ON THE REPLICATION ORIGIN SPECIFICITY OF YEAST SPECIES

Rebecca Martin

A dissertation
submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington
2021

Reading Committee:
Bonita J. Brewer, chair
M. K. Raghuraman
Maitreya Dunham
Suzanne Hoppins

Program Authorized to Offer Degree:
Molecular and Cellular Biology

University of Washington

Abstract

On the Replication Origin Specificity of Budding Yeast Species

Rebecca Marie Martin

Chair of the Supervisory Committee:
Bonita J. Brewer
Department of Genome Sciences

DNA replication is an essential process that is highly conserved in form and function throughout eukaryotes. However, budding yeast differ from higher eukaryotes in that their origins of replication are defined by sequence. This unique feature has long been a tool in the study of replication, but what drives origin sequence-specificity remained unclear until cryo-EM structures of the *Saccharomyces cerevisiae* Origin Recognition Complex (ORC) revealed base-specific contacts between origin DNA and ORC subunits 1, 2, and 4. Orc4 makes the most substantial contact through a long unstructured loop capped by an alpha helix termed the insertion helix (IH). I tested the hypothesis that the Orc4-IH drives sequence specificity of budding yeast origins by replacing the Orc4-IH in *S. cerevisiae* with the corresponding sequence in *Lachancea waltii* and by deleting the alpha helix of the Orc4-IH. I found that the chimeric Orc4-IH enabled recognition of *L. waltii* specific plasmids while the Orc4-IH deletion reduced plasmid retention. Both strains exhibited altered genome-wide origin usage and pre-replication complex loading. In addition, I made a complementary chimeric mutation in Orc1 and tested its impact on origin sequence specificity alone and in combination with the chimeric Orc4-IH. I found that on its own, the Orc1 chimera is unable to recognize *L. waltii* plasmids but does show altered origin usage. However,

when in conjunction with the Orc4-IH chimera, *L. waltii* plasmid retention was even greater, as was *S. cerevisiae* plasmid retention and origin function. It appears that while Orc4 alone is sufficient to alter origin specificity, complementary changes to Orc1 may stabilize overall ORC binding.

In addition to this work, I also helped obtain evidence for intermediates of the origin dependent inverted repeat amplification (ODIRA) model in a yeast strain engineered to select for interchromosomal amplification events. We found that genetic alterations consistent with ODIRA did occur and that induction of double stranded breaks did not increase these events, further suggesting that these events were caused by a replication error rather than through repair of double-stranded DNA breaks.

Acknowledgements

I must first thank my incredible mentors, Bonny and Raghu, for giving me such a supportive environment in which to learn. You gave me a place to explore, fail, and succeed safely. I have never been afraid to approach you with whatever I was struggling with or whatever novel idea I had the night before. You two helped reignite my love of science, and I am eternally grateful for your mentorship.

Likewise, thank you to Liz and Gina for being the best labmates I could have asked for. I appreciate every piece of advice, every whiteboard musing, every Thai Tom lunch, and even every Tuesday spent in lab with both of you.

To my MCB family, thank you to Rich Gardner, Katie Peichel, and Maia Lowe for giving me the resources and support I needed to find a suitable scientific home at UW. To my cohort, thank you for all the wonderful game nights, rotation and thesis commiserations, and for all the Gradsgivings spent around my table. MCB 2014 is an amazing science family.

Thank you to my committee, Suzanne Hoppins, Rich Gardner, Alan Herr, and Maitreya Dunham for committee meetings that I looked forward to every year. I deeply appreciate your enthusiasm for my project and ideas.

To the people who encouraged my interest in science throughout high school and college, Pam Bell and Steve Heidemann, thank you for encouraging me to continue to pursue my education and convincing me that I was qualified for the opportunities I wished to undertake.

And finally to my friends, family, and partner, thank you for unconditional love and support in every imaginable way. Christopher, thank you for your consistent and continual help despite ambiguous deadlines and moving goals. I would like to add a special thanks for your Wi-Fi hotspot that allowed me to finish this thesis. To my parents, particularly my mother Connie, thank you for always encouraging me regardless of the endeavor – and for enthusiastically listening to me explain my work diagrammed on the back of napkins.

Table of Contents

List of Figures

List of Tables

Chapter I. Introduction

DNA replication is an essential process required before every mitotic cell division. The objective of each round of DNA replication is to fully and faithfully duplicate an organism's entire genome, ensuring that an identical copy is passed on to daughter cells. The importance of this universal process is underscored by the multitude of redundancies built into the replication system and the functional conservation of that system across taxa. Despite many millions of years of divergence, organisms as dissimilar as humans and yeast share functional and even structural conservation of DNA replication processes and machinery.

All eukaryotes, with large genomes spanning multiple linear chromosomes, must initiate replication at many sites throughout the genome. These sites are called origins of replication. Origins are under tight spatiotemporal control as failure to replicate even a portion of the genome once *and only once* can have deleterious or lethal consequences for the daughter cell inheriting the imperfectly replicated DNA.[1]

While eukaryotic origins share many features and requirements, budding yeast origins differ from metazoans in a small but significant way: the sites where replication starts are DNA sequence-specific.[2] This key difference has been exploited in past replication research to enable probing of specific sites of replication,[3,4] but these interrogations were limited to small-scale changes in origin distribution and location. The exact mechanisms of origin sequence specificity in yeast remained elusive until recent cryo-EM structures revealed unique features of the budding yeast Origin Recognition Complex (ORC) that contacts DNA in a potentially sequence-dependent manner.[5,6] The most prevalent contact between DNA and ORC is a region of the *S. cerevisiae* Orc4 winged helix domain, where a long unstructured loop capped by an alpha helix (termed the insertion helix or Orc4-IH) extends into the DNA binding channel and makes direct contact with origin DNA. The implications of this contact are discussed more in section 1.4.3 of this chapter. My primary objective was to determine if the Orc4-IH did drive sequence specificity of yeast origins and to examine what the consequences of global origin alterations (if achieved) would be.

1.1 Features of eukaryotic origins

To replicate every chromosome completely, origins must be dispersed throughout the genome, but this distribution is not uniform. Origin density varies with the genomic landscape. Origin location is highly conserved within unicellular species and within cell types of multicellular species,[7] even though most multicellular organisms lack origin sequence specificity. However, to what extent origin location genome-wide matters for cell/organismal viability remains unclear, as the tools to globally alter origin location have been lacking. What is clear is that DNA context helps determine where origins are located and which origins are primarily utilized during replication.[8] While many features interact to determine the prime locations of origins, two fundamental DNA features characterize most eukaryotic origins, regardless of sequence specificity: open chromatin and minimal risk of conflict with transcription.[9,10]

One of the main requirements for replication origins is DNA accessibility. The first step in replication is origin licensing, which requires the binding of the Origin Recognition Complex (ORC) to origin DNA, followed by loading head-to-head MCM2-7 hexamers, the replicative helicases.[11–13] ORC has a DNA footprint of around 30 bp[14] while the footprint of the MCM hexamers is approximately 62 bp.[11] Thus, origin DNA must be accessible enough to load complexes with a combined footprint of ~100bp. While research has shown that nucleosomes shift to accommodate these complexes (or that these complexes shift nucleosomes),[15,16] DNA that is already accessible tends to be enriched for replication origins. In other words, euchromatin has more replication origins than heterochromatin. Therefore, in metazoans where euchromatic and heterochromatic regions are defined by cell type, origin enrichment is also cell-type specific.

In many species, from humans to *Arabidopsis* to fission yeast, replication origins are prevalent in A/T-rich DNA.[17,18] This A/T bias often helps fulfill both requirements for origin location. A/T regions tend to be intergenic, meaning that origins in these locations are typically excluded from gene bodies and transcribed regions. Moreover, A/T stretches form looser helical structures than G/C-rich regions and melt more easily—a requirement for replication initiation. Likewise, long A/T stretches also serve

as nucleosome exclusion signals, potentially by their tendency to preclude tight winding around nucleosomes.[19,20]

However, in metazoans, replication origins are also enriched at CpG islands near transcription start sites of highly transcribed genes.[21,22] While this association may seem counterintuitive at first, initiating replication at transcription start sites (TSS) also targets euchromatin in organisms with varying cell types where different regions of the genome are accessible in a cell type-dependent manner. Moreover, enrichment at TSSs is preferable to replication within gene bodies where collisions between replication and transcription machinery may result in DNA damage and delayed replication.[23,24]

1.2 Budding yeast origins of replication

Yeast origins are called autonomously replicating sequences (ARS) and were named for their ability to enable plasmid maintenance when introduced into yeast.[2] According to OriDB, there are 628 confirmed and likely origins, but only about 400 of these origins fire in a given cell cycle.[25] Every ARS contains a conserved 11-17 bp motif known as the ARS Consensus Sequence (ACS), which is necessary, but not sufficient for origin activity.[26] The ACS, bound by ORC, is the "A element" which is accompanied by a varying number of "B elements" that differ in sequence between ARSs (Figure 1.1).[12,27] In *ARS1*, the most studied ARS, the B1 element has been shown to have a short TTT stretch amidst a weakly defined motif that is generally T-rich and may contribute in part to ORC binding.[5,28] The B2 element has been shown to overlap with the MCM2-7 helicase position and thus was thought to be the MCM binding site.[29] However, Wilmes and Bell (2002) observed that the B2 element resembles a degenerate inverse of the ACS,[30] a finding that complements work which shows that bidirectional MCM helicase

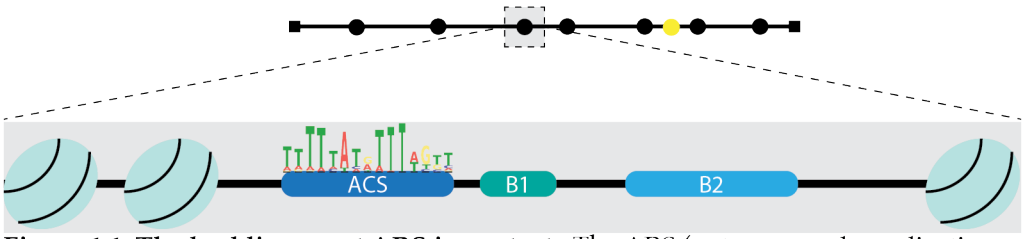


Figure 1.1: The budding yeast ARS in context. The ARS (autonomously replicating sequence) is composed of the 11-17 bp ACS (ARS consensus sequence) followed by the B1 element (typically a 3-5 bp A/T rich motif), and the B2 element that resembles a degenerate ACS on the opposite strand from the primary ACS.

loading requires two ORCs bound in opposite orientation.[31] Together, these two findings suggest that B2 is the binding site of a second ORC. (This hypothesis is somewhat controversial and will be discussed further in section 1.4.) Meanwhile, the B3 element, present in only a subset of ARSs, is the only element not to bind a factor attached to the replisome. Rather, it acts as a binding site of ARS binding factor 1 (Abf1), which serves to position nucleosomes around origins.[15,32]

While budding yeasts have sequence requirements for origins, sequence is not enough to define yeast origins of replication. There are far more matches to the ACS in the yeast genome than there are licensed origins.[33] As previously mentioned, the chromatin context helps define which ORC-ACS matches serve as origins. Nucleosome positioning around the ORC-ACS is important in defining function ARSs. As depicted in Figure 1.1, nucleosome positioning around the ARS is asymmetrical; the ACS is closer to its flanking nucleosome than the B2 element (downstream on the opposite strand) is to its flanking nucleosome. In cases where both the ACS and B2 element are good matches to the ideal ACS, this nucleosome positioning may help determine which sequence serves as the primary binding site for ORC.[33]

Though they share common motifs and sequence requirements, individual yeast ARSs are unique in sequence with the exception of the origin in the ribosomal DNA (rDNA) of yeast, which is present in every rDNA repeat. The rDNA origin is of particular importance because as the most abundant origin sequence, it is the most likely site at which origin sequence may be under selective pressure. Moreover, the rDNA locus often serves as a barometer of replication stress, shrinking in response to defects in replication protein mutations and alterations.[34,35] *S. cerevisiae* has between 90 and 300 copies (roughly 180 copies in wild-type laboratory strain S288C) of the 9.1 kb rDNA array. Each repeat contains a copy of the rDNA ARS (rARS) that lies in the non-transcribed spacer between the 5′ ends of the 35S and 5S transcription units. The rARS is a relatively weak origin with an imperfect match to the ACS.[36] Only 30-40 of these origins are active in a given S phase, but because the rARS is the only available sequence to facilitate replication within a 1.6 Mb region, compromising or removing the rARS ACS eliminates initiation within the rDNA locus leads to substantial reduction of

the rDNA array (~10 copies) and severely impacts cell fitness.[35] Upon long-term culturing of yeast without the rARS ACSs, origin activity returns to the rDNA locus, allowing slow re-expansion of the rDNA array. Under these conditions, the B2 element of the rARS appears to function as the ACS.[35] Conversely, strengthening the rARS by replacing it with an origin with a better match to the ACS can also impair yeast fitness and shorten cell lifespan, potentially by sequestering limiting replication factors away from the rest of the genome, leading to delayed and/or incomplete genomic replication.[36]

1.3 Origin Licensing

The process by which ORC binds to DNA and loads the MCM helicases is referred to as origin licensing. The two MCM helicases loaded at the origin are called the prereplication complex (pre-RC). Only sites at which the pre-RC has assembled can serve as origins during replication. To ensure that replication occurs only once in a given cell cycle, pre-RC formation is temporally separated from DNA replication. While in yeast, ORC binding can be detected throughout the cell cycle, MCM loading only occurs during late M phase and G1 when active CDK is absent and several of the additional proteins that help load MCM2-7 are only expressed during these phases.

The formation of the pre-RC is a dynamic process where the same interactors participate in a multitude of steps and undergo several conformational changes. While *in vitro* assays have helped determine which proteins are involved at which stage, clear mechanistic details of every step have not been fully determined. Figure 1.2 lays out the canonical model for ORC loading of the pre-RC and is detailed below.

1.3.1 Loading of the first MCM2-7 helicase

After assembly off of DNA in yeast, ORC binds as an open ring to origin DNA in an ATP dependent manner (Figure 1.2). Notably, ORC binding alone bends the DNA out of normal B conformation.[6] Next, Cdc6 binds to the open gap between Orc1 and Orc2, forming a tighter, salt-stable ring around DNA (Figure 1.2-2) that has increased sequence specificity.[37–39] Cryo-EM structure of ORC-Cdc6 alone around DNA does not

yet exist at sufficient resolution to see DNA specific contacts. Therefore, we do not know if this additional binding specificity is due to contacts between Cdc6 and DNA that are specific to this conformation or due to alterations of proximity between ORC and DNA, enabling new hydrogen bonding between ORC and DNA. Cdc6 is capable of binding and hydrolyzing ATP; however, the interaction between origin DNA and Cdc6 prevents this ATP hydrolysis.[38] If ORC and Cdc6 bind in the absence of DNA, Cdc6 immediately hydrolyzes ATP, causing its release from ORC, thus freeing the complex to bind origin DNA. Likewise, while ORC alone may be capable of binding to an imperfect ACS match, the increased sequence specificity of Cdc6 binding may trigger ATP hydrolysis and Cdc6 release, demonstrating that ORC presence alone on DNA may not be an adequate measure of DNA's proclivity to promote origin licensing.[38]

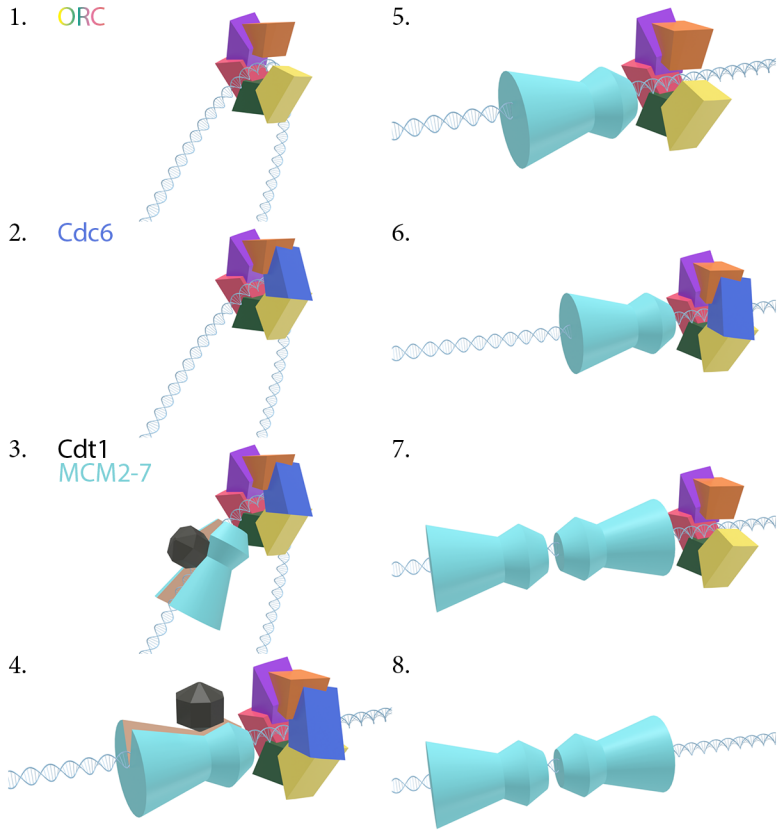


Figure 1.2: Mechanism of MCM2-7 loading via ORC. 1) ORC binds origin DNA as an open ring. Shown are ORC1-5 proteins; ORC6 is not an integral component of the ORC ring. 2) Cdc6 binds between Orc1 and Orc5, closing the gap and tightening the ring around DNA. 3) & 4) Cdt1 helps load the first MCM2-7 hexamer onto the DNA, which results in straightening of the DNA following MCM loading. 5) Cdc6 and Cdt1 are released, closing MCM2-7 around DNA. 6) Another Cdc6 molecule binds to ORC, priming it to load another Cdt1-MCM2-7. 7) A second MCM2-7 hexamer is loaded head to head with the first MCM2-7 molecule. 8) ORC hydrolyzes ATP and is released from the origin, leaving the mobile MCM helicase at the origin.
Figure made by Charles Martin

After Cdc6 binds to ORC on origin DNA, it recruits Cdt1, which is bound to an MCM2-7 hexamer, resulting in the establishment of the OCCM intermediate.[5,40] The

cryo-EM structure of the OCCM shows that at this stage, DNA loses its pronounced bend, although it is somewhat deformed through the central binding channel of ORC[5]. Notably, the only DNA base-specific interactions that are observed in the OCCM are from the Orc4-IH and a small region on Orc2. Following this loading, Cdc6 and MCM hydrolyze ATP, resulting in release of both Cdc6 and Cdt1 from ORC-MCM2-7.[41,42] The release of Cdt1 triggers a conformational change in MCM2-7, closing the gap between MCM2 and 5 through which the hexamer was loaded onto DNA and leaves just ORC and MCM2-7 (the OM intermediate) on the DNA.[43]

1.3.2 Loading of the second MCM2-7 hexamer

The proposed mechanisms by which the second MCM2-7 hexamer is loaded have been somewhat controversial. Much of that controversy revolves around the number of ORC binding events that are required to complete pre-RC formation. The canonical view of the process has been that a single ORC bound to the ACS loads both hexamers. For a single ORC to load a second MCM2-7 hexamer in the opposite orientation as the first, it must utilize different contacts and mechanisms than the first MCM2-7 loading. Recently, *in vitro* assays have shown that the double hexamer formation requires binding of a second ORC molecule to load the second hexamer, essentially repeating steps 1-5 in the opposite orientation.[43] This mechanism would also be consistent with use of the ARS's B2 element as a second ORC binding site.[31] While ORC's capability to bind to the B2 element makes this mechanism a possibility, it is still somewhat unclear if it is the primary mechanism by which the pre-RC is formed. However, as the same intermediaries are required, regardless of the number of ORCs, to load the second hexamers, the following discussion presents loading of the second hexamer without commenting on the binding location of ORC.

To load the next hexamer, a second Cdc6 must first bind to the existing ORC-MCM to form the OCM intermediate.[13] The OCM intermediate then recruits a second MCM2-7 hexamer with Cdt1 and loads it onto the DNA head-to-head with the first MCM2-7 hexamer. As with the loading of the first MCM2-7, origin DNA must drastically bend out of typical B form to facilitate second hexamer loading. This further bending is

another opportunity for new or altered DNA contacts to be made between ORC-Cdc6 and origin DNA, illustrating that complete understanding of DNA sequence specificity may require high resolution imaging of these steps *in vitro* to determine base-specific contacts between ORC and the origin.

Cdc6 ATP hydrolysis again causes release of Cdc6 and Cdt1, and results in the second MCM2-7 hexamer closing around the double stranded DNA. Next, hydrolysis of ATP by ORC causes the release of ORC from origin DNA, leaving the two MCM2-7 hexamers head-to-head at the origin.[42] Once bound to DNA, the MCM hexamers are free to move along the DNA in an ATP-independent manner;[11] however, their movement *in vivo* is limited by features of chromatin structure such as histones.[44] This movement restriction means that on a population level, MCMs in G1 are largely found at known origins of replication.[14,45]

Ultimately, it is important to note the complexities of ORC-DNA interactions that may take place during pre-RC formation. As DNA bends and ORC undergoes conformational change, the strength of DNA-residue contacts may change. Moreover, ORC may bind outside the canonical ORC-ACS during pre-RC formation, so its impact on sequence-specificity may extend beyond just the initial binding site.

1.4 Origin Recognition Complex structure

1.4.1 General Structure

In all eukaryotes, ORC is comprised of six subunits that bind as an open ring to DNA. Although there is sequence variation, ORC is predominantly functionally and structurally conserved throughout eukaryotes, as shown by the structural alignment of ORC in different species (Figure 1.3).[6,46,47] (Note: the gap in *D. melanogaster* ORC where the other species have an orange protein domain contacting the yellow domain is due to unresolvable structure in the cryo-EM complex, resulting in a truncation of that domain in the final model, but not in the structure itself.[47]) When these complexes are aligned in panel D, the structural similarities are even more apparent.

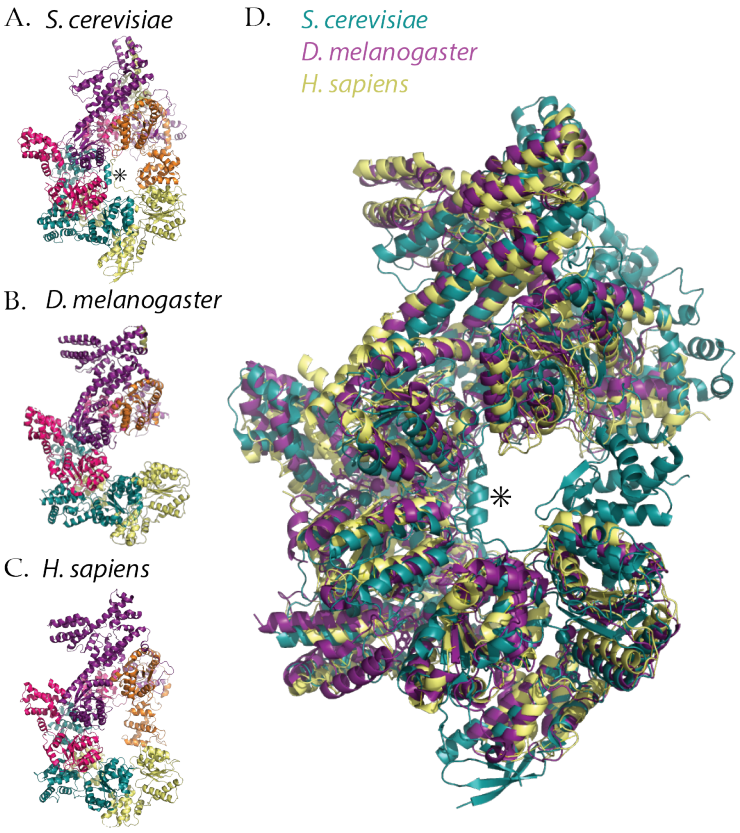


Figure 1.3: ORC structure is conserved throughout eukaryotes
Ribbon diagrams of A) *S. cerevisiae* ORC, B) *D. melanogaster* ORC, and C) *H. sapiens* ORC. D) The overlay of all three species' ORCs shows that the structure is predominantly conserved in the heterohexamer, but *S. cerevisiae* ORC extends more into the DNA binding channel than the other species. The * denotes *S. cerevisiae* Orc4 insertion helix.

This conservation highlights just how essential these proteins are and makes it easier to identify the slight differences between species that might explain the differences in origin recognition methods. For example, while predominantly conserved, panel D shows that *S. cerevisiae* ORC does have more regions extending into the DNA binding channel than either *D. melanogaster* or *H. sapiens*.[47,48] The Orc4-IH is denoted by asterisks in panels A and D.

1.4.2 Subunit Conservation and Divergence

The individual ORC subunits are also structurally conserved with one another. Eukaryotic ORC subunits 1-5 and Cdc6 descend from a single archaeal subunit, Orc1/Cdc6, and all share the same basic structure: an AAA+ (ATPase Associated with diverse cellular Activities) domain bound to a winged helix domain (WHD) (Figure 1.4B).[5,49] Parallels have been drawn between the general assembly of domains in ORC and the clamp loader subfamily: a set of proteins that load open replication processivity clamps onto dsDNA and hydrolyze ATP to close the clamp, which closely mirrors ORC's role in loading MCM2-7 onto dsDNA.[50] In general, AAA+ domain proteins are a superfamily of motor molecules that hydrolyze ATP to enact conformational changes.[51] In the case of ORC-Cdc6, this ATP hydrolysis involves conformational changes

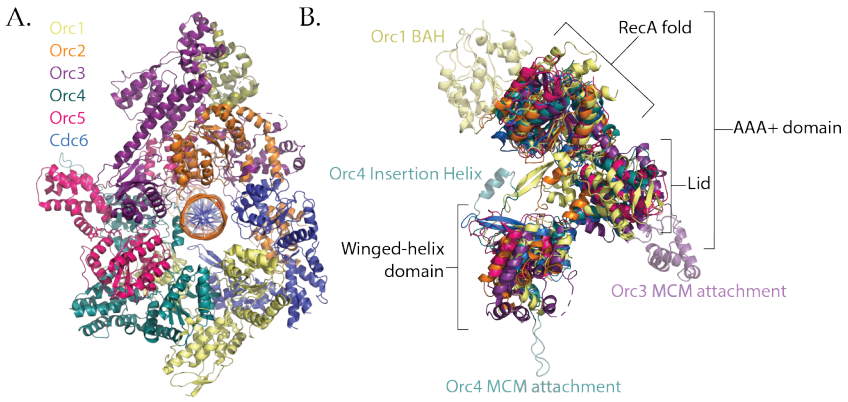


Figure 1.4: Conservation of subunit structure in ORC-Cdc6.
A) A ribbon diagram of *S. cerevisiae* ORC-Cdc6 surrounding origin DNA.[5] B) Orc1-5 and Cdc6 structural alignment shows the conservation of structure among the six proteins. All contain a AAA+ domain attached to a WHD. Differences in structure include two subunit-specific attachments to MCM2-7, the N-terminal Orc1 BAH domain that binds histone, and the Orc4-IH that contacts origin DNA in the major groove.

triggering the proteins' release.[52] Winged helix domains are well-known DNA contacting domains that often contain helix-turn-helix motifs that make direct contact with major groove DNA via the outermost helix, typically in sequence-specific manners.[53] These alignments highlight the structural conservation present within ORC and across eukaryotes, and, like interspecies alignments facilitate recognition of species-specific features, this subunit alignment enables identification of subunit-specific features. Two of these visible subunit-specific features are the Orc1 BAH and the Orc4-IH.

The Orc1-bromo-adjacent homology domain (BAH) is shared among eukaryotes.[54,55] This domain interacts with histones to enable ORC binding adjacent to nucleosomes. In this way, even without sequence specificity, ORC is able to target a subset of genome locations via chromatin interaction. While the BAH domain of its paralog, Sir3, cannot bind H4K16 acetylated histones, limiting its presence to heterochromatin, the Orc1 BAH is equally capable of binding to histones with or without this silencing mark.[56] This lack of specificity allows Orc1-BAH to bind to histones in heterochromatin and euchromatin regions, ensuring that replication origins are adequately distributed throughout active and inactive chromatin domains.

The second subunit-specific region of interest is the Orc4 insertion helix (Orc4-IH). As shown in Figure 1.3, this region extends into the central DNA binding channel of ORC and makes contact in the major groove of origin DNA. This region was the first of any ORC subunit that was shown to make direct contact with DNA bases (enabling

sequence-specific contact), and, as shown in Figure 1.5, is unique to budding yeast just like origin sequence-specificity.[5]

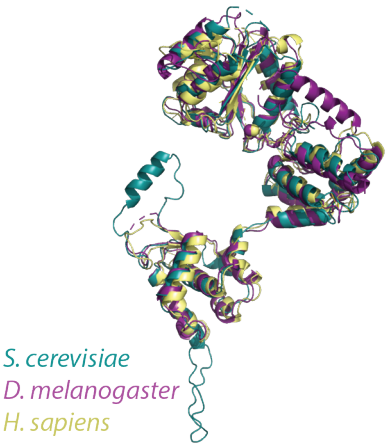


Figure 1.5: Comparison of the structure of Orc4 in yeast, fly, and human. The Orc4-IH is unique to yeast, despite otherwise strong conservation of Orc4 structure between species.

1.4.3 Ancestry of budding yeast ORC sequence specificity

Winged helix domains are known for their sequence-specific DNA binding. The ancestral archaeal Cdc6/Orc1 WHD serves this purpose; up to six Orc1/Cdc6 subunits bind DNA, making DNA specific contact via the WHD and secondary contact in the minor grove via the AAA+ domain.[57] However, this DNA contacting feature has clearly been lost in most eukaryotes, and even within most ORC subunits within budding yeast. The nearly ubiquitous loss of sequence-specificity in eukaryotic ORC subunits raises the question of whether the Orc4-IH's sequence specificity is a preservation of or reversion to the ancestral state. Figure 1.6 shows a single subunit of *S. solfactaricus* Cdc6/Orc1 with archaeal origin DNA; *S. cerevisiae* Orc4 is structurally aligned to it.[6,49] Overall, the *S. cerevisiae* Orc4-WHD has a less compact structure than *S. solfataricus's* WHD, with longer unstructured turns linking the alpha helices. This pattern remains when examining each protein's DNA contacting alpha helix. The *S. cerevisiae* Orc4-IH looks like a loosened version of the *S. solfactaricus* Cdc6/Orc1 DNA contacting helix. Whether this loosening allowed preservation of DNA-specific contact throughout the evolution of Orc4 or restored DNA specificity of this subunit specifically in budding yeast remains an open question.

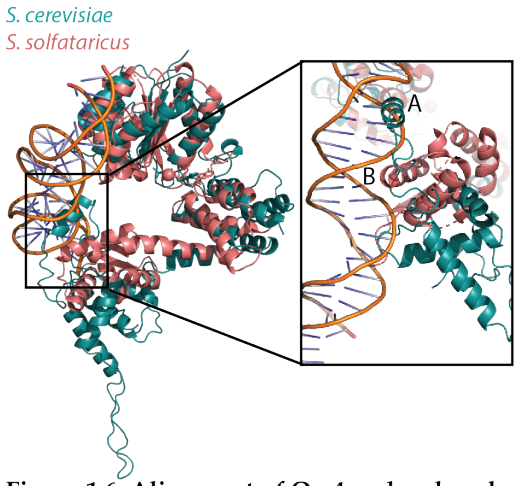


Figure 1.6: Alignment of Orc4 and archaeal Cdc6/Orc1. DNA-bound *S. cerevisiae* Orc4 aligned to *Saccharolobus solfataricus* Cdc6/Orc1 bound to DNA shows conservation of the AAA+ domain and the 'loosening' of the WHD. In the closeup view, the *S. cerevisiae* Orc4-IH (A) appears to be the helix that contacts DNA while one of the helices (B) in the helix-turn-helix motif of the *S. solfataricus* Cdc6/Orc1 contacts DNA.

1.5 Origin Firing

Not all origins that are licensed are used during S phase. In yeast, more than 700 origins are licensed,[58] but only roughly 400 will actually serve as sites of replication initiation in a given S phase.[59] The exact collection of origins utilized differs from cell to cell,[60] but some origins are used in more cells than other origins. How many cells within a population use a specific origin is referred to as "origin efficiency". The time when an origin fires in S phase is referred to as "origin timing".

There are some general trends for genome-wide replication timing. In eukaryotes, euchromatin replicates earlier than heterochromatin, and telomeric regions tend to be the latest replicating.[61] Another conserved timing trend is the early replication of centromeric regions. In fact, centromeres actually help determine this early timing as moving a centromere to a different location causes origins in that location to advance in replication timing.[3] Origin firing throughout S phase is controlled by a variety of factors. Replication factors Sld2, Sld3, Dpb11, Dbf4, Sld7, and Cdc45 (known as the limiting factors SSDDSC) are required for origin firing, but are in such limiting supply that they must be recycled from fired origins to create a cascade of origin firing across S phase.[62] Stoichiometry of these limiting factors is crucial: overexpression of 4 of these factors does not change origin activity, but overexpression of all six of these limiting factors ablates the temporal replication program, causing origins to fire synchronously.[34,62] However, depletion of a single limiting factor decreases replication efficiency genome-wide rather than narrowing the subset of origins utilized. Meanwhile, different cyclins are also required for origin activity throughout S phase. Clb6 and Clb5 are primarily responsible for enabling firing throughout S phase, but Clb6 is much shorter lived than Clb5, which is the major form of CDK for late origin firing.[63] The deletion of *CLB5* prevents late origins from firing, requiring those sites to be passively replicated by origins activated previously.[64]

While there are a greater number of high-efficiency origins in early S phase versus late S phase, it is possible to have late, efficient origins.[64] The ongoing debate between some individuals in the replication field is whether DNA replicates via a temporal program or stochastic firing. In the temporal model, DNA replication follows a

temporal order where regions of the genome replicate at different times and replication timing is a property of origins, separate from origin efficiency. In a stochastic firing model, timing is random and the perceived timing profile of a population results from each origin's set probability of firing. Origins with higher probabilities of firing (more efficient), on average, fire in more cells and before origins with lower probabilities of firing (less efficient), thus creating the appearance of a population-wide temporal program.[65]

Another open question is what determines an origin's efficiency. In yeast, stronger matches to the ACS tend to result in more efficient origins, and increasing the strength of an origins ACS can increase its efficiency at its endogenous location.[36] (Notably, strengthening the ACS of the rDNA origin does *not* increase efficiency.[35]) Some hypothesize and show data that this increase in efficiency is due to increased MCM loading at that origin.[66] This correlation would suggest that the efficiency of an origin should be predicted by the MCM signal at that origin in G1. In support of this model, deletion of Sir2-driven MCM distribution does cause build-up of MCMs in later replicating regions, increasing the origin activity of these regions.[67] Nevertheless, the link between MCM loading and efficiency is still a matter of debate and ongoing research and will be discussed briefly in the results of chapter 2.

Chapter 2. ORC-driven origin specificity

2.1 Introduction

While eukaryotic replication is highly conserved across species in both function and structure, budding yeast are unique in that they have sequence-specific origins of replication. This feature has proven invaluable to researchers interrogating the impact of genomic location on origin function or the impacts of deleting specific origins or sets of origins, but what drives sequence specificity remains unknown.[3,4] However, recent cryo-EM structures of the yeast Origin Recognition Complex (ORC) bound to DNA revealed a small region of *S. cerevisiae* Orc4 that extends into the DNA binding lumen and makes direct contact with origin DNA.[5,6] This region is called the insertion helix (IH) and it is composed of a long unstructured loop capped with an alpha helix that reaches into the major groove of DNA. This interaction between Orc4-IH and DNA has been documented in two separate stages of origin licensing: when ORC alone is bound to DNA and when ORC is bound to DNA with Cdc6, Cdt1, and MCM2-7 as part of the OCCM.[5,6] This persistence of contact in both conformations lends further credence to the hypothesis that this region may be the primary driver of sequence specificity in budding yeast origins. Moreover, structural comparisons between Orc4 from humans and *Drosophila* reveal that this Orc4-IH is unique to budding yeast, just like yeast's origin sequence specificity. This observation led to the hypothesis that the Orc4-IH may be the primary driver of origin specificity in yeast.

2.1.1 DNA-specific contacts between ORC and origin DNA in S. cerevisiae

The ORC-DNA structure provides sufficient resolution to identify specific contacts between ORC subunits and origin DNA in this single conformation. Between the two cryo-EM conformations, three ORC subunits contact DNA in a base-specific manner: Orc1, Orc2, and Orc4. The Orc4-IH is the most substantial of these contacts as the insertion helix sits squarely within the major groove of DNA, an interaction apparently preserved through MCM2-7 loading. However, the Orc4-IH DNA interaction only impacts part of the ACS. Figure 2.1 shows the *S. cerevisiae* ACS and the positions that

have base-specific interactions with ORC subunits. The Orc1 basic patch (Orc1-BP) influences two bases near the beginning of the ACS at positions T5 and T7. The Orc2 initiator specific motif (Orc2-ISM) and Orc4-IH both have base-specific interactions with T9, but the Orc4-IH also interacts with T10, T11, and with the cytosine paired to guanine at position G13.[6]

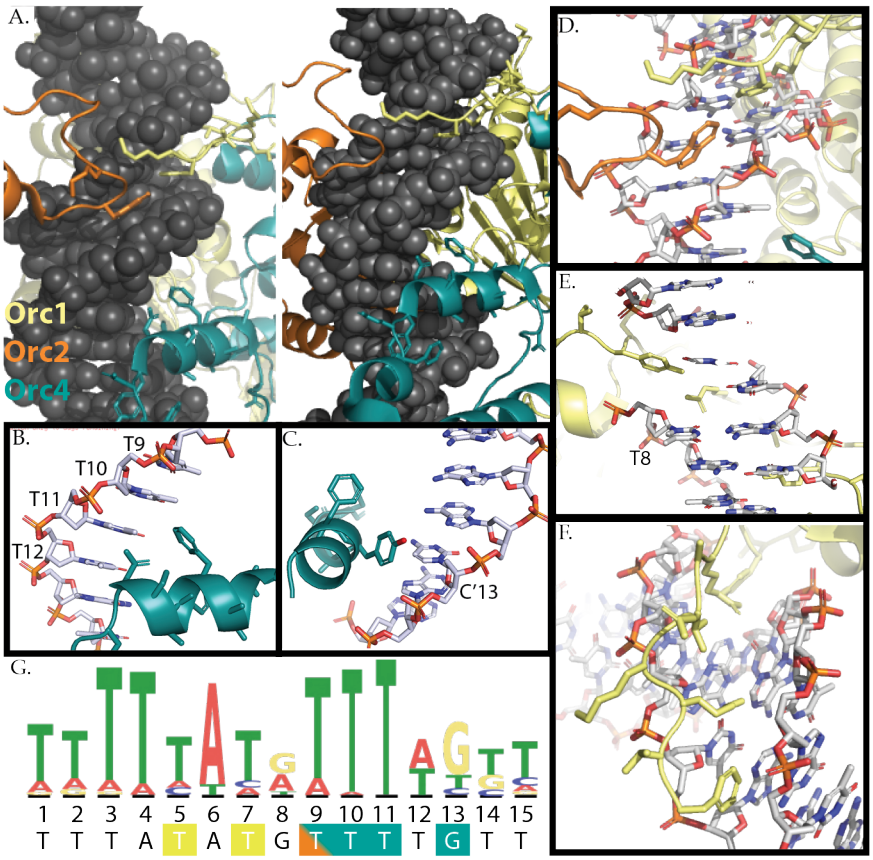


Figure 2.1: ORC DNA-specific contacts at the ACS. A) Two views of ORC contacts around origin DNA. Orc1, 2, and 4 all have base-specific contacts within the major (Orc4) or minor (Orc1 and 2) groove. B) Interaction between F485 of Orc4-IH with thymine stretch. C) Interaction between Y486 and a cytosine on the opposite strand of the ACS. D) Orc2 W396 in the minor groove. E) Orc1 R367 and Y372 interact with thymine. F) Orc1-BP contacts with minor groove origin DNA. G) *S. cerevisiae* ACS labeled by position. Below is the ACS of *ARS305*, the origin shown with the ORC structures. Bases in *ARS305* are highlighted according to the subunits that have sequence-specific interactions with ORC at that base.

Origin sequence determination may be driven by multiple ORC subunit contacts with DNA; however, not all subunits show the same extensive interaction as the Orc4-IH nor are the other subunit contacts as structurally independent as the Orc4-IH. Much of the appeal of the Orc4-IH lies in the potential for its manipulation outside the secondary and tertiary structure of Orc4. Because ORC subunits must bind to each other, DNA, and MCM in several conformational states, even mild manipulation of embedded residues may disrupt ORC stability at large.

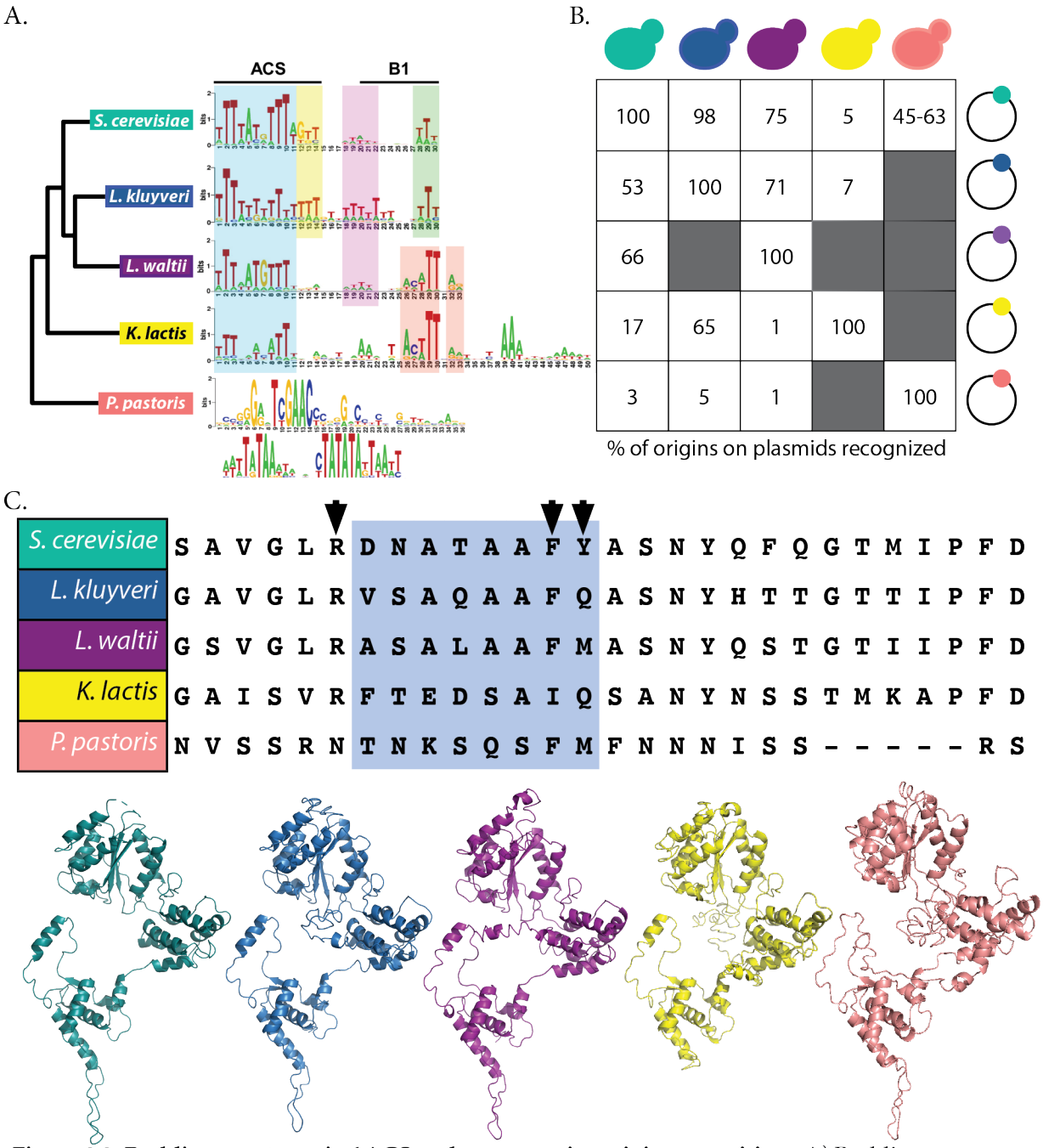


Figure 2.2: Budding yeast species' ACS and cross-species origin recognition. A) Budding yeast phylogeny showing the ACS of each species. ACS diversity increases as species divergence increases. *P. pastoris* appears to recognize two separate sequences, the top of which may be a transcription factor binding site. B) Cross-species origin recognition showing what percentage of origins from the species on the right are able to be recognized by the species on top (when the species on top is transformed with origins on plasmids). Grey squares indicate no data available for the combination. Adapted from Raghuraman and Liachko 2006. C) Amino acid alignment of budding yeasts Orc4 reveals sequences corresponding to the Orc4-IH in *S. cerevisiae*. The first two DNA-contacting residues in *S. cerevisiae* (indicated by black arrows) are predominantly conserved across species while the third DNA-contacting residue diverges in the other species. Differences in the Orc4-IH sequence result in differences in predicted secondary structure within the other species' putative Orc4-IH as modeled from I-TASSER structures.[72]

2.1.2 Leveraging budding yeast clade to test ORC-driven origin sequence-specificity

The Orc4-IH provides a unique opportunity to potentially manipulate sequence specificity independent of ORC assembly or MCM2-7 loading. In further support of the hypothesis that the Orc4-IH determines sequence specificity, the Orc4-IH, while not found in metazoans, is present in other budding yeasts that also have sequence-specific origins. The ACS is conserved across origins within a budding yeast species, but there are differences between species (Figure 2.2 A).[68] Most share a degenerate A/T motif, suggesting that there is functional overlap between ARSs of different species. Both the characterization of other species' ACSs and the cross-recognition of ARSs in yeast species has been undertaken in a broad range of budding yeast species (Figure 2.2 B).[69-72] Collectively, these works found that yeast species recognize varying numbers of neighboring species' ARSs with varying affinity. For example, *S. cerevisiae* can recognize two thirds of ARSs from *L. waltii* (one third with high maintenance when on plasmids, one third with low maintenance when on plasmids).[69]

The Orc4-IH sequences of budding yeasts also diverge as ACSs diverge between species. (Figure 2.2 C). These divergent amino acid sequences lead to differences in predicted secondary structure for these regions, particularly in the terminal alpha helix.[73] This covariance means that even when DNA-contacting amino acids (as defined by contacts between *S. cerevisiae* Orc4-IH and *S. cerevisiae* ACS) are conserved between species (as is the case with *L. kluyveri* and *L. waltii*), the orientation of amino acid side chains are changed, potentially altering contacts with DNA. Interestingly, the Orc4-IH has a higher percent conservation between the *Saccharomyces* and *Lachancea* budding yeast genera than Orc4 at large between these genera, suggesting that this region is under selective pressure, potentially to maintain similar ACS-Orc4 contacts (Supplemental Table 2.1).

To test the hypothesis that the Orc4-IH drives sequence specificity of yeast origins, I decided to leverage the existing diversity within the budding yeast clade to make chimeric Orc4 proteins in a *S. cerevisiae* haploid. By replacing the Orc4-IH sequence of *S. cerevisiae* with the insertion helices from the species shown in Figure 2.3, I could potentially alter which *S. cerevisiae* origins were recognized by the chimeric

proteins. I hypothesized that functional origins in the chimeric strains would be biased towards the ACS of the species from which the insertion helix sequence was taken. To determine the optimal ACSs, ARS-Seq assays are used. In this assay, genomic DNA from a desired species is sheared and ligated into origin-less plasmids.[69–71] The yeast cells are then transformed with these plasmids and grown in culture to determine which genomic sequences could support plasmid replication. ARS consensus sequences were determined from the plasmids supporting replication. These ARS-Seq libraries provide ready-made tools to test whether the Orc4-IH amino acid changes results in altered maintenance of ARS plasmids from other species. Moreover, cross-species ARS-Seq assays provide information both on which foreign origins *S. cerevisiae* could and could not recognize as well as how many *S. cerevisiae* origins other species could recognize, providing ample information to predict which chimeric strains might be viable. Additionally, the percentage of *S. cerevisiae* origins recognized ranged from 5% to 95%, allowing me to potentially narrow down a threshold of the number of origins required to complete genome replication.

2.2 The role of the Orc4 insertion helix in sequence specificity

2.2.1 Chimeric strain construction

I constructed chimeric strains via transformation of haploid BY4741 cells with PCR template and a CRISPR/Cas9-gRNA plasmid,[74] selecting for the Cas9-gRNA plasmid marker, *URA3*. By selecting for plasmid maintenance, I selected for cells in which Cas9 could no longer cut at the gRNA target site in *ORC4*, indicating that the *ORC4* gene had been successfully mutated, or less likely, that Cas9 had been silenced in transformed colonies. Based on the percentage of *S. cerevisiae* origins each "donor" species could recognize, I hypothesized that *L. waltii* and *L. kluyveri* templates would produce viable chimeric strains with little defects as both of these donor species recognize 75%+ of *S. cerevisiae* origins. In contrast, I expected that *K. lactis* templates would produce no viable mutant, or that any transformants produced would be extraordinarily slow growing as *K. lactis* only recognizes 5% of *S. cerevisiae* origins.[70] I hypothesized that *P.*

pastoris would be more likely to produce viable mutants than *K. lactis* because it recognizes approximately 50% of *S. cerevisiae* origins, but that any mutants produced would be considerably sicker than those produced from *Lachancea* templates.[70,72]

After multiple initial transformation attempts, only the *L. waltii* template transformation produced a viable chimeric strain. This result was somewhat perplexing as *L. kluyveri* recognizes more *S. cerevisiae* origins than *L. waltii,* so failure to adequately replicate the *S. cerevisiae* genome did not appear to be a compelling reason that constructing a *L. kluyveri* chimera failed. However, there is a technical reason that the *L. kluyveri* and other strain transformations could have failed that did not truly reflect on a chimeric Orc4's ability to support genome replication: by selecting for plasmid maintenance, I also required that any mutant Orc4 produced could recognize the *S. cerevisiae* 2µm origin present on the Cas9-gRNA plasmid. To circumvent this issue, I tested a version of the CRISPR/Cas9-gRNA plasmid with the panARS, an origin specifically engineered to function in *S. cerevisiae* and the four template strains.[75] Changing the ARS did not improve transformation results; no further chimeric strains were produced. However, transformation with the *K. lactis* template did produce two independent mutants, each bearing the same six amino acid deletion within the Orc4-IH alpha helix (Figure 2.3). This mutation

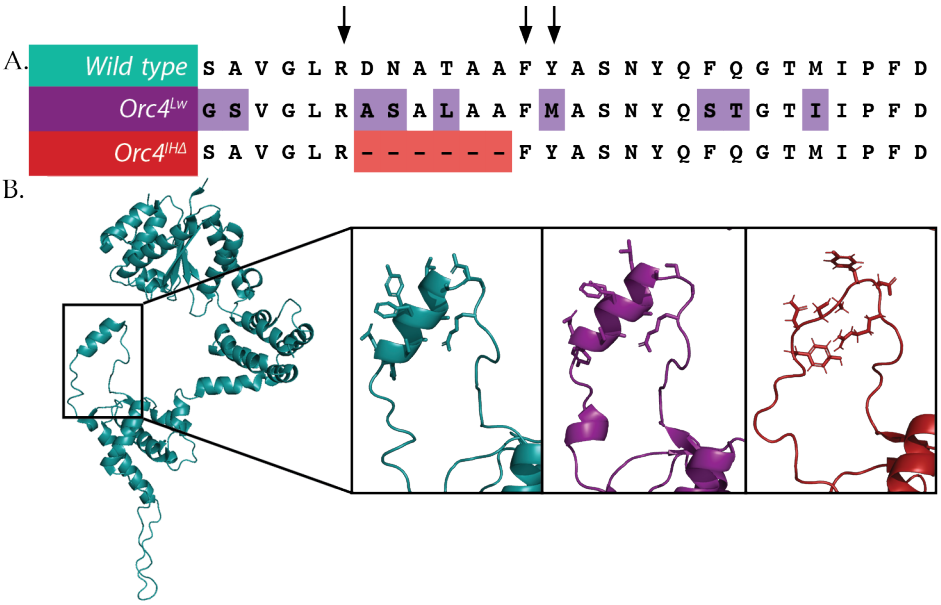


Figure 2.3 Insertion helix sequence and predicted structures in Orc4-IH mutants. A) Amino acid alignment of wild-type Orc4-IH and mutants. The chimera has 9 differences from wild type, including 1 out of 4 DNA contacts while the deletion has 6 amino acids missing, putting 2 DNA contacting residues adjacent to each other. B) The cryo-EM structure of wild-type Orc4-IH and the predicted structures of the chimera and deletion show slight alterations in orientation of DNA-contacting residues, particularly in the deletion.

essentially shortened the loop overall and removed most of the alpha helix that interacts with the DNA in the major groove. Therefore, I had two instances in which to investigate Orc4-IH and origin specificity: the chimera, in which I predicted specificity to be altered to that of *L. waltii*, and the deletion, in which I expected specificity to be reduced due to the shortening of the loop that contacts DNA.

2.2.2 Orc4-IH mutant viability and plasmid maintenance.

The chimeric strain had a population doubling time essentially indistinguishable from wild type. Wild type had a doubling time of 92.9 ± 0.9 minutes, the chimera 91.4 ± 0.6 minutes based on five biological replicates. Meanwhile, the deletion was somewhat slower with a doubling time of 97.9 ± 1.3 minutes (Supplemental Figure 2.1). This slower growth was also accompanied by slight changes in S phase progression as shown via arrest and release flow cytometry analysis (Supplemental Figure 2.2). The deletion appears to be slightly slower at entering S phase compared to either wild type or the chimera. It also appears to take longer to exit S phase and spends longer in G2 than either other strain.

As a first pass at testing the origin specificity of the mutant Orc4 strains, I transformed different strains with a small set of plasmids with origins from either *L. waltii* or *S. cerevisiae*. The *L. waltii* plasmids had been isolated from *L. waltii* ARS-Seq by Sara DiRienzi and were not capable of supporting replication in *S. cerevisiae*. As a reciprocal control, I used a plasmid containing *ARS228*, an inefficient origin in *S. cerevisiae* distal to the gene *SUL1* that cannot be maintained in *L. waltii*.[76] I predicted that the Orc4^{lw} chimera would be capable of recognizing the *L. waltii* plasmids, but not the *S. cerevisiae ARS228* plasmid. I also hypothesized that the Orc4$^{IH\Delta}$ would have reduced/relaxed origin specificity and thus be able to recognize all plasmids but with worse transformation efficiency or slower colony growth than either wild type or chimera because the less-specific Orc4-IH would bind origins with weaker affinity. In transformations, I found that, as expected, wild-type *S. cerevisiae* could be transformed with the *ARS228* plasmid, but not with the *L. waltii* plasmids. In contrast, the Orc4-IH chimera could maintain both the *ARS228* plasmid as well as the *L. waltii* origin

plasmids. Upon initial transformation, it appeared that the chimera had poorer transformation efficiency of the *ARS228* plasmid, suggesting that origin specificity had been altered away from *S. cerevisiae* as it had altered toward *L. waltii* (Supplemental Figure 2.3). Repeated transformations demonstrated consistent, relatively high transformation efficiency of the *L. waltii* plasmids into the chimera and would occasionally produce a single transformant from wild type and/or the deletion. However, repeated transformations also showed that transformation efficiency was too variable to determine differences in the *ARS228* plasmid transformations, and thus required loss rate experiments to determine plasmid maintenance competency between strains.

The initial plasmid loss rates of the *L. waltii ARSIV-582* plasmid (Supplemental Figure 2.4) showed that the chimera did have a much lower loss rate than either the deletion or wild type; however, all loss rates were high and likely due to the lack of functional *S. cerevisiae* CEN on the plasmid to help ensure stable plasmid segregation. Moreover, to have viable wild-type and deletion transformants with the *L. waltii* plasmid, I had to wait up to 8 days for streaked out colonies to grow up to the point where single colonies could be used to inoculate 5 mL cultures. While 5 mL cultures for the chimera bearing the *L. waltii* plasmid grew to near saturation overnight, I had to wait 3-5 days before the 5 mL cultures of wild type or deletion with *L. waltii* plasmid were dense enough to preserve as freezer stocks. This propagation method and extra time allotted for these cells means that I was stringently selecting for wild-type and deletion cells that were capable of maintaining the plasmid prior to conducting plasmid loss rates. The true loss rate of the *L. waltii* plasmids in wild type and deletion are likely much higher than what I was able to measure.

2.2.3 μARS317-Seq shows greater intolerance for ACS changes in Orc4 mutants

Cryo-EM images of ORC alone bound to origin DNA were at high enough resolution to determine that the Orc4-IH contacts origin DNA in the four bases following the ATG motif in the *S. cerevisiae* ACS.[6] This base pair resolution motivated me to interrogate if the chimera and deletion could tolerate different mutations at this

region in the ACS compared to wild type. To evaluate differential tolerance of specific mutations in the ACS, I performed mutARS317-Seq (μARS317-Seq), a deep mutational scan of the *S. cerevisiae* origin *ARS317 (*Figure 2.4 A). In its native context, *ARS317* is a low efficiency origin that is an imperfect match to the *S. cerevisiae* ACS, namely it has an A instead of G at position 8, and a G at position 14 rather than a T.[77] In addition to the expected B1 and B2 elements downstream of the ACS, *ARS317* also possesses Forkhead 1 (Fkh1) and repressor-activator protein 1 (Rap1) binding sites.[78]

Forkhead 1 is a transcription factor that has many roles, one of which is facilitating early origin activation.[79] Fkh1 binding sites often overlap with the ACS of origins and in such cases it is not possible to make mutations in the Fkh1 binding site without also altering the ORC binding site. In the case of *ARS317* the Fkh1 binding site overlaps with what is annotated as the ACS in OriDB, but not what previous μARS317-Seq studies have annotated as the origin in *ARS317*.[80]

Rap1, a DNA binding protein that functions in transcriptional and telomeric silencing, binds to and tethers plasmid to the inner nuclear membrane during segregation, ensuring their passage to daughter cells.[6,46,47] Because the μARS317-Seq plasmid library lacks a centromere to ensure stable transmission at cell division, I would expect the Rap1 region to be as intolerant of mutations as the ACS since the ACS is needed for replication while the Rap1 binding site is essential for segregation. Because Rap1 binding influences plasmid maintenance independent of ORC binding, the Rap1 binding site can serve as an inter-strain control for the μARS317-Seq assay. Regardless of which strain is analyzed, all strains should show relatively similar intolerance of mutations in this region and identify the known Rap1 binding motif. Differences in this region could indicate incomplete coverage of the library either in transformation or in sequencing.

Typically, μARS-Seq experiments involve a competitive growth culture. I first started by comparing *ARS317* variants present in strains after initial transformation with the μARS317 library, but prior to competitive growth. I hypothesized that the tolerance for mutations would be weakest in the ACS and Rap1 binding sites for all

strains examined, but that the chimera would be better able to tolerate mutations to the ACS that matched the *L. waltii* ACS. For the deletion, I hypothesized that it would be better able to tolerate mutations in the ACS compared to wild type as the deletion of the IH would reduce DNA binding specificity, but that there would be no preference for specific changes.

As expected, all strains show lower general tolerance for mutations in the ACS and Rap1-binding site than in undefined regions (Figure 2.4 B). Examining the region of interest—the ACS—the data show wild-type cells can tolerate some mutations in the ACS, but both the chimera and deletion were more intolerant of mutations in this region with a few notable exceptions. The first example is tolerance for the 8A>G mutation in the ACS. This change in tolerance/preference at this site mimics the *L. waltii* ACS which has a stronger preference for G at this site compared to *S. cerevisiae*, although *S. cerevisiae* also shows a mild preference for G in the overall ACS. The deletion also shares this preference. Interestingly, the 8th position of the ACS is actually one base-pair upstream of the origin DNA shown to be contacted by the Orc4-IH when ORC alone is bound to the origin. However, not enough structural information is known about ORC/DNA configurations during the entire process of ORC binding and loading MCMs to say that there is no sequence-specific contact between the Orc4-IH and the 8th base of the ACS during origin licensing. In fact, these data suggest that this ACS position does interact with Orc4-IH or changes the nature of Orc4-IH interaction with origin DNA in some way, possibly by altering the structure or malleability of the DNA.

The other tolerated ACS mutation in the chimera, 12A>T, again brings the origin sequence more inline the with *L. waltii* consensus sequence as there is a slight, but definite preference/tolerance for T over A at this position in the ACS. Again, the Orc4$^{IH\Delta}$ shares this tolerance/preference to a lesser extent and additionally shows tolerance of a C>A mutation at the subsequent site. Worth noting, when the Orc4$^{IH\Delta}$ is modeled in ORC binding to origin DNA, new DNA contacts between Orc4$^{IH\Delta}$ and the ACS are predicted at 13C. This position is one base pair beyond where the wild-type Orc4-IH is shown to contact origin DNA when only ORC is bound to DNA.

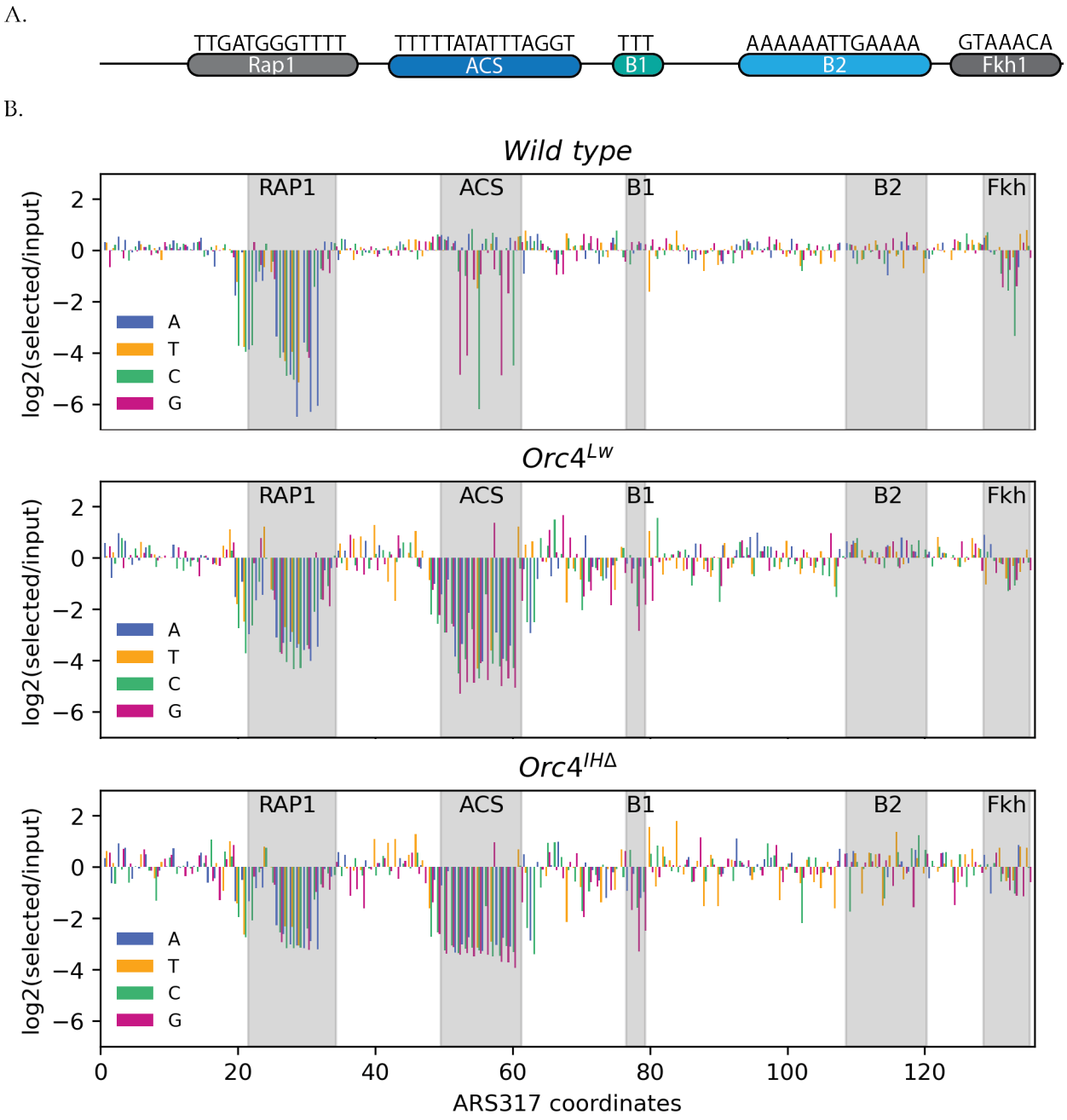


Figure 2.4: µARS317-Seq data for entirety of *ARS317* for Orc4 mutants. A) A diagram of *ARS317* and the conserved elements within the origin. The Rap1 binding site is crucial for plasmid segregation while the ACS, B1, and potentially the B2 elements are important for ORC recognition. B) Relative frequency of variants across *ARS317*. All strains show expected intolerance of mutations in the Rap1 binding site, but the chimera and deletion are more intolerant of ACS mutations with a few exceptions that closely match the *L. waltii* ACS.

It is important to note that existing cryo-EM structures show ORC in two snapshots of the origin licensing process: bound alone to origin DNA and bound as a part of the OCCM (ORC-Cdc6-Cdt1-MCM2-7). ORC binding and MCM loading are

dynamic processes where the DNA is bent and deformed from typical B form DNA to load large complexes as open rings before ATP hydrolysis in various AAA+ domains within ORC, Cdc6, and MCM2-7 cause conformational changes closing the ring.[13,39,41,42] In fact, earlier studies have shown that Cdc6 binding to ORC on DNA tightens ORC around DNA and has greater sequence specificity than ORC alone.[38] These changes could be due to Cdc6 contacts with DNA or result from the increased proximity between ORC and DNA, enabling new hydrogen bonding between ORC and DNA. It is possible that other regions of the Orc4$^{IH\Delta}$ or the other members of ORC come into contact with origin DNA in sequence-specific manners in the steps of origin licensing for which there are not cryo-EM structures, thus altered specificity outside of the four bases in direct contact with the Orc4-IH are possible upon mutation of Orc4. However, it is also possible that by altering the insertion loop at large, I have changed the orientation of key amino acids and thus changed DNA contacts between the chimeric or deletion Orc4-IH that are unique to the mutants investigated here.

2.2.4 Genome-wide early origin usage in Orc4 mutants shows variable origin activity

Because I hypothesized that mutating the Orc4-IH would change ORC binding and thus the location of potential origins, I went on to assay genome-wide origin activity. I focused on genome-wide origin usage in early S phase by performing the single-stranded DNA (ssDNA) assay developed in our lab.[81] This synchronized arrest and release into 200mM HU enables cells to begin replication but prevents replenishing of dNTP pools by inhibiting ribonucleotide reductase, and thus creates regions of ssDNA around active origins. Single-stranded DNA in samples collected in S phase and G1 are differentially labeled with fluorescent nucleotides and co-hybridized to a microarray. Plotting S/G1 ratios along chromosomal coordinates reveals where ssDNA has accumulated, thus labeling peaks at active origins. However, as forks move away from an origin, the ssDNA at the origin is filled in, creating split peaks where the local minimum between the two peaks is the area around the origin where replication has completed and the peaks on either side represent the forks moving away from the origin. Thus, in comparing of origin activity between strains, it is important to consider

both peak height and peak area to determine if a given origin is differentially impacted between two strains.

I expected to see a mix of lost and decreased activity origins in the chimera and even greater loss of activity in the deletion given that the deletion had a longer doubling time than wild type. Because the chimera had such a similar S phase progression and doubling time when compared to wild type, I hypothesized that losses in origins would be accompanied by potential gain in new origin sites, especially since the chimera was able to recognize plasmids that wild-type cells could not. In the deletion, I hypothesized that reduced specificity would decrease the overall efficiency of most origins (shown by a decrease in peak height) but also result in new low efficiency origins dispersed throughout the genome.

Biological replicate assays of strains showed agreement between called peaks as well as peak heights, although the deletion strain did show more variability in peak height and area than wild type and the chimera (Figure 2.5 A). Overall, 128 peaks/origins were identified in wild type, 121 in the chimera, and 88 in the deletion (Figure 2.5 B). The chimera and deletion each had 14 origins that were not active in wild type (7/14 of these origins were the same between the chimera and the deletion), but none were at new sites (sites not documented as origins in other studies). Based on single-stranded DNA analysis alone, I cannot say whether these changes demonstrate an increase in efficiency or just a change in timing. The best way to compare overall efficiency of these increased origins would be via asynchronous 2D gel electrophoresis.[82]

Many origins identified in the mutant strains showed differential activity/efficiency compared to wild type. To compare origins that were significantly up or down compared to wild type, I plotted the log2 value of mutant/wild type as a histogram and highlighted origins that were outside 3 standards of deviation from variation between both wild type and mutant replicates. Separately, I compared origins using a student's two-tailed t-test with a p-value cut off of 0.05. Lastly, I also highlighted origins that were outside the 95% prediction interval of wild type and mutant replicates when mutant peak heights were plotted against wild type peak heights (Figure 2.5 C).

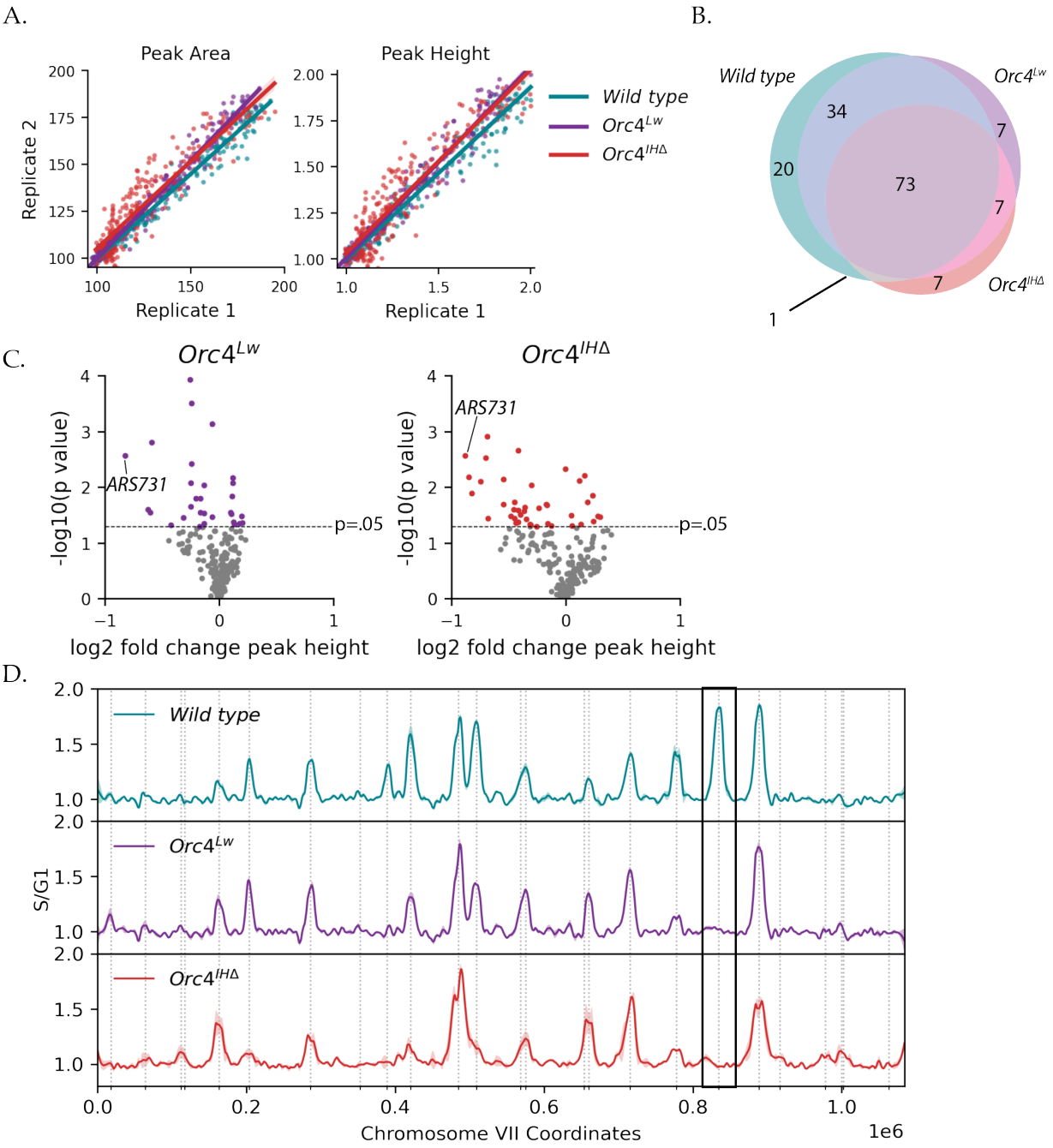


Figure 2.5: Early origin usage in Orc4 mutants. A) Both peak area and peak height were in good agreement between replicates in all assays. B) Venn diagram of active origins in the wild type, Orc4^{Lw} and Orc4$^{IH\Delta}$ strains. A total of 21 origins were unique to the chimera and/or the deletion, but these origins represent changes in timing or efficieny of existing origins. C) Volcano plots of significantly changed origins in the chimera and deletion when compared to wild type. The chimera had 30 significantly altered origins (21 down, 9 up in activity) while the deletion had 42 altered origins (32 down, 10 up). D) Early origin activity across chromosome VII. Mean values of two replicates are shown by opaque lines, while transparent regions represent range of values from individual replicates. The box shows ARS731, a highly efficiencient origin in wild type that has lost activity in both Orc4 mutants.

The first method showed that the chimera had 2 origins that were significantly increased compared to wild type and 29 origins that were significantly decreased. Meanwhile the deletion had 4 origins significantly increased and 28 significantly decreased. Using the student's two-tailed t-test, I found that compared to wild type, the chimera had 20 significantly decreased origins and 10 significantly increased origins while the deletion had 31 origins significantly decreased and 11 origins significantly increased. The third method proved more conservative: the chimera had 4 down origins and 0 up origins while the deletion had 16 down origins and 0 up origins. The following analysis and discussion uses numbers from the first method, as this method had intermediate results between all three statistical approaches utilized.

2.2.5 Early origin fork progression

Because both the chimera and deletion showed greater decreases in origin activity compared to increases and because these changes were not accompanied by large changes in doubling time, I next investigated whether peaks were on average broader in the mutant Orc4 strains. The overall peak broadening, particularly if accompanied by peak splitting as described earlier, could be evidence of population-level changes in average distance that individual replication forks were able to move in the presence of HU. Greater fork movement in strains with fewer origins firing is consistent with slower depletion of the dNTP pool. If fewer origins are firing in a given cell during HU treatment, then there is less immediate depletion of dNTPs meaning that forks should be able to travel farther than they can in a population with more origin firing. With fewer origins active in a given cell cycle, forks must travel farther to ensure complete genome replication, and if cell doubling time remains relatively unchanged (as it does in the deletion and chimera), then fork speed may need to increase to compensate for the lower number of active origins.

While the chimera does have 29 origins that are significantly decreased, these decreases are relatively small, and the overall number of active origins (121) is comparable enough to the 128 in wild type that I would not expect to see much, if any, change in fork speed in the chimera; however, as the deletion has 40 fewer active

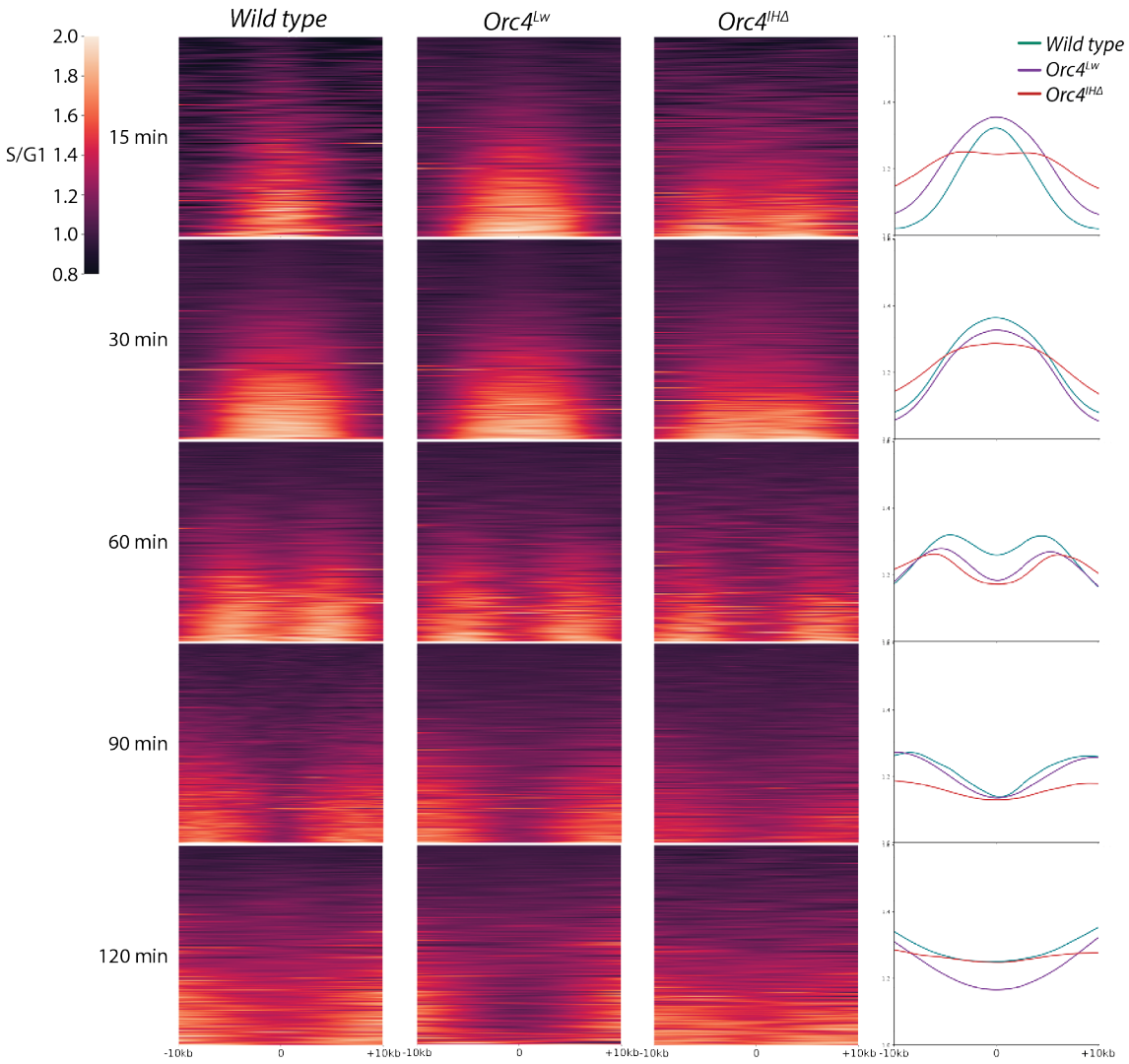


Figure 2.6: Peak widths of active early origins. Forks move away from the origin causing S/G1 peaks to broaden and then split. Heat maps and average peak shape line plots show that fork progression moves at different speeds in different strains. The deletion shows this splitting as early as the 15 minute sample. While the kinetics of peak splitting in the chimera are much more similar to wild type, the average peak shape as shown by the line graph is still significantly different from wild type at every timed sample (Kruskal-Wallis with post-hoc Wilson Cox).

origins than wild type, I hypothesized that some forks are definitely traveling farther in the deletion than in wild type. Figure 2.4 shows heatmaps of peak heights +/- 10 kb from the peak midpoint. As expected, little/no change is visible between the chimera and wild type, but the chimera shows a flattening and broadening of peak height/width characteristic with lower overall activity and greater fork movement.

Additionally, in the fifteen-minute sample, the sum of peak heights at each value shows a cumulative split peak, which also is the result of forks moving away from an active origin. Peak splitting is not visible at thirty minutes, possibly because a greater number of origins are now firing, obscuring the progression of the first few earliest firing origins. Kolmogorov-Smirnoff tests of the cumulative curves shown in the right most panel indicate that for both the deletion and the chimera, the curves differ significantly from the wild-type curve for all sample times. Thus, even the slighter changes in origin activity in the chimera cause a small, but significant increase in the distance forks move (Statistic in Supplemental Table 2.2).

2.2.6 MCM2-ChEC Seq reveals genome-wide ORC binding by proxy

To follow up on early origin activity in the single stranded assay, I proceeded to perform MCM2 Chromatin Endogenous Cleavage Sequencing (ChEC-seq).[83] The unique head-to-head positioning of the MCM2-7 hexamers enables cleavage of the entire MCM footprint simply by tagging a single subunit with MNase. Because stable binding of ORC is required for MCM loading, sequencing the location of MCM footprints in G1 prior to DNA replication initiation reveals locations adjacent to or overlapping with ORC binding sites. Note: while MCMs have been shown to be able to freely slide along DNA, this movement is often impeded by nucleosomes and previous assays of MCM locations reveal most are found in the close vicinity of where the double hexamer was initially loaded at the ACS.[14,44]

First examination of the resulting data for all strains showed that there is no discernable correlation between MCM2 peak height and single stranded assay peak height. Figure 2.7A shows early origins in wild type, chimera, and deletion strains with a single replicate of MCM2-ChEC Seq results plotted on the same X-axis. From this single chromosome, one can see that early origin efficiency does not match the height of MCM2 peaks. This result suggests that MCM2-ChEC seq peak heights do not serve as proxies of origin activity in HU. I refrain from using the term MCM2 occupancy to describe the measurements from the ChEC Seq because at this time I am unable to determine if the magnitudes in signal reflect true MCM occupancy at those sights,

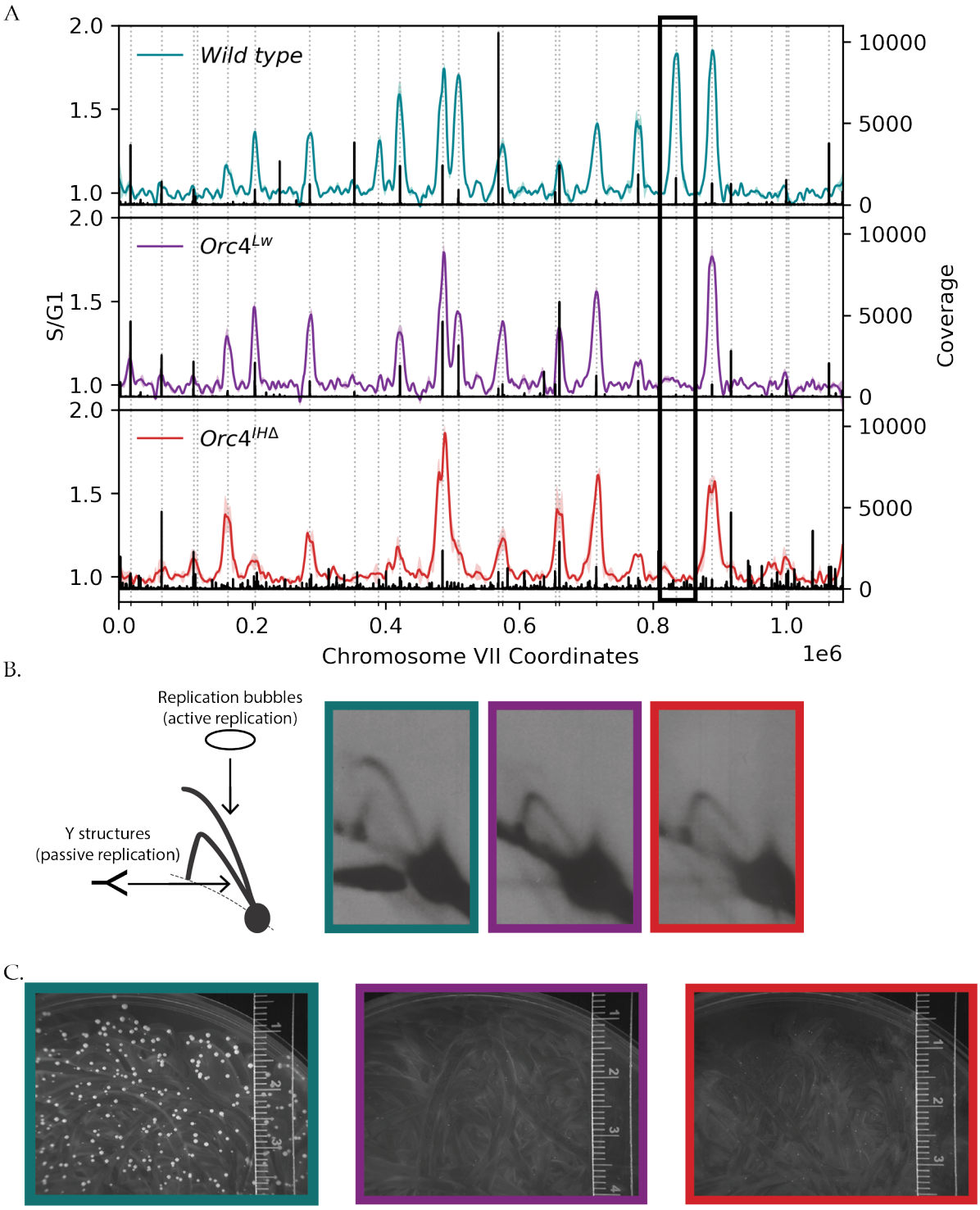


Figure 2.7: MCM2-ChEC signal compared to early origin activity. A) MCM2-ChEC signal shows MCMs at early origins, but there is no visible correlation between MCM signal and peak height. *ARS731* does show less MCM signal in the chimera and deletion where it is inactive. B) Asynchronous 2D gel analysis of *ARS731* shows that *ARS731* remains inactive throughout S phase in the chimera and deletion while it is actively replicated in wild type. C) *ARS731* plasmid transformations show robust colony formation in wild type by day three but only produces pin-prick sized colonies in the chimera and deletion.

biased PCR amplification during sequencing, or access of nuclear compartments to the influx of calcium. The precise nature of ChEC-Seq cleavage means that signal resolution is finer; in fact, many of the fragments generated are between 60-70 base pairs, the size of the MCM footprint. However, this precision also makes it difficult to differentiate between clonal signal resulting from MCMs occupying the same footprint in multiple cells versus PCR bias during amplification. In typical sequencing analysis with DNA sheared via sonication or other less precise methods, PCR duplicates are eliminated by removing reads that start and end at the exact same location. For MCM2-ChEC Seq and methods such as MNase-ChIP or ATAC-Seq that generate cleavage via precise enzymatic means, eliminating reads that start and end at the exact same location is not justified and removes most of the signal.

To distinguish between biologically relevant clonal cleavage sites and PCR bias, I will need to use unique molecular identifiers (UMI) in future library preps. If UMIs are added during library prep, I can uniquely mark every DNA molecule created from MNase cleavage prior to PCR amplification of the library. After sequencing the library, I can filter the resulting data so that UMIs are only present once, eliminating reads that result from sequencing multiple PCR products of the same original molecule. This method will allow me to still retain the biological clonal signal that results from cleavage of the MCM footprint. However, it does not resolve the question of calcium accessibility of different portions of the nuclei during cleavage.

Although I could not directly compare MCM signal to origin peak height, I then compared MCM signals between strains to see if differentially active origins also had relatively different MCM signals at those origins. Looking at *ARS731* in Figure 2.7A (boxed in black), it is clear that the signal is reduced at the origin in both the chimera and deletion. This lack of MCM signal suggests that *ARS731* is not utilized in these strains at any point during DNA replication.

To confirm that the chimera and deletion have lost the ability to recognize *ARS731* and utilized it as an origin, we performed two assays: 2D gel electrophoresis of *ARS731* from asynchronously growing cultures and transformation assays using *ARS731* as the plasmid's primary origin. [82] Both assays (Figure 2.7 B and C)

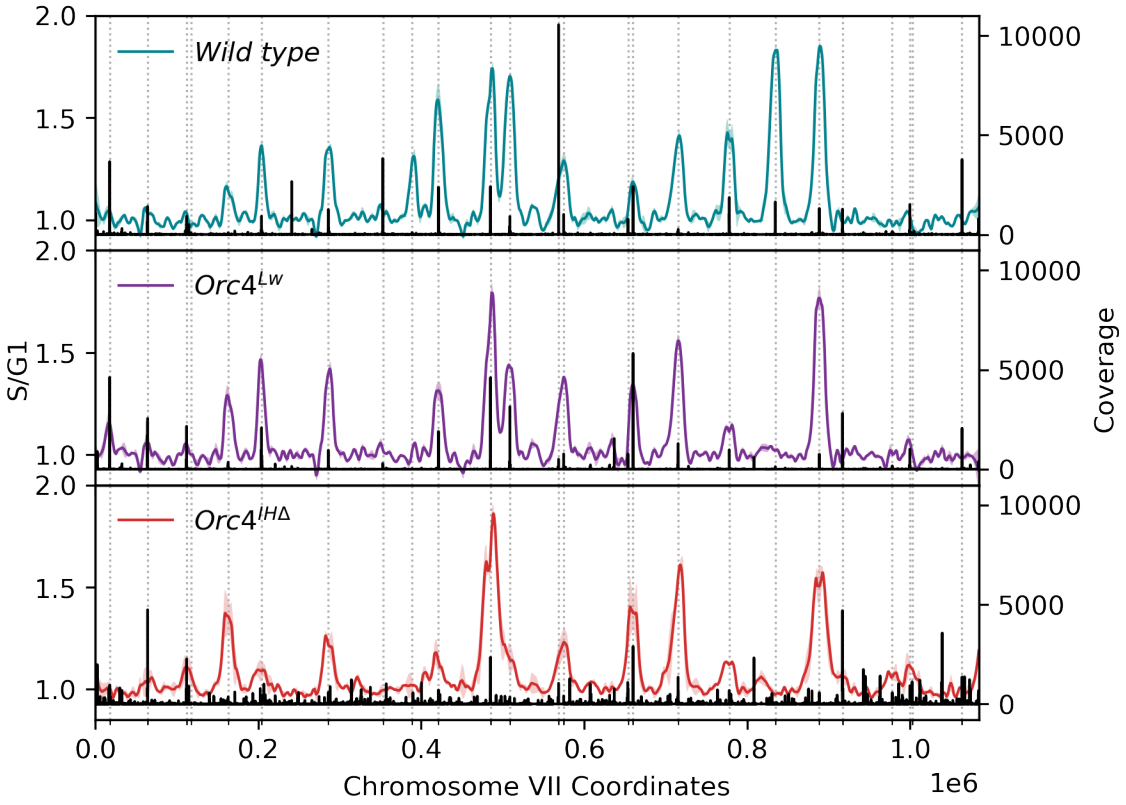


Figure 2.8: Increased number of sites of loaded MCM in the deletion relative to wild type or chimera. The ssDNA profiles from 30 minutes are overlaid with the MCM-ChEC signals. All strains have been down-sampled to the same number of sequencing reads per sample (per genome). The deletion strain's MCM signal is distributed across far more sites than either wild type or the chimera.

demonstrate that the chimera and the deletion do not use this origin of replication at any point in S phase nor can they recognize the origin sequence on a plasmid. Another noticeable trend in the sequencing data, is the greater amount of pervasive MCM2-ChEC signal in the deletion compared to wild type and even more pronounced in the chimera. Figure 2.8 shows MCM signal on chromosome VII, where the greater number of sites of signal can be easily seen. Even though the data plotted have been down-sampled so that each sample has the same number of sequencing reads genome-wide, the deletion had far more places with visible MCM signal. This pervasive signal could be due to reduced specificity of ORC binding that enables loading of MCMs at more sites in the genome. However, these sites of MCM loading do not fire in early S phase of HU treatment and may not fire at all in cells. An S-phase time course of 2D gels would help determine if these sites act as origins during normal or HU-treated S phase.

Because these data were acquired recently, there is much analysis that remains to be done. I plan to conduct much more thorough motif analysis of MCM sites from this dataset, which I hypothesize will show that the top sites in the chimera produce a motif more similar to the *L. waltii* ACS, particularly at the bases contacted by the Orc4-IH. I anticipate that the deletion MCM sites will produce a weaker motif that shows less specificity than either wild type or the chimera. Meanwhile, I hypothesize that the wild type will recreate the *S. cerevisiae* ACS. I also would like to make full use of the MCM2-ChEC dataset that includes larger reads indicative of nucleosomes surrounding the ACS, given that nucleosome positioning helps determine which origins are used. I hypothesize that looking at the MCM signal in conjunction with the nucleosome signal may be a better predictor of origin activity than MCM alone. Specifically, I hypothesize that MCMs loaded at sites where the ACS is asymmetrically positioned between two adjacent nucleosomes will correspond with origin efficiency.

2.3 The role of Orc1 in sequence specificity

Orc4 only contacts bases 9-11 and 13 of the *S. cerevisiae* ACS when just ORC is bound to DNA.[6] Both Orc1 and Orc2 make contacts with DNA bases at or upstream of the Orc4 contacts (Figure 2.1). However, the regions that contact DNA are not independent regions of the protein separate from the larger structure (limited in interaction/contribution to the tertiary structure) as the Orc4-IH loop is. This organization means that partial deletion of these DNA contacting regions (as was achieved in the Orc4$^{IH\Delta}$ mutant) is more likely to have effects on the tertiary structure of the protein, potentially impacting its stability and ability to bind with ORC and MCM complex members. As I wanted to isolate the DNA contacting ability as much as possible, I aimed to make complementary *L. waltii* chimeric mutations to the DNA contacting regions of Orc1 (the DNA contacting residues of Orc2 are conserved between *L. waltii* and *S. cerevisiae*.) Of the 3 DNA bases within the major groove of origin DNA only F360 is altered in *L. waltii* where it is a tyrosine, but between F360 and Y372, where the major groove contacting bases in Orc1 are found, there are two additional differences between *L. waltii* and *S. cerevisiae*. One is a serine at the region

corresponding to N366 and the other is an alanine at the site corresponding to *S. cerevisiae*'s K371. The Orc1 chimera I created has the N366S change, but not the K371A change. In addition, I generated a double chimera mutant containing both the Orc1 and Orc4 *L. waltii* mutations.

2.3.1 Orc1 and Orc1/4 strain viability and plasmid maintenance

Both the Orc1 and Orc1/4 chimeras had doubling times comparable with wild type. The Orc1 chimera had a doubling time of 93.7 ± 2.1 minutes while the double chimera had a doubling time of 90.5 ± 1.1 minutes, based on five biological replicates. While the difference in doubling time between wild type and the chimera is not statistically significant (student t-test P value of 0.13), there is a general decreasing trend of doubling time in the Orc4 chimera and the Orc4 double chimera that I and others in lab have repeatedly noticed while working with these strains. S phase progression data for the Orc1 chimera and double chimera are similar to wild type (Supplemental Figure 2.2).

Repeating the same plasmid maintenance assays as before, I found that the double chimera had lower loss rates than wild type as well as the Orc4 chimera for the *L. waltii* plasmid while the Orc1 chimera loss rates were not significantly different from wild type. This outcome was somewhat unexpected because the *L. waltii* plasmid ACS has the largest divergence from the *S. cerevisiae* ACS at residues 7 and 8 (ATs instead of TAs), the region contacted by Orc1. In the cryo-EM structure of ORC around *ARS305,* the 8th residue of the *S. cerevisiae* ACS is contacted by F360, which is changed to tyrosine while T7 is contacted by R366, which remains unchanged in the Orc1 chimera. However, this residue is adjacent to the A365S mutation in the chimera, which could alter its positioning within the major groove. Additionally, while the Y372 residue, which also contacts T7, is conserved between species, the preceding residue is not. The lysine at position 371 in *S. cerevisiae* is perpendicular to the DNA backbone, positioning the tyrosine in the major groove. It is possible that the *L. waltii* DNA specificity at the 7th residue of the ACS relies on the preceding alanine to change the tyrosine's position. Because the Orc1 chimera I made lacks this change, it is possible that the tyrosine

remains in a position that favors the *S. cerevisiae* ACS contact over the *L. waltii* ACS contact.

2.3.2 μARS317-Seq reveals potential B2 binding of Orc1 chimera:

I performed μARS317-Seq on both the Orc1 chimera and the double chimera to determine their tolerance for mutations in the ACS and compare these results to both wild type and the Orc4 chimera. Based on the cryo-EM data, I expected the Orc1 chimera to potentially show tolerance of adenine at position 7 in the ACS (T in *S. cerevisiae* and A in *L. waltii*) as this region is contacted by the Orc1-BP upon initial ORC binding. Likewise, I expected the double chimera to share this tolerance with the Orc1 chimera as well as show tolerance to the same mutations tolerated by the Orc4 chimera. However, my hypotheses were proved incorrect on both counts.

In contrast to all other strains analyzed, the Orc1 chimera had high, somewhat indiscriminate tolerance for mutations across the ACS with only a few changes being intolerant: T3G, A8C, A10G, and T13C were the only changes not tolerated across the ACS (Figure 2.9 A). The other mutations showed even greater enrichment than the sequence surrounding the ACS, which was not a part of any specific element. This finding led me to hypothesize that perhaps the canonical *ARS317* ACS was not being utilized as the primary ORC binding site in the Orc1 chimera. Consistent with this hypothesis, the B2 element of *ARS317* and the surrounding areas showed far greater intolerance of mutations (Figure 2.9 B). However, the region in which mutations are not tolerated is so large that it is hard to identify where in or around the B2 element that ORC may be binding in the Orc1 chimera. There are many slight matches to an ACS on the T rich negative strand of the B2 element, so it could be that ORC is capable of binding promiscuously along this area. Additionally, while one region under selection is the Forkhead1 binding site, this region is also the predicted ACS in Nieduszynski et al (2006), so selective pressure there could be coming from its use as an ACS rather than or in addition its use as a Fkh1 binding site.[84] To easily investigate this in future work, I

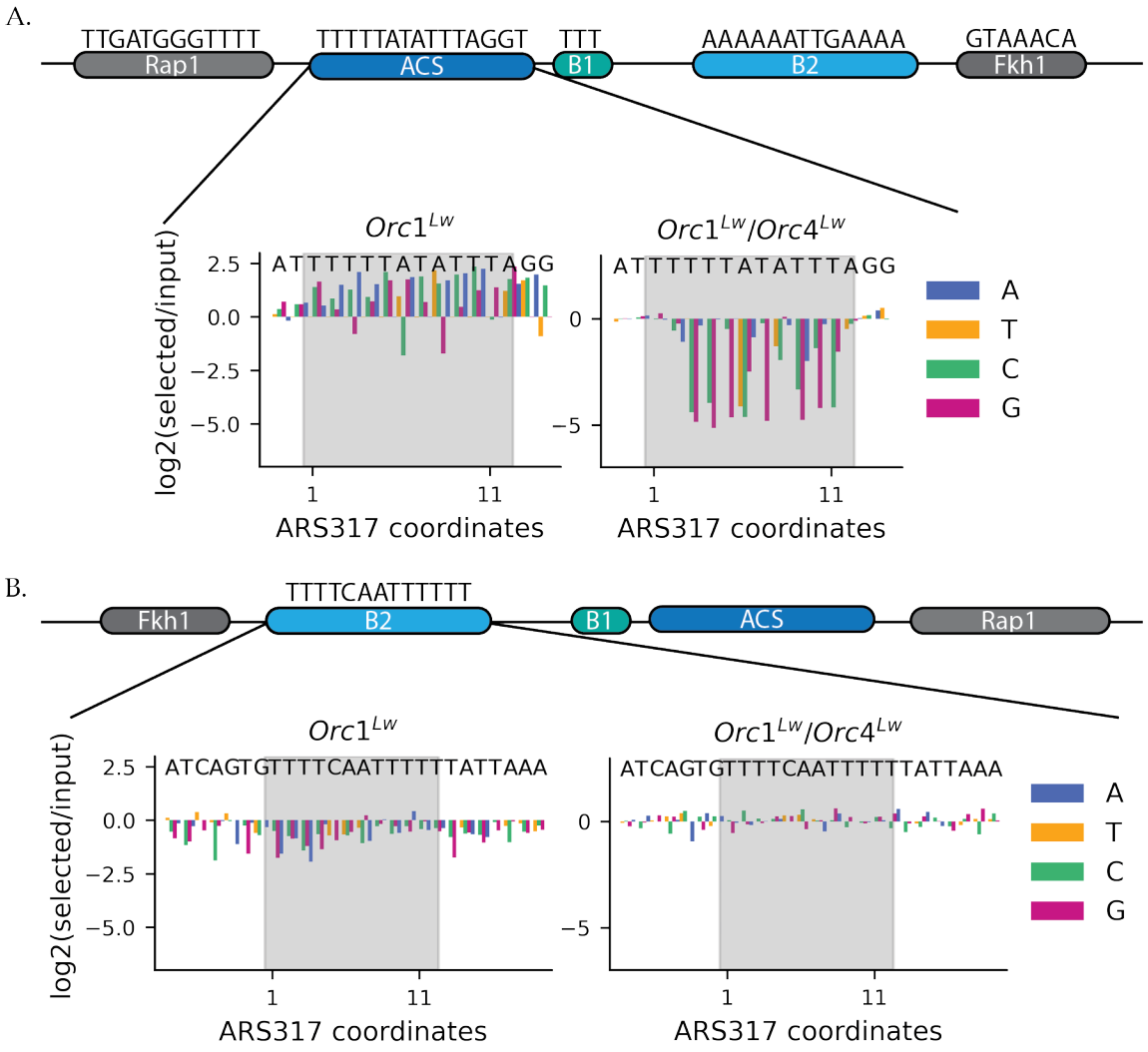


Figure 2.9: µARS317-Seq of Orc1 chimera and Orc1/4 double chimera. A) The Orc1 chimera shows wide tolerance of mutations at the ACS while the double chimera is intolerant of ACS mutations. B) The Orc1 chimera does not tolerate mutations to the B2 element, suggesting that this may be where ORC is binding. The double chimera tolerates B2 mutations.

could simply repeat this experiment in a Fkh1 deleted strain and see if mutational tolerance in this region increases.

In contrast to the Orc1 chimera, the double chimera was intolerant of mutations at the ACS and tolerant of mutations at the B2 element. Unexpectedly, the double chimera also did not show the same level of tolerance of the A10G mutation as the Orc4 chimera. Because the double chimera shows lower loss rates of *L. waltii*-specific plasmids, I expected that tolerance in the double chimera would match if not exceed that of the Orc4 chimera for changes that mimicked the *L. waltii* ACS. However, overall, the

double chimera was more tolerant of ACS mutations than the Orc4 chimera as the magnitudes of variant depletion ACS was smaller than in the Orc4 chimera (full graphs in supplemental figure 2.3).

2.3.3 Analysis of early origins in Orc1 chimera strains

Given the results of plasmid maintenance in the Orc1 chimera strains, I hypothesized that the Orc1 chimera alone would have ssDNA profiles similar to those of wild type, or at least more similar than that of the Orc4 chimera that was capable of maintaining new origins. Because the double chimera exhibited lower plasmid loss rates and no change in doubling time, I expected the double chimera to show more altered origin usage than the Orc4 chimera alone, including potentially using new origins. Biological replicates of the strains showed less agreement between replicates than either wild type or the chimera and were more similar to the variability between assays observed with the deletion. However, because I only count origins that passed thresholds in both strains and calculate significant peak differences using values from both replicates, these differences in replicates should only decrease my statistical power to identify differentially active peaks but should not result in false positives being reported.

I found that the Orc1 chimera had 108 origins surpassing the threshold while the double chimera had 174 (Figure 2.10). All but two of the Orc1 origins were shared with wild type while 95 of the origins were shared with the double chimera. There were no origins in the Orc1 chimera that were not shared with either wild type or the double chimera. However, in the double chimera, 13 of the 174 peaks occurred at sites that are not annotated origins in previous work (as documented by OriDB).

Because the Orc1 chimera had an intermediate number of active origins compared to the deletion and the Orc4 chimera, I hypothesized that the extent of fork movement in this strain (as shown via peak with and average peak shape) would be between that of the Orc4 chimera and the deletion. I thought that there would be more peak broadening and splitting than in the Orc4 chimera, but not as severe as seen in the Orc4-IH deletion. Meanwhile, because the double chimera had such an abundance of active

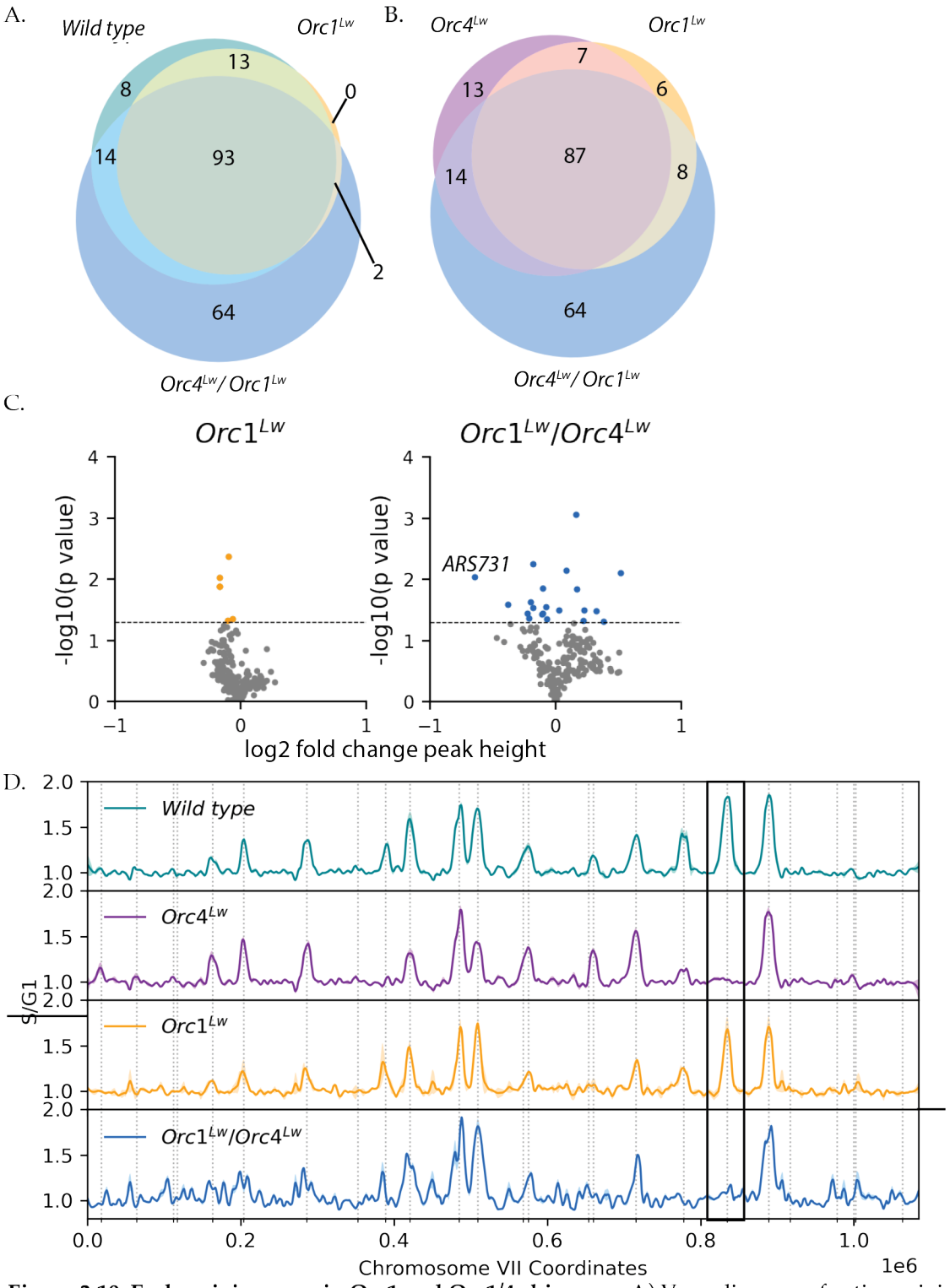


Figure 2.10: Early origin usage in Orc1 and Orc1/4 chimeras. A) Venn diagram of active origins in wild type, Orc1 chimera, and double chimera. No origins are unique to the Orc1 chimera. B) Venn diagram of active origins in chimeric strains. C) Significantly altered early origins in Orc1 chimeras. There were 5 decreased origins in the Orc1 chimera; the double chimera had 21 significantly altered origins (13 down, 8 up). D) Early origin activity in chromosome VII across chimeric strains. The Orc1 chimera retains early firing of *ARS731*, while the double chimera does not.

origins, but only a negligibly faster doubling time, I anticipated that fork distances in this strain would be decreased, especially considering the limited pool of dNTPs available for continued replication.

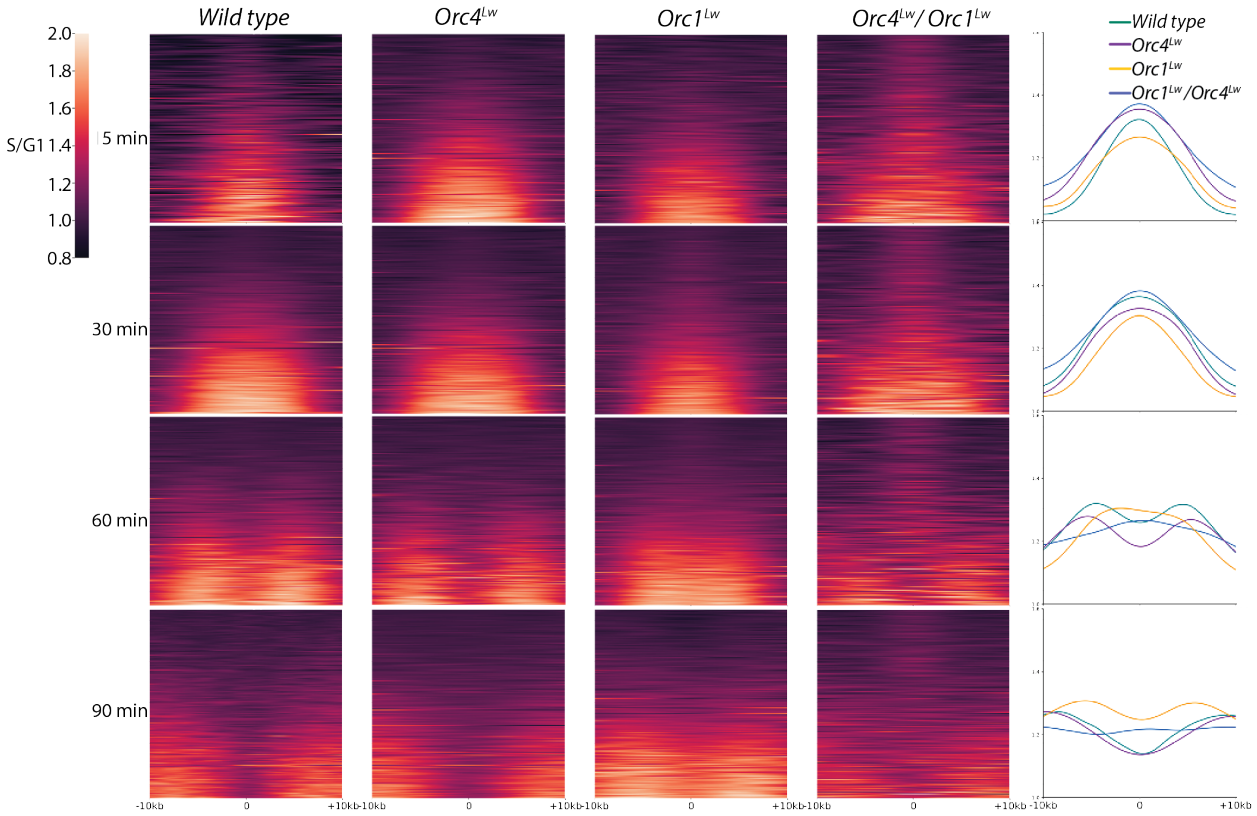


Figure 2.11: Peak widths in Orc1 and Orc4 chimeric strains. The Orc1 chimera exhibits lower overall peak height, suggesting origin efficiency is reduced; moreover, these peaks are slow to split, meaning forks progression is slower in this strain compared to others. The double chimera appears to have two sets of peaks; the brighter (taller) peaks that split over time and dimmer, more narrow peaks that persist throughout the timecourse.

In opposition to my prediction, the Orc1 chimera heatmaps and average peak shape show little to no expansion of replication compared to either the Orc4 mutants or to wild type (Figure 2.11). In fact, peak broadening seems to be delayed by at least 30 minutes. The Orc1 chimera's 90-minute sample looks much closer to the 60-minute sample in wild type and the Orc4 chimera. However, in order to confidently identify this difference as a pattern of the Orc1 chimera, the experiment needs to be repeated several more times. This kind of profile may just indicate a slow release into S phase rather than a true slow progression through early S phase in the Orc1 chimera. While the same quantities of alpha factor (the mating hormone used to arrest budding yeast in

G1) and pronase (used to degrade alpha factor and thus trigger cells' release into S phase) are used in each assay, slight batch-to-batch variations in reagents may cause slower release. That said, untreated controls are released in tandem with treated controls during this assay to ensure that S phase progresses as expected in these strains, and these untreated controls showed no delay. This normal untreated release could indicate that the Orc1 chimera has an increased sensitivity to HU treatment.

Meanwhile, in the double chimera, the heatmaps suggest that there are two classes of origins—one where forks progress over the time-course producing split peaks and the second one where replication initiates but the forks remain stationary. The cumulative peak shape produced by these two classes of origins is different from what was seen in any other strains (Figure 2.11, statistics in Supplemental Table 2.3). To better dissect the changes occurring in the double chimera, I repeated the peak width analysis looking at only the top 100 ssDNA peaks in each strain (Figure 2.12, statistics in Supplemental Table 2.4).

With this new threshold, I see that the double chimera peaks broaden and split over time as expected, however the shape of the average curve is still quite different than the other chimeric strains or wild type at certain time points. Overall, the initial peaks are broader and shorter than the Orc4 chimera or wild type. At 60-minutes, this trend is accompanied by a bimodal peak somewhat consistent with a split peak, but with a shallower valley around the initiation site than seen in the Orc4 chimera or wild type. Given the class of unmoving peaks and this shallow valley in the 60-minute sample, it is distinctly possible that these differences in the double chimera result from technical noise that is higher in these samples than in other samples. Greater noise would decrease the overall magnitude of ssDNA peaks called and potentially result in an increase in false positive peaks being called. If the same pattern of noise persists in both replicates, then these false positive peaks would pass cross-replicate comparisons designed to remove replicate-specific technical noise. Consistent with this hypothesis, unlike replicates for the other strains, the replicates for the double chimera were collected, prepped, and processed concurrently using array slides from the same batch.

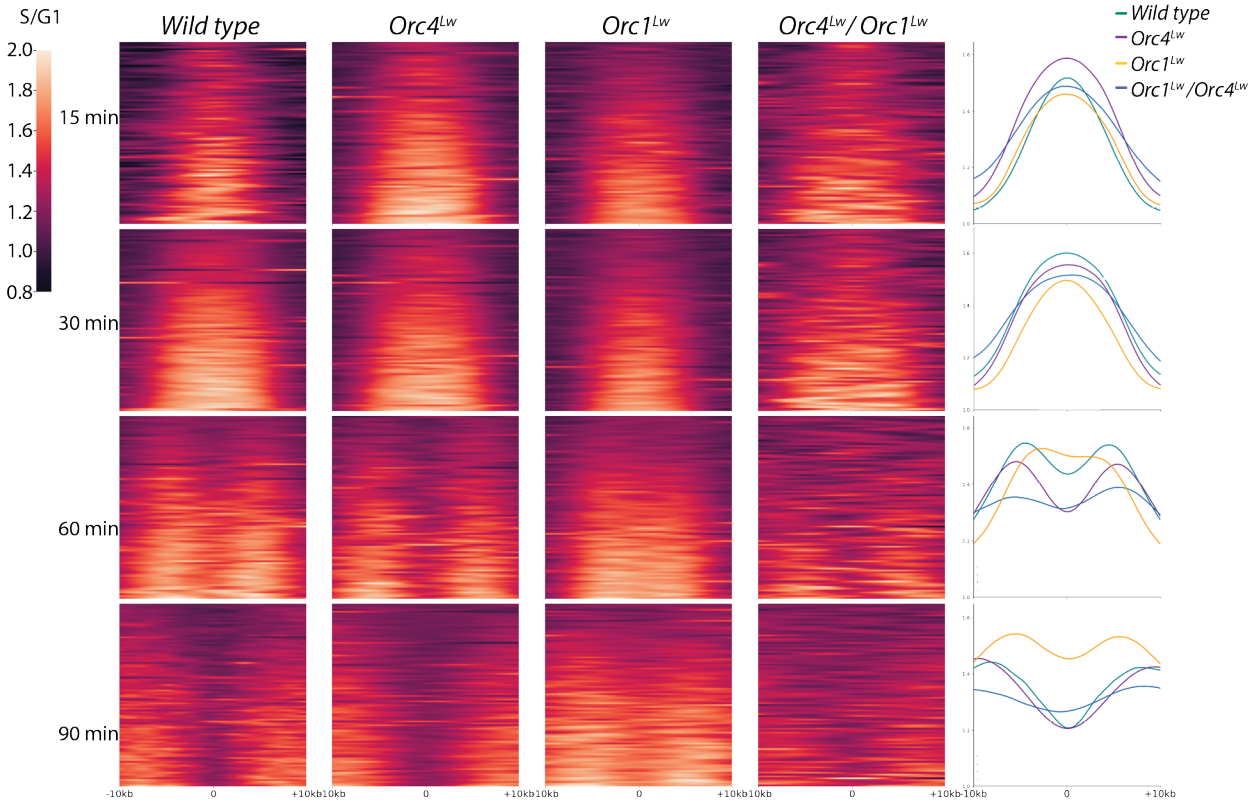


Figure 2.12: Top 100 peaks in Orc1 and Orc4 chimeric strains. The top 100 origin peaks determined by area preserve the lack of movement observed in the Orc1 chimera while they only show the class of peaks in the double chimera that split; however, the average peak shape of the top 100 peaks is still significantly different between the double chimera and wild type.

2.3.4 MCM2-ChEC Seq in Orc1 and Orc1/4 double chimeras

Preliminary analysis of the MCM2-ChEC Sequencing of the Orc1 chimera and the double chimera was limited due to low read depth on the Orc1 chimera, but the same general trends were observed in the Orc1 and double chimera as the other strains. To account for the low read depth of the Orc1 chimera, I multiplied the coverage by 2 in Figure 2.13 A. For final analyses and subsequent publications, I will re-sequence the libraries to achieve better coverage.

Examining *ARS731,* the Orc1 chimera shows MCM signal at the site, while the double chimera has no visible signal. In asynchronous 2D gels at the origin; however, the double chimera does appear to have some activity at *ARS731*. This activity is significantly reduced compared to wild type. Likewise, the plasmid transformations

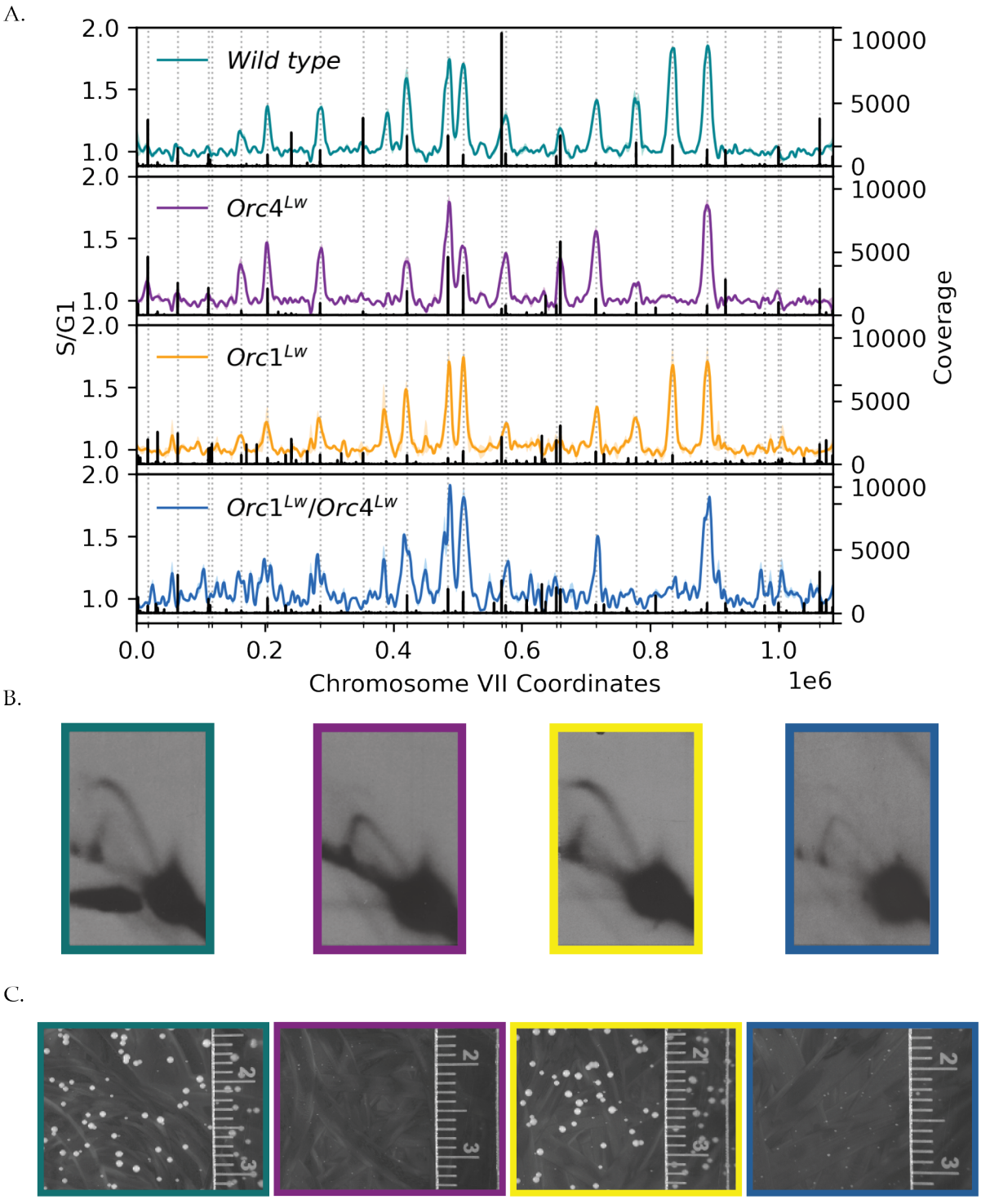


Figure 2.13: MCM ChEC vs early origin activity and lost origin verification. A) MCM2-ChEC signal aligned with early origin activity shows no correlation between the two assays in the Orc1 or Orc1/4 chimeras. B) Asynchronous 2D gel of *ARS731* shows active origin initiation in Orc1 chimera and barely detectable origin activity in the double chimera. C) Transformations with the *ARS731* plasmid show robust colony growth in wild type and the Orc1 chimera. The double chimera has slightly larger colonies than the Orc1 chimera.

appear to show minor restoration of *ARS731* recognition in the double chimera because after 3 days of growth, the double chimera transformant colony size is slightly larger than that of the Orc4 chimera transformants.

2.4 Discussion

The results of these experiments clearly demonstrate that the Orc4-IH does drive sequence specificity to some degree in *S. cerevisiae* and likely in other budding yeasts with sequence-specific origins. With the initial plasmid transformations alone, I determined that replacing the *S. cerevisiae* Orc4-IH with the putative *L. waltii* Orc4-IH was sufficient to enable recognition of *L. waltii* specific origins, demonstrating both that the region provides sequence specificity in *S. cerevisiae* and that the corresponding region in *L. waltii* enabled recognition of the *L. waltii* ACS. These results indicate that the Orc4-IH likely helps drive sequence specificity of *L. waltii* origins as well.

The emergence of two identical deletions from the *K. lactis* transformation attempts suggest that this deletion arose either via microhomology on the *K. lactis* template provided or potentially that this deletion was one of few viable ways of ablating the Cas9 cut site. While there appear to be subtle differences in the orientation of the DNA contacting amino acids, the DNA contacting residues are retained. Moreover, a full turn of an alpha helix is approximately 3.6 residues, so deleting the 6 amino acids between the DNA contacting residues approximately 120 degrees under-rotated. Cryo-EM of the Orc4-IH deletion in complex with ORC around DNA would be enlightening as to how these bases now interact with DNA in absence of the alpha helix structure and whether the orientation of the side chains has changed in this new configuration. µARS317-Seq results clearly indicate sequence specificity akin to that of *S. cerevisiae* and *L. waltii* at the bases typically contacted by the residues as a part of the alpha helix.

Subsequent experiments and results revealed that sequence specificity is more complex than a stable interaction between the Orc4-IH and the four bases in the ACS contacted by the Orc4-IH in the cryo-EM structure of ORC on *ARS305*. In *Drosophila,* DNA topology helps define origin location, and it could be that DNA topology also influences yeast ORC binding specificity as well.[85] Alternatively, during the dynamic

process of Cdc6 binding and subsequent helicase loading, the exact location of ORC on DNA may shift, driving contacts with neighboring bases. Either DNA topology and/or additional DNA contacts with ORC and pre-RC components must influence the ACS as the ORC-DNA structure only shows contacts with a subset of ACS nucleotides. Given that the ACS is not a gapped motif, other factors must drive the specificity of the bases in between the contacts shown in Figure 2.1.

While the Orc1 chimera was created primarily with the intention of creating the double chimera so that all DNA contacting residues would be changed to that of *L. waltii*, the Orc1 chimera revealed unexpected and novel results in both the μARS317-Seq and ssDNA assay experiments. The pattern of tolerated variants in the μARS317-Seq experiment suggests that the Orc1 chimera may prefer the B2 element as its primary binding site. Similarly, MCM2-ChEC Seq showed that the average MCM peak maximum overlapped with the best match for the ACS at each MCM peak site, suggesting that ORC is binding adjacent to the ACS and loading MCMs over the ACS. Furthermore, the ssDNA assay showed little origin usage loss or change, but the time course showed that fork progression had been significantly slowed compared to all other strains analyzed. While further repetitions would be necessary to confirm that this result is true and not due to batch-to-batch variation in alpha factor or pronase used during arrest and release, there is also a potential biological explanation for this reduced distance that forks move. Previous research showed that ACS matches in the *S. cerevisiae* genome that served as origins had characteristic nucleosome placement directly upstream of the ACS.[33] If the Orc1 chimera is using a different binding site within the canonical ARSs, it may be that the altered binding of ORC changes the location of ORC compared to surrounding nucleosomes and prevents the nucleosome remodeling around origins that is needed for efficient initiation.[15] Other recently published research on the Orc4-IH found that a complete deletion of the Orc4-IH altered binding specificity of ORC and eliminated the nucleosome repositioning expected around ORC binding sites.[86]

An unexpected finding in this work was the restoration of origin activity in the Orc1/Orc4 chimera to origins that were inactive in the Orc4 chimera, particularly

ARS731. It could be that matching the DNA contacting regions to be of the same species has a stabilizing effect on ORC binding. Understanding through what means this stabilizing occurs would require imaging of the Orc1/Orc4 chimera structure. A second explanation could be rooted in the results of the Orc1 μARS317-Seq where it appeared that the Orc1 chimera utilized part of the B2 region as its primary binding site rather than the canonical *ARS317* ACS. Perhaps the Orc1 chimera's use of *ARS731* is enabled by a secondary binding site, potentially the B2 element of *ARS317*. Because the ssDNA assay does not have sufficient resolution to determine where in the region that replication initiates, the results would be indistinguishable from wild-type *ARS731* utilization. In the case of the Orc1/Orc4 chimera, therefore, the explanation would be that Orc1 enables binding and loading of the pre-RC at the B2 element, which is actually destabilized or decreased by the Orc4 chimeric mutation. This hypothesis would explain the reduced efficiency of the double chimera compared to the Orc1 chimera. This hypothesis may be testable by examining the MCM2-ChEC Seq data for an ORC footprint adjacent the MCM footprint on the opposite side that it may be found in wild type. Given the spacing and opposite orientation of the B2 element and the ACS, binding to either site would load MCMs in the same general space, but occasionally resolution is high enough to see a roughly 30 bp gap between the MCM footprint and adjacent nucleosome footprints that is the appropriate size of ORC. Alternatively, I could generate a μARS731-Seq library to determine where mutations are tolerated in either species.

Ultimately, this work leaves many open avenues of exploration. I did show that the Orc4-IH is sufficient to alter origin sequence specificity, but further investigation showed that multiple ORC subunits contribute to origin sequence specificity in ways that are not fully explained by DNA-ORC contacts. Moreover, an entirely new question was raised by the MCM2-ChEC Seq experiments: is MCM occupancy actually a proxy for origin efficiency? The data here contradict recent findings and suggest that there are instances where MCM occupancy and origin efficiency are uncoupled. Probing this matter further could provide new biological answers and inform the limits of MCM assays.

2.5 Materials and Methods

Single Mutation Strain Construction: Chimeric Orc4, Orc1 and Orc4$^{IH\Delta}$ strains were constructed from BY4741 transformed with a plasmid encoding Cas9 and a guide RNA targeting the region of interest along with 1-3 μg of DNA template generated from PCR using the donor strain genomic DNA as a template. 8x50 μL PCR reactions were pooled, precipitated with ethanol, and resuspended 20 μL TE 0.1 (10mM Tris, 0.1mM EDTA, pH 8) for use in transformation. Cells were grown in log phase for 2+ doublings before LiAc transformation with ~100 ng plasmid and the 1-3 μg template. Transformants were plated onto -URA and allowed to grow for 3-5 days before streaking to single colonies. Transformant sequence was confirmed first via PCR and restriction enzyme digest and then by Sanger sequencing. A list of all strains used in this study can be found in Supplemental Table 2.4.

Double chimera strain construction: Double chimeric Orc4/Orc1 was made by first mating the chimeric Orc1 strain with BY4742 MATα to produce Orc1 chimera MATα. The MATα chimeric Orc1 was then mated with the chimeric Orc4 strain. PCR, restriction enzyme digest, and Sanger sequencing were used to verify the Orc1/4 chimeric strain. Only a single double chimera was created and used for the experiments described.

Species-specific plasmid transformations: Plasmid transformations were carried out in the same conditions used to make the initial strains. 100 mL yeast cells were grown in log phase for at least two doublings before harvested at ODs between 0.5 and 1.0. Each individual transformation was done with the volume of culture where ODxVolume = 5 (10 mLs for a culture at OD = 0.5 and 5 mLs for a culture at OD = 1.0). Cells were spun down and washed 2x with TE-LiAc, and then resuspended in transformation buffer containing 200 ng of target plasmid. For plasmids with KanMX selectable markers, cells were plated on D and grown for 12-18 hours before replica plating on D+G418. For plasmids with *URA3* selectable markers, cells were plated directly onto -ura plates. Transformants were imaged at days 3 and 5 of growth on selectable plates before being streaked to single colonies for freezer stocks and plasmid loss assays.

Plasmid Loss Assays: Initial plasmid loss assay for LwARSVI-582 (data not shown) were conducted on plasmid transformed strains by growing cells to log phase in selective media (D+G418) before cells were spun down, washed twice with non-selective media and then started in 40mL cultures at an OD of 0.1. At the initial timed sample, and all subsequent timed samples, 250 cells were plated in triplicate on D plates and cells were collected for future Southern blot analysis. Culture samples were taken every six hours where cultures were again cut back to OD = 0.1 to remain in log growth throughout the experiment. After 1-2 days of growth on D plates, colony number was counted before replica plating onto D+G418. Plates were incubated for 2 days and then colony counts were attempted; however, replica plating obscured adjacent colonies too much for accurate counting, so colony count data (important for plasmid segregation analysis) was foregone, and Southern blots were used at the determinant of plasmid replication.

Single-stranded DNA assay: Performed as described in Feng et al. 2006.[81] Log phase cells (~12-18 hours of growth from liquid culture) were arrested with 3 mM alpha factor (1:1000 ratio) for 1.5-2 doublings until microscopy revealed 0% new buds and <5% large buds. 200 mL cultures were collected as the G1 sample, 1 mL was taken for a FACS sample, and an additional 10-50 mL was transferred into a new flask for an untreated control examination of S phase progression. Next, HU was added to primary culture to concentration of 200 mM. The culture was left for 15' for HU to go into solution. Both untreated and treated cultures were then release with pronase. 10' FACS samples were taken for the untreated control while 15', 30',60' ,90', and 120' samples were collected from the treated culture for microarrays and FACS analysis. Two biological replicates for each strain were collected and processed.

ssDNA analysis: Microarrays were processed using the ArrayScripts GUI developed in our lab. Then, data were normalized according to the mode value across all chromosomes with a custom R script written by MK Raghuraman. Early active origins were identified using the SciPy script find_peaks with a minimum height of 1.1,

prominence of 0.16, and a minimum distance between peaks of 10kb.[87] Origins/peaks were separately called in each replicate and in the mean of the replicates. Peaks that were within 6.5 kb of each other were matched as the same peak. Only peaks identified in both replicates were used in subsequent analysis. For matching peaks/origins between strains, the same 6.5 kb window was used to identify overlapping peaks. If a peak was missing in the strain, then the max value within +/- 6.5 kb of the mean peak midpoint was recorded as 'peak height' for volcano plots and heat maps. Volcano plot p-values represent p-values taken from student's t test comparison of both replicates of ORC mutant versus both replicates of wild type; for simplicity, the points plotted are the mean values of each strain.

For peak width heatmaps, peaks were centered at the peak midpoint. The peak midpoint was defined as the center value between the left and right bounds of the peak at half prominence. This method better centers split peaks that would otherwise be skewed to one side, depending on which point of the split peaks was higher. The efficacy of this strategy can be seen in the time course of heatmaps that shows the local minima of split peaks to be roughly centered on the heatmap. Values plotted are of the mean of biological replicates. For strains without identified origins/peaks at certain sites, the heatmap is centered on the mean coordinates of peak midpoints in strains where that peak has been called. Summary peak shape plots were made by summing the heatmaps vertically and dividing by the number of rows (peaks) to produce an average curve shape for all sites in the heatmap. Analysis for the top 100 peaks of each strain was conducted in the same fashion, excluding the lower value peaks at each sampled time.

ssAssay plotting: ssAssay plotting was done primarily with the python package matplotlib.pyplot[88]. Comparison scatter plots of ssAssay replicates were made with seaborn.regplot with a 95% confidence interval highlighted.

ssAssay statistical analysis: Statistical analysis was performed using the Scikit.stats python package. Students' two tailed t-test was used for identifying significantly up or

down peaks. Kolmogorov-Smirnoff tests were used for a comparison between wild type and Orc4-IH deletion summary peak shapes while a Kruskal-Wallace with posthoc Conover's was used for a multiway comparison of wild type and the Orc4-IH, Orc1, and Orc1/4 chimeras.

Protein structure predictions and alignments: Structural predictions of Orc4 in the budding yeast clade were done via I-TASSER.[73] Sequence alignments were based on structural alignments in PyMOL. All protein structure images were also taken from PyMOL visualizations. For chimeric Orc4 and Orc1 mutations in PyMOL, the orientation of new amino acid side chains were residues that minimized conflict with the *ARS305* DNA from 5zr1 (ORC-DNA) in PDB.[6] Because the Orc4-IH deletion shrank the Orc4-IH, this modeling was not possible as the predicted structure intersected the DNA, thus interactions between the Orc4-IH deletion and origin DNA could not be predicted.

μARS317 Sequencing: Frozen competent yeast strains were transformed with a CEN-less *μARS317*-Seq library.[80] *ARS317* is approximately 135 bp long, so mutating every base would produce a library of (135 x 3) + 1, or 406 variants. However, with Poisson modeling, about half of the library is made up of wild-type sequences, so 1x coverage of the library would require about 800-900 individual colonies. Approximately 150,000-200,000 colonies were produced for each strain, representing library coverage ranging from 150-200x. Initial transformation plates were replica plated to remove/reduce excess input plasmid and inviable single cell transformations from being pooled. Then, plates were scraped and diluted into freezer stocks such that each tube would bring a 50 mL culture to an OD of 0.3. Freezer stocks were considered time 0.

50 mL cultures in -ura media were inoculated with freezer stocks and grown continuously for 36-48 hours with samples collected every four hours for 24 hours, then every 8 hours until 48 hours. DNA was isolated by the smash and grabs protocol and was then used as template in 5 separate PCR reactions to amplify the μARS317 plasmid with custom sequencing primers designed by Ivan Liachko.[89] These reactions were then pooled and purified via Zymo select-a-size columns. The starting library, and

times 0, 8, 16, and 24 were sequenced on a Next-Seq, but coverage was only high enough on the 0 time sample for initial analysis. There are two biological replicates for wild type, the Orc4 chimera, and the Orc4-IH deletion whose reads were pooled in subsequent analysis. Analysis was conducted using the Enrich2 GUI and scripts I adapted from custom scripts from Liz Kwan.[90] Variant plotting was performed in python using matplotlib.[88]

MCM2-ChEC Seq: MCM2-ChEC strains were made using constructs provided by the Bedalov lab and their wild-type MCM2-ChEC strain yAB16747.[83] Plasmid bearing the MCM2-MNase-KanMX construct was used as template in a PCR reaction that was then diluted and used in a second round of PCR reactions. Product was then gel purified and concentrated using Zymo clean and concentrate. PCR product was transformed into log phase Orc4 chimera and Orc4-IH deletion cells (like previous transformations) and plated on D-plates for 12-18 hours before replica plating onto D+G418 plates. MCM2-MNase integration and integrity was verified via PCR and Sanger Sequencing. For generation of MCM2-MNase Orc1 chimeras, yAB16747 was mated to BY4742 and sporulated to produce yRX9 (*MAT*α MCM2-MNase). This strain was then mated to the Orc1 chimera and separately to the Orc1/4 double chimera and sporulated to produce MCM2-MNase Orc1 and Orc1/4 *MATa* chimeras.

ChEC Seq was performed according to the protocol described in Henikoff and Foss papers[91,83]. Briefly, cells were grown in log phase for several generations then diluted back to 0.3-0.5 OD before alpha factor arrest. After 2-3.5 hours (when budding stopped and large buds disappeared) cells were permeabilized with digitonin for 5' before MNase was activated with 200 mM $CaCl_2$. Cutting was stopped at 30", 1' ,5', and 10' by adding SDS/stop buffer at the indicated times. DNA was then immediately isolated with phenol/chloroform (after proteinase K digestion) and precipitated with ethanol overnight at -20 or -80C. DNA was resuspended with 40 uL 0.1 TE and 1.3 µL 10 mg/mL RNase and allowed to go into suspension over 1-3 days at 4˚C. Efficacy of MNase digestion was evaluated via agarose gel electrophoresis with EtBr stain.

DNA was library prepped using custom library prep protocol from the Bedalov lab or via TruSeq ChIP library prep kits (Single index TruSeq adapters were used in both cases). DNA was end-repaired and A-tailed before ligating adapters and purifying 2x via Ampure beads using a 1.8x bead ratio. 15 cycles of PCR amplification were used to amplify libraries. Because the MCM2-7 DH fragment is ~62 bp and the next largest fragment, nucleosomes, is approximately 130 bp, PCR amplification alone should bias library toward ideal sized fragments without gel purification. Libraries were then analyzed via TapeStation and chosen based on minimization of adapter dimers as well as maximization of MCM2 fragment: nucleosome fragment ratio. DNA was sequenced on Illumina HiSeq using paired end sequencing.

Sequence alignment and quality control: Paired demultiplexed Fastq files were adapter trimmed using the trim_galore wrapper for CutAdapt before aligning to *S. cerevisiae* genome with Bowtie2.[92,93(p2)] A custom version of the *S. cerevisiae* genome (SacCer2) that has 2 rDNA repeats and no mitochondrial genome was used as the reference genome. Sam files were then converted to bam files using samtools before bamtools filtering for reads with a map quality score over 20. Bam files were sorted and indexed then written to coverage files for plotting. To isolate MCM footprint fragments, which are 62 bp in size, I filtered for insert sizes between 50 bp and 100 bp.

Peak calling: Homer was used to call peaks on all strains/replicates.[94] Tag directories were made using a fragment length of 75. To account for differences in read depth, tag directories were downsized to 1E6 tags with lengths between 50 and 100 bp. Peak size was set to 100 bp with a minimum distance between peaks of 200, and a minimum tag threshold of 100. To determine differential peaks between mutant strains and wild type. I used the Homer command getDifferentialPeaks. The top 5-6 differential peaks in the Orc4 chimera were chosen to test for replication efficiency on plasmids.

Chapter 3. Testing the ODIRA model

The work described here was done by me, Claudia Espinosa, Bonny Brewer, Samantha Paskvan, and Madison Miller. Claudia Espinosa and Bonny Brewer designed and constructed the original split *URA3* strain and characterized clones via CHEF gel and ArrayCGH. I continued the project by analyzing clones via CHEF gel and ArrayCGH and designed CRISPR/Cas9 experiments to distinguish between ODIRA and DSB. I, along with Samantha Paskvan and Madison Miller, analyzed clones from the CRISPR experiments. The writing and most figures in this chapter are based on an unpublished manuscript drafted by Bonny Brewer.

3.1 Introduction

While DNA replication is a tightly controlled process with multiple safeguards for correcting mistakes and ensuring that DNA replicates completely, replication errors do still occur. These errors include those that occur during normal genome replication as well as errors that arise during repair of broken or damaged DNA. They can range from mismatches and small insertions/deletions to large scale deletions, inversions, and duplications.[95,96] Often, certain genomic features such as inverted repeats or short tandem repeats can predispose a locus to instability.[97–99] Errors, particularly large-scale insertions, deletions, and copy number variants are implicated in a number of human developmental disorders and cancers.[97,100,101] However, errors in DNA replication and repair also serve as an important source of genetic diversity upon which evolution can act.[102,103,104] For example, when *S. cerevisiae* are grown in sulfate-limited media, yeast that have amplified the sulfate transporter locus (*SUL1*) sweep the population.[105,106] By producing more sulfate transporters, these yeast outcompete other yeast in the sequestration of the limited sulfate, enabling faster growth.

The predominant form of *SUL1* amplicon in these chemostat evolutions is one in which the region containing *SUL1*, including the adjacent origin of replication, is triplicated with the central copy in inverted orientation on an otherwise intact

chromosome.[102,106] The exact size of each amplicon differs, but the amplicons do retain some common features:

1. The amplified segment contains at least one origin of replication.
2. The amplicons are invariably odd in number (often 3 copies) with the center copy in inverted orientation.
3. The junctions denoting the boundary of amplifications are at short, interrupted repeats that span less than the distance of an Okazaki fragment.
4. The distal end of chromosome II remains intact.

Models of repair based on double stranded breaks in DNA as the causal event do not adequately explain the formation of these internal inverted amplicons, particularly the preservation of the distal end of the chromosome.[107] The Brewer/Raghuraman lab has previously proposed and tested a model (called Origin-Dependent Inverted-Repeat Amplification or ODIRA) to explain this specific amplification.[108]

In this model (Figure 3.1), replication proceeds from an origin and through the adjacent gene; fork pause and fork regression expose inverted repeats with homology to the lagging strand template. This free end anneals to its complement on the opposite strand, enabling replication to proceed back along the opposite strand until it is ligated to an Okazaki fragment. If a similar error occurs on the other side of the origin, the replication forks are now both 'closed' and a self-complementary loop of newly replicated DNA is the result. A replication fork from the adjacent replicon extrudes the closed loop that can then anneal into a 'dog bone' intermediate. As this dog bone intermediate possesses an origin, it is capable of replication in a subsequent S phase where it is converted into a double stranded circle of DNA with two copies of the gene and adjacent origin in inverted orientation. This circle can then be reintegrated at the original locus through homologous recombination, resulting in three copies of the gene with the middle copy inverted. ODIRA is also possible with an error occurring at a single replication fork (typically the centromere proximal fork).[108] A single error would occur more commonly than cases of two errors occurring in the same fashion in the same replication cycle. A single ended ODIRA event would produce a hairpin intermediate extending to the telomere that could replicate into an extrachromosomal

palindromic linear duplex where both ends of the linear are capped by telomeres. Recombination at the original locus would not necessitate reintegration but could result in swapping of telomeric arms between the original locus and the extrachromosomal duplex.[109] While this extrachromosomal linear could provide cells with the same selective advantage as the circular integration, this duplex would lack a centromere, making it less stable and thus less heritable.[97] A second ODIRA event on the palindromic linear could then generate a structure capable of recombining with chromosome II to generate the inverted triplication. This inherent instability of the linear duplex would explain why chemostat growth only showed the ostensibly rarer but more stable events of integration onto chromosome II.

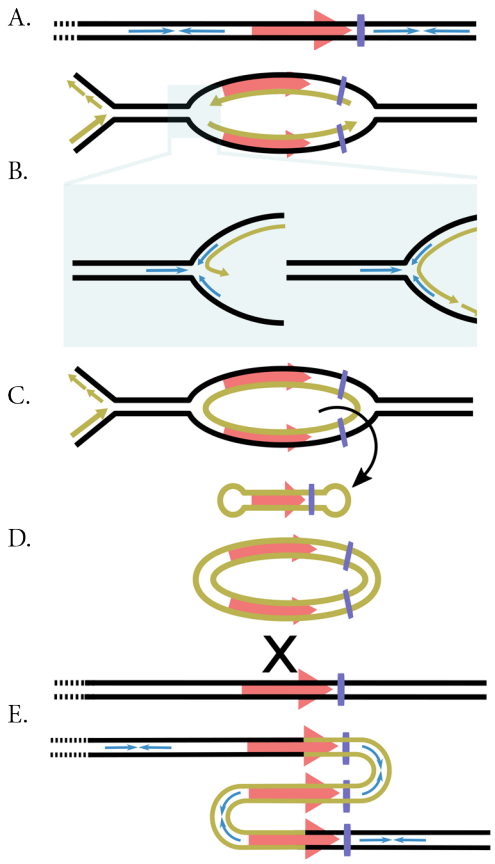


Figure 3.1: Origin Dependent Inverted Repeat Amplification. A) Replication starts at an origin of replication (vertical blue bar). B) The leftward moving fork proceeds through the inverted repeat (blue arrows) and then regresses, allowing the leading strand to anneal with the complementary sequence on the gapped lagging strand. Ligation of the leading strand with the Okazaki fragment on lagging strand creates a "closed" fork that is incapable of resuming replication. C) The same replication error at the rightward moving fork results in a closed loop that is extruded from template DNA to form 'dog bone' intermediate. D) The dog bone intermediate replicates to form a double stranded circular DNA that contains inverted copies of the sequences between the two sets of inverted repeats. E) The circular intermediate reintegrates via recombination at the original locus producing an inverted triplication.

While previous work published in the lab demonstrates that the intermediates of ODIRA are capable of being formed,[110] we set out to create a strain in which ODIRA

events could be directly selected for by requiring that the either a hairpin or dog bone intermediate replicates and then reintegrates on a separate chromosome to produce a functional, full copy of *URA3*.

3.2 Results

3.2.1 Strain construction and rationale

To isolate ODIRA events, Claudia Espinosa constructed a strain where she replaced most of *SUL1* with the 5′ end of *URA3* (*ura)* in the same transcriptional direction (Figure 3.2 A). She left a small portion of the 3′ end of *SUL1* intact at the beginning of the *URA3* fragment that could be easily probed in subsequent Southern blots. Next, she cloned in the 3′ end of *URA3* (*ra3*) (with a substantial portion of the gene overlapping with the *URA3* fragment on chromosome II) into an intergenic region on the left arm of chromosome IX. For a functional copy of *URA3* to be produced, recombination would either have to take place directly between the two chromosomes or via an intermediate such as the replicated versions of the ODIRA dog bone or hairpin intermediates (Figure 3.2 B and C).

We assumed that direct recombination between the two chromosomes would results in an unstable and lethal dicentric chromosome if it occurred in G1. Nevertheless, we recovered many events that had evidence of direct recombination between the two chromosomes that had undergone many different secondary rearrangements to produce genetically stable monocentric chromosomes, resulting in varying degrees of aneuploidy (Figure 3.2 D). One common and simple secondary recombination is the deletion of one of the two centromeres on the resulting dicentric chromosome, but cycles of breakage-fusion-bridge during mitosis could be resolved via addition of a neo-telomere or via break-induced repair with other chromosomes at homologous regions producing stable, secondarily translocated chromosomes.

Figure 3.2 E shows hypothetical pulsed-field gel analysis (ethidium bromide stains and Southern blots) for the original split *ura3* strain and Ura+ clones. A key feature that differentiates ODIRA events from direct recombination is the 2:1 ratio of 5′ *SUL1* probe

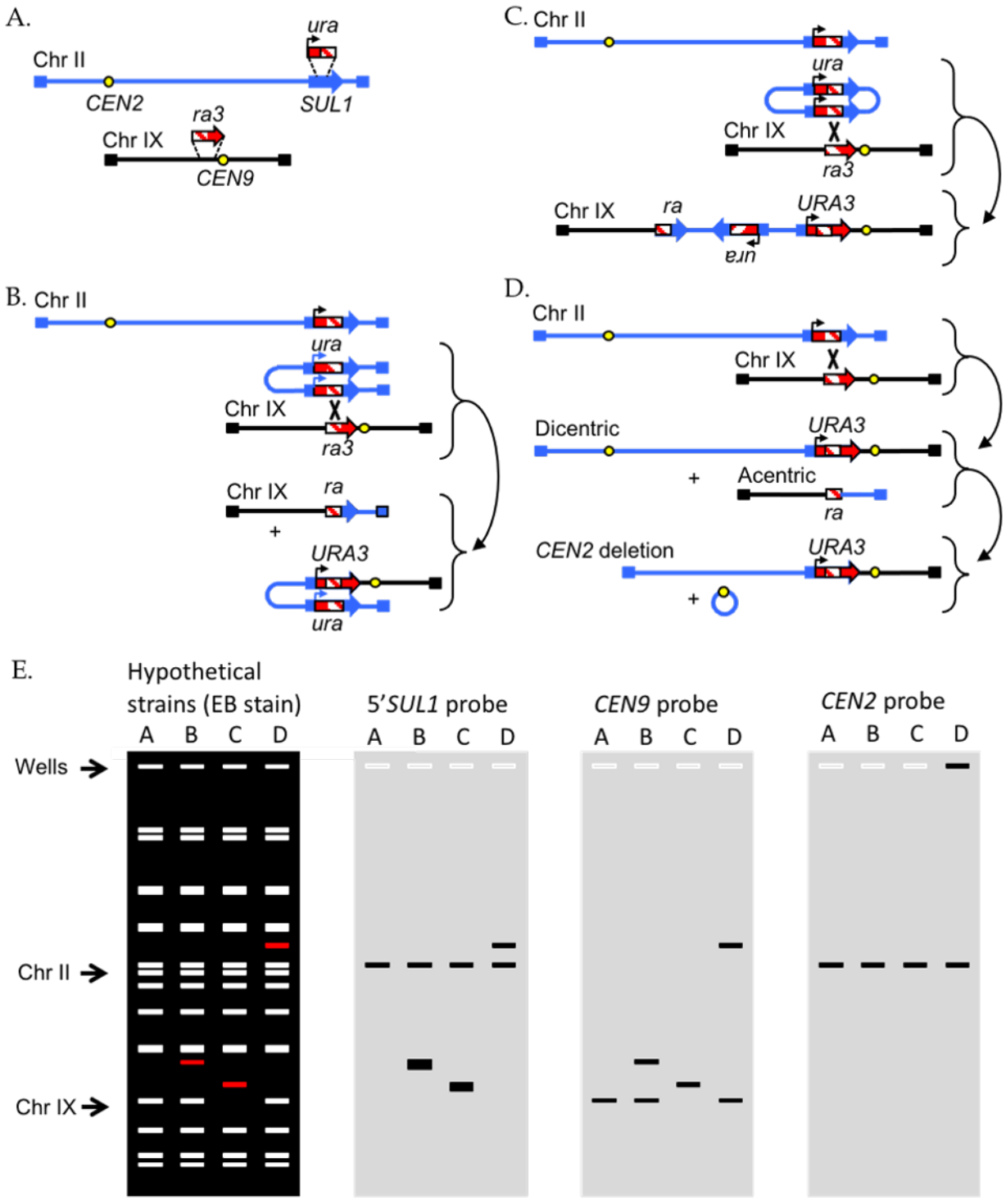


Figure 3.2 Potential rearrangements of Ura+ clones. A) Creation of the split *ura* strain. B) Integration of a hairpin ODIRA intermediate resulting from a single template switching error at the centromere proximal replication fork. C) Integration of canonical ODIRA dog bone intermediate D) Direct recombination between the two chromosomes and dicentric resolution. E) Hypothetical CHEF gels and Southern blots from scenarios A-D. Only ODIRA events produce 2:1 ratios of *5'SUL1* signal in the first Southern blot.

at a new locus compared to the single copy on chromosome II. This ratio results from the replication of a replication-competent intermediate (either the dog bone or hairpin described previously) prior to integration at the new locus. While both hairpin and dog bone intermediates would generate the 2:1 ratio, only the dog bone intermediate would result in no new chromosomes and instead increase the size of chromosome IX, causing an upward shift in the EtBr-stained gel.

We then examined the frequency of Ura+ clones and used contour-clamped homogenous electric field (CHEF) gels, Southern blotting, array comparative genome hybridization, and Sanger sequencing to separate clones into two broad classes: clones that underwent direct recombination of chromosomes (reciprocal translocation) and clones that had dog bone/hairpin structures compatible with both ODIRA and dsDNA break repair via end-resection and fold-back.

3.2.2 Characterization of Ura+ clones:

Single colonies of the Ura+ strain were grown to saturation in 1 mL culture before plating 10^8 cells on plates lacking uracil. Each colony>culture>plating produced at least one Ura+ clone. A total of 27 initial events were analyzed via CHEF gel electrophoresis and Southern blotting. We then performed array Comparative Genomic Hybridization (aCGH) and sequenced the junctions on a subset of the 27 events that were representative of variations seen in CHEF gel and Southern blot results; clones undergoing aCGH follow-up represent both classes of recombination products.

We observed five categories of gross chromosomal rearrangements in the 27 Ura+ clones. Four of these classes have been described previously in literature and are consistent with dicentric chromosomes undergoing secondary rearrangements. The secondary rearrangements were:

1. Deletion of *CEN2* (4 clones) or *CEN9* (2 clones) from dicentric chromosomes.
2. Neotelomere generation at new breakpoints between the two centromeres on a dicentric chromosome (2 clones)

3. Break-induced repair that acquired a telomere segment from another chromosome that occurred at a long terminal repeat (solo delta) between *ra3* and *CEN9*.
4. An inverted hybrid chromosome resulting from bridge-breakage-fusion events (3 clones).

Three more clones contained more than one of the above events. All of these clones were inconsistent with ODIRA models of rearrangement. However, the fifth class (5 clones) had aCGH profiles where part of chromosome II, beginning between *CEN2* and the *SUL1* locus and extending to the right telomere, was increased to three copies. Importantly, there was not a clean breakpoint between 1 and 3 copies, but a steady rise in copy number proximal to the centromere. This 'waterfall' or sloping pattern of copy number increase is consistent with an inverted junction. These five clones possess linear chromosomes that are partially palindromic and contain a *ura-sul1* inverted duplication with varying sized flanking regions from chromosome II. The partially palindromic linear chromosomes have one telomere from the right arm of chromosome II and the other telomere from the right arm of chromosome IX.

The fifth class of clones is consistent with the ODIRA or DSB models of repair. In contrast to the neotelomeres in the 3rd class which had junctions at Ty and delta sequences, these junctions occurred at interrupted palindromic repeats very similar to those first identified in the *SUL1* chemostats and that are capable of folding back after DSB to form a repaired hairpin.

Figure 3.3 shows five of the analyzed ODIRA clones, three of which have a 2:1 ratio of *5'SUL1,* consistent with a hairpin ODIRA intermediate. The clone A1 was chosen for subsequent aCGH, which revealed a duplication of the right arm and centromere of chromosome IX as well as a triplication of the ODIRA-prone locus on chromosome II. The right telomere of chromosome II was present in two copies, suggesting that the neochromosome seen in the CHEF gel is composed of the right arm and centromere of chromosome IX, a duplicated inverted ODIRA locus, and the right telomere of chromosome II. The predicted size of the neochromosome on the CHEF gel is 410 kb,

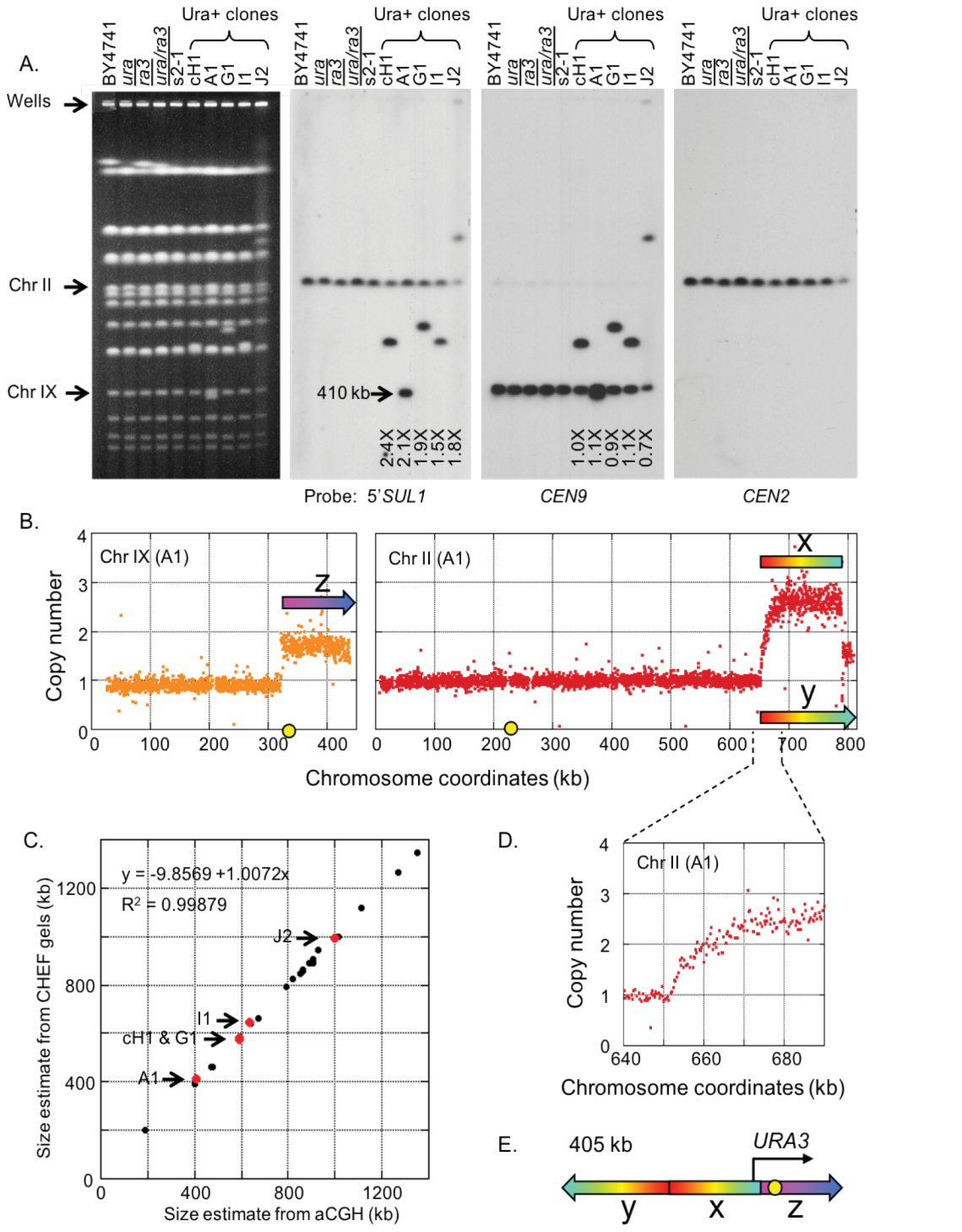


Figure 3.3: Analysis of a Ura+ clones consistent with ODIRA hairpin intermediate. A) CHEF gel and Southern blots of 5 Ura+ clones generated by recombination between a hairpin intermediate and chromosome IX. B) aCGH of A1 shows duplication of the right arm and centromere of chromosome IX as well as triplication of ura-sul1 locus from chromosome II. C) Size of A1 neochromsome estimated to be 410kb from CHEF gel. D) "Waterfall" increase in copy number beginning at ~651 kb on chromosome II is consistent an inverted junction on the neochromosome. E) Reconstruction of neochromosome of A1 based on CHEF gel and aCGH. Figure by Bonny Brewer, used with permission.

while aCGH shows aneuploidies resulting in 405 kb of additional sequence. Supplemental Figure 3.1 shows another example of an ODIRA-like event.

3.2.3 Distinguishing between DSB and ODIRA with CRISPR/Cas9

The hairpin ODIRA intermediate observed in the Ura+ clones we isolated could be produced by either ODIRA or double strand breaks (DSBs) followed by end resection and foldback repair at the same inverted repeats that facilitate ODIRA template switching (Figure 3.4). To distinguish between these two possibilities, we targeted CRISPR/Cas9 to cut chromosome II upstream of some of the common junctions seen in aCGH of ODIRA/DSB-like Ura+ clones, thus inducing double stranded breaks. Should DSBs prove to be a mechanism of hairpin formation, we would expect to see a greater number of Ura+ clones overall and an increase in frequency of palindromic neochromosomes with 2:1 ratios of 5'*SUL1* in Southern blots. The three sites chosen for Cas9 cleavage were at chromosome II coordinates 650 kb, 720 kb, and as a control, a cut site at 794 kb; we designed two gRNAs to each location. The 794 kb cut site should be a negative control for ODIRA-like events as it induces a break distal to the *ura-sul1* locus and potentially facilitates direct recombination by creating a sticky end that bears homology to the *ra3* cassette on chromosome IX.

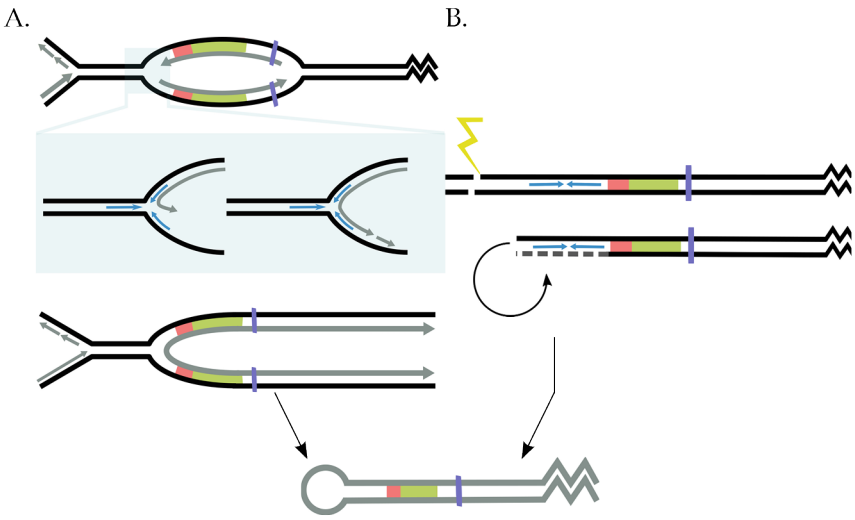


Figure 3.4: ODIRA or dsDNA breaks can form hairpin intermediates. A) The hairpin intermediate is formed from a replication error at the regressed fork (ODIRA). B) Hairpins resulting from DSBs occur when a chromosome breaks upstream of a short inverted repeat. End resection and fold-over then creates a hairpin intermediate.

Because DSBs are often lethal events and continuous cleavage would keep necessitating repair and recombination, we put Cas9 under control of a *GAL10*

promoter so that Cas9 expression and cutting could be briefly pulsed. We initially plated cells on -leucine plates supplemented with glucose, ensuring that the Cas9 was repressed. We then selected single colonies and resuspended them in -leucine medium lacking a carbon source. This suspension was then split between two 1 mL cultures containing either galactose or glucose to have a cut and uncut control, respectively, originating from the same single colony. After 24 hours, glucose was added to the galactose cultures to stop Cas9 expression and facilitate growth of the cells. Once cultures were at saturation density, we plated 10^8 cells on -uracil plates. There were no increases in overall frequencies of Ura+ clones in either of the 650 kb or 720 kb sites, but there was approximately a 20-fold increase in Ura+ events in the 794 kb site. This increase is to be expected with erosion at the cut site revealing the *ura-sul1* locus that could then invade the *ra3* site on chromosome IX for repair. The lack of increase in overall Ura+ events at the other two cut sites suggests that DSBs are not the drivers of the hairpin intermediate that produces the palindromic neochromosomes.

To confirm that this lack of increase in events was accompanied by no change in frequency of the characteristic 2:1 ratio of *5'SUL1*: *CEN9*, we characterized 33 of these clones from the 650 and 720 kb sites via CHEF gel electrophoresis, Southern blot, and Sanger sequencing. In total, we found 3 events that exhibited the 2:1 ratio expected for ODIRA-like events. Analysis of the 794 kb site clones revealed events consistent with the chromosome IX strand invasion as expected. We also used Sanger sequencing at the cut sites to ensure that Cas9 was cutting. Sequencing across the sites revealed PAM or gRNA sequence nullifying mutations that would prevent further Cas9 targeting/cutting, suggesting that these sites had been successfully cut and repaired. However, we also found these mutations in the glucose-repressed clones, but not in the parent strain used in the transformation, showing that the GAL promoter was allowing for Cas9 expression even in the presence of glucose.

3.3 Discussion

The engineered split *ura* strain enabled detection of palindromic Ura+ chromosomes consistent with ODIRA model. These palindromes could not be easily explained via

direct recombination or dicentric formation, and our recovery of multiple instances of direct recombination enabled comparison to ODIRA-like events, showing many dissimilarities. In our analysis, however, we recovered no clones with evidence of a dog bone intermediate formed from template switching errors at both ends of the replication bubble. Two such errors would be far rarer than the single error required to produce a hairpin intermediate. By not selecting for Ura+ clones under competitive growth, we were able to detect these more common, but less stable events that were not seen in the sulfate limited chemostats.

To test whether the hairpin intermediates were not the results of double stranded breaks, we specifically introduced double stranded breaks via CRISPR/Cas9 upstream of some common junctions to look for an increase in Ura+ clones. As we aimed to introduce breaks ideally at intergenic regions, but at the very least within non-essential genes, the sites at which were able to direct Cas9 were limited. We saw no increase in frequency of Ura+ events nor of events with hairpin intermediates that would suggest DSBs are causing the hairpin intermediate.

Our reconstruction of ODIRA-consistent neochromosomes depended on size information gained from CHEF gels and copy number increases detected by aCGH. To truly confirm these structures, additional analysis would be needed. PacBio sequencing may provide long reads spanning the inversion that could confirm the turn-around, but the inverted structure of the chromosome may actually prevent accurate sequencing. A less expensive, but more work-intensive method may even be to design single primers that fall within the copy number increase shown by aCGH but point upstream toward the breakpoint. PCR with intact DNA and a single primer would only produce a PCR product if there were an intact inversion. Given the somewhat limited resolution of the aCGH, it would take much trial and error to find a suitable primer sequence close enough to the inversion to produce an amplifiable PCR product.

3.4 Materials and Methods

Strain construction: The split *ura* strain was constructed in two isogenic strains of opposite mating type (BY4741). First, complete *URA3* genes were targeted to the two

target loci using PCR primers with homology to the flanking region (one locus targeted per strain). Ura+ transformants were selected on –uracil plates. A subsequent transformation was conducted with a truncated *URA3* sequence and loss of *URA3* was selected for on 5-FOA plates. The two strains were then mated together, the diploid sporulated, and a haploid with both partial *URA3* sequences identified.

Identifying Ura3+ clones: Haploid cells with both partial *URA3* sequences were streaked out on synthetic complete media plates. Individual colonies were selected for overnight growth in liquid synthetic complete media and subsequently plated on *-ura* plates. Resulting colonies were then analyzed via CHEF gel.

CHEF gel analysis: Cells were embedded in agarose plugs according to Argueso protocol.[45] Electrophoresis was carried out in 0.8% LE agarose gels in 0.5x TBE for 60 hours ramping from 47 to 170 seconds at 165 volts at 14°C. Gels were then stained with Ethidium Bromide and UV imaged.

Southern blotting: Gels were transferred to GeneScreen Nylon membranes using Southern Blot protocols. Southern blots were subsequently probed with a ^{32}P 5′*SUL1* fragment, *CEN2*, or *CEN9* adjacent fragments (blots were stripped and re-probed). Bio-Rad Personal Molecular Imager and QuantityOne software were used to quantify blots.

CRISPR/Cas9 plasmid construction: gRNAs were designed to cut at sites near common breakpoints in aCGH that were not in any essential genes. The 794 kb gRNA was designed as a positive control based on the hypothesis that creating a 3′ single stranded overhang at the site of *URA3* homology would enable frequent recombination between *URA3* fragments on chromosome II and IX. gRNAs were designed using tools in Laughery et al.[74] and cloned into a modified version of their Cas9 plasmid with a galactose inducible promoter in front of *CAS9* and a *LEU2* selectable marker (made by Joseph Sanchez).

CRISPR/Cas9 experiments: Cells were grown on plates or in liquid synthetic complete media supplemented with 2% glucose (for repressed Cas9) or galactose (for induced

Cas9) and grown to saturation before plating on -Ura-Leu plates. A subset of resulting colonies was then analyzed via CHEF gel to look for 2:1 ratios of *5'SUL1:CEN9*. Some of these colonies were also Sanger sequenced at the breakpoints to ensure mutation at the PAM site/gRNA location (successful Cas9 cutting).

Chapter 4. Conclusions and Future Directions

In my thesis work, I primarily investigated the role of ORC subunits in budding yeast origin specificity. My ultimate goal was to develop multiple chimeric Orc4 strains in which I could examine how different amounts of global origin alteration impacted cell fitness and genome architecture. However, despite multiple attempts and troubleshooting, I was only able to produce a single chimeric strain, with the Orc4-IH sequence borrowed from *Lachancea waltii*, a budding yeast species whose ACS is marginally different from that of *S. cerevisiae.* Nevertheless, this chimera and a deletion created serendipitously gave me two strains in which to analyze ORC specificity. My work was bolstered by the release of a second cryo-EM structure of ORC bound to origin DNA released during my investigations of the Orc4-IH that revealed additional DNA-specific contacts in Orc1 and Orc2. This new structure allowed me to complement the Orc4 chimera with an Orc1 chimera and change all known DNA-specific contacts in *S. cerevisiae* to that of *L. waltii*. While I have not yet been able to examine the long-term fitness consequences of altered origin location, my rotation project work, described in chapter 3, on looking for ODIRA intermediates gave me the opportunity to examine how origin location and replication can impact genome architecture.

4.1 ORC determination of origin sequence specificity

With the first transformations of *L. waltii* plasmid into the Orc4-IH chimera, I was able to show that the Orc4-IH does drive sequence specificity to some extent as altering this region is sufficient to enable recognition of a new origin. My work on Orc4-IH driven DNA specificity of yeast origins now comes in the context of two publications arriving at similar conclusions through overlapping yet complementary methods. The first published work tested a suite of one or two amino acid substitution mutations in the Orc4-IH and used mutated ARS libraries, ChIP-Seq, and replication profiles to show that mutations to the DNA contacting bases did alter DNA specificity. While neither I nor the authors of the study,[111] were able to produce viable complete deletions of the Orc4-IH, another research group did succeed.[86] Their work found that ORC binding

and origin initiation was pervasive throughout the genome and accompanied by failure to adjust nucleosomes at typical sites of initiation. All three works, mine included, demonstrate that the Orc4-IH does have major influence over sequence specificity.

Understanding the interplay between origin specificity and ORC requires understanding where ORC is binding. A growing body of research suggests that the B2 element is utilized for ORC binding during pre-RC formation. However, thus far, ChIP data on ORC binding does not have the resolution to definitively say where ORC is binding during pre-RC loading. If the B2 serves as a loading site for ORC, it removes a layer of complexity from the mechanisms of origin licensing but does add a complexity to understanding ORC sequence specificity. Because the B2 element match to the ACS occurs on the opposite strand of the ACS, ORC binding and loading an MCM from each location would result in head-to-head hexamers without requiring two different sets of ORC/MCM contacts and conformational changes to explain the asymmetric loading from a single location.

However, because the two ORC binding site model results in the same MCM loading pattern as the one ORC model, the high resolution MCM2-ChEC data I generated cannot distinguish between the two models. ORC or Cdc6 MNase ChIP seq (or ChEC seq) may provide the resolution necessary to determine all the potential sites of ORC binding (and loading). ORC on its own loosely binds DNA without much sequence specificity, but Cdc6 binding increases sequence-specificity as it creates a more stable complex for pull-down.[38,112] I would like to perform high resolution Cdc6 ChIP sequencing to see if that resolution can reveal ORC-Cdc6 binding at the B2 element. ORC-Cdc6 bound at the B2 would confirm that endogenous B2 binding can support MCM loading.

Moreover, these data would give more complete information about ORC binding specificity in the context of origin loading and potentially allow us better insight into ORC-DNA interactions and help explain why certain origins were lost, lessened, or improved in the ORC mutants I generated for this work. These data could also confirm my hypothesis that the Orc1 chimera preferentially binds to the B2 element of certain origins.

The addition of the Orc1 chimera to my work complicates the picture of ORC binding and ORC mutant viability. μARS317-Seq in the Orc1 chimera suggested that the Orc1 chimera was now binding primarily to the B2 element in *ARS317,* and preliminary MCM2-ChEC Seq analysis shows that MCM signal is centered on the ACS, not adjacent, which also seems to suggest that the B2 element may be the primary binding site of the Orc1 chimera. To test this possibility, I am interested in replacing the rDNA ARS with a stretch of GCs in the Orc1 chimera. Previous work in the lab suggests that origin function returned to such an rARS mutant during continuous selection for improved growth, potentially because ORC shifted to utilizing the B2 element as the ACS.[35] If the Orc1 chimera does predominantly bind to the B2 element, I hypothesize that the Orc1 chimera would restore origin function to a CRISPRed rARS faster than wild type or the Orc4 chimera.

When the Orc1 chimera was combined with the Orc4 chimera, the double chimera rescued origin function at many of the origins lost in the Orc4-IH chimera alone. This result prompts me to question the relationship between Orc1 DNA contacts and Orc4-IH DNA contacts. On its own, the Orc1 chimera is slower growing than the Orc4 chimera or wild type, but in conjunction with the Orc4 chimera, doubling time is reduced. These results raised the question: does matching contacts between species improve viability? As a future direction, I would like to re-attempt creating the other Orc4 chimeras concomitantly with or after making the matching Orc1 chimera. I hypothesize that by changing the DNA contact residues together may create viable Orc4-IH chimeras for the other attempted species.

Additionally, I and others in the lab have observed that the double chimera does appear to grow marginally faster than wild-type *S. cerevisiae*. While five growth curve replicates did not reach significance, even a slight difference in doubling time can be observed in head-to-head competitions between strains where one is tagged with GFP.[113] I would like to compete the double chimera against wild type in these conditions to see the double chimera does have a growth advantage over wild type.

4.2 The impact of origin location on genome integrity

At the outset of my research in the Brewer/Raghuraman lab, I sought to investigate how globally altered origin location impacted genome integrity. I did not have time to pursue this question in my PhD, but I am interested in utilizing mutated Orc4-IH strains as a tool to investigate the essentiality of origin location conservation. With the Orc4-IH partial deletion strain that I created, I saw reduced rDNA copy number and increased doubling time while other studies show increased HU and DNA damage sensitivity in Orc4-IH mutation strains that have altered specificity. I would like to evolve some of the more severely impacted mutants, especially the total Orc4-IH deletion discussed previously,[86] in turbidostats and analyze samples for mutation rate and potential chromosomal rearrangements.

Because late replicating regions, particularly where replication terminates, are often subject to high mutation rates and greater frequency of breakage,[114] I hypothesize that altering origin location genome-wide would change mutational patterns in cells. While this change would not be measurable genome-wide at the population level in turbidostats, I can sample over time to look for canavanine resistance resulting from loss or mutations in the gene for the arginine transporter (*CAN1*). By altering the location of the *CAN1* gene to late or early replicating regions, I could then compare mutation frequencies at different *CAN1* sites between Orc4-IH mutant strains.

As a tie-in to my rotation work on ODIRA, where the location and potentially late initiation of *ARS228* enables gene amplification, I would like to evolve the Orc4-IH total deletion in a sulfate-limited chemostat. With pervasive initiation, the Orc4-IH total deletion may not initiate replication at a locus conducive to ODIRA events in *SUL1*, which could produce different styles of amplicons. Alternatively, the pervasiveness of initiation could cause more fork stalling and increase the frequency of ODIRA-like events.

4.3 Evolutionary history of ORC sequence specificity

Origin sequence specificity appears to be the ancestral state of replication origins, but whether sequence specificity in yeast results from maintenance of the ancestral state or reversion to the ancestral state is somewhat unclear. Budding yeast are also unique

from higher eukaryotes because they have point centromeres versus regional or pericentromeres. Previously published work on Orc4-IH posits that these features are related.[111]

Given the uniqueness of sequence-specific origins, point centromeres, and even the timing of centromeres replication, it seems possible that these features are evolutionarily linked. Deepening the initial investigation in Hu et al. (2020) by looking at the ratio of nonsynonymous to synonymous mutations (dN/dS ratios) within a greater variety of budding yeast species or within *sensu stricto* species may illuminate similar rates of evolution between proteins essential for point centromeres and sequence specific origins.[111] Looking at dN/dS ratios of DNA contacting regions may also reveal co-evolution of these regions, which would harken back to my previous idea that matching DNA contacting residues to that of the same species improves strain viability.

Conserved origin sequence appears to be a means of restricting replication initiation to intergenic sites of low conflict. However, in multicellular species, regions of low conflict and open chromatin change with cell types, thus origin plasticity is required to reduce conflict. Engineering a metazoan cell line with sequence-specific ORC would be a long-term and potentially fruitless effort, but it would be very interesting to make a sequence specific ORC pluripotent cell line and attempt to differentiate the cells. If origin plasticity were required for multicellularity, I'd hypothesize that cells would die during the differentiation process. As a control, it may also be possible to engineer a sequence specific ORC that recognizes a similar binding site to an active transcription factor in the cell type. Doing so may help target ORC to the sites that non-sequence specific ORC would bind to and thus make sure that the impact on cell viability is due to altered origin location and not due to general decrease in ORC integrity due to the gross changes necessary to make a metazoan sequence-specific ORC.

4.4 Additional tests for ODIRA intermediates

With my work on the ODIRA model, I helped demonstrate that the clones characterized from the engineered *ura3* strain were likely the result of ODIRA and not DSBs as introducing DSBs via CRISPR/Cas9 did not increase the frequency of inverted amplification events. To show that ODIRA fork pausing is required for the hairpin formation intermediate, we could knock out *EXO1* in the engineered strain. *EXO1* is required for fork pausing, so deleting this gene should eliminate the possibility of ODIRA events in subsequent clones. As a complement, we could also potentially target a dead Cas9 protein (dCas9) to the regions flanking *SUL1*. Targeting dCas9 to these regions should increase the frequency of fork stalling by adding a physical block to the site. In this scenario, I would expect the frequency of ODIRA-like events to increase. We could target dCas9 to both the centromere proximal and telomere proximal regions of *SUL1*. Targeting to the telomere proximal region may cause breakage that could just increase the frequency of microhomology mediated recombination that we see when Cas9 is targeted to the telomere proximal end; however, it may finally allow us to capture evidence of the dog bone intermediate that is at the heart of the ODIRA model of gene amplification.

References

1. Chunduri NK, Storchová Z. The diverse consequences of aneuploidy. Nature Cell Biology. 2019;21(1):54–62. doi:10.1038/s41556-018-0243-8

2. Stinchcomb DT, Struhl K, Davis RW. Isolation and characterisation of a yeast chromosomal replicator. Nature. 1979;282(5734):39–43. doi:10.1038/282039a0

3. Pohl TJ, Brewer BJ, Raghuraman MK. Functional Centromeres Determine the Activation Time of Pericentric Origins of DNA Replication in Saccharomyces cerevisiae. PLOS Genetics. 2012;8(5):e1002677. doi:10.1371/journal.pgen.1002677

4. Dershowitz A, Snyder M, Sbia M, Skurnick JH, Ong LY, Newlon CS. Linear Derivatives of Saccharomyces cerevisiae Chromosome III Can Be Maintained in the Absence of Autonomously Replicating Sequence Elements. Molecular and Cellular Biology. 2007;27(13):4652–4663. doi:10.1128/MCB.01246-06

5. Yuan Z, Riera A, Bai L, Sun J, Nandi S, Spanos C, Chen ZA, Barbon M, Rappsilber J, Stillman B, et al. Structural basis of Mcm2–7 replicative helicase loading by ORC–Cdc6 and Cdt1. Nature Structural & Molecular Biology. 2017;24(3):316–324. doi:10.1038/nsmb.3372

6. Li N, Lam WH, Zhai Y, Cheng J, Cheng E, Zhao Y, Gao N, Tye B-K. Structure of the origin recognition complex bound to DNA replication origin. Nature. 2018;559(7713):217–222. doi:10.1038/s41586-018-0293-x

7. Hansen RS, Thomas S, Sandstrom R, Canfield TK, Thurman RE, Weaver M, Dorschner MO, Gartler SM, Stamatoyannopoulos JA. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proceedings of the National Academy of Sciences. 2010;107(1):139–144. doi:10.1073/pnas.0912402107

8. Devbhandari S, Jiang J, Kumar C, Whitehouse I, Remus D. Chromatin Constrains the Initiation and Elongation of DNA Replication. Molecular Cell. 2017;65(1):131–141. doi:10.1016/j.molcel.2016.10.035

9. Bell SP, Kobayashi R, Stillman B. Yeast origin recognition complex functions in transcription silencing and DNA replication. Science. 1993;262(5141):1844–1849. doi:10.1126/science.8266072

10. Brewer BJ. Intergenic DNA and the sequence requirements for replication initiation in eukaryotes. Current Opinion in Genetics & Development. 1994;4(2):196–202. doi:10.1016/S0959-437X(05)80045-0

11. Remus D, Beuron F, Tolun G, Griffith JD, Morris EP, Diffley JFX. Concerted Loading of Mcm2-7 Double Hexamers Around DNA during DNA Replication Origin Licensing. Cell. 2009;139(4):719–730. doi:10.1016/j.cell.2009.10.015

12. Bell SP, Stillman B. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature. 1992;357(6374):128–134. doi:10.1038/357128a0

13. Fernández-Cid A, Riera A, Tognetti S, Herrera MC, Samel S, Evrin C, Winkler C, Gardenal E, Uhle S, Speck C. An ORC/Cdc6/MCM2-7 Complex Is Formed in a Multistep Reaction to Serve as a Platform for MCM Double-Hexamer Assembly. Molecular Cell. 2013;50(4):577–588. doi:10.1016/j.molcel.2013.03.026

14. Xu W, Aparicio JG, Aparicio OM, Tavaré S. Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S. cerevisiae. BMC Genomics. 2006;7(1):276. doi:10.1186/1471-2164-7-276

15. Lipford JR, Bell SP. Nucleosomes Positioned by ORC Facilitate the Initiation of DNA Replication. Molecular Cell. 2001;7(1):21–30. doi:10.1016/S1097-2765(01)00151-4

16. Deshpande AM, Newlon CS. The ARS consensus sequence is required for chromosomal origin function in Saccharomyces cerevisiae. Molecular and Cellular Biology. 1992;12(10):4305. doi:10.1128/MCB.12.10.4305

17. Segurado M, de Luis A, Antequera F. Genome-wide distribution of DNA replication origins at A+T-rich islands in Schizosaccharomyces pombe. EMBO reports. 2003;4(11):1048–1053. doi:10.1038/sj.embor.7400008

18. Wheeler E, Brooks AM, Concia L, Vera DL, Wear EE, LeBlanc C, Ramu U, Vaughn MW, Bass HW, Martienssen RA, et al. Arabidopsis DNA Replication Initiates in Intergenic, AT-Rich Open Chromatin. Plant Physiology. 2020;183(1):206. doi:10.1104/pp.19.01520

19. Nelson HC, Finch JT, Luisi BF, Klug A. The structure of an oligo(dA).oligo(dT) tract and its biological implications. Nature. 1987;330(6145):221–226. doi:10.1038/330221a0

20. Anderson JD, Widom J. Poly(dA-dT) Promoter Elements Increase the Equilibrium Accessibility of Nucleosomal DNA Target Sites. Molecular and Cellular Biology. 2001;21(11):3830–3839. doi:10.1128/MCB.21.11.3830-3839.2001

21. Karnani N, Taylor CM, Malhotra A, Dutta A. Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. Molecular Biology of the Cell. 2010;21(3):393–404. doi:10.1091/mbc.e09-08-0707

22. Cayrou C, Coulombe P, Vigneron A, Stanojcik S, Ganier O, Peiffer I, Rivals E, Puy A, Laurent-Chabalier S, Desprat R, et al. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. Genome Research. 2011 Jul 12:gr.121830.111. doi:10.1101/gr.121830.111

23. García-Muse T, Aguilera A. Transcription–replication conflicts: how they occur and how they are resolved. Nature Reviews Molecular Cell Biology. 2016;17(9):553–563. doi:10.1038/nrm.2016.88

24. Hamperl S, Bocek MJ, Saldivar JC, Swigut T, Cimprich KA. Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. Cell. 2017;170(4):774-786.e19. doi:10.1016/j.cell.2017.07.043

25. Siow CC, Nieduszynska SR, Müller CA, Nieduszynski CA. OriDB, the DNA replication origin database updated and extended. Nucleic Acids Research. 2012;40(D1):D682–D686. doi:10.1093/nar/gkr1091

26. Broach JR, Li Y-Y, Feldman J, Jayaram M, Abraham J, Nasmyth KA, Hicks JB. Localization and Sequence Analysis of Yeast Origins of DNA Replication. Cold Spring Harbor Symposia on Quantitative Biology. 1983;47:1165–1173. doi:10.1101/SQB.1983.047.01.132

27. Theis JF, Newlon CS. Domain B of ARS307 contains two functional elements and contributes to chromosomal replication origin function. Molecular and Cellular Biology. 1994;14(11):7652–7659. doi:10.1128/MCB.14.11.7652

28. Rao H, Stillman B. The origin recognition complex interacts with a bipartite DNA binding site within yeast replicators. Proceedings of the National Academy of Sciences of the United States of America. 1995;92(6):2224–2228. doi:10.1073/pnas.92.6.2224

29. Chang F, May CD, Hoggard T, Miller J, Fox CA, Weinreich M. High-resolution analysis of four efficient yeast replication origins reveals new insights into the ORC and putative MCM binding elements. Nucleic Acids Research. 2011;39(15):6523–6535. doi:10.1093/nar/gkr301

30. Wilmes GM, Bell SP. The B2 element of the Saccharomyces cerevisiae ARS1 origin of replication requires specific sequences to facilitate pre-RC formation. Proceedings of the National Academy of Sciences. 2002;99(1):101–106. doi:10.1073/pnas.012578499

31. Coster G, Diffley JFX. Bidirectional eukaryotic DNA replication is established by quasi-symmetrical helicase loading. Science. 2017;357(6348):314–318. doi:10.1126/science.aan0063

32. Buchman AR, Kimmerly WJ, Rine J, Kornberg RD. Two DNA-binding factors recognize specific sequences at silencers, upstream activating sequences, autonomously replicating sequences, and telomeres in Saccharomyces cerevisiae. Molecular and Cellular Biology. 1988;8(1):210–225. doi:10.1128/MCB.8.1.210

33. Eaton ML, Galani K, Kang S, Bell SP, MacAlpine DM. Conserved nucleosome positioning defines replication origins. Genes & Development. 2010;24(8):748–753. doi:10.1101/gad.1913210

34. Lynch KL, Alvino GM, Kwan EX, Brewer BJ, Raghuraman MK. The effects of manipulating levels of replication initiation factors on origin firing efficiency in yeast. PLOS Genetics. 2019;15(10):e1008430. doi:10.1371/journal.pgen.1008430

35. Sanchez JC, Ollodart A, Large CRL, Clough C, Alvino GM, Tsuchiya M, Crane M, Kwan EX, Kaeberlein M, Dunham MJ, et al. Phenotypic and Genotypic Consequences of CRISPR/Cas9 Editing of the Replication Origins in the rDNA of Saccharomyces cerevisiae. Genetics. 2019;213(1):229–249. doi:10.1534/genetics.119.302351

36. Kwan EX, Foss EJ, Tsuchiyama S, Alvino GM, Kruglyak L, Kaeberlein M, Raghuraman MK, Brewer BJ, Kennedy BK, Bedalov A. A Natural Polymorphism in rDNA Replication Origins Links Origin Activation with Calorie Restriction and Lifespan. PLOS Genetics. 2013;9(3):e1003329. doi:10.1371/journal.pgen.1003329

37. Mizushima T, Takahashi N, Stillman B. Cdc6p modulates the structure and DNA binding activity of the origin recognition complex in vitro. Genes & Development. 2000;14(13):1631–1641.

38. Speck C, Chen Z, Li H, Stillman B. ATPase-dependent, cooperative binding of ORC and Cdc6p to origin DNA. Nature structural & molecular biology. 2005;12(11):965–971. doi:10.1038/nsmb1002

39. Sun J, Kawakami H, Zech J, Speck C, Stillman B, Li H. Cdc6-induced Conformational Changes in ORC Bound To Origin DNA Revealed by Cryo-Electron Microscopy. Structure(London, England:1993). 2012;20(3):534–544. doi:10.1016/j.str.2012.01.011

40. Zhai Y, Cheng E, Wu H, Li N, Yung PYK, Gao N, Tye B-K. Open-ringed structure of the Cdt1–Mcm2–7 complex as a precursor of the MCM double hexamer. Nature Structural & Molecular Biology. 2017;24(3):300–308. doi:10.1038/nsmb.3374

41. Coster G, Frigola J, Beuron F, Morris EP, Diffley JFX. Origin Licensing Requires ATP Binding and Hydrolysis by the MCM Replicative Helicase. Molecular Cell. 2014;55(5):666–677. doi:10.1016/j.molcel.2014.06.034

42. Randell JCW, Bowers JL, Rodríguez HK, Bell SP. Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. Molecular Cell. 2006;21(1):29–39. doi:10.1016/j.molcel.2005.11.023

43. Miller TCR, Locke J, Greiwe JF, Diffley JFX, Costa A. Mechanism of head-to-head MCM double-hexamer formation revealed by cryo-EM. Nature. 2019;575(7784):704–710. doi:10.1038/s41586-019-1768-0

44. Rodriguez J, Lee L, Lynch B, Tsukiyama T. Nucleosome occupancy as a novel chromatin parameter for replication origin functions. Genome Research. 2017;27(2):269–277. doi:10.1101/gr.209940.116

45. Gros J, Kumar C, Lynch G, Yadav T, Whitehouse I, Remus D. Post-licensing Specification of Eukaryotic Replication Origins by Facilitated Mcm2-7 Sliding along DNA. Molecular Cell. 2015;60(5):797–807. doi:10.1016/j.molcel.2015.10.022

46. Jaremko MJ, On KF, Thomas DR, Stillman B, Joshua-Tor L. The dynamic nature of the human origin recognition complex revealed through five cryoEM structures Subramaniam S, Kuriyan J, Grant T, editors. eLife. 2020;9:e58622. doi:10.7554/eLife.58622

47. Schmidt JM, Bleichert F. Structural mechanism for replication origin binding and remodeling by a metazoan origin recognition complex and its co-loader Cdc6. Nature Communications. 2020;11(1):4263. doi:10.1038/s41467-020-18067-7

48. Tocilj A, On KF, Yuan Z, Sun J, Elkayam E, Li H, Stillman B, Joshua-Tor L. Structure of the active form of human origin recognition complex and its ATPase motor module Kuriyan J, editor. eLife. 2017;6:e20818. doi:10.7554/eLife.20818

49. Dueber ELC, Corn JE, Bell SD, Berger JM. Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. Science (New York, N.Y.). 2007;317(5842):1210–1213. doi:10.1126/science.1143690

50. Bleichert F. Mechanisms of replication origin licensing: a structural perspective. Current Opinion in Structural Biology. 2019;59:195–204. (Catalysis and Regulation ● Protein Nucleic Interactions). doi:10.1016/j.sbi.2019.08.007

51. J S, G T, Wa H. The AAA+ superfamily of functionally diverse proteins. Genome Biology. 2008;9(4):216–216. doi:10.1186/gb-2008-9-4-216

52. Tsakraklides V, Bell SP. Dynamics of pre-replicative complex assembly. The Journal of Biological Chemistry. 2010;285(13):9437–9443. doi:10.1074/jbc.M109.072504

53. Harami GM, Gyimesi M, Kovács M. From keys to bulldozers: expanding roles for winged helix domains in nucleic-acid-binding proteins. Trends in Biochemical Sciences. 2013;38(7):364–371. doi:10.1016/j.tibs.2013.04.006

54. Noguchi K, Vassilev A, Ghosh S, Yates JL, DePamphilis ML. The BAH domain facilitates the ability of human Orc1 protein to activate replication origins in vivo. The EMBO Journal. 2006;25(22):5372–5382. doi:10.1038/sj.emboj.7601396

55. De Ioannes P, Leon VA, Kuang Z, Wang M, Boeke JD, Hochwagen A, Armache K-J. Structure and function of the Orc1 BAH-nucleosome complex. Nature Communications. 2019;10(1):2894. doi:10.1038/s41467-019-10609-y

56. Dueber ELC, Corn JE, Bell SD, Berger JM. Replication Origin Recognition and Deformation by a Heterodimeric Archaeal Orc1 Complex. Science. 2007;317(5842):1210–1213. doi:10.1126/science.1143690

57. Ausiannikava D, Allers T. Diversity of DNA Replication in the Archaea. Genes. 2017;8(2):56. doi:10.3390/genes8020056

58. Wyrick JJ, Aparicio JG, Chen T, Barnett JD, Jennings EG, Young RA, Bell SP, Aparicio OM. Genome-Wide Distribution of ORC and MCM Proteins in S. cerevisiae: High-Resolution Mapping of Replication Origins. Science. 2001;294(5550):2357–2360. doi:10.1126/science.1066101

59. Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, Lockhart DJ, Davis RW, Brewer BJ, Fangman WL. Replication Dynamics of the Yeast Genome. Science. 2001;294(5540):115. doi:10.1126/science.294.5540.115

60. Müller CA, Boemo MA, Spingardi P, Kessler BM, Kriaucionis S, Simpson JT, Nieduszynski CA. Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads. Nature Methods. 2019;16(5):429–436. doi:10.1038/s41592-019-0394-y

61. Stevenson JB, Gottschling DE. Telomeric chromatin modulates replication timing near chromosome ends. Genes & Development. 1999;13(2):146–151.

62. Mantiero D, Mackenzie A, Donaldson A, Zegerman P. Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. The EMBO Journal. 2011;30(23):4805–4814. doi:10.1038/emboj.2011.404

63. Schwob E, Nasmyth K. CLB5 and CLB6, a new pair of B cyclins involved in DNA replication in Saccharomyces cerevisiae. Genes & Development. 1993;7(7A):1160–1175. doi:10.1101/gad.7.7a.1160

64. McCune HJ, Danielson LS, Alvino GM, Collingwood D, Delrow JJ, Fangman WL, Brewer BJ, Raghuraman MK. The Temporal Program of Chromosome Replication: Genomewide Replication in clb5Δ Saccharomyces cerevisiae. Genetics. 2008;180(4):1833–1847. doi:10.1534/genetics.108.094359

65. Rhind N, Gilbert DM. DNA Replication Timing. Cold Spring Harbor Perspectives in Biology. 2013 [accessed 2021 Apr 10];5(8). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3721284/. doi:10.1101/cshperspect.a010132

66. Dukaj L, Rhind N. The capacity of origins to load MCM establishes replication timing patterns. PLOS Genetics. 2021;17(3):e1009467. doi:10.1371/journal.pgen.1009467

67. Hoggard T, Müller CA, Nieduszynski CA, Weinreich M, Fox CA. Sir2 mitigates an intrinsic imbalance in origin licensing efficiency between early- and late-replicating euchromatin. Proceedings of the National Academy of Sciences. 2020;117(25):14314–14321. doi:10.1073/pnas.2004664117

68. Raghuraman MK, Liachko I. Sequence Determinants of Yeast Replication Origins. In: Kaplan DL, editor. The Initiation of DNA Replication in Eukaryotes. Cham: Springer International Publishing; 2016. p. 123–141. https://doi.org/10.1007/978-3-319-24696-3_7. doi:10.1007/978-3-319-24696-3_7

69. Di Rienzi SC, Lindstrom KC, Lancaster R, Rolczynski L, Raghuraman MK, Brewer BJ. Genetic, genomic, and molecular tools for studying the protoploid yeast, L. waltii. Yeast (Chichester, England). 2011;28(2):137–151. doi:10.1002/yea.1826

70. Liachko I, Bhaskar A, Lee C, Chung SCC, Tye B-K, Keich U. A Comprehensive Genome-Wide Map of Autonomously Replicating Sequences in a Naive Genome. PLOS Genetics. 2010;6(5):e1000946. doi:10.1371/journal.pgen.1000946

71. Liachko I, Tanaka E, Cox K, Chung SCC, Yang L, Seher A, Hallas L, Cha E, Kang G, Pace H, et al. Novel features of ARS selection in budding yeast Lachancea kluyveri. BMC Genomics. 2011;12(1):633. doi:10.1186/1471-2164-12-633

72. Liachko I, Youngblood RA, Tsui K, Bubb KL, Queitsch C, Raghuraman MK, Nislow C, Brewer BJ, Dunham MJ. GC-rich DNA elements enable replication origin activity in the methylotrophic yeast Pichia pastoris. PLoS genetics. 2014;10(3):e1004169. doi:10.1371/journal.pgen.1004169

73. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols. 2010;5(4):725–738. doi:10.1038/nprot.2010.5

74. Laughery MF, Hunter T, Brown A, Hoopes J, Ostbye T, Shumaker T, Wyrick JJ. New Vectors for Simple and Streamlined CRISPR-Cas9 Genome Editing in Saccharomyces cerevisiae. Yeast (Chichester, England). 2015;32(12):711–720. doi:10.1002/yea.3098

75. Liachko I, Dunham MJ. An Autonomously Replicating Sequence for use in a wide range of budding yeasts. FEMS yeast research. 2014;14(2):364–367. doi:10.1111/1567-1364.12123

76. Di Rienzi SC, Lindstrom KC, Mann T, Noble WS, Raghuraman MK, Brewer BJ. Maintaining replication origins in the face of genomic change. Genome Research. 2012;22(10):1940–1952. doi:10.1101/gr.138248.112

77. Poloumienko A, Dershowitz A, De J, Newlon CS. Completion of Replication Map of Saccharomyces cerevisiae Chromosome III. Molecular Biology of the Cell. 2001;12(11):3317–3327.

78. Lieb JD, Liu X, Botstein D, Brown PO. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein–DNA association. Nature Genetics. 2001;28(4):327–334. doi:10.1038/ng569

79. Knott SRV, Peace JM, Ostrow AZ, Gan Y, Rex AE, Viggiani CJ, Tavaré S, Aparicio OM. Forkhead Transcription Factors Establish Origin Timing and Long-Range Clustering in S. cerevisiae. Cell. 2012;148(1):99–111. doi:10.1016/j.cell.2011.12.012

80. Hoggard T, Liachko I, Burt C, Meikle T, Jiang K, Craciun G, Dunham MJ, Fox CA. High Throughput Analyses of Budding Yeast ARSs Reveal New DNA Elements Capable of Conferring Centromere-Independent Plasmid Propagation. G3: Genes|Genomes|Genetics. 2016;6(4):993–1012. doi:10.1534/g3.116.027904

81. Feng W, Collingwood D, Boeck ME, Fox LA, Alvino GM, Fangman WL, Raghuraman MK, Brewer BJ. Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. Nature Cell Biology. 2006;8(2):148–155. doi:10.1038/ncb1358

82. Brewer BJ, Fangman WL. The localization of replication origins on ARS plasmids in S. cerevisiae. Cell. 1987;51(3):463–471. doi:10.1016/0092-8674(87)90642-8

83. Foss EJ, Gatbonton-Schwager T, Thiesen AH, Taylor E, Soriano R, Lao U, MacAlpine DM, Bedalov A. Sir2 suppresses transcription-mediated displacement of Mcm2-7 replicative helicases at the ribosomal DNA repeats. PLOS Genetics. 2019;15(5):e1008138. doi:10.1371/journal.pgen.1008138

84. Nieduszynski CA, Knox Y, Donaldson AD. Genome-wide identification of replication origins in yeast by comparative genomics. Genes & Development. 2006;20(14):1874–1879. doi:10.1101/gad.385306

85. Remus D, Beall EL, Botchan MR. DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding. The EMBO journal. 2004;23(4):897–907. doi:10.1038/sj.emboj.7600077

86. Lee CSK, Cheung MF, Li J, Zhao Y, Lam WH, Ho V, Rohs R, Zhai Y, Leung D, Tye B-K. Humanizing the yeast origin recognition complex. Nature Communications. 2021;12(1):33. doi:10.1038/s41467-020-20277-y

87. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods. 2020;17(3):261–272. doi:10.1038/s41592-019-0686-2

88. Hunter JD. Matplotlib: A 2D Graphics Environment. Computing in Science Engineering. 2007;9(3):90–95. doi:10.1109/MCSE.2007.55

89. Kwan EX, Wang XS, Amemiya HM, Brewer BJ, Raghuraman MK. rDNA Copy Number Variants Are Frequent Passenger Mutations in Saccharomyces cerevisiae Deletion Collections and de Novo Transformants. G3 (Bethesda, Md.). 2016;6(9):2829–2838. doi:10.1534/g3.116.030296

90. Rubin AF, Gelman H, Lucas N, Bajjalieh SM, Papenfuss AT, Speed TP, Fowler DM. A statistical framework for analyzing deep mutational scanning data. Genome Biology. 2017;18(1):150. doi:10.1186/s13059-017-1272-5

91. Zentner GE, Kasinathan S, Xin B, Rohs R, Henikoff S. ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. Nature Communications. 2015;6(1):8733. doi:10.1038/ncomms9733

92. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17(1):10–12. doi:10.14806/ej.17.1.200

93. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012;9(4):357–359. doi:10.1038/nmeth.1923

94. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular Cell. 2010;38(4):576–589. doi:10.1016/j.molcel.2010.05.004

95. Ganai RA, Johansson E. DNA Replication—A Matter of Fidelity. Molecular Cell. 2016;62(5):745–755. doi:10.1016/j.molcel.2016.05.003

96. Kunkel TA, Bebenek K. DNA Replication Fidelity. Annual Review of Biochemistry. 2000;69(1):497–529. doi:10.1146/annurev.biochem.69.1.497

97. Lu S, Wang G, Bacolla A, Zhao J, Spitser S, Vasquez KM. Short Inverted Repeats Are Hotspots for Genetic Instability: Relevance to Cancer Genomes. Cell Reports. 2015;10(10):1674–1680. doi:10.1016/j.celrep.2015.02.039

98. Preston BD, Albertson TM, Herr AJ. DNA replication fidelity and cancer. Seminars in Cancer Biology. 2010;20(5):281–293. (Mutator Phenotype in Cancer: Origins and Consequences). doi:10.1016/j.semcancer.2010.10.009

99. Carvalho CMB, Ramocki MB, Pehlivan D, Franco LM, Gonzaga-Jauregui C, Fang P, McCall A, Pivnick EK, Hines-Dowell S, Seaver LH, et al. Inverted genomic segments and complex triplication rearrangements are mediated by inverted repeats in the human genome. Nature Genetics. 2011;43(11):1074–1081. doi:10.1038/ng.944

100. Paulson H. Chapter 9 - Repeat expansion diseases. In: Geschwind DH, Paulson HL, Klein C, editors. Handbook of Clinical Neurology. Vol. 147. Elsevier; 2018. p. 105–123. (Neurogenetics, Part I). https://www.sciencedirect.com/science/article/pii/B9780444632333000099. doi:10.1016/B978-0-444-63233-3.00009-9

101. Carvalho CMB, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. Nature Reviews Genetics. 2016;17(4):224–238. doi:10.1038/nrg.2015.25

102. Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, DeSevo CG, Botstein D, Dunham MJ. The Repertoire and Dynamics of Evolutionary Adaptations to Controlled Nutrient-Limited Environments in Yeast. PLoS Genetics. 2008 [accessed 2021 Apr 15];4(12). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2586090/. doi:10.1371/journal.pgen.1000303

103. Croll D, Zala M, McDonald BA. Breakage-fusion-bridge Cycles and Large Insertions Contribute to the Rapid Evolution of Accessory Chromosomes in a Fungal Pathogen. PLOS Genetics. 2013;9(6):e1003567. doi:10.1371/journal.pgen.1003567

104. Gorkovskiy A, Verstrepen KJ. The Role of Structural Variation in Adaptation and Evolution of Yeast and Other Fungi. Genes. 2021;12(5):699. doi:10.3390/genes12050699

105. Payen C, Di Rienzi SC, Ong GT, Pogachar JL, Sanchez JC, Sunshine AB, Raghuraman MK, Brewer BJ, Dunham MJ. The Dynamics of Diverse Segmental Amplifications in Populations of Saccharomyces cerevisiae Adapting to Strong Selection. G3: Genes|Genomes|Genetics. 2013;4(3):399–409. doi:10.1534/g3.113.009365

106. Araya CL, Payen C, Dunham MJ, Fields S. Whole-genome sequencing of a laboratory-evolved yeast strain. BMC Genomics. 2010;11:88. doi:10.1186/1471-2164-11-88

107. Llorente B, Smith CE, Symington LS. Break-induced replication: What is it and what is it for? Cell Cycle. 2008;7(7):859–864. doi:10.4161/cc.7.7.5613

108. Brewer BJ, Payen C, Raghuraman MK, Dunham MJ. Origin-Dependent Inverted-Repeat Amplification: A Replication-Based Model for Generating Palindromic Amplicons. PLOS Genetics. 2011;7(3):e1002016. doi:10.1371/journal.pgen.1002016

109. Yu S, Graf WD. Telomere capture as a frequent mechanism for stabilization of the terminal chromosomal deletion associated with inverted duplication. Cytogenetic and Genome Research. 2010;129(4):265–274. doi:10.1159/000315887

110. Brewer BJ, Payen C, Rienzi SCD, Higgins MM, Ong G, Dunham MJ, Raghuraman MK. Origin-Dependent Inverted-Repeat Amplification: Tests of a Model for Inverted DNA Amplification. PLOS Genetics. 2015;11(12):e1005699. doi:10.1371/journal.pgen.1005699

111. Hu Y, Tareen A, Sheu Y-J, Ireland WT, Speck C, Li H, Joshua-Tor L, Kinney JB, Stillman B. Evolution of DNA replication origin specification and gene silencing mechanisms. Nature Communications. 2020;11(1):5175. doi:10.1038/s41467-020-18964-x

112. Sánchez H, McCluskey K, van Laar T, van Veen E, Asscher FM, Solano B, Diffley JFX, Dekker NH. DNA replication origins retain mobile licensing proteins. Nature Communications. 2021;12:1908. doi:10.1038/s41467-021-22216-x

113. Kwan EX, Alvino GM, Lynch KL, Levan PF, Amemiya HM, Wang XS, Johnson SA, Sanchez JC, Miller MA, Croy M, et al. A minimal rDNA array replicates early, delays genome replication, and uncouples anaphase entry from S phase completion. bioRxiv. 2021 Feb 25:2021.02.25.432950. doi:10.1101/2021.02.25.432950

114. Miotto B, Ji Z, Struhl K. Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. Proceedings of the National Academy of Sciences. 2016;113(33):E4810–E4819. doi:10.1073/pnas.1609060113

Appendix

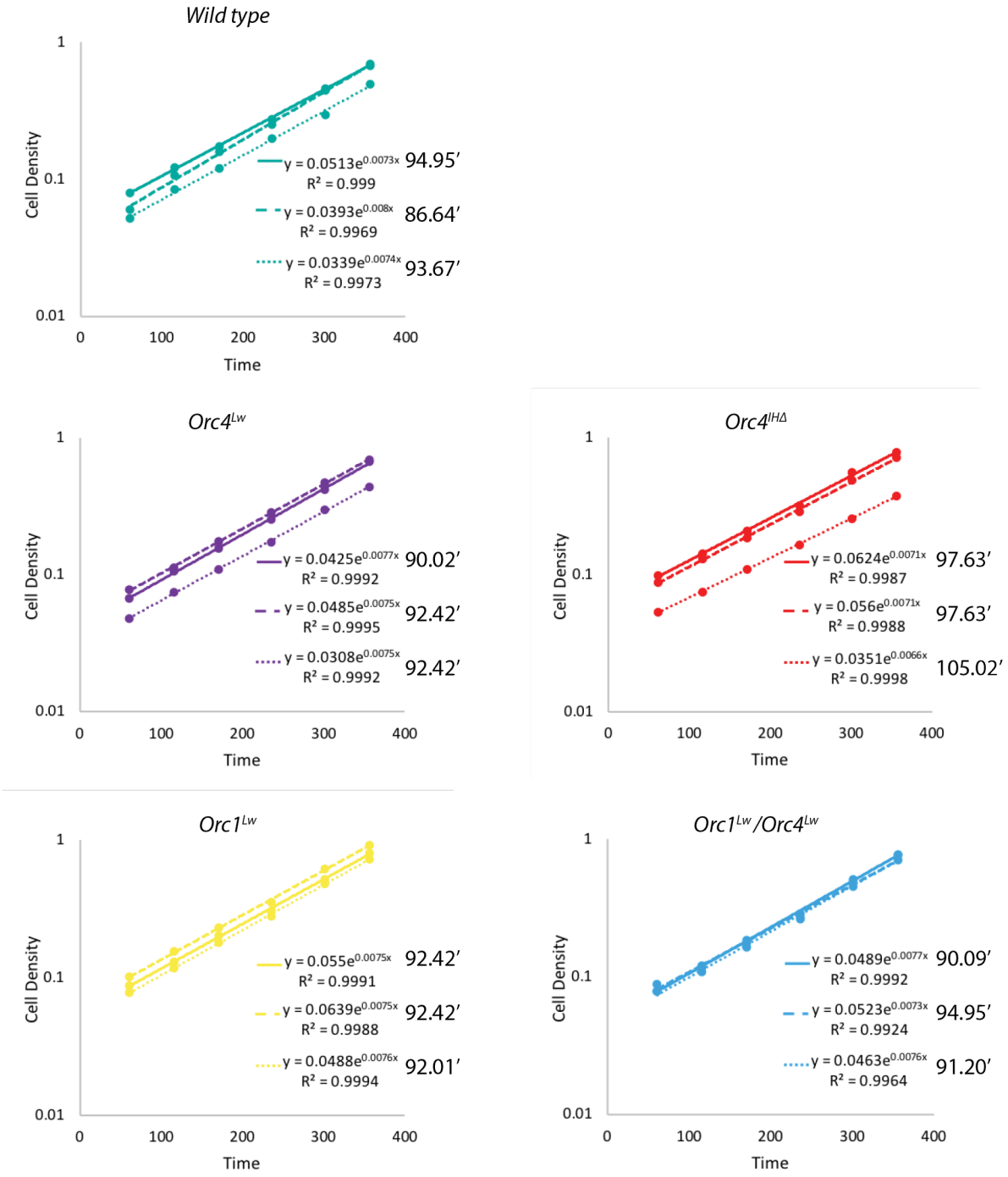
Table A: Percent Identity of Orc4-IH between species

	S. cerevisiae	*L. kluyveri*	*L. waltii*	*K. lactis*	*P. pastoris*
S. cerervisiae	100%				
L. kluyveri	67.86%	100%			
L. waltii	67.86%	75%	100%		
K. lactis	28.57%	35.71%	32.14%	100%	
P. pastoris	10.71%	7.14%	14.29%	14.29%	100%

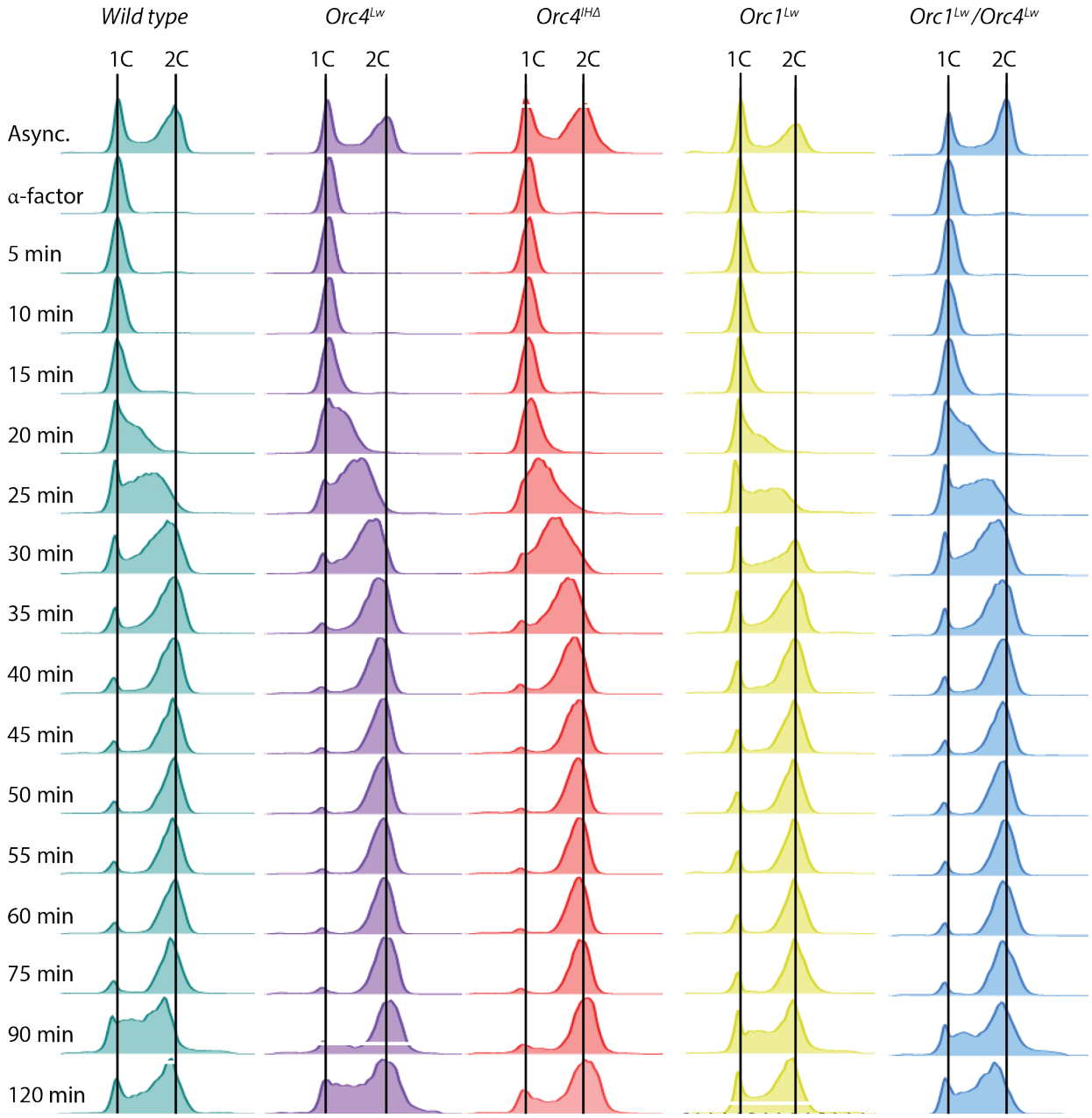
Table B: Percent Identity of Orc4 between species

	S. cerevisiae	*L. kluyveri*	*L. waltii*	*K. lactis*	*P. pastoris*
S. cerervisiae	100%				
L. kluyveri	51.32%	100%			
L. waltii	49.15%	58.83%	100%		
K. lactis	38.56%	38.19%	37.97%	100%	
P. pastoris	19.73%	22.22%	22.92%	20.50%	100%

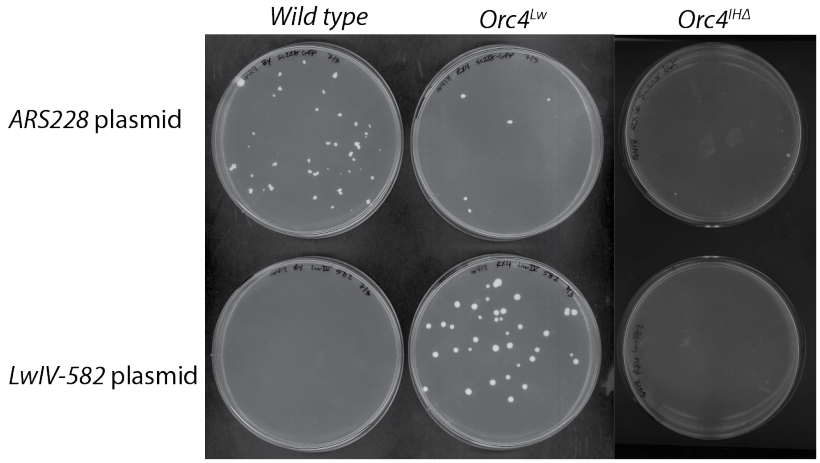
Supplemental Table 2.1: Percent identity between budding yeasts' Orc4-IH (A) and Orc4 (B). Percent identity is higher between *Lachancea* and *Saccharomyces* Orc4-IH than between *Lachancea* and *Saccharomyces* Orc4.



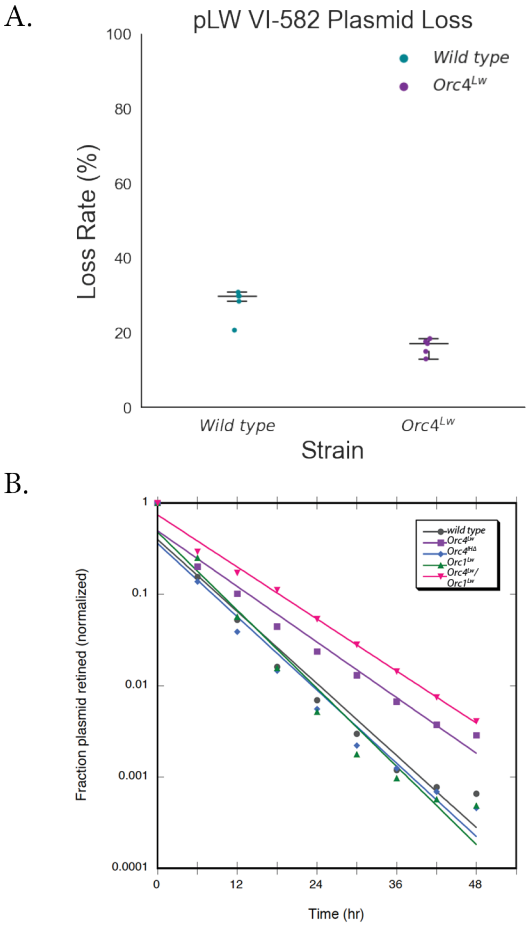
Supplemental Figure 2.1: Growth curves for wild type and mutant ORC strains. Each graph depicts 3/5 of the growth curve replicates used to determine doubling time. Exponential lines are fitted to each set of data points and the corresponding doubling time is displayed to the right. Data show that all strains except the deletion have relative comparable doubling times (average between 90 and 93 minutes) while the deletion is 5-8 minutes slower. Growth curves conducted in complete synthetic medium.



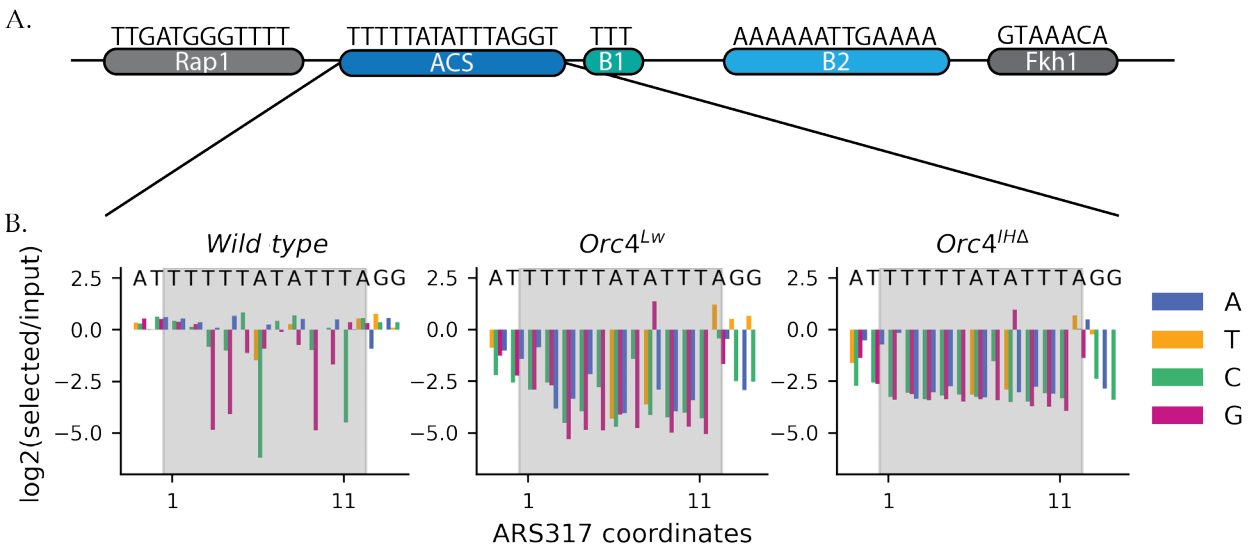
Supplemental Figure 2.2: S-phase progression for wild-type and mutant ORC strains. 1C line denotes one copy of genomic DNA, while 2C denotes two copies of genomic DNA. Changes between mutant strains and wild type are subtle. The Orc4 chimera appears somewhat slower moving through S phase (20 min - 45 min) as does the Orc4$^{IH\Delta}$. The deletion appears to also spend more time in G2. Changes in the Orc1 chimera or double chimera are also subtle and may be partially obscurred by the population of cells that are slower to release from G1 (an artifact of the assay/collection, not necessarily of the strain).



Supplemental Figure 2.3: Initial plasmid transformations with *ARS228* and *LwIV-582* origin plasmids
Initial transformations show that the *S. cerevisiae* origin plasmid does transform into wild type and the chimera, but with seemingly lower transformation efficiency (alternatively, colonies in the chimera may be too small to be seen), but on two colonies are seen in the deletion (circled in red). In contrast, the *L. waltii* origin plasmid does not transform into wild-type *S. cerevisiae* (as expected), but does produce colonies in the chimera. No colonies are visible in the deletion transformed with the *L. waltii* origin plasmid.



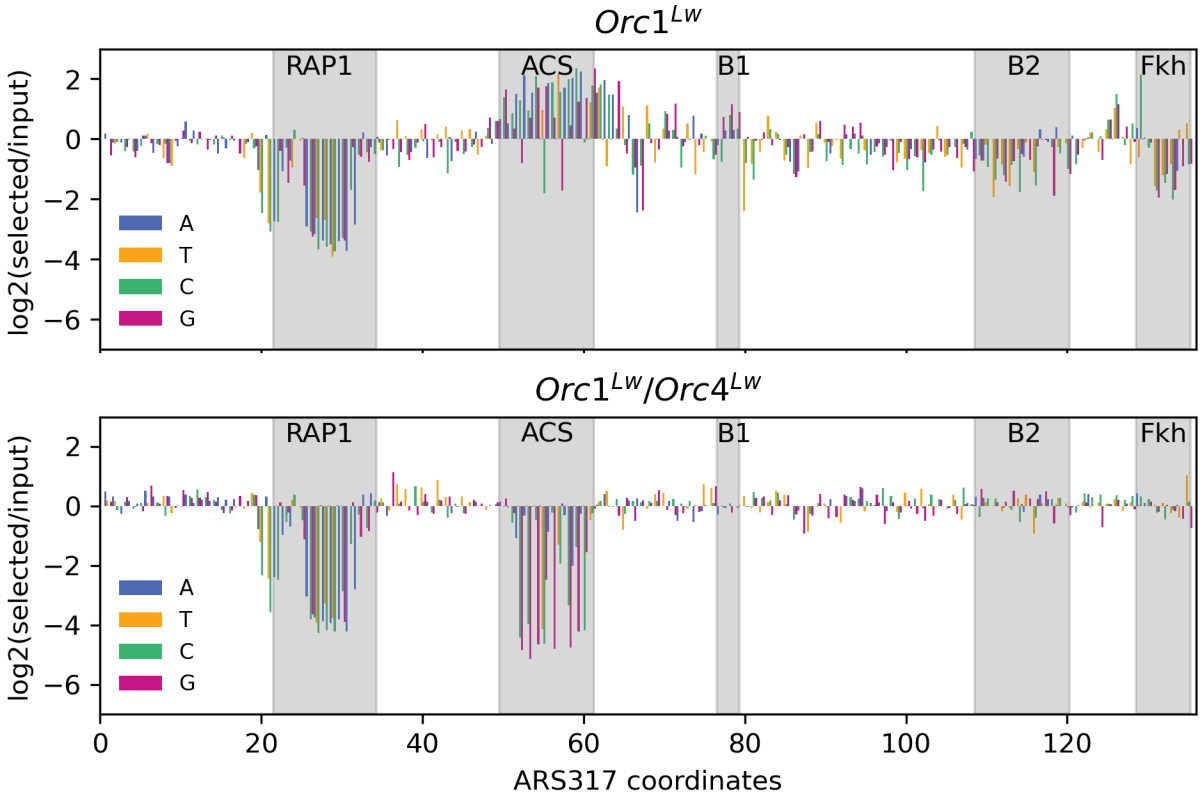
Supplemental Figure 2.4: Plasmid loss rates for *L. waltii* origin plasmid in all strains. A) Five replicates each of wild-type and chimera transformations were grown to stationary in non-selective media and show that the chimera loss rate is lower than wild type. B) Log phase growth of all five strains in non selective media with time points taken every six hours show that both the Orc4 and Orc1/4 chimeras have lower loss rates than the other strains. The double chimera marginally outperforms the Orc4 chimera.



Supplemental Figure 2.5: µARS317-Seq reveals different tolerances to ACS mutations across strains. A) A diagram of *ARS317* and the conserved elements within the origin. The Rap1 binding site is crucial for plasmid segregation while the ACS, B1, and potentially the B2 elements are important for ORC recognition. B) Relative frequency of ACS variants after transformation with µARS317-Seq library. Wild-type *S. cerevisiae* shows mild tolerance for a range of mutations in the ACS, while the chimera and deletion strains are more intolerant of ACS mutations with a few exceptions that more closely match the *L. waltii* ACS.

15min
Wt vs wt: KstestResult(statistic=0.0, pvalue=1.0)
Orc4 chimera vs wt: KstestResult(statistic=0.3333333333333333, pvalue=2.749213992556676e-10)
Orc4-IHdel vs wt: 6 KstestResult(statistic=0.5223880597014925, pvalue=2.118921073212328e-25)
30min
Wt vs wt: KstestResult(statistic=0.0, pvalue=1.0)
Orc4 chimera vs wt: KstestResult(statistic=0.2885572139303483, pvalue=8.851120895960871e-08)
Orc4-IHdel vs wt: 6 KstestResult(statistic=0.4129353233830846, pvalue=1.0043417118769988e-15)
60min
Wt vs wt: KstestResult(statistic=0.0, pvalue=1.0)
Orc4 chimera vs wt: KstestResult(statistic=0.48756218905472637, pvalue=4.974565763226831e-22)
Orc4-IHdel vs wt: 6 KstestResult(statistic=0.6467661691542289, pvalue=6.321647948233494e-40)
90min
Wt vs wt: KstestResult(statistic=0.0, pvalue=1.0)
Orc4 chimera vs wt: KstestResult(statistic=0.15920398009950248, pvalue=0.012151438844154684)
Orc4-IHdel vs wt: 6 KstestResult(statistic=0.7014925373134329, pvalue=1.0153137588476315e-47)
120min
Wt vs wt: KstestResult(statistic=0.0, pvalue=1.0)
Orc4 chimera vs wt: KstestResult(statistic=0.7014925373134329, pvalue=1.0153137588476315e-47)
Orc4-IHdel vs wt: 6 KstestResult(statistic=0.3781094527363184, pvalue=3.4565155601662736e-13)

Supplemental Table 2.2: Kolmogorov Smirnoff test statistics of wild-type average early origin peak shape versus Orc4 mutants (pairwise comparison)



Supplemental Figure 2.6: µARS317-Seq data for entirety of *ARS317* for Orc1 chimera and Orc1/4 double chimera. Both strains show intolerance of Rap1 mutations as expected, but the Orc1 chimera tolerates/prefers almost all ACS mutations while showing intolerance for mutations along the B2 element and surrounding sequence.

15min				
KruskalResult(statistic=113.68880645442505, pvalue=1.763433658540674e-24)				
	Wt	Orc1/4	Orc4	Orc1
Wt	1.000000e+00	2.047536e-19	7.811817e-13	5.869607e-01
Orc1/4	2.047536e-19	1.000000e+00	4.987387e-02	1.815517e-17
Orc4	7.811817e-13	4.987387e-02	1.000000e+00	3.035168e-11
Orc1	5.869607e-01	1.815517e-17	3.035168e-11	1.000000e+00
30min				
KruskalResult(statistic=103.956240029112, pvalue=2.191157689058762e-22)				
	Wt	Orc1/4	Orc4	Orc1
Wt	1.000000e+00	1.200440e-02	1.482846e-03	1.837980e-14
Orc1/4	1.200440e-02	1.000000e+00	1.620892e-08	1.520006e-23
Orc4	1.482846e-03	1.620892e-08	1.000000e+00	4.521103e-06
Orc1	1.837980e-14	1.520006e-23	4.521103e-06	1.000000e+00
60min				
KruskalResult(statistic=117.31788420463727, pvalue=2.91758122392146e-25)				
	Wt	Orc1/4	Orc4	Orc1
Wt	1.000000e+00	2.535767e-23	3.200701e-22	4.544164e-11
Orc1/4	2.535767e-23	1.000000e+00	7.797355e-01	3.493030e-04
Orc4	3.200701e-22	7.797355e-01	1.000000e+00	9.699395e-04
Orc1	4.544164e-11	3.493030e-04	9.699395e-04	1.000000e+00
90min				
KruskalResult(statistic=377.8631984233825, pvalue=1.3797524691968608e-81)				
	Wt	Orc1/4	Orc4	Orc1
Wt	1.000000e+00	4.342090e-04	9.895975e-05	9.724386e-67
Orc1/4	4.342090e-04	1.000000e+00	7.042400e-01	1.832689e-87
Orc4	9.895975e-05	7.042400e-01	1.000000e+00	9.575598e-90
Orc1	9.724386e-67	1.832689e-87	9.575598e-90	1.000000e+00

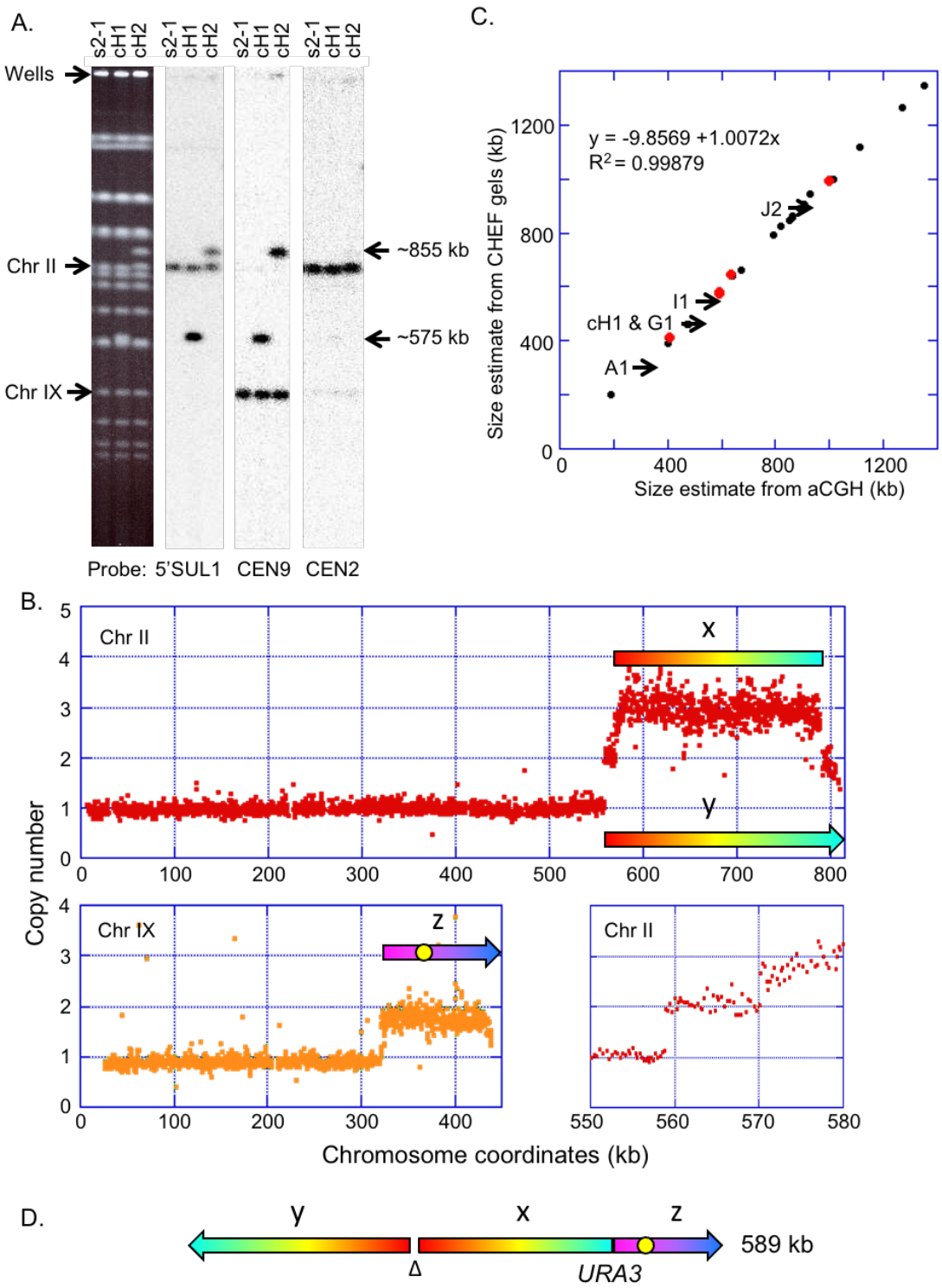
Supplemental Table 2.2: Kruskal Wallace and Conover's statistics for average peak shape of active early origins in wild type and chimeric strains.

15min				
KruskalResult(statistic=113.68880645442505, pvalue=1.763433658540674e-24)				
	Wt	Orc1/4	Orc4	Orc1
Wt	1.000000e+00	2.047536e-19	7.811817e-13	5.869607e-01
Orc1/4	2.047536e-19	1.000000e+00	4.987387e-02	1.815517e-17
Orc4	7.811817e-13	4.987387e-02	1.000000e+00	3.035168e-11
Orc1	5.869607e-01	1.815517e-17	3.035168e-11	1.000000e+00
30min				
KruskalResult(statistic=103.956240029112, pvalue=2.191157689058762e-22)				
	Wt	Orc1/4	Orc4	Orc1
Wt	1.000000e+00	1.200440e-02	1.482846e-03	1.837980e-14
Orc1/4	1.200440e-02	1.000000e+00	1.620892e-08	1.520006e-23
Orc4	1.482846e-03	1.620892e-08	1.000000e+00	4.521103e-06
Orc1	1.837980e-14	1.520006e-23	4.521103e-06	1.000000e+00
60min				
KruskalResult(statistic=117.31788420463727, pvalue=2.91758122392146e-25)				
	Wt	Orc1/4	Orc4	Orc1
Wt	1.000000e+00	2.535767e-23	3.200701e-22	4.544164e-11
Orc1/4	2.535767e-23	1.000000e+00	7.797355e-01	3.493030e-04
Orc4	3.200701e-22	7.797355e-01	1.000000e+00	9.699395e-04
Orc1	4.544164e-11	3.493030e-04	9.699395e-04	1.000000e+00
90min				
KruskalResult(statistic=377.8631984233825, pvalue=1.3797524691968608e-81)				
	Wt	Orc1/4	Orc4	Orc1
Wt	1.000000e+00	4.342090e-04	9.895975e-05	9.724386e-67
Orc1/4	4.342090e-04	1.000000e+00	7.042400e-01	1.832689e-87
Orc4	9.895975e-05	7.042400e-01	1.000000e+00	9.575598e-90
Orc1	9.724386e-67	1.832689e-87	9.575598e-90	1.000000e+00

Supplemental Table 2.3: Kruskal Wallace and Conover's statistics for average peak shape of active early origins in wild type and chimeric strains.

Strain Name	Strain ID	Strain Background	Genotype
	BY4741	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0
	BY4742		MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0;
Orc4-IH Lw	yRX4	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0
Orc4-IH del	yRX6	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0
Orc4-IH del	yRX7	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0
Orc1 Lw	yRX8	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0
Orc1 Lw/Orc4-IHlw	yRX25	BY4741/2	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0
Orc4-IH lw MCM2-MNase	yRX4M	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 Kan
Orc1 Lw MCM2-MCM2-MNase	yRX27	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 Kan
Orc1 Lw MCM2-MNase	yRX28	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 Kan
Orc4-IH del MCM2-MNase	yRX32	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 Kan
Orc1/Orc4-IH lw MCM2-MNase	yRX33	BY4741/2	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 Kan ?
Orc1/Orc4-IH lw MCM2-MNase	yRX34	BY4741/2	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 Kan ?
MCM2-MNase	yAB16747	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 Kan

Supplemental Table 2.5: Strains used in chapter 2.



Supplemental Figure 3.1 Another Clone consistent with ODIRA hairpin intermediate. A) cH-1 has a neochromosome with characteristic 2:1 *5'SUL1* ratio at neochromosome versus chromosome II. The neochromosome contains *CEN9*. B) aCGH shows the characteristic waterfall increase in copy number preceding the triplication of the ODIRA *ura* locus and surrounding region. The jump from 1 to 2 copies preceding the 'waterfall' suggests a small deletion of that region. C) CHEF gel analysis of size predicts a neochromosome of ~575 kb. D) The proposed assembly of the neochromosome.
Figure by Bonny Brewer, used with permission.

Strain ID	Strain Name	Strain Background	Genotype
Split *ura* strain	s2-1	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ
Ura+ clone of split ura strain	cH-1	BY4741	MATα his3Δ1 leu2Δ0 lys2Δ0
Ura+ clone of split ura strain - 2:1 5′SUL1:CEN9	A1	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0
Ura+ clone of split ura strain	G1	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0
Ura+ clone of split ura strain	I1	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0
Ura+ clone of split ura strain	J2	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0
Ura+ clone of split ura strain	cH-2	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0

Supplemental Table 3.1: Chapter 3 (ODIRA) strain list