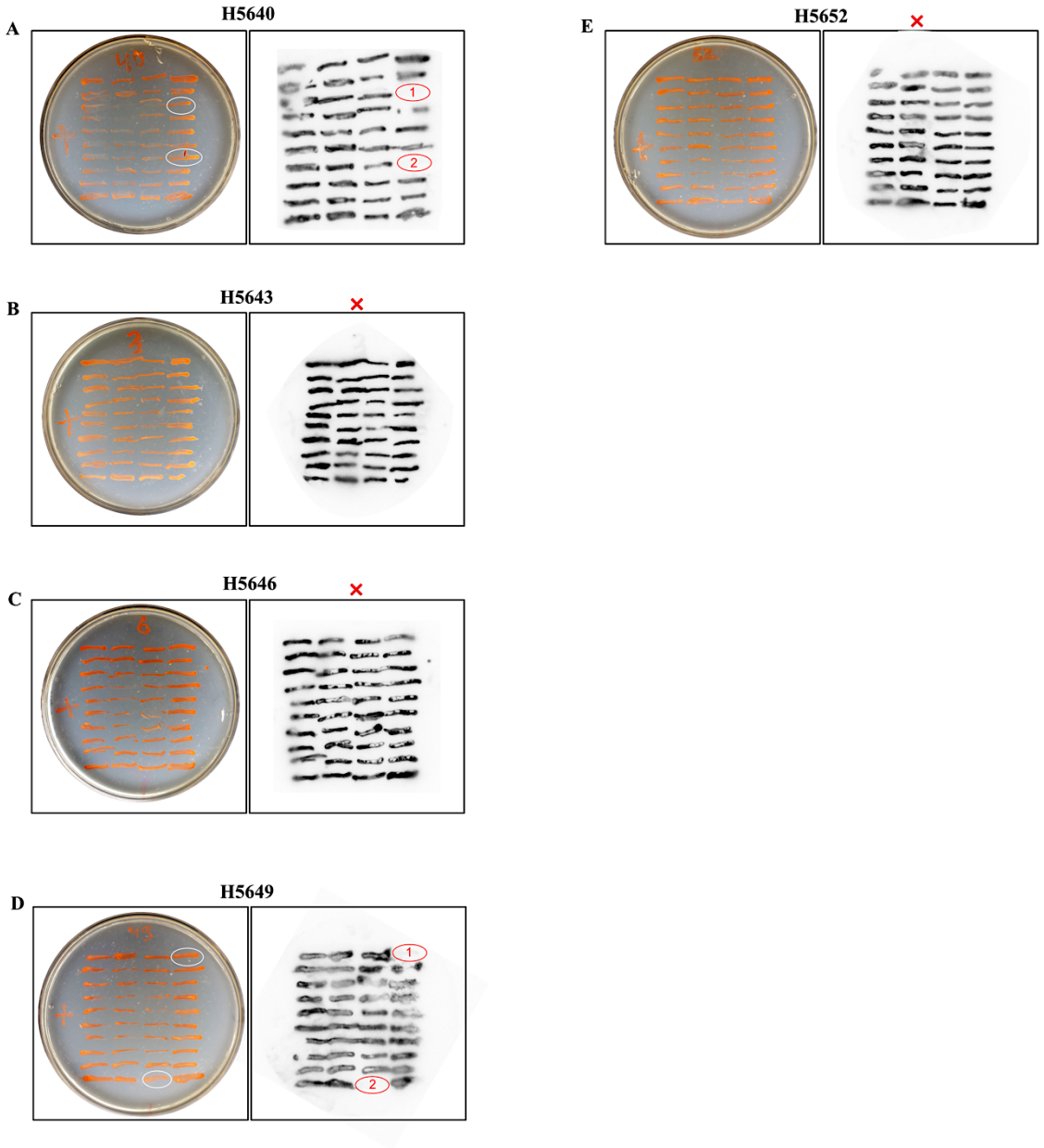


Figure 29. Colony hybridisation as a confirmation for Δori-pHV3, Δorc6 in Hel308 deletion and point mutation strains, using the ori-pHV3 sequence as a control probe.

7.3 Future work

Following the successful deletion of the ori-pHV3, provided that the pHV3 is episomal through pulsed-field gel electrophoresis, the origin-deleted strains will be further characterised. The growth rates will compare the wild-type against the mutant strains, as well as test the strains using DNA damage assays (UV, Mitomycin C etc.,) Finally, the Hi-C technology will be used to investigate the correlation between genomic rearrangements induced by gene deletion, and the genotype.

Bibliography

1. Gaillard, H., García-Muse, T. & Aguilera, A. Replication stress and cancer. *Nat. Rev. Cancer* **15**, 276–289 (2015).
2. Avery, O. T., MacLEOD, C. & McCarthy, M. Studies on the chemical nature of the substance inducing transformation. *J. Exp. Med.* **79**, 137–158 (1943).
3. De Silva, R. T. *et al.* Improved growth and morphological plasticity of Haloferax volcanii. *Microbiology* **167**, (2021).
4. Stent, G. S. Prematurity and Uniqueness in Scientific Discovery. *Sci. Am.* (1972).
5. Mirsky, A. E. & Pollister, A. W. CHROMOSIN, A DESOXYRIBOSE NUCLEOPROTEIN COMPLEX OF THE CELL NUCLEUS. *J. Gen. Physiol.* **30**, 117–148 (1946).
6. Chargaff, E. Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia* **6**, 201–209 (1950).
7. Levene, P. A. & La Forge, F. B. On Chondrosamine. *Proc. Natl. Acad. Sci.* **1**, 190–191 (1915).
8. Hershey, A. D. & Chase, M. INDEPENDENT FUNCTIONS OF VIRAL PROTEIN AND NUCLEIC ACID IN GROWTH OF BACTERIOPHAGE. *J. Gen. Physiol.* **36**, 39–56 (1952).
9. Wyatt, H. V. How history has blended. *Nature* **249**, 803–805 (1974).
10. Watson, J. D. & Crick, F. H. C. General implications of the structure of deoxyribonucleic acid. **171**, 964–967 (1953).
11. Franklin, R. & Gosling, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature* **171**, 740–741 (1953).
12. Stokes, A. R. & WIlson, H. R. Molecular Structure of Deoxypentose Nucleic Acids. *Nature* **171**, (1953).
13. Pauling, L. & Corey, R. B. A Proposed Structure For The Nucleic Acids. *Proc. Natl. Acad. Sci.* **39**, 84–97 (1953).
14. Watson, J. D. & Crick, F. H. C. A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, 737–738 (1953).

15. Bloch, D. P. A possible mechanism for the replication of the helical structure of desoxyribonucleic acid. *Proc. Natl. Acad. Sci.* **41**, 1058–1064 (1955).

16. Meselson, M. & Stahl, F. W. THE REPLICATION OF DNA IN ESCHERICHIA COLI. **44**, 671–682 (1958).

17. Holmes, F. The DNA replication problem. *Cell* **23**, 117–120 (1998).

18. Lehman, I. R., Bessman, M. J., Simms, E. S. & Kornberg, A. Enzymatic Synthesis of Deoxyribonucleic Acid. PREPARATION OF SUBSTRATES AND PARTIAL PURIFICATION OF AN ENZYME FROM ESCHERICHIA COLZ. *J. Biol. Chem.* **233**, 163–170 (1958).

19. Bessman, M. J., Lehman, I. R., Simms, E. S. & Kornberg, A. Enzymatic Synthesis of Deoxyribonucleic Acid. *J. Biol. Chem.* **233**, 171–177 (1958).

20. Lieberman, I., Kornberg, A. & Simms, E. S. ENZYMATIC SYNTHESES OF PYRIMIDINE AND PURINE NUCLEOTIDES. II.1 OROTIDINE-5'-PHOSPHATE PYROPHOSPHORYLASE AND DECARBOXYLASE. *J. Am. Chem. Soc.* **76**, 2844–2845 (1954).

21. Cori, G. T. & Cori, C. F. CRYSTALLINE MUSCLE PHOSPHORYLASE. *J. Biol. Chem.* **151**, 57–63 (1943).

22. Brutlag, D. & Kornberg, A. Enzymatic Synthesis of Deoxyribonucleic Acid. *J. Biol. Chem.* **247**, 241–248 (1972).

23. Delarue, M., Poch, O., Tordo, N., Moras, D. & Argos, P. An attempt to unify the structure of polymerases. *Protein Eng. Des. Sel.* **3**, 461–467 (1990).

24. Joyce, C. M. & Steitz, T. A. FUNCTION AND STRUCTURE RELATIONSHIPS IN DNA POLYMERASES. *Annu. Rev. Biochem.* **63**, 777–822 (1994).

25. Bebenek, K. & Kunkel, T. A. Functions of DNA Polymerases. in *Advances in Protein Chemistry* vol. 69 137–165 (Elsevier, 2004).

26. Braithwaite, D. K. & Ito, J. Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.* **21**, 787–802 (1993).

27. Beese, L. S., Derbyshire, V. & Steitz, T. A. Structure of DNA Polymerase I Klenow Fragment Bound to Duplex DNA. *Science* **260**, 352–355 (1993).

28. Morin, J. A. *et al.* Active DNA unwinding dynamics during processive DNA replication. *Proc. Natl. Acad. Sci.* **109**, 8115–8120 (2012).

29. Bérut, A. *et al.* Experimental verification of Landauer's principle linking information and thermodynamics. *Nature* **483**, 187–189 (2012).

30. Otsuka, J. & Nozawa, Y. Self-reproducing System can Behave as Maxwell's Demon: Theoretical Illustration under Prebiotic Conditions. *J. Theor. Biol.* **194**, 205–221 (1998).

31. Forterre, P., Filée, J. & Myllykallio, H. Origin and Evolution of DNA and DNA Replication Machineries. in *The Genetic Code and the Origin of Life* 145–168 (Springer US, Boston, MA, 2004).

32. Rothwell, P. J. & Waksman, G. Structure and mechanism of DNA polymerases. in *Advances in Protein Chemistry* vol. 71 401–440 (Elsevier, 2005).

33. Steitz, T. A. DNA- and RNA-dependent DNA polymerases. *Curr. Opin. Struct. Biol.* **3**, 31–38 (1993).

34. Bębenek, A. & Ziuzia-Graczyk, I. Fidelity of DNA replication—a matter of proofreading. *Curr. Genet.* **64**, 985–996 (2018).

35. Ekundayo, B. & Bleichert, F. Origins of DNA replication. *PLOS Genet.* **15**, e1008320 (2019).

36. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).

37. Ryter, A., Hirota, Y. & Jacob, F. DNA-Membrane Complex and Nuclear Segregation in Bacteria. *Cold Spring Harb. Symp. Quant. Biol.* **33**, 669–676 (1968).

38. Morange, M. What history tells us XXXI. The replicon model: Between molecular biology and molecular cell biology. *J. Biosci.* **38**, 225–227 (2013).

39. Jacob, F., Brenner, S. & Cuzin, F. On the Regulation of DNA Replication in Bacteria. *Cold Spring Harb. Symp. Quant. Biol.* **28**, 329–348 (1963).

40. Huberman, J. A. & Riggs, A. D. On the mechanism of DNA replication in mammalian chromosomes. *J. Mol. Biol.* **32**, 327–341 (1968).

41. Novick, R. P. Plasmid Incompatibility. *Microbiol. Rev.* **51**, 381–395 (1987).

42. Kohiyama, M., Hiraga, S., Matic, I. & Radman, M. Bacterial Sex: Playing Voyeurs 50 Years Later. *Science* **301**, 802–803 (2003).

43. Yasuda, S. & Hirota, Y. Cloning and mapping of the replication origin of Escherichia coli. *Proc. Natl. Acad. Sci.* **74**, 5458–5462 (1977).

44. Mackiewicz, P. Where does bacterial replication start? Rules for predicting the oriC region. *Nucleic Acids Res.* **32**, 3781–3791 (2004).

45. Chakraborty, T., Yoshinaga, K., Lother, H. & Messer, W. Purification of the E.coli dnaA gene product. *EMBO J.* **1**, 1545–1549 (1982).

46. Fujita, M. Q., Hiroshi, Y. & Ogasawara, N. Structure of the dnaA region of Pseudomonas putida: conservation among three bacteria, Bacillus subtilis, Escherichia coli and P.putida. *Mol. Gen. Genet. MGG* **215**, 381–387 (1989).

47. Chan, C. S. & Tye, B. K. Autonomously replicating sequences in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci.* **77**, 6329–6333 (1980).

48. Broach, J. R. *et al.* Localization and Sequence Analysis of Yeast Origins of DNA Replication. *Cold Spring Harb. Symp. Quant. Biol.* **47**, 1165–1173 (1983).

49. Bell, S. P. & Stillman, B. AlP.dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357**, (1992).

50. Gavin, K. A., Hidaka, M. & Stillman, B. Conserved Initiator Proteins in Eukaryotes. *Science* **270**, 1667–1671 (1995).

51. Taylor, J. H. Increase in DNA replication sites in cells held at the beginning of S phase. *Chromosoma* **62**, 291–300 (1977).

52. Hamlin, J. L. *et al.* A revisionist replicon model for higher eukaryotic genomes. *J. Cell. Biochem.* **105**, 321–329 (2008).

53. diCenzo, G. C. & Finan, T. M. The Divided Bacterial Genome: Structure, Function, and Evolution. *Microbiol. Mol. Biol. Rev.* **81**, e00019-17 (2017).

54. Harrison, P. W., Lower, R. P. J., Kim, N. K. D. & Young, J. P. W. Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends Microbiol.* **18**, 141–148 (2010).

55. Krawiec, S. & Riley, M. Organization of the Bacterial Chromosome. *MICROBIOL REV* **54**, 502–539 (1990).

56. Suwanto, A. & Kaplan, S. Physical and Genetic Mapping of the Rhodobacter sphaeroides 2.4.1 Genome: Presence of Two Unique Circular Chromosomes. *J BACTERIOL* **171**, 5850–5859 (1989).

57. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci.* **74**, 5088–5090 (1977).

58. Wu, Z., Liu, J., Yang, H. & Xiang, H. DNA replication origins in archaea. *Front. Microbiol.* **5**, 1–7 (2014).

59. Myllykallio, H., Lopez, P., Lopez-Garcia, P. & Heilig, R. Bacterial Mode of Replication with Eukaryotic-Like Machinery in a Hyperthermophilic Archaeon. *Science* **288**, 2212–2215 (2000).

60. Matsunaga, F., Forterre, P., Ishino, Y. & Myllykallio, H. *In vivo* interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. *Proc. Natl. Acad. Sci.* **98**, 11152–11157 (2001).

61. Matsunaga, F., Norais, C., Forterre, P. & Myllykallio, H. Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin. *EMBO Rep.* **4**, 154–158 (2003).

62. Robinson, N. P. *et al.* Identification of Two Origins of Replication in the Single Chromosome of the Archaeon Sulfolobus solfataricus. *Cell* **116**, 25–38 (2004).

63. Lundgren, M., Andersson, A., Chen, L., Nilsson, P. & Bernander, R. Three replication origins in *Sulfolobus* species: Synchronous initiation of chromosome replication and asynchronous termination. *Proc. Natl. Acad. Sci.* **101**, 7046–7051 (2004).

64. She, Q., Peng, X., Zillig, W. & Garrett, R. A. Gene capture in archaeal chromosomes. *Nature* **409**, 478–478 (2001).

65. Robinson, N. P. & Bell, S. D. Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proc. Natl. Acad. Sci.* **104**, 5806–5811 (2007).

66. Zhang, R. & Zhang, C.-T. Multiple replication origins of the archaeon Halobacterium species NRC-1. *Biochem. Biophys. Res. Commun.* **302**, 728–734 (2003).

67. 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**, 478 (2012).

68. Hawkins, M. S. DNA Replication Origins in Haloferax volcanii. (University of Nottingham, 2009).

69. Dyall-Smith, M. L. *et al.* Haloquadratum walsbyi : Limited Diversity in a Global Pond. *PLoS ONE* **6**, e20968 (2011).

70. Rozgaja, T. A. *et al.* Two oppositely oriented arrays of low-affinity recognition sites in oriC guide progressive binding of DnaA during Escherichia coli pre-RC assembly: DnaA-oriC interaction at arrayed sites. *Mol. Microbiol.* **82**, 475–488 (2011).

71. Matsui, M., Oka, A., Takanami, M., Yasuda, S. & Hirota, Y. Sites of dnaA protein-binding in the replication origin of the Escherichia coli K-12 chromosome. *J. Mol. Biol.* **184**, 529–533 (1985).

72. Shimizu, M. *et al.* Near-atomic structural model for bacterial DNA replication initiation complex and its functional insights. *Proc. Natl. Acad. Sci.* **113**, (2016).

73. Erzberger, J. P. The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation. *EMBO J.* **21**, 4763–4773 (2002).

74. Erzberger, J. P. & Berger, J. M. Evolutionary Relationships and Structural Mechanisms of AAA+ Proteins. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 93–114 (2006).

75. Kowalski, D. & Eddy, M. J. The DNA unwinding element: a novel, cis-acting component that facilitates opening of the Escherichia coli replication origin. *EMBO J.* **8**, 4335–4344 (1989).

76. Wegrzyn, K. & Konieczny, I. Toward an understanding of the DNA replication initiation in bacteria. *Front. Microbiol.* **14**, 1328842 (2024).

77. Robinson, N. P. & Bell, S. D. Origins of DNA replication in the three domains of life: Origins of DNA replication. *FEBS J.* **272**, 3757–3766 (2005).

78. Schwob, E. Flexibility and governance in eukaryotic DNA replication. *Curr. Opin. Microbiol.* **7**, 680–690 (2004).

79. Segurado, M., Gómez, M. & Antequera, F. Increased Recombination Intermediates and Homologous Integration Hot Spots at DNA Replication Origins. *Mol. Cell* **10**, 907–916 (2002).

80. Gilbert, D. M. *et al.* Space and Time in the Nucleus: Developmental Control of Replication Timing and Chromosome Architecture. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 143–153 (2010).

81. Labib, K. How do Cdc7 and cyclin-dependent kinases trigger the initiation of chromosome replication in eukaryotic cells? *Genes Dev.* **24**, 1208–1219 (2010).

82. O'Donnell, M., Langston, L. & Stillman, B. Principles and Concepts of DNA Replication in Bacteria, Archaea, and Eukarya. *Cold Spring Harb. Perspect. Biol.* **5**, a010108–a010108 (2013).

83. Hartwell, L. H. Sequential function of gene products relative to DNA synthesis in the yeast cell cycle. *J. Mol. Biol.* **104**, 803–817 (1976).

84. Hofmann, J. F. X. & Beach, D. cdt 1 is an essential target of the Cdc 1 O/Sct 1 transcription factor: requirement for DNA replication and inhibition of mitosis. *EMBO J.* **13**, 425–434 (1994).

85. Nishitani, H., Lygerou, Z., Nishimoto, T. & Nurse, P. The Cdt1 protein is required to license DNA for replication in Fission yeast. *Nature* **404**, 625–628 (2000).

86. Schmidt, J. M. *et al.* A mechanism of origin licensing control through autoinhibition of S. cerevisiae ORC·DNA·Cdc6. *Nat. Commun.* **13**, 1059 (2022).

87. Randell, J. C. W., Bowers, J. L., Rodríguez, H. K. & Bell, S. P. Sequential ATP Hydrolysis by Cdc6 and ORC Directs Loading of the Mcm2-7 Helicase. *Mol. Cell* **21**, 29–39 (2006).

88. Costa, A. & Diffley, J. F. X. The Initiation of Eukaryotic DNA Replication. *Annu. Rev. Biochem.* **91**, 107–131 (2022).

89. Ilves, I., Petojevic, T., Pesavento, J. J. & Botchan, M. R. Activation of the MCM2-7 Helicase by Association with Cdc45 and GINS Proteins. *Mol. Cell* **37**, 247–258 (2010).

90. Zegerman, P. & Diffley, J. F. X. Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature* **445**, 281–285 (2007).

91. Greci, M. D. & Bell, S. D. Archaeal DNA Replication. *Annu. Rev. Microbiol.* **74**, 65–80 (2020).

92. Lopez, P., Philippe, H., Myllykallio, H. & Forterre, P. Identification of putative chromosomal origins of replication in Archaea. *Mol. Microbiol.* **32**, 883–886 (1999).

93. Liu, J. *et al.* Structure and Function of Cdc6/Cdc18: Implications for Origin Recognition and Checkpoint Control. *Mol. Cell* (2000).

94. Matsunaga, F. *et al.* Genomewide and biochemical analyses of DNA-binding activity of Cdc6/Orc1 and Mcm proteins in Pyrococcus sp. *Nucleic Acids Res.* **35**, 3214–3222 (2007).

95. Matsunaga, F. *et al.* Localized melting of duplex DNA by Cdc6/Orc1 at the DNA replication origin in the hyperthermophilic archaeon Pyrococcus furiosus. *Extremophiles* **14**, 21–31 (2010).

96. 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 (2007).

97. Grainge, I. Biochemical analysis of components of the pre-replication complex of Archaeoglobus fulgidus. *Nucleic Acids Res.* **31**, 4888–4898 (2003).

98. Samson, R. Y. *et al.* Specificity and Function of Archaeal DNA Replication Initiator Proteins. *Cell Rep.* **3**, 485–496 (2013).

99. Dueber, E. L. C., Corn, J. E., Bell, S. D. & Berger, J. M. Replication Origin Recognition and Deformation by a Heterodimeric Archaeal Orc1 Complex. *Science* **317**, 1210–1213 (2007).

100. Grainge, I. *et al.* Biochemical Analysis of a DNA Replication Origin in the Archaeon Aeropyrum pernix. *J. Mol. Biol.* **363**, 355–369 (2006).

101. Ausiannikava, D. & Allers, T. Diversity of DNA Replication in the Archaea. *Genes* **8**, 56 (2017).

102. Akita, M. *et al.* Cdc6/Orc1 from *Pyrococcus furiosus* may act as the origin recognition protein and Mcm helicase recruiter. *Genes Cells* (2010) doi:10.1111/j.1365-2443.2010.01402.x.

103. Kelman, L. M., O'Dell, W. B. & Kelman, Z. Unwinding 20 Years of the Archaeal Minichromosome Maintenance Helicase. *J. Bacteriol.* **202**, (2020).

104. Kasiviswanathan, R., Shin, J.-H. & Kelman, Z. Interactions between the archaeal Cdc6 and MCM proteins modulate their biochemical properties. *Nucleic Acids Res.* **33**, 4940–4950 (2005).

105. Samson, R. Y., Abeyrathne, P. D. & Bell, S. D. Mechanism of Archaeal MCM Helicase Recruitment to DNA Replication Origins. *Mol. Cell* **61**, 287–296 (2016).

106. Samson, R. Y. & Bell, S. D. MCM Loading—An Open-and-Shut Case? *Mol. Cell* **50**, 457–458 (2013).

107. Mohammed Khalid, A. A., Parisse, P., Medagli, B., Onesti, S. & Casalis, L. Atomic Force Microscopy Investigation of the Interactions between the MCM Helicase and DNA. *Materials* **14**, 687 (2021).

108. Kelman, Z., Lee, J.-K. & Hurwitz, J. The single minichromosome maintenance protein of *Methanobacterium thermoautotrophicum* ΔH contains DNA helicase activity. *Proc. Natl. Acad. Sci.* **96**, 14783–14788 (1999).

109. Haugland, G. T., Shin, J.-H., Birkeland, N.-K. & Kelman, Z. Stimulation of MCM helicase activity by a Cdc6 protein in the archaeon Thermoplasma acidophilum. *Nucleic Acids Res.* **34**, 6337–6344 (2006).

110. Pérez-Arnaiz, P., Dattani, A., Smith, V. & Allers, T. *Haloferax volcanii* —a model archaeon for studying DNA replication and repair. *Open Biol.* **10**, 200293 (2020).

111. Cairns, J. The bacterial chromosome and its manner of replication as seen by autoradiography. *J. Mol. Biol.* **6**, 208–213 (1963).

112. Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K. & Sugino, A. Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc. Natl. Acad. Sci.* **59**, 598–605 (1968).

113. Kelman, L. M. & Kelman, Z. Archaea: an archetype for replication initiation studies?: Initiation of archeal DNA replication. *Mol. Microbiol.* **48**, 605–615 (2003).

114. Georgescu, R., Langston, L. & O'Donnell, M. A proposal: Evolution of PCNA's role as a marker of newly replicated DNA. *DNA Repair* **29**, 4–15 (2015).

115. Leipe, D. D., Aravind, L. & Koonin, E. V. Did DNA replication evolve twice independently? *Nucleic Acids Res.* **27**, 3389–3401 (1999).

116. Forterre, P. Why Are There So Many Diverse Replication Machineries? *J. Mol. Biol.* **425**, 4714–4726 (2013).

117. Lujan, S. A., Williams, J. S. & Kunkel, T. A. DNA Polymerases Divide the Labor of Genome Replication. *Trends Cell Biol.* **26**, 640–654 (2016).

118. Iyer, L. M. Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucleic Acids Res.* **33**, 3875–3896 (2005).

119. Kreuzer, K. N. Recombination-dependent DNA replication in phage T4. *Trends Biochem. Sci.* **25**, 165–173 (2000).

120. Mosig, G., Gewin, J., Luder, A., Colowick, N. & Vo, D. Two recombination-dependent DNA replication pathways of bacteriophage T4, and their roles in mutagenesis and horizontal gene transfer. *Proc. Natl. Acad. Sci.* **98**, 8306–8311 (2001).

121. Belanger, K. G. & Kreuzer, K. N. Bacteriophage T4 Initiates Bidirectional DNA Replication through a Two-Step Process. *Mol. Cell* **2**, 693–701 (1998).

122. Syeda, A. H., Hawkins, M. & McGlynn, P. Recombination and Replication. *Cold Spring Harb. Perspect. Biol.* **6**, a016550–a016550 (2014).

123. Yang, S., VanLoock, M. S., Yu, X. & Egelman, E. H. Comparison of bacteriophage T4 UvsX and human Rad51 filaments suggests that RecA-like polymers may have evolved independently11Edited by M. Belfort. *J. Mol. Biol.* **312**, 999–1009 (2001).

124. Liu, J. & Marians, K. J. PriA-directed Assembly of a Primosome on D Loop DNA. *J. Biol. Chem.* **274**, 25033–25041 (1999).

125. Jones, J. M. & Nakai, H. The φX174-type primosome promotes replisome assembly at the site of recombination in bacteriophage Mu transposition. *EMBO J.* **16**, 6886–6895 (1997).

126. Epum, E. A. & Haber, J. E. DNA replication: the recombination connection. *Trends Cell Biol.* **32**, 45–57 (2022).

127. Anand, R. P., Lovett, S. T. & Haber, J. E. Break-Induced DNA Replication. *Cold Spring Harb. Perspect. Biol.* **5**, a010397–a010397 (2013).

128. He, L. *et al.* Interaction of human HelQ with DNA polymerase delta halts DNA synthesis and stimulates DNA single-strand annealing. *Nucleic Acids Res.* **51**, 1740–1749 (2023).

129. Malkova, A. & Ira, G. Break-induced replication: functions and molecular mechanism. *Curr. Opin. Genet. Dev.* **23**, 271–279 (2013).

130. Ravoitytė, B. & Wellinger, R. Non-Canonical Replication Initiation: You're Fired! *Genes* **8**, 54 (2017).

131. Kogoma, T. & Lark, K. G. Characterization of the replication of Escherichia coli DNA in the absence of protein synthesis: Stable DNA replication. *J. Mol. Biol.* **94**, 243–256 (1975).

132. Masai, H., Asai, T., Kubota, Y., Arai, K. & Kogoma, T. Escherichia coil PriA protein is essential for inducible and constitutive stable DNA replication. *EMBO Rep.* **13**, 5338–5345 (1994).

133. Crossley, M. P., Bocek, M. & Cimprich, K. A. R-Loops as Cellular Regulators and Genomic Threats. *Mol. Cell* **73**, 398–411 (2019).

134. Masai, H. Replicon hypothesis revisited. *Biochem. Biophys. Res. Commun.* **633**, 77–80 (2022).

135. 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 (2013).

136. Kogoma, T. Recombination by Replication. *Cell* **85**, 625–627 (1996).

137. Mc Teer, L. *et al.* Cooperation between two modes for DNA replication initiation in the archaeon *Thermococcus barophilus*. *mBio* e03200-23 (2024) doi:10.1128/mbio.03200-23.

138. Rosenshine, I., Tchelet, R. & Mevarech, M. The Mechanism of DNA Transfer in the Mating System of an Archaebacterium. *Science* **245**, 1387–1389 (1989).

139. Gehring, A. M. *et al.* Genome Replication in Thermococcus kodakarensis Independent of Cdc6 and an Origin of Replication. *Front. Microbiol.* **8**, 2084 (2017).

140. Yang, H. *et al.* Activation of a dormant replication origin is essential for Haloferax mediterranei lacking the primary origins. *Nat. Commun.* **6**, 8321 (2015).

141. Kuhn, T. S. *The Structure of Scientific Revolutions*. (University of Chicago Press, Chicago, 1970).

142. Stanier, R. Y. & Niel, C. B. The concept of a bacterium. *Arch. F□ r Mikrobiol.* **42**, 17–35 (1962).

143. Whitman, W. B. The Modern Concept of the Procaryote. *J. Bacteriol.* **191**, 2000–2005 (2009).

144. Brock, T. D. The Value of Basic Research: Discovery of *Thermus aquaticus* and Other Extreme Thermophiles. *Genetics* **146**, 1207–1210 (1997).

145. Van De Vossenberg, J. L. C. M., Driessen, A. J. M. & Konings, W. N. The essence of being extremophilic: the role of the unique archaeal membrane lipids. *Extremophiles* **2**, 163–170 (1998).

146. Fox, G. E. & Woese, C. R. Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proc Nat Acad Sci USA* **74**, 4537–4541 (1977).

147. Lanyi, J. K. Salt-Dependent Properties of Proteins from Extremely Halophilic Bacteria. *BACTERIOL REV* 272–290 (1974).

148. Darland, G., Brock, T. D., Samsonoff, W. & Conti, S. F. A Thermophilic, Acidophilic Mycoplasma Isolated from a Coal Refuse Pile. *Science* **170**, 1416–1418 (1970).

149. Pace, N. R., Sapp, J. & Goldenfeld, N. Phylogeny and beyond: Scientific, historical, and conceptual significance of the first tree of life. *Proc. Natl. Acad. Sci.* **109**, 1011–1018 (2012).

150. Sanger, F., Brownlee, G. G. & Barrell, B. G. A two-dimensional fractionation procedure for radioactive nucleotides. (1965).

151. Albers, S.-V., Forterre, P., Prangishvili, D. & Schleper, C. The legacy of Carl Woese and Wolfram Zillig: from phylogeny to landmark discoveries. *Nat. Rev. Microbiol.* **11**, 713–719 (2013).

152. Zillig, W. *et al.* The phylogenetic relations of DNA-dependent RNA polymerases of archaebacteria, eukaryotes, and eubacteria. *Can. J. Microbiol.* **35**, 73–80 (1989).

153. Zillig, W. Comparative biochemistry of Archaea and Bacteria. *Curr. Opin. Genet. Dev.* **1**, 544–551 (1991).

154. Woese, C. R. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Nat Acad Sci USA* **87**, 4576–4579 (1990).

155. Spang, A., Caceres, E. F. & Ettema, T. J. G. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* **357**, eaaf3883 (2017).

156. Cavicchioli, R. Archaea — timeline of the third domain. *Nat. Rev. Microbiol.* **9**, 51–61 (2011).

157. Woese, C. R. A New Biology for a New Century. *Microbiol. Mol. Biol. Rev.* **68**, 173–186 (2004).

158. Kim, J. Making Sense of Emergence. *Philos. Stud.* **95**, 3–36 (1999).

159. Delker, R. K. & Mann, R. S. From Reductionism to Holism: Toward a More Complete View of Development Through Genome Engineering. in *Precision Medicine, CRISPR, and Genome Engineering* (ed. Tsang, S. H.) vol. 1016 45–74 (Springer International Publishing, Cham, 2017).

160. Kitano, H. Looking beyond the details: a rise in system-oriented approaches in genetics and molecular biology. *Curr. Genet.* **41**, 1–10 (2002).

161. Strange, K. The end of "naïve reductionism": rise of systems biology or renaissance of physiology? *Am. J. Physiol.-Cell Physiol.* **288**, C968–C974 (2005).

162. Woese, C. R. The archaeal concept and the world it lives in: a retrospective. *Photosynth. Res.* **80**, 363–372 (2004).

163. Satam, H. *et al.* Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology* **12**, 997 (2023).

164. Trieselmann, B. A. & Charlebois, R. L. Transcriptionally active regions in the genome of the archaebacterium Haloferax volcanii. *J. Bacteriol.* **174**, 30–34 (1992).

165. 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. *Appl. Environ. Microbiol.* **70**, 943–953 (2004).

166. Lestini, R., Duan, Z. & Allers, T. The archaeal Xpf/Mus81/FANCM homolog Hef and the Holliday junction resolvase Hjc define alternative pathways that are essential for cell viability in Haloferax volcanii. *DNA Repair* **9**, 994–1002 (2010).

167. Cline, S. W. & Doolittle, W. F. Efficient transfection of the archaebacterium Halobacterium halobium. *J. Bacteriol.* **169**, 1341–1344 (1987).

168. Cline, S. W., Lam, W. L., Charlebois, R. L., Schalkwyk, L. C. & Doolittle, W. F. Transformation methods for halophilic archaebacteria. *Can. J. Microbiol.* **35**, 148–152 (1989).

169. 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. *J. Bacteriol.* **185**, 772–778 (2003).

170. 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. *J. Biol. Chem.* **254**, 4876–4883 (1979).

171. Johnsen, U. *et al.* d-Xylose Degradation Pathway in the Halophilic Archaeon Haloferax volcanii. *J. Biol. Chem.* **284**, 27290–27303 (2009).

172. Rados, T., Andre, K., Cerletti, M. & Bisson, A. A sweet new set of inducible and constitutive promoters in Haloferax volcanii. *Front. Microbiol.* **14**, 1204876 (2023).

173. Niessen, N. & Soppa, J. Regulated Iron Siderophore Production of the Halophilic Archaeon Haloferax volcanii. *Biomolecules* **10**, 1072 (2020).

Appendices

7.4 A Brief Exploration into the Rules of Life

"Algebra applies to the clouds, the radiance of the star benefits the rose – no thinker would dare to say that the perfume of the hawthorn is useless to the constellations. Who could ever calculate the path of a molecule? How do we know that the creations of worlds are not determined by falling grains of sand?"

– Victor Hugo, *Les Misérables*

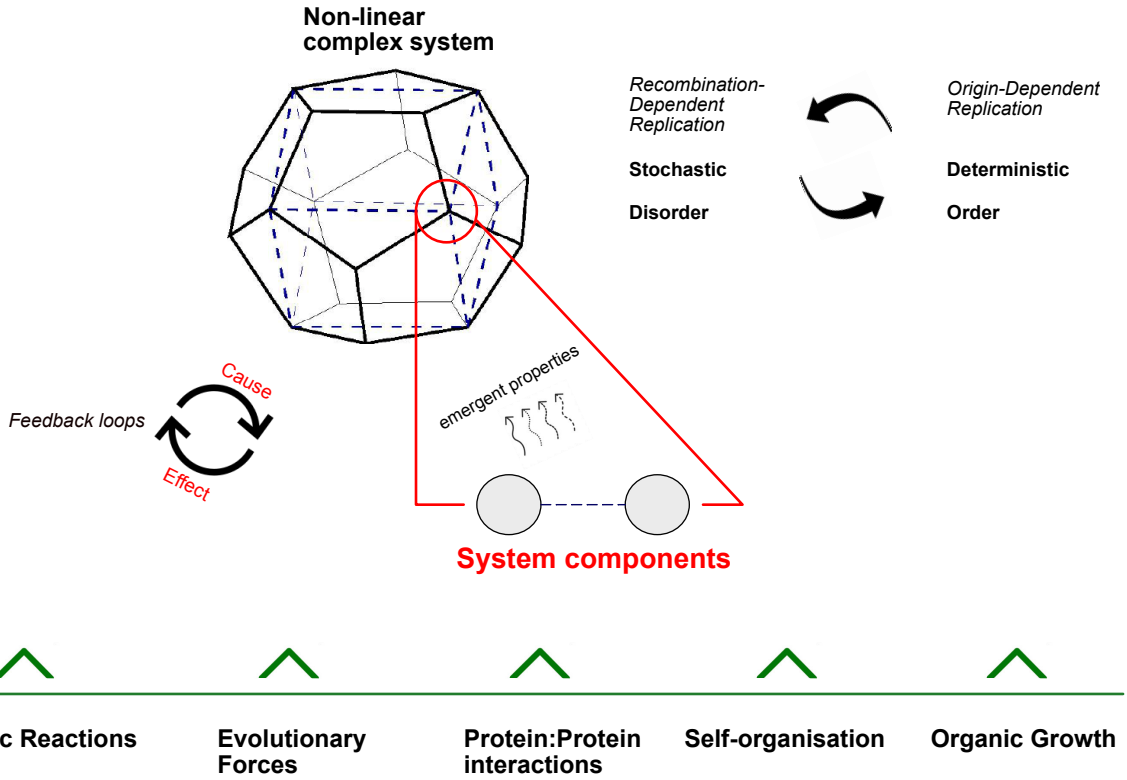


Figure 30. Conceptual model showing the relations between the components of a living organism (i.e., non-linear complex system) with reference to replication processes of *H.volcanii*.

A living organism can be classed as a dissipative structure – that is, a system whose conditions are far from thermodynamic equilibrium, and is therefore shaped through its **dynamic** interactions with the external environment.

A biological entity's full scope of qualities can only be truly appreciated through the complexity theory approach.
If the causally linked components form a network within a system – a relational structure – then it is the effects these components exert, and the operations within that structure rather than nature of the isolated components themselves which should direct our scientific inquiry.

7.5 Project Development Roadmap

7.5.1 Project strand 1:

Main goal: construction of *rnh* gene knockout strains (*rnhABCE* quadruple deletions) in WT and main chromosome origin-deleted backgrounds

RNaseH gene deletions	Strains	Confirmation	Outstanding experiments	Comments
	H5617/H5618 H5619/H5620 H5621/H5622 H5623/H5624	Colony hybridisation	**Short-term:** a. Final confirmation of genotype through Southern blotting b. Construction of plasmids from the gene-deleted strains c. Transformation with Δori-pHV3 and/or Δorc6 plasmids (p1801/p1802/p1803 **Long-term:** Investigate whether R loops prime replication directly by putting RadA recombinase under a tryptophan inducible promoter and determining the frequency and location of R loops through genomic techniques such as DRIP-seq	

7.5.2 Project strand 2:

Main goal: construction of Δori-pHV3 deleted strains through Hel308 deletion or inactivating point mutations

Hel308 mutation	Strains	Confirmation	Future Experiments	Comments

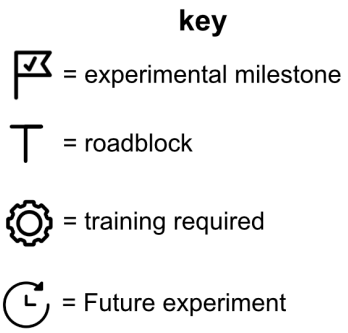
	H5638 H5665 H5666 H5667	Colony hybridisation	(a) Final confirmation of origin deletion through Southern blotting and assign new strain numbers (b) Confirmation of episomal state of pHV3 through pulsed gel electrophoresis

7.5.3 Project strand 3:

Main goal: construction of ori-pHV3 deleted strains where the *adh2* gene is under a tryptophan inducible promoter (i.e., ptnaA)

	Strains	**Confirmation**	**Future Experiments**	
		Colony PCR	Final confirmation of genotype through Southern blotting	
Δ*adh2*		Colony hybridisation	N/A	Contingency plan if strain construction with tryptophan-inducible promoter does not work

Contingency plan: If Δori-PHV3 is proven to be unsuccessful or has integrated into the main chromosome through either of the 3 project strands, then a strain combined for all the above mutations will be constructed

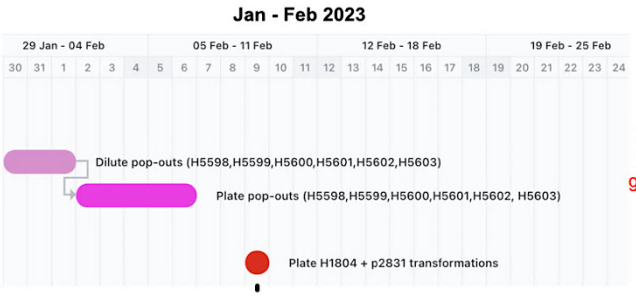


Gantt chart links (ClickUp™ project manager) -> for annotated schematics *see below*

1) https://sharing.clickup.com/26461449/g/h/6-900500338739-7/3b248f734b5a787
(ΔrnhA strain construction)

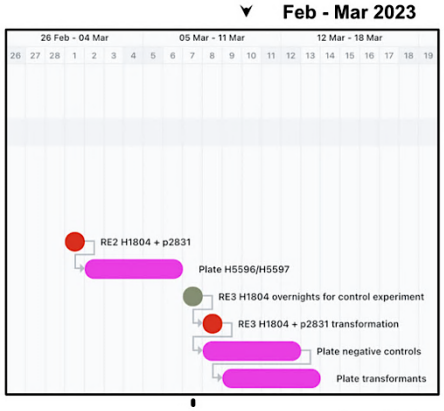
2) https://sharing.clickup.com/26461449/g/h/6-900500380667-7/4ad1ba4932cebb8
(ΔrnhC plasmid construction)

3) https://sharing.clickup.com/26461449/g/h/6-900500433535-7/023bfa84475656d
(pHV3 origin deletion in Hel308 mutation background strains)

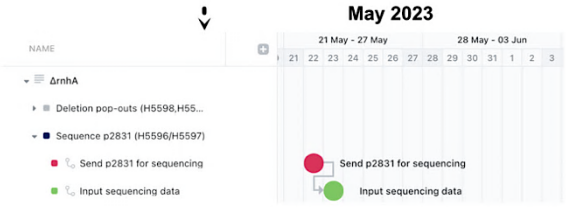


gene-deleted colonies

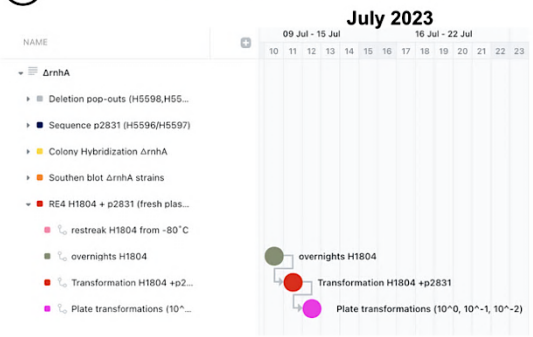
proceed to confirm deletions



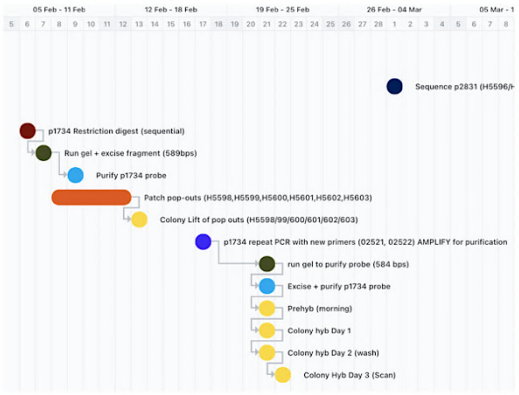
transformation
did not work for one of the strains



plasmid confirmed through sequencing - repeat transformation in July

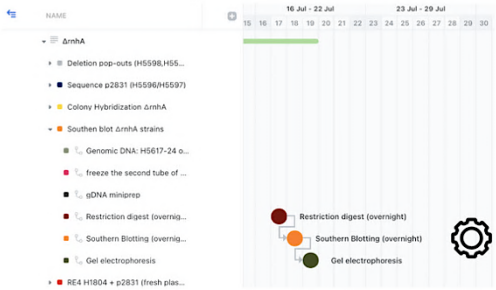


Colony hybridisation: Δ*rnhA* confirmed



Short term aims :

- Confirm deletions through Southern Blotting (scheduled for mid-July)



Southern blot :
local training required

Medium term aims :

- Transform confirmed strains with Δori-pHV3, Δorc6 plasmids (p1801/p1802/p1803)
- **Confirm deletions**: colony hybridisation followed by southern blotting

Contingency: attemp transformations starting with quadruple *rnh* deletions and narrow down to triple

Long term aims :

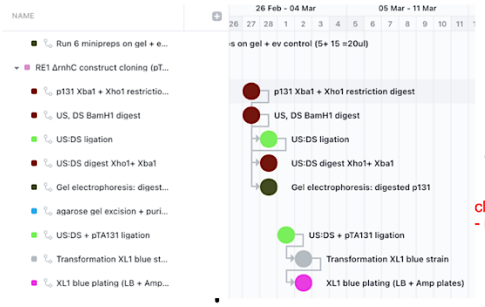
- DRIP-seq to determine frequency and location of R loops
- Hi-C to correlate the frequency of R loops with integration into the main chromosome

ΔrnhC plasmid construction through cloning

Feb 2023 - Plasmid construction: cloning attempt 1/3

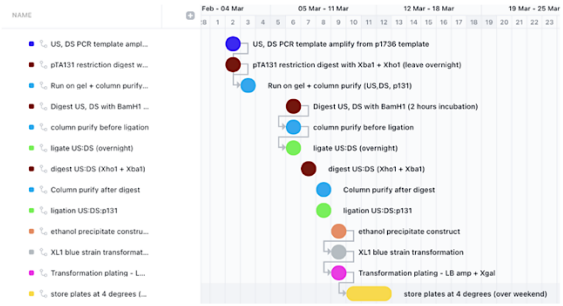
cloning unsuccessful: incorrect insert

Feb 2023 - Plasmid construction: cloning attempt 2/3

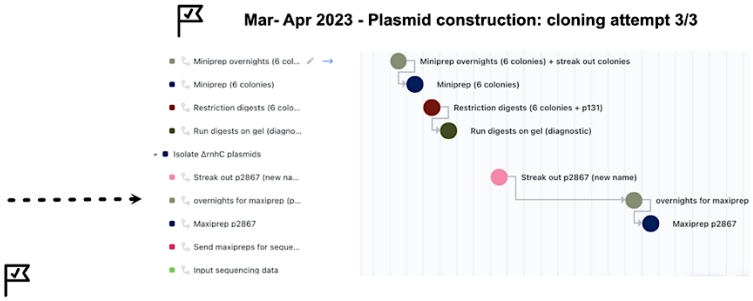


cloning unsuccessful at electroporation stage
- remove remaining salt through ethanol precipitation

Mar 2023 - Plasmid construction: cloning attempt 3/3



Mar- Apr 2023 - Plasmid construction: cloning attempt 3/3



Successful E.coli transformation