THE WING HELIX DOMAIN OF ORC4 IS THE PRIMARY DETERMINANT OF DNA BINDING SPECIFICITY OF THE ORIGIN RECOGNITION COMPLEX

A Dissertation

Presented to the Faculty of the Graduate School

Of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

By

Michael William Dowicki

May 2015

THE WING HELIX DOMAIN OF ORC4 IS THE PRIMARY DETERMINANT OF DNA BINDING SPECIFICITY OF THE ORIGIN RECOGNITION COMPLEX

Michael William Dowicki, Ph. D.

Cornell University 2015

The process of DNA replication is regulated to ensure that the entire genome is replicated only once during every cell division cycle. Eukaryotic DNA replication begins with the binding of the Origin Replication Complex (ORC) to multiple replication origins or Autonomously Replicating Sequences (ARSs) on every chromosome. The ORC machinery is conserved from fungal to mammalian systems, however the ARSs to which the ORC binds have diverged significantly. In the budding yeast *S. cerevisiae* the ORC binds a well-defined 17bp ACS, conversely in the fission yeast *S. pombe* ORC binds to AT rich sequences in a stochastic manner.

Previously the replication origins of the yeast *K .lactis* have been identified as a 50bp sequence necessary and largely sufficient for replication. Through testing of plasmids constructed to contain either *S. cerevisiae* or *K. lactis* ARSs, it was found that each species is largely unable to replicate ARSs from the other species, indicating that the replication machinery has significantly diverged to specifically recognize its own origin sequence. In this thesis, I am examining the role subunits of the ORC complex play in determining the binding specificity of the ORC complex.

The ORC proteins contain a DNA binding Winged Helix Domain in their C-termini. I have constructed *S. cerevisiae* strains containing chimeric ORC proteins which interact with the *S. cerevisiae* machinery while containing the *K. lactis* WHD. The chimeric ORC4 and ORC5 proteins fail to substitute for their respective endogenous proteins but the ORC4 chimera results in a dominant loss of silencing at the HMR locus, which is mediated by ORC. ChIP-Seq analysis showed that the chimeric strain binds to a *K. lactis* ACS at this locus and at several other distinct sites in the *S. cerevisiae* genome including the centromeres. Additionally the chimeric ORC4 demonstrates the ability to replicate plasmids containing a *K. lactis* ARS, unlike the wild type *S. cerevisiae* strain and *S. cerevisiae* containing a chimeric ORC5. This study suggests that the DNA binding specificity for the *S. cerevisiae* ORC, *K. lactis* ORC and most likely ORCs in other fungi is primarily determined by the WHD of Orc4.

BIOGRAPHICAL SKETCH

Michael Dowicki was born in 1985 in southern California. He grew up in Westminster, California. He attended Fairmont High school and graduated in 2003. He then attended Amherst College, graduating with honors in 2007 with a BA in Biology. While there he joined the lab of Prof. Dominic Poccia, researching the mechanisms of membrane formation immediately following fertilization in sea urchin eggs. Following graduation he spent a year tutoring students in biology. In 2008 Michael enrolled at Cornell and in 2009 joined the lab of Prof. Bik Tye. During his time in the Tye lab, he spent a year and a half continuing his research at HKUST.

ACKNOWLEDGEMENTS

I have many people I would like to thank. First and foremost I would like to thank Bik for taking me into the lab, making this possible. Her patience and advice have helped me immensely helpful and I am grateful for all of her support. I would also like to thank all the members of the lab who helped me. I would like to thank Xin and Chanmi for helping me get started in the lab, and helping me learn yeast genetics and techniques that I have used. I also need to thank Ivan, for his work on the previous project, without which my work would not have been possible. He has also provided me with advice over the years which has been invaluable.

There have been many others at Cornell who I am very appreciative of. I would like to thank my committee, Prof. Robert Weiss and Prof. Zhenglong Gu for their advice and support. I would also like to thank the Gu lab, especially Xuepeng and Brandon for helping me get set up and understand the computational work. I would also like to thank our collaborator Prof. Uri Keich for his work both with me and Ivan on the computational analysis of our data. I would also like to thank the Fox lab for their support, especially Leah who gave me a lot of good advice.

I am grateful for the help many people gave me during the part of my graduate career spent at HKUST. I would like to thank Prof. Chun Liang and HKUST as a whole for the use of their lab space and expertise. Dr. Rentian Wu also provided me with a great deal of advice and assistance in the lab. I would like to especially thank Dr. Yuanliang Zhai for all his help with the project, along with his great advice. I learned a great deal from him.

I would like to thank my friends and cohort here at Cornell for all their advice and support. Finally I would like to thank my family, especially my parents. Their constant support and advice have helped me immensely.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Chapter 3

LIST OF ACRONYMS USED

AAA+: ATPases Associated with diverse cellular Activities

ACS: ARS Consensus Sequence

ARS: Autonomously Replicating Sequence

ATP: Adenosine Triphosphate

BAH: Bromo Adjacent Homology

ChIPSeq: Chromatin Immunoprecipitation Sequencing

DDK: Dbf4 Dependent Kinase

EM: Electron Microscope

HML: Hidden MAT Left

HMR: Hidden MAT Right

ISM: Initiator Specific Motif

Kl: Kluyveromyces lactis

MAT: Mating Type

MCM: Mini Chromosome Maintenance

ORC: Origin Recognition Complex

RIF2: Rap1p Interacting Factor 2

Sc: Saccharomyces cerevisiae

WGD: Whole Genome Duplication

WHD: Winged Helix Domain

WT: Wild Type

Chapter 1: Introduction:

Evolution of the Origin Recognition Complex (ORC) and Its Binding Sites

DNA replication is a highly conserved and regulated cellular process (Blow and Dutta, 2005). Much of this regulation occurs at the earliest stage of DNA replication, the licensing and initiation of DNA replication (Blow 1993). The general features of the replication process are conserved throughout eukaryotes. In eukaryotes, DNA replication begins at multiple locations called origins throughout the genome requiring regulation to ensure initiation only occurs once per cell cycle. A six subunit complex called the origin recognition complex or ORC binds to these origins serving to recruit the other proteins involved in the initiation of DNA replication and does so both *in vivo* and *in vitro* (Takashi and Diffley 2000, Rowles and Blow, 1997). These sites are well defined and conserved within a species. In yeast, origins are defined in a sequence dependent manner while in higher eukaryotes more complex strategies of defining origins are used (Bryant et al. 2001). Yeast origin sequences, when placed in a plasmid, allow for the initiation of replication of that plasmid independently from the yeast genome. These sequences are therefore referred to as Autonomously Replicating Sequences or ARSs (Brewer and Fangman, 1987). In species such as *S. cerevisiae*, the ARSs contain a well-defined consensus sequence that is essential for ARS function. This sequence is referred to as an ARS Consensus Sequence or ACS.

It has not been determined previously how the binding of the ORC complex is specified. In *S. cerevisiae* the individual subunits of the ORC do not show strong binding ability so the binding specificity of individual subunits cannot be easily assessed (Lee and Bell, 1997). Due to the essential nature of the proteins, disruption of the complex is not especially informative as a lethal phenotype could be indicative of many different possibilities. A mutation which disrupts

the binding specificity of the ORC complex may yield a lethal phenotype, however a lethal phenotype would also occur if a mutant subunit is unstable or if it disrupts interactions with the complex. Additionally the lack of crystal structures of the fungal ORC complex as a whole makes analysis of how the complex interacts with DNA difficult. However, some structural studies in combination with protein-DNA crosslinking studies provided some information on the general orientation of the ORC subunits relative to the origins allowing us to narrow down which subunits are nearest to the ACS (Lee and Bell, 1997, Chen et al. 2008).

Figure 1.1

Fig1.1 A model of the association of Orc1, Orc4 and Orc5 with the *S. cerevisiae* ACS. Orc1, Orc4 and Orc5 are most closely associated with the ACS while Orc5 also associates with the B1 element. This model is based upon the experiments of Lee and Bell, 1997 and Chen et al. 2008. Lee and Bell carried out DNA binding and cross-linking assays to identify the locations of the bound ORC subunits. Chen et al. fused maltose binding protein (MBP) to the N or C terminal end ORC1, ORC2, ORC3, ORC4 and ORC5 constructing nine different complexes (ORC5 tagged at the C-terminal end was not used) and by using electron microscopy with each MBP tagged subunit were able to localize each subunit within the complex.

DNA replication begins with the binding of the ORC complex at genomic loci identified as replication origins (Feng and Kipreos 2003). Origins vary greatly among fungi. The budding yeast *S. cerevisiae* have a well-defined 17bp sequence. Higher eukaryotes do not have consensus sequences with predictive power for replication origins and appear to initiate replication over

large replication zones rather than at specific loci suggesting that DNA topology, rather than sequence, may play a significant role (Remus et al. 2004). Many of the factors that define these regions as origins are still not well understood (Vaughn et al. 1990). While some origins in higher eukaryotes have been well defined to the base pair level, they do not conform to a defined sequence throughout the genome (Bielinsky and Gerbi, 2001). Despite the lack of conformity in the consensus sequences for ORC binding the overall ORC complex is well conserved in structure and in function. The six ORC subunits, Orc1, -2, -3, -4, -5, -6, named in descending order of their sizes, are highly conserved in sequence and structure from species to species throughout the eukaryotes.

The ORC complex, while essential for the first step for DNA replication, requires the recruitment of other proteins to initiate replication. Beginning in G1 phase CDC6 and Cdt1 are recruited to the ORC complex followed by the recruitment of the hexameric MCM2-7 complex in multiple copies per origin. This assembly of the pre-replication complex (pre-RC) during G1 phase is referred to as the licensing of replication origins. The MCM2-7 complex acts as a helicase and serves to unwind DNA and remains with the replication fork during replication elongation. Origin activation requires the recruitment of several other proteins, most notably Cdc45, after the phosphorylation of the MCM2-7 complex by the Dbf4 dependent kinase (DDK), and then GINS, following the action of the cyclin-dependent kinase (CDK) (Diffley et al 1995, Randell et al., 2010). Mcm10 is required for the final step in the activation of the origin and it remains associated with the MCM2-7 complex during replication. The activation culminates in the recruitment of the DNA polymerases, and the beginning of DNA replication.

Even among closely related fungal species, there is dramatic sequence divergence in the ACSs and the number of ORC binding sites throughout the genome. For example, the ACS of *S.*

cerevisiae is 17 bp while that of the related species *K. lactis* is 50 bp (Liachko et al., 2010). *S. cerevisiae* contains between 300 and 400 active origins while *K. lactis* contains only about 150 origins even though these two yeasts have roughly similar genome sizes. In order for genomes to replicate completely in a timely manner within S phase, replication origins have to be distributed throughout the genome without leaving large regions devoid of replication origins. Deletion studies suggest that a single origin may be sufficient for an entire chromosome, however it is unclear if this would be stable on a genome wide level (Desrshowitz et al., 2007). Given that ORC have diverged rapidly in its recognition sequences in a short evolutionary time frame, an obvious consequence of this process is the reassignment of replication origins in a genome-wide scale based on the evolving specificity of ORC. This scenario begs the question of whether the positioning of replication origins plays an important role in the DNA replication cycle. If so, how is the precise positioning of replication origins achieved with an evolving ORC. Several studies implicated a role for DNA replication in gene activation by temporal association (Mechali et al 2013) in that early replication origins are associated with actively transcribed genes (Shor et al., 2009). Other studies showed that ORC invariably binds near transcriptional start sites but the level of transcription regulates origin selection and replication timing suggesting that the establishment of replication origins is passive to chromatin rearrangements (Dellino GI et al, 2013). If the sole function of ORC is to provide a platform for the initiation of DNA replication at sites throughout the genome, then the locations of replication origins within a genome is irrelevant but the number and frequency of replication origins throughout the genome may be important. In support of this notion, comparative genomic analysis of related yeast species showed that there is no conservation of the syntenic locations of the replication origins relative to nearby genes (Liachko et al., 2010). Whether conservation of the positions of

replication origins is important for cell survival or not in fungi, the co-evolution of new binding specificities of ORC and the emergence of a sufficient number of binding sites in the hundreds in intergenic regions is an intriguing process that defies odds.

In this thesis, I will address some of the fundamental problems that evolving species must face as the ORC complex diverged in specificity. The ORC complex consists of six subunits all of which are essential for the binding of ORC to its cognate sequence. I will investigate which of these subunits contribute to the specificity of ORC binding. I will dissect the protein domain(s) that confers the specificity of binding. I will investigate the effect of expressing a foreign ORC with altered binding specificity in the budding yeast to see how the landscape of replication origin selection may or may not be changed.

The Structure and Function of the ORC Complex

The ORC complex is a heterohexamer that is conserved throughout eukaryotes. ORC1 through ORC5 along with CDT1 share a similar core structure, which is found in the homologous proteins of the archeal origin binding proteins(Duncker et al, 2009). This basic structure consists of a WH domain, which have been shown to be involved in DNA binding, and a AAA+ domain which binds and hydrolyzes ATP and is essential for the recruitment and binding of the ORC complex. The binding of ORC to origin DNA takes place in an ATP dependent manner in *S. cerevisiae* (Makise et al, 2003). The requirement for ATP is not universally true however as recruitment of the ORC in *S. pombe* to the *S. pombe* origins occurs in an ATP independent manner which is attributed to the AT hook in the N-terminal of the *S. pombe* ORC4 (Lee, J.K. et al 2001). The *S. pombe* ORC4 contains 9 AT-hooks, each of which consists of basic amino acids that interact with the minor groove. The multiple AT-hooks, each interacting with the DNA, allow for sufficient interaction for binding. In the archeal models the AAA+ domain has been

shown to make contact with the DNA through the initiator specific domain (ISM) and so the AAA+ domain may be directly involved in the binding of the ORC complex to the origin DNA (Gaudier et al, 2007). Of equal importance as the AAA+ domain is the WHD found in ORC1-5. The WHD has been shown to be involved in DNA binding, and in the archeal homologs, makes direct contact to the archeal consensus sequences (Fig1.2) (Gaudier et al, 2007). In contrast to the highly conserved core structure, the N-terminal domains show significant differences. The most notable difference can be seen in ORC1. A well-conserved feature of ORC1 is the N-terminal Bromo-adjacent homology or BAH domain. This large domain makes ORC1 significantly larger than the other ORC subunits with ORC1 consisting of 914 amino acids while the other subunits are between 620 to 479 amino acids. The ORC1 BAH domain is involved in protein-protein interactions with nucleosomes (Eaton, et al. 2010) (Muller et al. 2010). This interaction is important as the ORC complex helps to establish regions of nucleosome free DNA around active origins so that other proteins involved in replication initiation can be successfully recruited. The BAH domain is not essential in yeast, but disruption of the BAH domain leads to a subset of origins becoming unused (Eaton et al. 2010). The importance of this domain is highly conserved as it can also be found in humans and disruptions in the BAH domain in human ORC1 has been linked to the development of Meier-Gorlin syndrome, a disorder which results in primordial dwarfism (Noguchi et al. 2006) (Kuo, et al, 2012).

Figure 1.2.

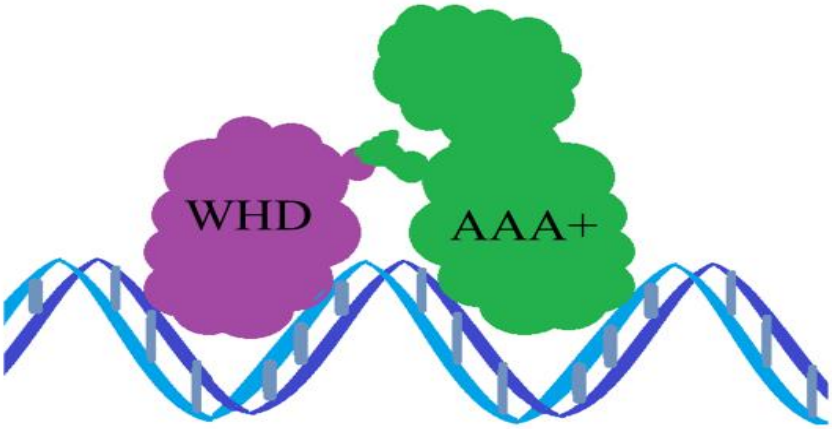

Fig1.2. Based on the work of Gaudier et al, Science, 2007, Model of the Archeal ORC complex. Both the WHD and the AAA+ domain make contact with the ORB sequence, but the majority of the contact is made through the WHD.

Distribution and Locations of Replication Origins

Replication origins are established at specific locations in the genome. They are not found in actively transcribed DNA presumably because bound ORC complex would interfere with transcription of the gene. Origins are not located in closed chromatin as compact DNA may not be easily accessible and reduces the chances of recruitment of the ORC complex and other essential proteins. As a result, replication origins are found exclusively at intergenic regions that are nucleosome free. Another important feature observed in the location of replication origins is their spacing throughout the genome to ensure that all parts of the genome are efficiently replicated. Computational methods to predict *S. cerevisiae* origins have found success by taking the known 12,000 ACSs in the *S. cerevisiae* genome and examining the flanking sequences for their effect on nucleosome placement (Eaton et al. 2010). *S. cerevisiae* ACSs located adjacent to an asymmetrical A rich region that have low affinity for nucleosomes are more likely to serve as origins than ACSs located at other DNA context (Breier et al. 2004).

The activity of origins is temporally regulated with some origins fire early in S phase in nearly all cell cycles, while other origins consistently fire later in S phase. The timing of the firing of origins appears to have important biological significance since the regional replication profiles of related yeast genomes appear to have similar schedules. Experiments using hybrid strains showed that much of the similarity may be explained by local contexts such as sequences which affect nucleosome positioning (Muller and Nieduszynski 2012). In addition to local effects on the regulation of the initiation of DNA replication, there are effects due to global chromatin organization, which determines the accessibility of large regions of the genome to replication proteins (Rhind and Gilbert 2013)

The Divergent DNA Binding Specificities of ORC

The initiation of DNA replication is dependent on the binding of the ORC to replication origins, however the target sequence that ORC binds is distinct to the species. In *S. cerevisiae* ORC binds to a well-defined 17 bp consensus sequence (Nieduszynski et al. 2007, Hamlin et al. 2008). Looking further out in the fungal tree of life (Fig1.3), the fission yeast *S. pombe* ORC lacks a specific consensus sequence for binding. In *S. pombe* ORC binds to larger replication zones that span 500bp to 1kb in size. These replication zones do not have a specific consensus sequence but are very asymmetrically AT rich (Lee et al. 2001, Wu and Nurse 2009). There is a correlation between AT-richness in a replication origin and early firing, but it is not as readily predictable as *S. cerevisiae* origin sequences. *Xenopus* origins do not show any consensus sequence and their ORC complexes do not show any specificity in their DNA binding. However *S. pombe* ORC does show some ability to localize to *Xenopus* origins, suggesting that *S. pombe* origins may have some similarity to the origin structure of metazoans (Kong et al. 2003). Going further out to higher eukaryotes we find a total lack of consensus sequences with replication

origins spanning large multiple KB zones (Vaughn et al. 1990). Higher eukaryotes differ in other ways as well, in that the ORC proteins themselves might not determine the sites of origins. (Cvetic and Walter 2005). This lack of stringency in DNA sequence specificity suggests the involvement of alternative recruitment mechanisms such as interaction with other DNA binding proteins. An example is the possible recruitment of ORC by E2F in *Drosophila melanogaster* during chorion amplification in follicle cells (Zhang and Tower 2004). Similar recruitment strategies can be found in the replication of the Epstein Bar virus where ORC is recruited to the OriP by interaction with EBNA1 (Moriyama et al 2012, Dhar et al, 2001). The promiscuity of ORC in DNA binding may also be used to its advantage during embryonic development in insects and amphibians when cells have to undergo rapid divisions and many more origins are needed for the short S phase (Mechali, 2001, Norio et al, 2005).

Figure 1.3.

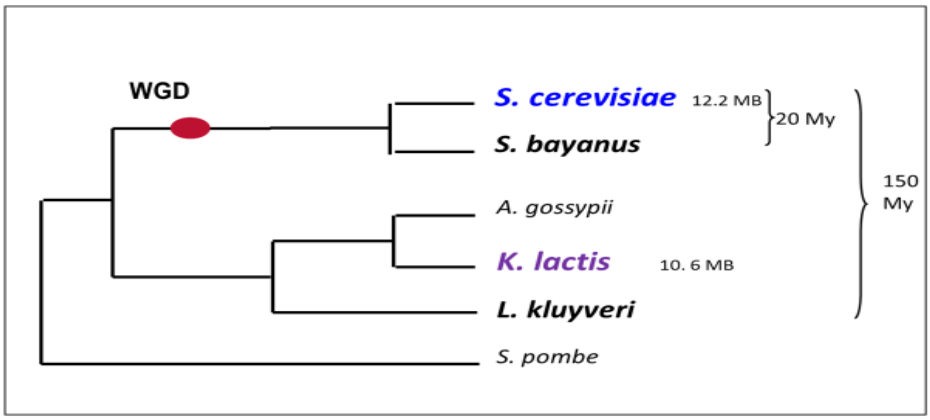

Fig1.3. A subset of species in the fungal tree of life. Those marked in bold are species for which the ACS has been identified. The two species of greatest interest are marked in blue (for *S. cerevisiae*) and purple (for *K. lactis)*. The genome sizes of these two species are shown in megabases. The whole genome duplication event is marked with a red ellipse.

In *S. cerevisiae,* a functional ARS is determined not just by the mere presence of an ACS. While the ACS is present approximately 14,000 times throughout the *S. cerevisiae* genome there are only 300 to 400 active origins (Breier et al. 2004,). This disparity indicates that there must be additional factors in determining whether a given ACS site can serve as an active origin. One of these factors are the three B elements, which are sequences adjacent to the ACS that show poor conservation. Deletion or disruption of one or two B elements results in a reduction of the origin activity while deleting all 3 elements will abolish ARS activity (Celniker et al. 1984).

A strategy to dissect the determining factors for ORC DNA binding specificity is to successively replace one ORC subunit at a time with a foreign Orc protein and look for altered DNA binding specificity in the chimeric ORC complex. An intuitive choice would be to pick the two ORC complexes most diverse in binding specificities such as those of the *S. cerevisiae* ORC and the *S. pombe* ORC for comparison. However, the undefined ACS of *S. pombe* ORC makes it difficult to predict the specificity of a chimeric ORC complex for analysis. Instead, we decided to focus on two well-studied yeast species with well-defined ACSs that are dramatically different for this study. By examining the ACSs in species closely related to *S. cerevisiae* subtler differences can be found, and with the availability of genomic sequences for dozens of fungal species spanning a large evolutionary time frame, this strategy may provide insight into the evolution of the mechanism for DNA replication initiation. The species of *S. cerevisiae*, *S. bayanus*, *L. kluyveri*, and *K. lactis* were examined to determine their ACS (Fig1.4). Looking at these species there are important genetic events that may have shaped the evolution of the ORC complex and the location and number of replication origins. The most important of these is a whole genome duplication event that occurred just before the emergence of *S. cerevisiae* and *S. bayanus* branch, separated from the emergence of *K. lactis* by about 150 My. This duplication event would double the number

of origins immediately after followed by extensive loss of much of the duplicated DNA to reestablish a streamlined genome for the new environment. Additionally the duplication of genes involved in DNA replication may play a role in the evolution of the DNA replication initiation system.

Figure 1.4.

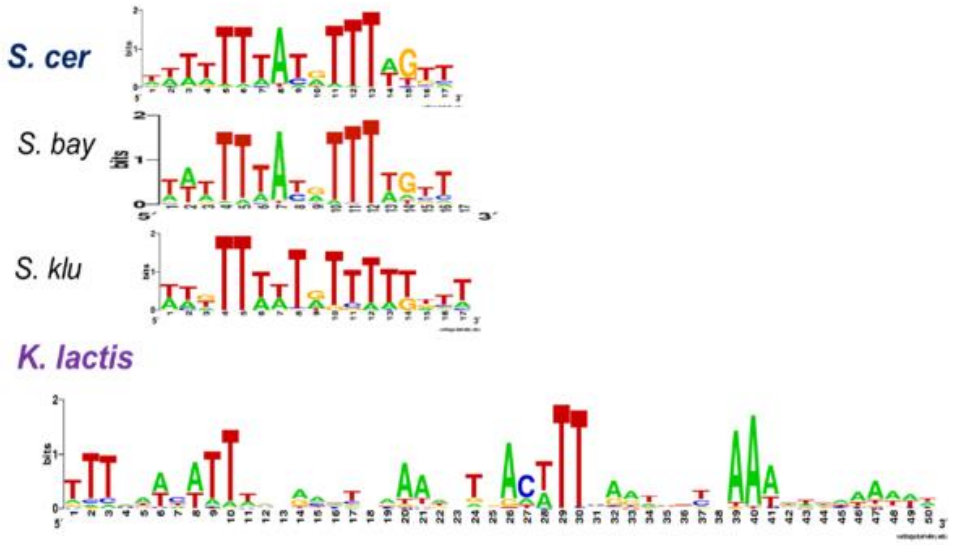

Figure1.4. From Liachko et al, Plos Genetics, 2010. The ACSs from the four yeast species. The ACS of *S. bayanus* is nearly identical to that of *S. cerevisiae* consisting of 17bp. The ACS of *S. kluy*veri is of similar size but is more T rich. The ACS of *K. lactis* is markedly different spanning a 50bp sequence, which resembles an inverted repeat.
(Liachko et al, Plos Genetics, 2010)

By using a plasmid based assay to identify functional ARSs and computational analysis to predict yet to be identified ARSs, Ivan Liachko in the Tye Lab was able to identify the entire repertoire of genomic ARSs and the ACSs of *S. bayanus*, *L. kluyveri*, and *K. lactis.* This assay was done by cloning genomic fragments from each yeast species into non-replicating vectors and looking for fragments that support the replication and maintenance of the vector in each species. In the case of *S. bayanus*, the ACS was determined to be a 17bp sequence that was nearly identical to *S. cerevisiae. L. kluyveri* ARSs also contain a 17bp ARS consensus sequence that is

more T rich, and sometimes occur in multiple copies in tandem. Most interesting of all is the ACS of *K. lactis*, which is 50bp long and appears to have what may be an inverted repeat structure. Furthermore, while the *S. cerevisiae* ACS requires additional B elements to function as an active origin, the *K. lactis* 50bp ACS is sufficient to function as a replication origin without additional sequences. Using an iterative computational approach the entire genomic repertoire of origins for *K. lactis* was determined and approximately 150 origins were found (Liachko et al. 2010).

The cross species function of these ARSs was then determined through a similar plasmid maintenance assay using vectors containing ARSs from each of the four species (Fig1.5). These experiments demonstrated that *S. cerevisiae* and *S. bayanus* are able to initiate replication at all of the ARSs tested from the other species. This result suggests that the sequence specificity of the ORC complex has not changed during the evolution of these two species and that they recognize the same bases. This explanation is consistent with the near identical sequences seen in their ACSs. *L. kluyveri* was able to initiate replication at a fairly high percentage of ARSs from all of the other three species. Conversely *S. cerevisiae* and *S. bayanus* were only able to initiate replication at roughly half of *L. kluyveri* ARSs. This result suggests that *L. kluyveri* ORC has a more promiscuous recognition mechanism for binding. *K. lactis* ARSs showed little ability to be replicated by the other species and similarly *K. lactis* is only able to initiate replication at a very small percentage of the ARSs from the other three species. Analysis of the ARSs that replicate in both *S. cerevisiae* and *K. lactis* showed that they all contain a match to the ACS of the host species. These serendipitous ACSs are what allowed these sequences to replicate in the heterologous species. These results suggest that the ORC from each species recognizes only its own cognate ACS. Because of the stringency of the *K.lactis* ACS, it is less likely to occur

serendipitously in the ARSs of other yeast species. The *K. lactis* ORC has diverged significantly from the other three species to recognize specifically the 50bp motif (Liachko et al. 2010).

The difference between the ACS recognized by *S. cerevisiae* and the ACS recognized by *K. lactis* provides an interesting avenue for study. The *K. lactis* ACS's larger size may be an indication that it is actually two ACSs in an inverted repeat. There is evidence that some ARSs in *S. cerevisiae* show a similar dimeric structure and they can be made more efficient by altering the spacing of the repeats suggesting that there are two ACSs at some ARSs and that both have the ability to recruit ORC complexes. These redundant ACSs may increase the recruitment potentials for ORC (Bolon and Bielinsky 2006).

Figure 1.5.

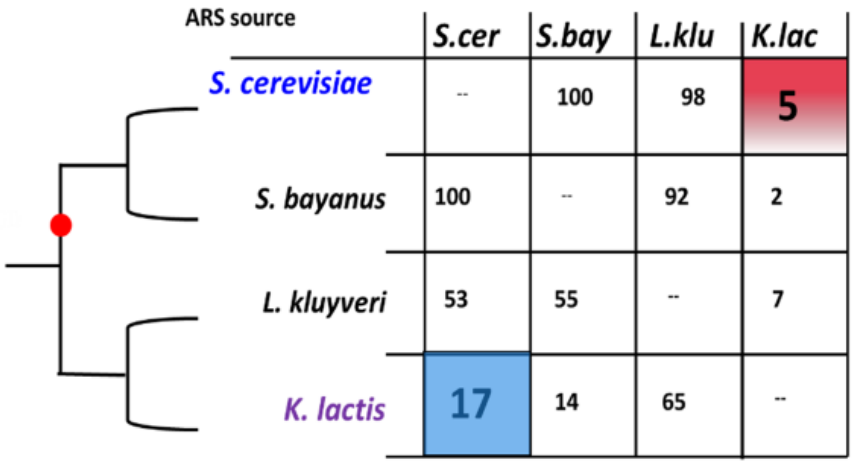

Fig1.5. The cross species function of ARSs. *S. cerevisiae* and *S. bayanus* ARSs share a common ACS that is very similar to that of *L. kluyveri* but significantly different from *K. lactis*. *K. lactis* appears to have the most stringent ACS (Liachko et al, Plos Genetics, 2010).

The *S. cerevisiae* and *K. lactis* species show a great difference in the sequences their ORC complexes recognize as a replication origin as well as a great specificity for only their own origin sequences. Modeling of the ORC and CDC6 complex shows its direct interaction with the ACS

with multiple subunits in *S. cerevisiae* suggesting that the sequence specificity is determined by one or more of the subunits of this complex. ORC1 through ORC5 all have a WHD DNA binding motif (Chen et al. 2008). ORC1, ORC4 and ORC5 show the closest proximity to the ACS and B elements suggesting that these subunits likely play a role in the sequence specificity of the binding of the ORC complex. Cross-linking studies also indicate that ORC1 and ORC4 are most closely associated with the ACS (Lee and Bell, 1997).

Examinations of *S. pombe* further narrow the possibilities of which subunits may direct the binding of the ORC complex to a given ACS of a species. The ORC4 subunit of *S. pombe* has been shown to localize to *S. pombe* origins independently, in an ATP independent manner (Lee et al. 2001). However, *S. pombe* ORC4 is significantly structurally different from the ORC4 subunits of other species. *S. pombe* ORC4 contains an AT hook domain at its very N-terminal region and this domain is absent in the ORC4 of other well studied yeast species and is in fact absent in metazoans including humans.

Following the whole genome duplication event there were two copies of each ORC subunit. Most of these duplicate gene copies were subsequently lost, however paralogs of ORC1 and ORC4 were maintained. The paralog of ORC1, SIR3, plays a role in interacting with histones to establish silent chromatin states, and is dependent upon RAP1 for its recruitment for silencing sites. Rif2, the paralog of ORC4, plays a role in the silencing of telomeres as well as controlling telomere length (Wotton and Shore 1997, Bryne and Wolfe 2005). The ORC complex plays a role in silencing at the telomeres, indicating that the presence of both ORC4 and RIF2 is a separation of function of the ancestral state rather than a novel phenotype of these paralogs. Given that these subunits were maintained unlike the paralogs of other subunits, this maintenance would indicate

that these subunits are more important to cellular functions, making them likely candidates for determining the sequence specificity of the ORC complex.

While no crystal structures of the *S. cerevisiae* ORC complex or individual subunits exist, studies have been made of the archeal homologs of the ORC subunits. The archaeal pre-replication proteins share a good deal of structural homology with the eukaryotic ORC complex subunits, as both sets contain both an AAA+ domain and a winged helix domain. However, while eukaryotes have six conserved ORC subunits, the number of ORC genes in archaeal species vary between one to three copies although there can be as many as 14 copies (Norais et al. 2007). The crystal structural of an ORC from the archaea *A. pernix* showed that the winged helix domain occupies a large segment of the replication origin and the AAA+ domain occupies a smaller adjacent segment (Fig1.2) (Gaudier et al, 2007). The direct contacts between the WHD and the ISM of the AAA+ with the conserved DNA element suggest that these domains determine the DNA binding of Orc. Additional studies with another archaeal species, *Sulfolobus solfataricus* showed that mutations in the winged helix domain and also in the ISM of AAA+ domain of archaeal ORC1 alter the binding specificity of the complex (Dueber et al 2010).

A recent publication has determined the crystal structure of the ORC complex of *D. melanogaster*. This structure determined that the ISMs of ORC2-ORC5 and the WHDs of ORC1 and ORC3-ORC5 both form rings around a central channel for DNA. Comparisons of the drosophila ORC4 shows a good deal of structural similarity to known archeal ORC structures (Bleichert et al. 2015). This structure seems to fit in very well with our analysis of the *S. cerevisiae* ORC complex.

While it is clear that DNA replication is a highly regulated and highly conserved system among eukaryotes, it is also clear that the manner in which the specificity of ORC binding is

determined has greatly diverged. This disparity raises several questions. One question is how the ORC complex's binding specificity is conferred. Is it conferred by the ORC subunits, and if so which one or ones? A second question is, how would altering the specificity of the ORC complex affect genome wide DNA replication initiation and other regulatory mechanisms such as ORC mediated silencing?

To investigate the manner in which the binding specificity of the ORC complex is conferred, I constructed chimeric ORC subunits with the AAA+ and WHD domain of the *K. lactis* ORC4 and ORC5 subunits and the N-terminal domain of the *S. cerevisiae* ORC4 and ORC5 subunits, respectively. The effect of the chimeric subunits on the initiation of DNA replication and ORC mediated silencing was investigated using a three-pronged approach. To facilitate studying the effect of the chimeric ORC complexes on silencing with a visual assay, we replaced the HMR locus with the *ADE2* gene. To investigate where the chimeric ORC complex binds genome-wide, a ChIP-Seq assay was carried out and the locations and levels of ORC binding between a WT *S. cerevisiae* strain and a chimeric strain were compared. As a direct functional assay of replication initiation at *K. lactis* ARSs, I took advantage of the library of *K. lactis* ARSs in our collection and used them in the plasmid stability assay.

Works Cited

Blow, J. J., and A. Dutta. "Preventing Re-Replication of Chromosomal DNA." Nature reviews.Molecular cell biology 6.6 (2005): 476-86.

Blow, J. J. "Preventing Re-Replication of DNA in a Single Cell Cycle: Evidence for a Replication Licensing Factor." The Journal of cell biology 122.5 (1993): 993-1002

Takashi, S. and Diffley, J "Stepwise assembly of initiation proteins at budding yeast replication origins *in vitro*" Proc Natl Acad Sci (2002) 97(6): 14115-14120

Rowles, A., Blow, J.J. "Chromatin proteins involved in the initiation of DNA replication" Current Opinions in Genetics and Development (1997) 7(2):152-7

Bryant, J.A. et al. "Origins and complexes: the initiation of DNA replication" Journal of experimental Botany (2001) 52(355): 193-202

Brewer, B.J., and Fangman, W.L. "The localization of replication origins on ARS plasmids in *S. cerevisiae*" Cell (1987) 6;51(3):463-71

Feng, H., and E. T. Kipreos. "Preventing DNA Re-Replication--Divergent Safeguards in Yeast and Metazoa." Cell cycle (Georgetown, Tex.) 2.5 (2003): 431-4

Vaughn, J. P., P. A. Dijkwel, and J. L. Hamlin. "Replication Initiates in a Broad Zone in the Amplified CHO Dihydrofolate Reductase Domain." Cell 61.6 (1990): 1075-87

Bielinsky, A. K., Gerbi, S. A. "Where it all starts: eukaryotic origins of DNA replication" Journal of Cell Science (2001) 114(Pt 4):643-51

Remus, D. et al. "DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding" Embo J (2004) 25;23(4):897-907

Liachko, I., et al. "A Comprehensive Genome-Wide Map of Autonomously Replicating Sequences in a Naive Genome." PLoS genetics 6.5 (2010): e1000946

Eaton, M. L., et al. "Conserved Nucleosome Positioning Defines Replication Origins." Genes & development 24.8 (2010): 748-53

Dershowitz, A., et al. "Linear Derivatives of *Saccharomyces cerevisiae* Chromosome III Can Be Maintained in the Absence of Autonomously Replication Sequence Elements" Mol Cell Biol. 2007 Jul;27(13):4652-63

Mechali, M. et al. "Genetic and Epigenetic Determinants of DNA Replication Origins, Position and Activation" Current Opinion in Genetics &Development (2013) 23:124-131

Shor et al. "The origin recognition complex interacts with a subset of metabolic genes tightly linked to origins of replication" Plos Genetics (2009) Dec;5(12):e1000755

Dellino G.I., et al," Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing" Genome Res 2013 Jan;23(1):1-11

Duncker, B.P., Chesnokov, I.N., McConkey, B. J., "The origin recognition complex protein family" Genome Biology (2009); 10(3):214

Makise et al. "Kinetics of ATP binding to the origin recognition complex of Saccharomyces cerevisiae" J Biol Chem. (2003) Nov 21;278(47):46440-5

Kuo, A.J. et al. "The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome" Nature 2012 Mar 7;484(7392):115-9

Diffley, J. F., et al. "Stepwise Assembly of Initiation Complexes at Budding Yeast Replication Origins during the Cell Cycle." Journal of cell science.Supplement 19 (1995): 67-72.

Randell, J.C., et al. "Mec1 is one of multiple kinases that prime the Mcm2-7 helicase for phosphorylation by Cdc7" Mol Cell 2010 Nov 12;40(3):353-63

Breier, A., S. Chatterji, and N. Cozzarelli. "Prediction of Saccharomyces Cerevisiae Replication Origins." Genome biology 5.4 (2004): r22

Muller, P., et al. "The Conserved Bromo-Adjacent Homology Domain of Yeast Orc1 Functions in the Selection of DNA Replication Origins within Chromatin." Genes & development 24.13 (2010): 1418-33

Noguchi, K., et al. "The BAH Domain Facilitates the Ability of Human Orc1 Protein to Activate Replication Origins in Vivo." The EMBO journal 25.22 (2006): 5372-82

Muller, C.A. and Nieduszynski, C.A. "Conservation of Replication Timing Reveals Global and Local Regulation of Replication Origin Activity" Genome Research (2012) Oct;22(10):1953-62

Rhind, N., and Gilbert, D.M. (2013). "DNA Replication Timing". Cold Spring Harbor perspectives in medicine *3*, 1-26

Nieduszynski, C. A., et al. "OriDB: A DNA Replication Origin Database." Nucleic acids research 35.Database issue (2007): D40-6

Hamlin, J. L., et al. "A Revisionist Replicon Model for Higher Eukaryotic Genomes." Journal of cellular biochemistry 105.2 (2008): 321-9

Lee, J. K., et al. "The Schizosaccharomyces Pombe Origin Recognition Complex Interacts with Multiple AT-Rich Regions of the Replication Origin DNA by Means of the AT-Hook Domains of the spOrc4 Protein." Proceedings of the National Academy of Sciences of the United States of America 98.24 (2001): 13589-94

Wu, P. Y., and P. Nurse. "Establishing the Program of Origin Firing during S Phase in Fission Yeast." Cell 136.5 (2009): 852-64

Kong, D., T. R. Coleman, and M. L. DePamphilis. "Xenopus Origin Recognition Complex (ORC) Initiates DNA Replication Preferentially at Sequences Targeted by Schizosaccharomyces Pombe ORC." The EMBO journal 22.13 (2003): 3441-50

Vaughn, J. P., P. A. Dijkwel, and J. L. Hamlin. "Replication Initiates in a Broad Zone in the Amplified CHO Dihydrofolate Reductase Domain." Cell 61.6 (1990): 1075-87

Cvetic, C. and Walter, J.C. "Eukaryotic origins of DNA replication:
could you please be more specific?" Semin Cell Dev Biol 2005 Jun;16(3):343-53

Takeda, D.Y. "Recruitment of ORC or CDC6 to DNA is sufficient to create an artificial origin of replication in mammalian cells" Genes Dev 2005 Dec 1;19(23):2827-36

Zhang and Tower "Sequence requirements for function of the Drosophila chorion gene locus ACE3 replicator and ori-beta origin elements" Development (2004) May;131 (9):2089-99

Moriyama, K., et al. "Epstein-Barr nuclear antigen 1 (EBNA1)-dependent recruitment of origin recognition complex (Orc) on oriP of Epstein-Barr virus with purified proteins: stimulation by Cdc6 through its direct interaction with EBNA1" J Biol Chem 2012 Jul 6;287(28):23977-94

Dhar, S.K., et al." Replication from oriP of Epstein-Barr virus requires human ORC and is inhibited by geminin" Cell (2001) Aug 10;106(3):287-96.

Celniker, S. E., et al. "Deletion Mutations Affecting Autonomously Replicating Sequence ARS1 of Saccharomyces Cerevisiae." Molecular and cellular biology 4.11 (1984): 2455-66

Bolon, Y.T. and Bielinsky, A.K. "The Spatial Arrangement of ORC Binding Modules Determines the Functionality of Replication Origins in Budding Yeast" Nucleic Acids Research (2006);34(18):5069-80

Chen, Z., et al. "The Architecture of the DNA Replication Origin Recognition Complex in Saccharomyces Cerevisiae." Proceedings of the National Academy of Sciences of the United States of America 105.30 (2008): 10326-31

Lee, D.G. and Bell, S.P. "Architecture of the Yeast Origin Recognition Complex Bound to Origins of DNA Replication" Molecular Cell Biology (1997) Dec17(12):7159-68

Wotton, D. and Shore, D. "A Novel Rap1p-interacting Factor, Rif2p, Cooperates with Rif1p to Regulate Telomere Length in Saccharomyces Cerevisiae" Genes & Development (1997) Mar15;11(6):748-60

Bryne, K.P. and Wolfe, K.H. "The Yeast Gene Order Browser: Combining Curated Homology and Syntenic Context Reveals Gene Fate in Polyploid Species" Genome Research 15(10):1456-61

Norais, C. et al. "Genetic and Physical Mapping of DNA Replication Origins in Haloferax Volcanii" PLoS Genetics (2007) May 18;3(5):e77

Gaudier, M. et al. "Structural Basis of DNA Replication Origin Recognition by an ORC Protein" Science (2007) Vol 317: 1213-6

Dueber, E.C. et al. “Molecular Determinants of Origin Discrimination by Orc1 Initiators in Archaea” Nucleic Acids Research (2011) 39(9) 3621-31

Bleichert, F. et al “Crystal structure of the eukaryotic origin recognition complex” Nature (2015) 10.1038/nature 14239

Chapter 2: Materials and Methods:

Strains and Plasmids

All *S. cerevisiae* yeast strains were from a W303 background with any modifications noted. The MWD 174 strain was constructed by mating the ILY 171 and ILY 172 strains. All chimeric strains were constructed by integration into the MWD 174 strain using the primers given. All yeast growth was performed at 30C. *K. lactis* ARS plasmids were taken from (Liachko et al 2010).

Integration of genes encoding the chimeric ORC subunits at their endogenous loci

Each *ORC4* or *ORC5* chimera was amplified by PCR using primers with 50-55 base tails homologous to the corresponding endogenous region of the *S. cerevisiae* genome. Each chimeric PCR product was transformed into strain MWD 174 using a lithium acetate protocol. 10ml of culture was grown to mid log phase overnight and pelleted down, washed and resuspended in 100ul of "one-step" transformation buffer (Lithium Acetate 0.2M, PEG 4000 40%, DTT 100mM, Carrier DNA 0.5mg/mL) with 20ul of the PCR product with primer tails added and mixed. Transformation mixtures were incubated at 42C for 1 hour. Transformed cells were then spun down, washed and resuspended in 3ml of liquid YPD and allowed to grow overnight. Overnight cultures were then spun down and resuspended in 300ul and spread onto plates containing 500 micrograms/ml g418 to select for the KANMX marker and allowed to grow at 30C.

ADE2::HMR Silencing Assay

To assay the loss of silencing at the *HMR* locus in the chimeric strains, colors of colonies of cells of the chimeric strains were compared directly to their WT counterparts (Strain

MWD174). Strains were streaked onto YPD and grown at 30C for 3 days and then placed at 4C for 3 days. Colonies were then visually assayed for color that measures the level of expression or silencing of Ade2. Construction of the isogenic strains is described in Chapter 3.

Dissection of Chimeric Diploid Yeast

Cells were sporulated on SPO media for 5-10 days until approximately 70-90% of cells had formed asci. Cells were resuspended in 90 µl of water with added 30 µl of 0.5mg/ml zymolyase. Cells were digested in a time course from 4 to 10 minutes and 30 microliter samples were added to 1ml of distilled water and placed on ice to stop the zymolase reaction. Twenty five µl were added to a YPD plate and dissection of tetrads was performed using a dissection scope.

Assay of *K. lactis* ARS activity on plasmids in *S. cerevisiae*

A subset of the plasmid library selected for their differential ability to replicate in *S. cerevisiae* was transformed into both the WT and chimeric strains. All plasmids share the same primary sequence derived from the pUC19 vector bearing *LEU2*, *URA3*, and *CEN5*, with various *K. lactis* ACSs cloned in additionally. Strains were grown to mid log phase overnight in YPD, spun down and washed. For each plasmid transformation mix, approximately 650ng of plasmid DNA was added to 4X 10^8 cells. Cells were transformed using the "one-step" buffer method as described above. Cells were rinsed and resuspended in 120 µl of water and plated onto CSM-leu-ura plates and allowed to grow for five to seven days before scoring for growth.

Assay of maintenance of *K. lactis* ARS plasmids in chimeric strains

S. cerevisiae transformed with plasmids containing *K. lactis* ARSs in the manner described above were selected from CSM-leu-ura plates and grown to saturation in 3 ml of CSM-leu-ura medium. 4 µl of this culture was diluted into nonselective YPD media and incubated at 30C overnight to allow for ~10 rounds of cell division. This culture was then diluted in water to 10^{-4} and 10^{-5} dilutions. 70 µl of these dilutions were then plated on YPD plates and incubated at 30C for 72 hours. These YPD plates were then scored and replica plated onto CSM-leu-ura plates and YPD plates. These plates were then incubated at 30C for 72 hours and scored for growth. Averages and standard deviations of maintenance were then calculated.

ChIP of HA Tagged ORC6

(Performed in collaboration with Dr. Yuangliang Zhai.)

ChIP-Seq was performed on both the heterozygous ORC4 chimeric strain and the corresponding WT strain. Cells were grown in one liter of YPD overnight at 30C to early log phase and then synchronized in G2/M phase in medium with 30 microgram/ml benomyl for two hours. Cells were then processed for ChIP using protocols modified from (Lefrancois, 2010) (Buffer recipes can be found in this publication). The 1 liter of cells were treated with 28ml of 37% formaldehyde for 15 minutes to crosslink the ORC complex to DNA. Cells were then treated with 54ml of 2.5 M glycine to quench the cross-linking reaction. Cells were collected via filtration and washed twice with Milli-Q water. Cells were then spun down, the supernatant removed and the weight measured. For cell lysis, 800 µl of cell lysis buffer was added to each 2ml tube and 1.2 g of zirconium beads were added. Cells were subjected to 10 cycles of bead

beating for one minute with two-minute pauses between each cycle using a FastPrep machine. Cells were visually assessed for 80-90% cell breakage before continuing.

Lysates were centrifuged for 3 minutes at 1500 rpm and the cell pellet resuspended and transferred to Eppendorf tubes for sonication. Cells were sonicated in 10-second pulses with settings at 50% at level 4 followed by 2-minute cooling in an ice bath for 36 to 40 cycles. Supernatants of the sonicated lysates were collected in an Eppendorf tube from two successive spins; the first for 5 minutes at 3,000rpm followed by a second for 10 minutes at 14,000rpm. The size of the sonicated DNA was analyzed by gel electrophoresis and confirmed to be of the desirable length of approximately 200-250 bp.

For immunoprecipitation, 400µl of magnetic beads were washed twice with lysis buffer and then resuspended in 600µl of lysis buffer with BSA. Twenty µl of anti-HA antibody were added and put on a rotator at 4C overnight. Beads were then centrifuged for 2 minutes at 1,000rpm, and washed twice with PBS with added BSA. Beads were then alliquoted into ten separate tubes and added to the sonicated chromatin, and rotated at 4C overnight. A magnetic grid was then used to remove the beads from the supernatant, and 20 µl of the supernatant was saved as the supernatant fraction.

The beads were thoroughly washed twice with 1ml of lysis buffer, twice with 1ml of lysis buffer with 360 mM NaCl, twice with 1ml of wash buffer and once with 1ml of TE. DNA-protein crosslinks were reversed by overnight incubation in 1% SDS in TE at 65C. Samples were treated with 40 microgram/ml proteinase K in TE for 2-4 hours. DNA was then extracted with a 25:24:1 ratio of phenol:chloroform:isoamyl alcohol three times, keeping the aqueous phase. DNA was then precipitated by adding 50 µl of 5M LiCl and 1ml of 100% ethanol to the last extracted aqueous phase. Samples were centrifuged for 20 minutes at top speed and washed with

1ml of 70% ethanol and resuspended in TE. The samples were then treated with RNase for 30 minutes at 37C. DNA was finally extracted by eluting from a standard PCR purification column.

High throughput DNA sequencing

Sequencing of the ChIP samples was carried out by the facilities at Peking University.

Sequence Analysis

(Performed in collaboration with Prof. Uri Keich)

Two separate protocols were used to analyze the raw data, one by myself and one by Prof. Uri Keich. Both protocols gave similar results. Figures will use the analysis done by Prof. Uri Keich.

Protocol performed by me: Samples were first filtered by quality score to a minimum quality score of 20 using the FASTX-Toolkit. Sequences were aligned using Bowtie2 mapped as paired reads using a fragment size of 100-500 bp and the output was converted into a BAM file. Sequences with more than 3 mismatches were excluded using custom perl scripts, and samtools was used to filter out any unmapped reads and PCR replicates.

Protocol performed by Prof Keich: The data from the G2/M phase experiment consisted of 101 bp paired end reads. All of the reads were aligned using Bowtie (BT) to the *S. cerevisiae* genome (February 2011 version), using mate pair information. The mate paired reads were similarly summarized using a Python script that took the mate pair information into account when computing average coverage in the same 25bp windows across the genome. We had a baseline 'input' control, which we used to find the relative enrichment in coverage for each of the 25bp windows. The normalized enrichment ratio was computed in two steps. First, for each window, i, the coverage ratio relative to the baseline was computed as the average coverage in the

considered control, a(i), divided by b(i), the average input (baseline) coverage for that window. Since the input control left some windows with a baseline count of b(i)=0, we adjusted this ratio by dividing instead of b(i) by the maximum of b(i) and 4% of the median coverage of the input control for that chromosome:

r(i) = a(i) / max(b(i), 0.04 * median chromosomal coverage).

At the second step these ratios were normalized for the overall difference in coverage between the experiment and input control. This was done by defining the enrichment for window i as

e(i) = r(i) / (median of r along that chromosome).

This definition guarantees that the median enrichment level across each chromosome is 1.

List of primers used for construction of chimeric strains

	Primers used to integrate *K. lactis* ORC genes into PUG6
ORC4 Forward	GTGGAAAATCTATCCCTACCATTAT
ORC4 Reverse	TATATATATAGTCGACAGCCATTAATAAAACCTAGCTGAA
ORC5 Forward	GGGGGCCCCCGACGTCTCTGAACTTAGCATGGTTGAATTC
ORC5 Reverse	TATATATATAGTCGACTATACCCTTCTGTTCCAGCACAT
	Primers to construct chimeric ORC subunits
ORC4 Reverse	AACCATTATCACTCATATTATTTTACTGTTATTCTGCTTGACCAC GAATGTACTGCTAATACTATACAATTTCTTTTAGGGAGACCGGC AGATCCGCGGCC
ORC4 Forward 207	TTAGATTCGACCACGAAGACAAGAAATGAAGATAGTGGTGAG GTTGACAGAGAGAGTATACACGTAACGCTAGTGTTCATATTTG ATG
ORC4 Forward 281	AAAGAGTAGATTTTCTCAAAGAGTGATTTATATGCCGCAAATA CAGAATCTACAATTCAAGCATGTATTCAAAGAGC
ORC4 Forward 167	ACTCAATTGGAACAGCAGTTGCAGAAAATTCATGGCAGTGAAG AAAAAATTGACGATACTGGTTTATCTGAAGGTACTTTGACAG
ORC4 Forward 200	TTCTTTTACTCTTAGATTCGACCACGAAGACAAGAAATGAAGA TAGTGGTCACTCACAGAAAGAAAAATCAGAGCACG
ORC4 Forward 212	CAAGAAATGAAGATAGTGGTGAGGTTGACAGAGAGAGTATAA CAAAGATAACAGTTTTCATATTTGATGAAATTGACAAATTTGCT GG
ORC5 Reverse	GTTCAGTCTCCTCATTCATTGTTTTTCCATACAGTAGGGCGCTTT TAATAGTTTAGGGAGACCGGCAGATCCGCGGCC
ORC5 Forward 127	ACGTTGCACAATATTTTTGTCCAATATGAATCTTTGCAAGAAAA GACTTGCCTTTATGTTATACTCGATGGGCTGGATG
ORC5 Forward 158	TAGACGCCGCACTGTTTAACAAATATATCAAACTAAATGAATT ACTTCCAAAAGCGATGATTCAACTTCGGTTAATAATCTCG
ORC5 Forward 212	CGAAGTTTCTACTATATTAGTGATGTCTAGATGTGGCGAACTCA TGGAAGATCCGCTGTTATTGGAAAAAATGG
ORC5 Forward 140	GCAAGAAAAGACTTGCTTGTTCTTGATATTGGATGGTTTCGATA GTTTACAAGATGTGAACTTGGAAACACCTACGC

Primers used to sequence and confirm integrations

ORC4 F External	AGGTGCTCAGTTTGGTAAGCCTCA
ORC4 F Internal	GTACACCCTTTAGAGGAGTCACAGTAA
ORC4 R Internal	CGCTGCAAGAAATGCGATTGACGA
ORC4 R External	AGTTAGTCTTCGCATCTCGCTGGA
ORC5 F External	ATAACGCGAGGTTCAATGGCCTCT
ORC5 F Internal	GAGAAACCTGGTGTTTGTGATACAAT
ORC5 R Internal	AAGCAAACCAGCGAATAACCCGAG
ORC5 R External	GCTCCTGGGACAAATGGAAGGTTT
ORC4 Sequence F	GAGTCATTCAGTAATTCTCGTGGG
ORC4 Sequence R	GTTCTTGTCATTACGATCGAACA
ORC5 Sequence F	TTGAAAACCCTATATCCAAACATTC
ORC5 Sequence R	TCATTTTGATCAGATGATCCCTTA

Sequences of *K. lactis* ARSs used in plasmid analyses.

Class 1:

Plasmid X: 098-67-7

AGAGGGGAATGAAATCCTTAAATCGCATATATGCTCTCCCGTCATACATTCCACCTT
CATTCTTTCATAATTCTTTGATTGGATGAACTAATTACGACATTGGCATTCTCATTAT
TCATGTACCACTTACAGCAGTCATAGCCAGTAATTGCATATACTGCACATTATTTTAT
AAAACCCTTTAACATCATGTAAATACCAATTTCCCTATATATATATATAATTGCACTT
AGCGCATAATACTTATCGTTTTGATATTCTCTTACTAGTTATCCCTACTTACCATATTT
AGCTCGGGATTGAAATTCATGCCAATGACAACGAATGTTTTCAGTAAGGGTCGATAT
GCTTTAGTAACTTATTCTGACCATCGTTGTCTTCTTAACATTGTGACATTGAAAATTA
CTTTTGAATGGAAAAAGCATAATTATGTACAAAATTCATCAATTGAAATATACTGAA
GATAAATTTCGAAGGACTTTTGCCAAGAGTTAAGAGACTATATCTACGACTTATCAC
TTCCGTACGACCCATTATAAACCATCCAATATTCCATGACATCTCCAGATTTGAGAA
AAGTATGTTTACCCCTGTTTTCCAGGAAAAATCTTATCAGAATTTAAATTACTAACTA
TAATATGATTCGTTGCAAACAGTACATGGACAAGAAACTCATTCTACAGCTTAACGG
CAATAGGAAAATTATTGGAATATTACGAGGATATGATGCATTTTTGAACTTAGTACT
GGATAATCCAATACAGATTTTCGAAGAAGGAGAGGTCCAAATGGGACCCCAGACTG
TTGTTAGAGGTAACTCCATTGTATCGATAGAACCTCTTGACAGTTTGTAAGAGCAGA
AGGGAGACAAATACAACTCAATGTGTCGGAACACTTTACTTAGTGACATTTACGTAC
AATGAATTATGCTAGTAGAAAGGAAGGATAGATATGATAGAGCGCACACTGGTTAC
AACATCATTTGTCGAATCTGATTTGCATTTCGACACCAGCTTTTTTCAATCCACATTT
CTCATAGAATCCAACATTACTTTCATCACAATCAAGAATAACTTTATAGCATCCATA
CTCGTTAGCTAATTTGTATAAATACTGAATTAATATCAATCCCAATTTCTTCCCCTGT
T

Plasmid R: 098-c51-7

TGATGATACTTTTCTCAAAGCGCTCACTTCAATTACACTTTTCAAAATCGCTTGTTGT
TACAATATTAGTATTTTGTATTGAAGGTTTTTCACGAACTTAGAATTTCGGTTCTTGG
AGGGTAGTGAATGAATAGAAGGAAACAAAAAAAAAGTCACGTGAATAAGTGAAGAT
GTTGCATATTCATCACTGCATGGAAAGTCTTTACCGCTGGCGCTTCAAAGAGTTAGA
ATTGGGGGGTGTATTCATTGGAAGAGATTGTTATTATTATATATTTTGAGTTCGTATG
CCATTAAAACCGAGATTTTTATTGTTTTGTTATAAGACAGTTCTATCTTACAGGTAAC
CCAGGTTAATGCGAAGTAACATCCATTTCATCATCCAAAATATTGCATGGAATCTAA
ATGTTATTTAAAATATGGTTGATGGTTTAAATATGATTTTATAAGACGTTCCGAAGTT
CTGTATTTTTGGATGTGTAAAGTACTTTGCAAGACCAAATAGTATAATTAAAGCAAT

ATGTTAAATGGAAAATGTTTTTATATTTATATAAACTATTGAGCTAAAAGTCAATTA
GAAGCCTGAGAGGTTAAATATATCATGAGATAATGAAATCCACACACGAAGAGTAA
TTTCGTTGAGGTTTTATACTTAAACAAAGCGATACTCCATAGTGTTTAGTTTTTCGCT
TCGAACCAGTTTCGTATTTTAGACGGTTTAGAATTCAAACAAAGAACCAAACGTTTG
GTACTTCAGTTTGTCTATAGAAATGTCAAAGAGTTATGAAGTAGTTTTGGTAGTTTTA
GGATGTTGGTTTAGTGAAATGTAACGCTTTTTCTGGAGATGGGGGGTAGAGCTTTTT
TTCTCTTTTGAACCTGCAAAAACTTGGAAACACACTCTCTATCTCACAATCATCTATA
AATATAAAACTGGACTTGATGGTGAACGGGAAGATTGTTAAGAACGCTCATCCTTG
AGTGAGAAACTTCTTACCTGATACATGTACTGGTAAT

Plasmid O: 303-19-1

CAAAGATACTAGATAATGTTTGTTCTTGATGGCAAGAAATACACAATGTCAACACCC
GCGCTTTCTCTTCGATTCATTAGTTGATATCATTAACATTATATTATACATACTTTATT
TCTCTTGGAAATTAAATATTTGATGATGGTTTATCTAAATATTGTCTCACTCATTTTC
CCTTTTTTTTTAAGATAGAATATCAGTTTGAGGCAACAAATGATAAATTATGATTATG
TATATAAAACAATTTTAATTTTATTAGCATAAAAAAGGGACAAGGCCTCTGCTAAAA
TATCGATAAAAAAAGAAATAATATGAAAGAC

Plasmid V: 098-11-7

CCATGAACTTAATTTTCATACTGCACTTCTACTGTTGAATCTCTCGAAGCATTCAATA
GTACCGATAATACGATATATAATAGATTTTATATTTATATAAGGAAAATAGTACAAT
AAGCATTAAGCAAAAGATGGATACAAGGTAAATTTAGGAAATTATGGGTAAGAGTA
ATGATATCGGTGATTTCAATCCACTTTGAGTTCATCTTTAACATGTCCAATTTTACGT
TTAGTTTCCAAAGTCAAACTTACTATCAAGTTCTAAACCCAAAGTTATTTTTCATTCT
AATATCTGTAAAAAGAGCAGGTTACTGTTACCCGCAACACATACATGGTGTAAATCC
GATCAAAAGCTCAGCGGCTATTGAAAAGAGTATAAAAGTATCCAGCTTGGAGACTT
TACGATTTTCTTCTTCGAATCATAAACCGCCTTACATAGAAACGGTCTATTA

Class 2

Plasmid N: Kl Predict 12

GGCAGATTCAAGAAACCAGTGCCAGTGGTGTTGACTGTTCTTGGTGGTCTGGGTACT
TGGTACTACTACAAGAAGTACAAGGAATTCTACCCATAAGAATACTACATAATGTG
AGTACATATCATTCGTTTTAACTGTTTTTTTTCTAAATCTGTATATAGGTCTAATGTTA

AAGTCGAATTCATGAAACTAATAGTAAATATTAATATTATAACAGGAAAAATGATA
TTGAATTGGTAAAATGCATATAATATAGAGTGAGAGAAATATAATAATATATGGTA
GTATAATATATTATTGAAAAAAGATAGGTGTTCGTTACAAGTAGTGAACGTAAGCA
GGGTTGATAGTCAAAAATTCCACTATCAGAAAATTAAAACCAGTTTAATCTTGTAAA
ATGCATGGGGTATGATATATTTATTGTATGGTTGTGTGGATGATTGAGTGATTA

Plasmid L: 098-c12-7
AAGGGTTCTGTAACGCAGCATCCTTTGATTCGAGATATAATTGGTACTGTTGACGGG
GATGGCTACCTAACTATATACAAAGGTTTTTATGGCAGAGACGGAGAGGTTGCAGA
ATCTGAGTAAGCTAAAGACTATATGTATGAATATGTTTATGTCCAAGTGTGTATTAT
TGAAATTTGTTAACACGGTCAAAATAAATGTTTATTACTCTTTTGTGAATATTTAAAC
ATTCACGGTCCAAATTCGTCTTTAAAGGGCAAAACAATAAATCTTCCCTTTAACATT
CGATGGCCTATCATAGCAAGAATAACACCGGCATCGATAAGCAATGAAGGATACGA
TAGAATAAAAGTGTACCTACCTACAACATGCATACTGCGACGATTATTGGCTTAATC
AACTTTGTCTCAGT

Plasmid P: 098-99-7
ATCCAGAGTAAGACATGTGAGAGAGAGGGGGAGAAGATGGCCTGTGCTACTGCTGT
TGTAGCATTTGCGAGTAGGCGAGTCATGACTGTTGTAAGCTATGTATATGTGCTAGG
TAATTATTCCGCTGTTAATTATCTTCTTTTTAATGTCAACTTTATACCTTTTGAAACAT
AACGTAAAATTAACTGTCCAAATATCATCAATCGATTGAAACAGTTTAAGTTAAAGA
TGGATTCTACTATCTAAATTACAGTTTTATTTACGAAGT

Class 3:
Plasmid T: 098-42-7
GTTTGTTCTCTCTCTCAACACTTGTAATAATGAAAGAGAAACAAATGTCCTCGTTGTT
GTTAATTTTATTATCCTGTCATCCACATTTTATTTATTACGAAATTTCCCTAACTTATC
CTAAACTTAACTTAACTTAACTTATCCTGTCAATTAAAAAAAGAACACCTCTTTTAA
AACCATAAAAAAAGTATCAACAAAAATTCATCATTTATCTTCATGTCTTTATGATAA
TAACTTATTATTTGAATTGTTTACCCCCCTGCTTTGCTCACTAGCTGAAGGAGAAATC
CATTGAACTGGTTCAAAAGTTCAACGCATCCACTTGTATTTCTTTTGCTTCATCATAT
TTATTATCACTGATTTGCCTATCTACTGCTCTTGCATTCCCGGTCCTACACATCATGC
GCGAATTAGATTTCTTTCTGTGCTCATTATTCTTCTCAAATATTCACTTTACTTTGTTT
ACAATCTTTGTTTGGAATATGCTTGGGAATATCGAAGCTGAGTAGCAATAGCACTTC
ATGGAATTGTGCAAGCAATGAAAACCAATTGGATTTAGATTATCTTCGGTTGTTTCA
CTCAAAGTTGGAAAGTGAGCTCAATTATATAGGAACTCTTCAACATCTTGAAATATA
GTTTGAACATGTATTTGTATATGATATCACGAATATATATACATATATATTTATAATA
TAACATAGTGTATTACAACATCTGAAGCTATTAACTACGAGGTTCGCATATTATTAC
GACCTAACACCTTCATACTTTGGGTTTAAGACTTCAAAGTTAACATTTAGGTAAATA
TTAATAAACCAAAGTGAAATTTGAATTGAATAAAAATATTCCAAGTTTGCAAAACG

GAAATGAATACTCACCAGACCCTCCAATCACTTGTAACAATCAGCCTTGTAGTTAAT
GGCGTTACCAATTATCCTATACTGAGGGTTCCCACAACATAAAGGTATGCAATCGTC
ATAATTATCACGTGACAGACTTACTAGCTCAGGTTGGTCCTTCTCGTTCTTGCCGTTT
CGTGTTATTACCCTTTTAGGAAAGCTAATGAATTTCTGAATCTTCTTTCTCAGTAAGG
AAAAGACTTTAAATAATTGCGGCATCTGAGTATTAATTTTGCAGTTGAAACCACTGT
TCACACTGCTGGTACCTCTTTCTAATAACTTCAAGTGAGGTTTTCTTACCGCTTTAGA
GGTGTAATTGAATTGGTACAAGTGAGGCATCTCCTTTTGTTGTAGTATTAACTTTTCG
AGACACGTTGCGACGTGTGGGATAAAATATCAAGTTGCTGAAGTTCCGTTCAAGTCA
TACTTCTTCTTTGCGGTTACATTC

Plasmid Z: 098-68-7
GTTGGCTGGTACATAAGTCCAGGAAGTCTTATACTCGGAATGTCAAAGATGTGGTAC
CTTGGCACAAAGTTGCCTCAAAATGTATCTCAAGAGAACCTTCCATGTCAATAGCCT
TCGGATAGCATGCATGTACTGTGTTTGATGCATATATATCTGTGAGAGACACGCTTT
TTTCATTTTTCATACAGCAGTTCGCGTTTCTGGGCCACGTGATTACGTCACGTGATTA
AGTCTTCACTGGACTGAAATCATCTAATGAATTGATGAAGTGACTGTGGTAGTTCAG
TAACTGGTCATTCTCTATTATAACTAGAATACATCAACACAATGTGTGAGTTTTTCTG
TTTCAGTATGAAGTCATCATTCATTCGTAAAACTAAAGATTAAAGTCCAAAGTCACA
CTGTAAATGATAAAGTAGAAAAAGG

Plasmid Q: 098-77-7
ATAAGGCCCAATGATATTTATTTTGGTTATCCATTCAGGCAATGAAGGATTTGCCGT
ACTTTCTTTTCACAATTAAAGCTTAGAAAGTTTTGGTTTATTTACAAGTTTTTACTCA
AGATACTGATTTCAATGTCAGAATTTCATTGATATTGTTTAACTTTTAAAGTTAGTAG
AAAATTTATACACTAATTCCTCATATTAAATATGGCGAATTGTGGTGCTTTTTACTTA
ACTGAATAGAAAAGATACGAGACACTTTTGCTAGTTTGCTTGAAATATGTTCTAATG
ATTTGTTATTGTTATTAATGTCGTGTGCCATAGGTCAATTAATAATATGCCAGCAATG
TTGAAACGGTTCATAAACTAT

Plasmid C: 098-c102-7
AAGGCGTCTAGTTATCCTTTTATATTACTGCTTTTTCTTTCATTTGTAACGGTTTTATG
ACATTTTGTGATTAAAAGTGATGTGCCAAGCAACAATTAATAAGAAGCATGAATTTG
CATATTATAAAAACACATTATCGTTACATAGCGCATTGTAGAACTAAGGAACATATA
TATATATATATATATATATATATATGTATATTAACTACTTTAAGTGGTATACATTTGTAT
ATTCTTGTAATTAGATTCGAAAAAACCCAAGGAA

Plasmid D: 098-114-7
GAAGAACAAAGTACATCCAACGTTATTGTCAGATTTAATAGTAGAAAAGATGTGTG
TCAGATACACATGCAGCTGTTATTGCGACTTATAACAACATGGCACCAACAACAGCT
AAACCGGCTCCGACAATGTTACTGTCCAAAATGCCAGCACCGTTTTCAGTGTAGGGG

ATTGGGGTAGTTTCATTGGTTGGGGCAAGGTGATGAGTAGTTACCGGTTTGGTTTCA
GTTGTTGGATAATCGGTAGGAATAACTGGTTTTGTTTCAGTAGTTGGACAGTCAGTA
GTAGTAATGATAGTTTTGTTTTTAGAAACCCAGGTATAAGTCTTTGGTGGCATTGGG
TAAGAAGAGGTACTGTTATGGTGTGGTGGTGGAGCTGGGTAAGTGGAAGTTTTGTTG
TAATGTGTCGGAACCTCAGAAGTAGTTTCGCAAGGAATCTCCGACGTAACGGTGTAA
GTCTGATGACTGGTTGTAACATTAGATGTAACTTTGTATTTGGTGGAGATTTCAATGC
TGCTTTTACCCTCAGTGGTGGTTAAGGTCTCAGTCTTGTTAGATGGATATTCAGACTT
CGTAGTTGGATATTCAGACTTCGTGGTCAAGGTCTCAGTCTTGTTTGTTGGATATTCA
GACTTCGTCGTTGGATATTCAGGCTTAGTGGTGGGGTACTTAGACTTCGTAGTGGAG
TACTTGGACTTGGTAGTTGAGTAATCAGACTTGGTAGTTGAGTAATCAGACTTGGTA
GTTGAGTAATCAGACTTGGTAGTTGAGTAATCAGACTTGGTGGTTGAAATTTCAGTG
GTCTTTTTGTGTTCCTTGTCATGTTTGGAACTGCTCCATTTTGGTTTACTGTAATCTTC
ATCATCATCAGAGTGCTTACCGGCACGCACAACACTCAAGGTGGTTAAAGCGACAG
CAGCAATAGTAGAAAATTTCATTATTTGGAATGGATAGATGGCCTTGTGTTTGGTTA
TTCGTAGTCACGAGAAGCTGTGCCACGAAAGTCCATTAGAATATCTTATAACTGTCA
TTAATGTAATGTAACCGAAAGTGAAGGGATTTATGCCCTAAAAATGTGTCAATTTAT
ATAAAATCAAAAAACAGCAATAATAGAGGTACTGCATTCCCTGCAATTGTACAAAA
AGTTATGCAATGTCATCACTTAAGAAACGGATTTCTTGTGTGTATTGCAATCAAATC
TGTTACGGCTTTTTAAGGTTTTCCTCTGACGCTGACACCGGTTTTCAATTTTTTTTTGT
TTTGGCAAATTTAAGAAACTGACACGCGACAGCAAAAAGAAAATAATCAACTGAGG
TCGATTAAGGGTTCAATTGCAAATAAATCTTGCACGAGACCAGATGGAAAGGAGAA
ACCGTCAGGACAACTTTTACTTCTTGTGAATTCAATTTTTTTGCTATAGACCGATTAT
TTTGAAACGTTACTATACTTTCAGTATTACTTCTAATGGAAGGCAGTCACGTCCAAA
GAAATGGAGGGTACTATGAGTCCGGTTTTGAAATCAAGTAAGCTCCATTATTAAACT
ACTTTCGTATTTCTCAGTATAAAGATTATGCGACTTCCAGTTGTTGTCATATTCATAT
AAACTGTTGTTCTTGGAGGGGAATAATTTACCTTATGAGCCTTGAAGCTTAATTTGA
GTTACAAAATATTAATGTAAAATTAAGAAAACAACCAATGATGAATTACATAAATC
ATAACGATTGTTCCAAATGTTAATGCATTGCGACGGAGACAACCAATCTGCTGGGAA
GCAGCATCTCGAAATACTTAATTGACAAAAGTGGTACC

Plasmid H: 098-c15-7
TTCAATGCTTGCAATTAAAATTGAGCTGTAAACCCAATGACTGTGTTCTTCATAGGTT
CAAATCTTATTCGCTTCAAGCATACAATAGTTTTTATGTGACCATTTTAATGATACTA
TCATGAGGATTATAAAATTTATATAATTTCATTCAACTTATACGTATTTTATCATACT
AATATAGGAATACTGTTTAATGGTAAAACTGTCACCTTTTAAGTATCCCACTCCAGT
TATTGTTTGCAGCCAGATTCGAAGCTATTTGACAGTAGAACGGCTTTCACTTGATAG
TATATAATTTTTCTTTGTGAAATTGAAACTGACATTAAGCTTACAATTTGTAAAATTG
AATTGAAATGTGGAAAGGAATCTCAACAATATGAGATATTTAAAAGGCATAAGATA
CTATCCTATCCTGACAATCCCTGGTGTACTATCAGACACTCATTGCATGCTCGTTAGA
GCTTATATTACCGCTAGTTCTGTTCTTTAAAAAGTTTCCCACCCGTACACCGCACTAA

ATTCGTTTCTTTTAGACCATTTTCCATATCAAAAAATTTTCAAAGGCAGTATAATCCA
TAAAGCTAAAATGAAGCTCATCTCACCGCATCGCATCGCATCGCATCACTTAATAAG
AGACTGCGTATCACTGCGTTTAGAATTAAGTTTACACCTTATAATAAACAGGTGAAA
GACCAACCAGAATAACAATGGACCATACCATACCACCACCAGATGAAGAGGAATTA
GTTCTTTCCAAATTAGTGTTCGGAGACAATAGTGACTTTTATAATGAATTGCAGAAT
ATCAATCTGGAGCTAAGTAGCGAAGAAGACGCTACTGAGTACGTACAGAATAGTGA
CTCCGAAGATTCAGAGAGCGACAATGAAATTGGAAAGGTTATGGATGACCAGCTCT
TTTTCGTTGATGATGGTGAACTAAATGTGGATGATGAAGATGCGATGGACATTGACG
ATGAGAACGAACAAGCAAGTGGGTCTGACTACAGCAACGAGTCTGAGGAAAATGA
ATCTGATGCATGGGTTGATTCTGACGACGACAGACTACAGATAACTGTAACAGAAA
ATAACA

Class K:
Plasmid K: 098-c51-7
TGATGATACTTTTCTCAAAGCGCTCACTTCAATTACACTTTTCAAAATCGCTTGTTGT
TACAATATTAGTATTTTGTATTGAAGGTTTTTCACGAACTTAGAATTTCGGTTCTTGG
AGGGTAGTGAATGAATAGAAGGAAACAAAAAAAAAGTCACGTGAATAAGTGAAGAT
GTTGCATATTCATCACTGCATGGAAAGTCTTTACCGCTGGCGCTTCAAAGAGTTAGA
ATTGGGGGGTGTATTCATTGGAAGAGATTGTTATTATTATATATTTTGAGTTCGTATG
CCATTAAAACCGAGATTTTTATTGTTTTGTTATAAGACAGTTCTATCTTACAGGTAAC
CCAGGTTAATGCGAAGTAACATCCATTTCATCATCCAAAATATTGCATGGAATCTAA
ATGTTATTTAAAATATGGTTGATGGTTTAAATATGATTTTATAAGACGTTCCGAAGTT
CTGTATTTTTGGATGTGTAAAGTACTTTGCAAGACCAAATAGTATAATTAAAGCAAT
ATGTTAAATGGAAAATGTTTTTATATTTATATAAACTATTGAGCTAAAAGTCAATTA
GAAGCCTGAGAGGTTAAATATATCATGAGATAATGAAATCCACACACGAAGAGTAA
TTTCGTTGAGGTTTTATACTTAAACAAAGCGATACTCCATAGTGTTTAGTTTTTCGCT
TCGAACCAGTTTCGTATTTTAGACGGTTTAGAATTCAAACAAAGAACCAAACGTTTG
GTACTTCAGTTTGTCTATAGAAATGTCAAAGAGTTATGAAGTAGTTTTGGTAGTTTTA
GGATGTTGGTTTAGTGAAATGTAACGCTTTTTCTGGAGATGGGGGGTAGAGCTTTTT
TTCTCTTTTGAACCTGCAAAAACTTGGAAACACACTCTCTATCTCACAATCATCTATA
AATATAAAACTGGACTTGATGGTGAACGGGAAGATTGTTAAGAACGCTCATCCTTG
AGTGAGAAACTTCTTACCTGATACATGTACTGGTAAT

Chapter 3: Results:

Predicted 3-D Structures of the Orc4 and Orc5 Subunits

In making chimeras of any two proteins, it is important to ensure that the fusion junction falls in a linker region of two folded domains to avoid disruption of the 3-D structure of the chimera. Three-dimensional structures inform functions and the residues involved in carrying out the functions. Such information will aid in identifying the relevant residues in comparative genomic analysis of different species because misalignments based solely on primary sequence are prevalent. A common strategy is to align amino acid sequences based on 3D domain structures so as to more accurately compare the conserved sequences within each domain. Unfortunately, there are no crystal structures for the ORC complex or individual ORC subunits for *S. cerevisiae* or *K. lactis*. However, crystal structures of a couple of archeal ORC proteins bound to their target DNA have been determined (Gaudier et al 2007). Using the I-TASSER software (Y. Zhang 2008, A. Roy et al. 2010), I built model 3D structures of Orc4 and Orc5 by multiple threading alignments based on the archeal ORC protein structures. Archeal species have a varying number of ORC proteins, so correlating the archeal ORC protein structures with the six different yeast ORC subunits turned out not to be a straightforward exercise. While all the subunits of the ORC complex are essential for function in DNA replication initiation, not all of the subunits make contact with DNA and likely to play a role in the sequence specificity of the complex. Crosslinking studies showed that Orc4 and Orc5 are most closely associated to the ACS (Lee and Bell 1997). We therefore focused our attention on these two Orc subunits as a start.

Predicted models of multiple alternate structures were generated using I-TASSER. While secondary structures generated for the Orc4 and 5 subunits for both *K. lactis* and *S. cerevisiae*

were very similar, the tertiary structures varied quite significantly between alternate structures. Predicted structures were compared to the crystal structures of archeal ORC to select the best fit. Figure 3.1.

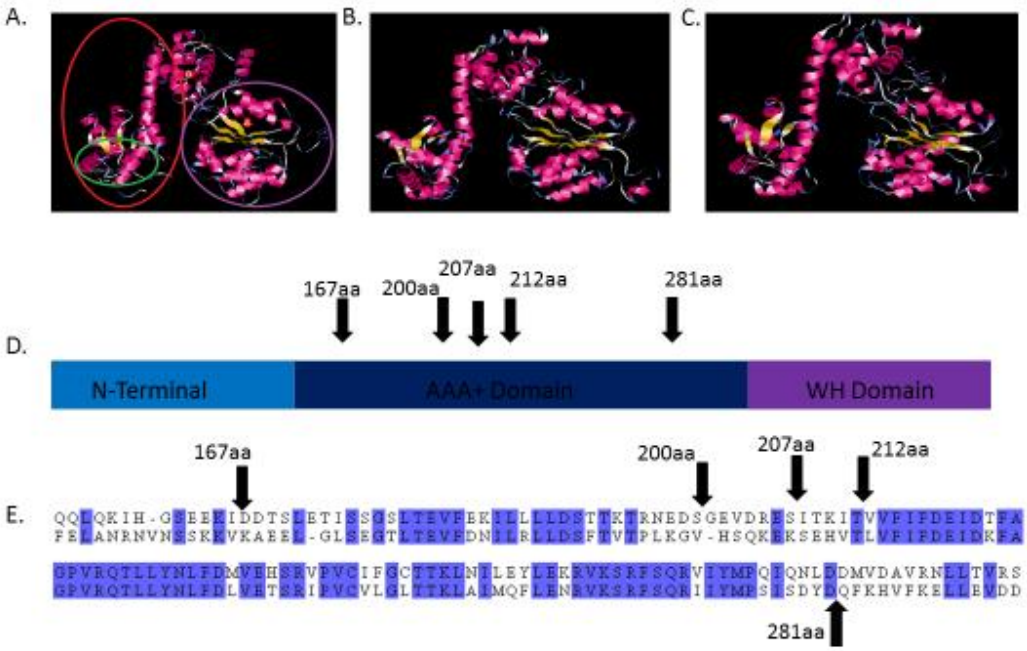

Fig3.1. A. The predicted structure of *S. cerevisiae* ORC4. The overall structure is based on the known crystal structures of archeal ORC proteins and chosen from models predicted by I-TASSER. The AAA+ domain is circled in purple, the winged helix domain is circled in red. The part of the WHD that likely makes contact with the DNA is the structure located in the bottom left of the figure. The region containing the 18 AA sequence that is not present in the *K. lactis* ORC4 is circled green. This region encompasses part of a small helix. The red arrows highlight the junction sites of two important chimeric constructs. Junction A at 207aa is the location of the junction in the chimeric used in the ChIP data set. Junction B at 281aa is the most C terminal junction that resulted in a functional chimeric ORC4 that disrupts *HMR* silencing in a diploid cell. B. The predicted structure of the *K. lactis* Orc4. The overall structural similarity to the *S. cerevisiae* ORC 4 can be clearly seen. C. The predicted structure for the chimeric Orc4. The junction in this chimeric is Junction A in figure 1A. The overall structural shape remains the same as both the *S. cerevisiae* and *K. lactis*. D. A schematic of the junctions of the various Orc4 chimeras tested. The position of the amino acid junctions given are the last amino acid from the *S. cerevisiae* Orc4. E. An alignment of a portion of the AAA+ of *S. cerevisiae* and *K. lactis* showing the junctions of the various Orc4 chimeras tested.

Figure 3.1.A shows the structure of the *S. cerevisiae* ORC4 protein. The purple circle highlights the AAA+ domain, while the red circle highlights the WH domain and the 18 amino

acid section of the protein which is present in the *S. cerevisiae* Orc4 structure but absent in the *K. lactis* Orc4 is highlighted in green. The junctions of the two chimeric ORC4 proteins most closely examined are marked with the red arrows. Arrow A marks the junction of the ORC4207 chimera while arrow B marks the junction of the ORC4281 chimera. Figure 3.1.B shows the predicted structure of the *K. lactis* ORC4 subunit while figure 3.1.C shows the predicted structure of the ORC4207 chimera. Initial junctions were selected based on the primary sequence of the protein and the predicted structures previously generated of the *S. cerevisiae* and *K. lactis* proteins. By generating models of these chimeras we could have some assurance that the methods for selecting junction sites would not lead to a disruption of the structure. As shown in the predicted structure for a chimeric Orc4207 in figure 3.1.C, the chimera does not result in a disruption of either the AAA+ or the WHD domain, nor does it appear to disrupt the overall predicted structure of the chimeric protein. The schematic in figure 3.1.D and E show the junctions of the various Orc4 chimeric strains tested. Models of the *S. cerevisiae* and *K. lactis* Orc5 were also generated in a similar manner and compared in figure 3.2 for two of the constructs.

Construction of Functional Chimeric ORC Subunits

It was unknown which subunit or subunits directed the binding of the ORC complex to the ACS. Previous work using the Orc5 of *Plasmodium falciparum*, which is very distant from yeast in evolution, showed that the C-terminal of an ORC subunit substituted for the *S. cerevisiae* subunit for function when it was joined to the N-terminal of the same ORC subunit (Gupta et al 2008). If the *S. cerevisiae* N-terminal region was not included in the construct then it did not to substitute. This report suggests that the N-terminal region is the interaction domain of

the subunits to form the ORC so the *S. cerevisiae* N-terminus must be retained in order to complex with other *S. cer* Orc subunits but the C-terminal region appeared to be exchangeable.

Figure 3.2.

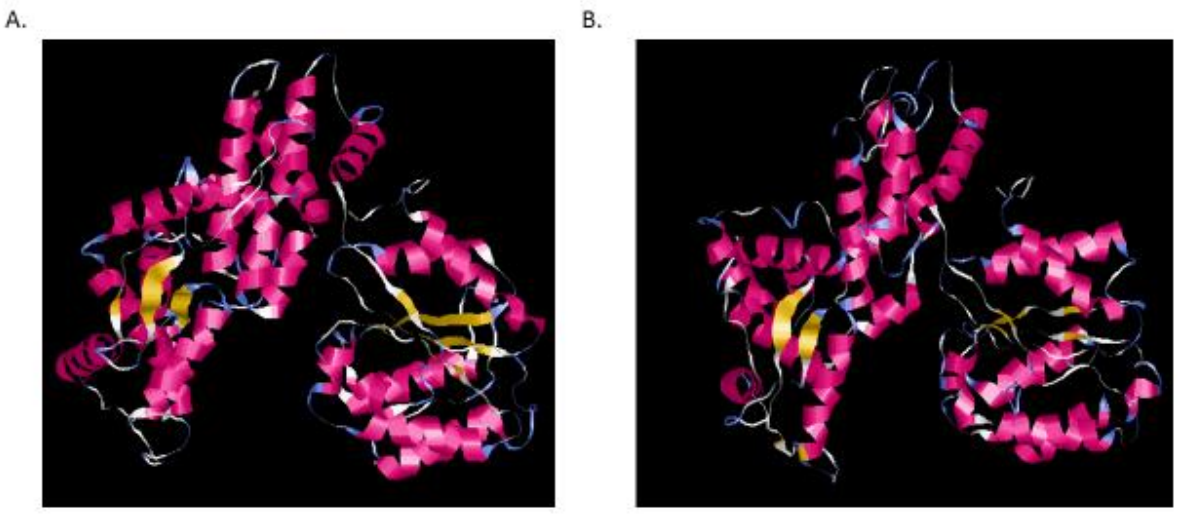

Fig3.2. The predicted structures of Orc5 in *S. cerevisiae* (left) and *K. lactis* (right). The overall structure of the subunits between the two species are very similar. Furthermore, the overall structures of the Orc5 subunit and the OrcRC4 subunit are conserved. The AAA+ domain can be seen in the lower right and the WHD in the lower left.

Each of the Orc4 and Orc5 subunit contains three functional domains; the complex forming protein-protein interaction domain located at the N-terminal half and the two DNA binding domains located at the C-terminal half, AAA+ and WHD (Gaudier et al. 2007). A functional chimeric ORC subunit would have to have an N-terminal domain that allows the incorporation of the chimeric subunit into the *S. cerevisiae* ORC complex and a C-terminal domain that binds DNA. Though the C-terminal DNA binding domain most likely determines the specificity of

binding, we cannot rule out that the N-terminal domain may also influence DNA binding or binding specificity through allosteric effects.

Previous studies of the structure of the ORC subunits showed that the C termini of Orc1-Orc5 contain an AAA+ and WHD that typically binds DNA (Duncker et al 2009). Additional studies have found that in archeal species the AAA+ and winged helix domains of their ORC proteins make direct contact with the DNA sequences specifying their replication origins (Gaudier et al 2007, Dueber et al 2011). Other studies involving chimeric constructions in *S. cerevisiae* showed that the N-terminal regions of the ORC subunits are required for subunit interactions for the complex formation (Gupta et al 2008). To investigate which sections of which ORC subunits may play a role in determining the specificity of the DNA binding of the ORC complex, chimeric constructs with varying N-C junctions were made containing the N-terminal portion of a *S. cerevisiae* subunit and the C-terminal portion of a *K. lactis* subunit. The basic design of these constructs are shown in figure 3.3.A. These constructs were integrated at the endogenous *ORC4* locus to yield a heterozygous diploid that contains both a chimeric and a wild type allele of *ORC4*. (Fig3.3.A). Viability of haploid spores containing only the chimeric allele could be easily tested by sporulation of the heterozygous diploid with direct comparison to the wild-type spores from the same ascus.

Several variations of each chimera were tested, each with a slightly different junction point in anticipation of the uncertainty of the extent of the N-terminal region essential for the formation of a stable complex as well as the uncertainty as to the stability of the chimeric proteins. Several methods were used to design the chimeric proteins. The primary sequences of the *S. cerevisiae* and *K. lactis* proteins were examined for regions of similarity where junctions would be least likely to disrupt the protein structures. Additionally since crystal structures for

neither individual ORC subunits nor the complex as a whole have been determined, predicted structures were used for modeling as shown in figure 3.1. By selecting for junctions in regions which were either unstructured in both species, or where the primary sequences were identical, multiple chimeras that are unlikely to disrupt the tertiary structure of the protein were constructed. Sporulation of the heterozygous diploids gave two live and two dead spores and none of the live spores showed G418 resistance indicating that none of the spores contained the chimeric Orc4 protein. These results showed that the chimeric subunits were unable to substitute for the wild-type subunit. (Fig3.3.B).

Figure 3.3.

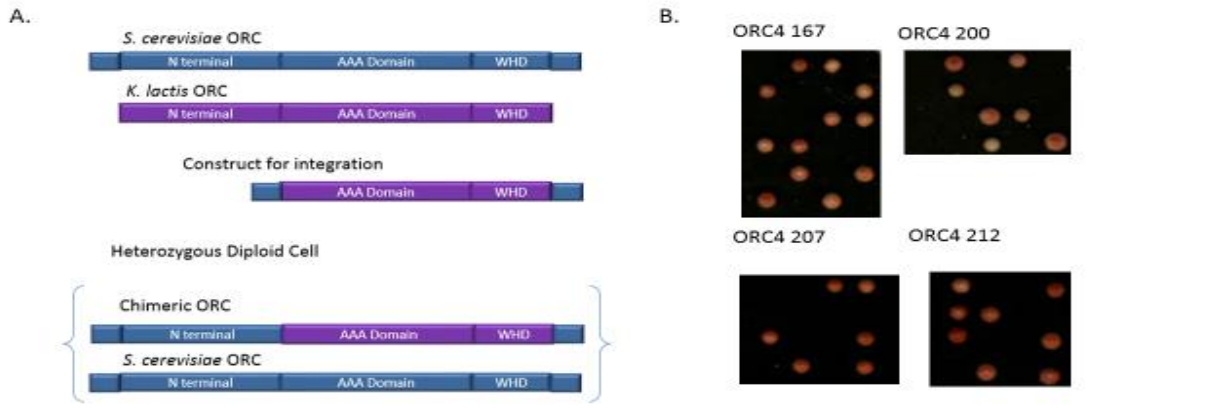

Fig3.3.A. Design of chimeric ORC constructs for genomic integration. Plasmids containing *K. lactis* ORC4 and ORC5 were engineered with a marker containing a KANMX allele confering G418 resistance 3' to the C-terminal end of the *K. lacits* ORC gene. These could then be integrated into the *S. cerevisiae* diploids through the use of PCR using primers with 50bp tails homologus to the regions of integration. The downstream integration site was distal to the endogenous ORC subunit, while the upstream integration site lied within the *S. cerevisiae* ORC gene. This DNA replacement would yield a chimeric ORC subunit that contained the N-terminal region of the *S. cerevisiae* Orc4 and the C-terminal region of the *K. lactis* Orc4. The KANMX selection marker at the 3' end of the *K. lactis* gene is not shown. Fig 3.3.B. Tetrad dissections from the tested ORC 4 chimeric strains showing 2 live and 2 dead spores in each tetrad, where none of the live spores contains the chimeric *ORC4* gene. This tetrad analysis indicates that the chimeric genes fail to substitute for the WT ORC4 gene. Numbers indicate the amino acid in *S. cerevisiae* where the chimeric junction occurs.

Using *ADE2* as a silencing reporter at the *HMRE* locus to assay for the chimeric ORC function

To assay directly whether the chimeric constructs altered the binding of the ORC complex to prevent it from binding to the *S. cerevisiae* ACS, a silencing assay at the *HMR-E* locus was used. Silencing at the *HMR* locus is dependent on two flanking elements, the E element and the I element. Each of these elements contains a *S. cerevisiae* ACS to which ORC binds and recruits the silencing machinery to the HMR region (Fox et al., 1995). To visually assay for the silencing of the HMR region an *ADE2* gene was integrated into the *MATa* locus of an *ade2* strain (Fig3.4.A). In a wild-type strain in which the silencing of the *HMR* locus is functioning properly, *ADE2* is not expressed resulting in the accumulation of a red intermediate that turns the colonies red. If a chimeric construct disrupts silencing, the *ADE2* gene will be expressed giving rise to white colonies (Sussel et al, 1993). By integrating each of the chimeric constructs into diploid *ade2* strains with *ADE2* inserted at one of the *HMR* alleles it is possible to quickly obtain a visual signal for a disruption of silencing caused by any of the chimeric constructs.

This tester strain was initially designed to assay for the disruption of ORC binding in a haploid strain expressing the chimeric ORC based on the assumption that the chimeric ORC could substitute for the *S. cerevisiae* ORC but had acquired a new specificity for binding (Gupta et al 2008). However, tetrad analysis indicated that spores inheriting the chimeric *ORC4* were inviable so the *ORC4* chimera was unable to substitute for the wild type *ORC4*. The inability of the chimeric subunits to substitute for a given subunit could be due to several different reasons. The chimeric subunit may fail to incorporate into a functional complex. Another possibility is

that the complex containing the chimeric subunit does properly interact with the *Sc*ORC but has acquired an altered binding specificity that binds infrequently throughout the genome to the extent that insufficient initiation events result in the incomplete replication of the genome. We were therefore left with the option of carrying out the silencing assay in the heterozygous diploid.

Figure 3.4.

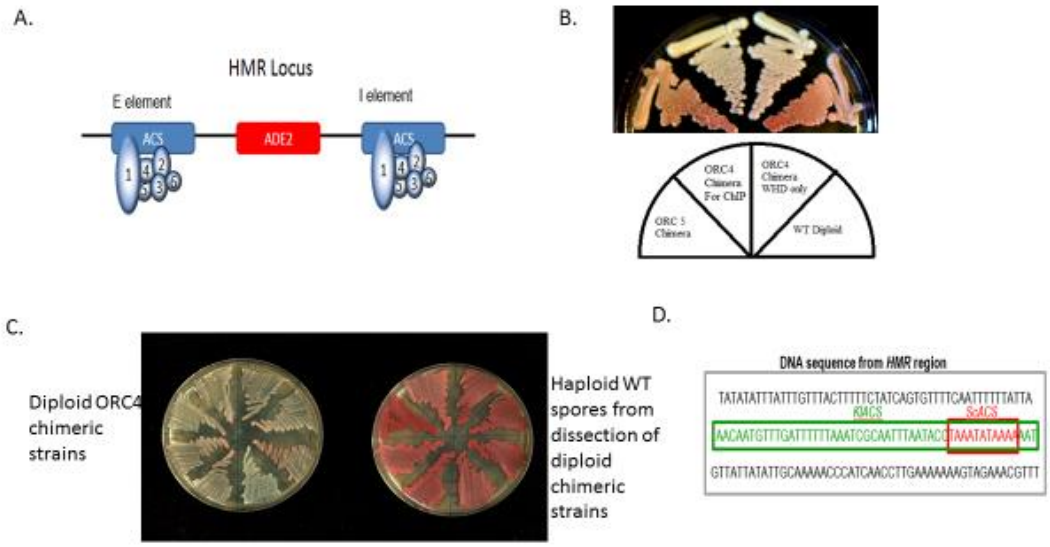

Fig3.4.A. Design of the HMR silencing assay. The *HMR MATa* is normally silenced by the binding of the ORC complex to the ARSs flanking the *MATa* gene. The *MATa* gene has been replaced by *ADE2*. Unperturbed ORC binding will silence *ADE2* resulting in a red colony phenotype, disrupted binding of the ORC complex will result in expression of the *ADE2* gene leading to a white colony phenotype. Fig3.4.B. Comparisons of diploid cells containing ORC5 and ORC4 chimeric subunits and the isogenic WT strain. The WT diploid cell is red showing that the expression of the *Ade2* allele at the HMR locus is repressed. The ORC5 chimera shows a very similar phenotype to the WT diploid suggesting that the ORC5 chimeric protein does not affect silencing. Strains containing the ORC4 chimeras show a white colony phenotype suggesting that silencing has been lost at the *HMR* locus. Fig3.4.C. Comparisons of ORC4 chimeric diploids (Left) with haploids sporulated from the same chimeric strain (Right). The white color indicating loss of silencing is well maintained, however there are hints of pink color that may result from some degree of *ADE2* silencing in some cells. Haploid cells containing only the *Sc*ORC4 show effective silencing of *ADE2*. Fig3.4.D. Sequence at the *HMR E* (essential) element. While both the *HMR* E and I are involved in the silencing of the *HMR* locus, disruption of the binding of ORC to the *HMR* E site will lead to a loss of silencing. Highlighted in red is the *S. cerevisiae* ACS and highlighted in green is the *K. lactis* ACS. These ACSs overlap but are slightly offset.

Chimeric constructs of Orc5 showed little to no effect on the color of the colonies compared to the isogenic homozygous ORC5 strain as shown in Figure 3.4.B. In contrast, some of the ORC4 chimeric constructs turned the color of the colonies from red to white or a very light pink suggesting a dominant loss of silencing (Fig 3.4.B). This observation suggests that the AAA+ and WHD domains of *K. lactis* Orc4 interfere with the establishment of silencing at the *HMR* locus in a dominant manner. Due to the presence of both the Wild-type *S. cerevisiae* ORC4 and the chimeric Orc4 in these cells, the loss of silencing cannot simply be due to a loss of Orc binding at the *HMR-E* locus, but rather must be due to a more direct interaction of the chimeric Orc4 subunits with the *HMR-E* locus that interfered with the wild-type ORC binding. Consistent with this interpretation, tetrad analysis of the heterozygous diploid showed that viable spores, which do not contain the chimeric *ORC4* allele, reestablish silencing, suggesting that the loss of silencing is dependent on the presence of the chimeric Orc4 protein (Fig3.4.C). This result further suggests that these chimeric Orc4 subunits are able to form a functional complex with the other ORC subunits. The inability of the chimeric Orc4 to fully substitute for the wild-type *S. cerevisiae* Orc4 is likely due to an inability to initiate replication, or an inability to initiate replication at a sufficient number of loci in the *S. cerevisiae* genome.

Examination of the sequence around the *HMR E* element showed a strong match to the *K. lactis* ACS overlapping the *S. cerevisiae* ACS as shown in figure 3.4.D. A likely explanation for the de-silencing of *ADE2* is the binding of the chimeric ORC to the serendipitous *K. lactis* ACS that overlaps the *S. cerevisiae* ACS leading to the exclusion of the *Sc*ORC and that the altered structure or slightly offset positioning of the chimeric ORC prevents the establishment of the silencing machinery.

ChIP-seq results show increased ORC binding at *HMR E* as well as several select origins

To investigate the genome-wide effects of the chimeric Orc4 on the binding of the ORC complexes a ChIP-Seq analysis was performed. Both *ORC6* alleles in the diploid strain were tagged with the 6HA epitope at their C-terminal ends to allow both sets of wild-type and chimeric ORC complexes to be pulled down by chromatin immuoprecipitation under identical conditions. Based on the predicted structure of ORC, the Orc6 C-terminus lies far away from the sites of interaction between DNA and the ORC complex, so the tags should not interfere with the binding of the complex (Chen et al., 2008).

Figure 3.5 A

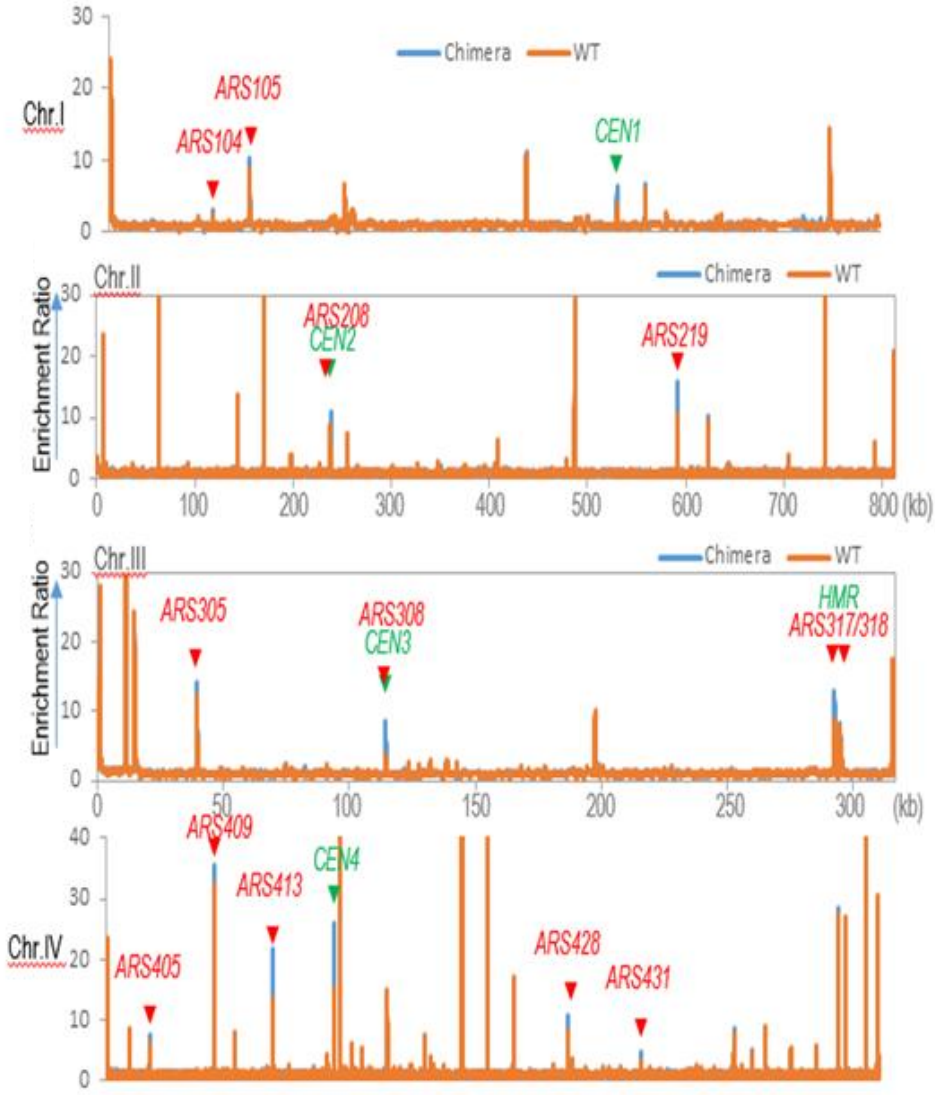

Fig3.5.A. The ChIP enrichment across chromosomes I-IV of *S. cerevisiae*. The ChIP signal in the WT strain is depicted in orange and the chimeric strain in blue. Overall, signals in the majority of the genome are very similar. There are not any ARS loci where the chimeric strain has significantly decreased the enrichment signals suggesting that the chimeric ORC4 construct does not interfere with the binding of the WT *S. cerevisiae* complex. Additionally there is not any significant signal enrichment in the chimeric strain that does not show some enrichment in the WT *S. cerevisiae* strain. Locations of increased enrichment in the chimeric strain correspond to locations where there is also some enrichment in the WT.

Figure 3.5 B

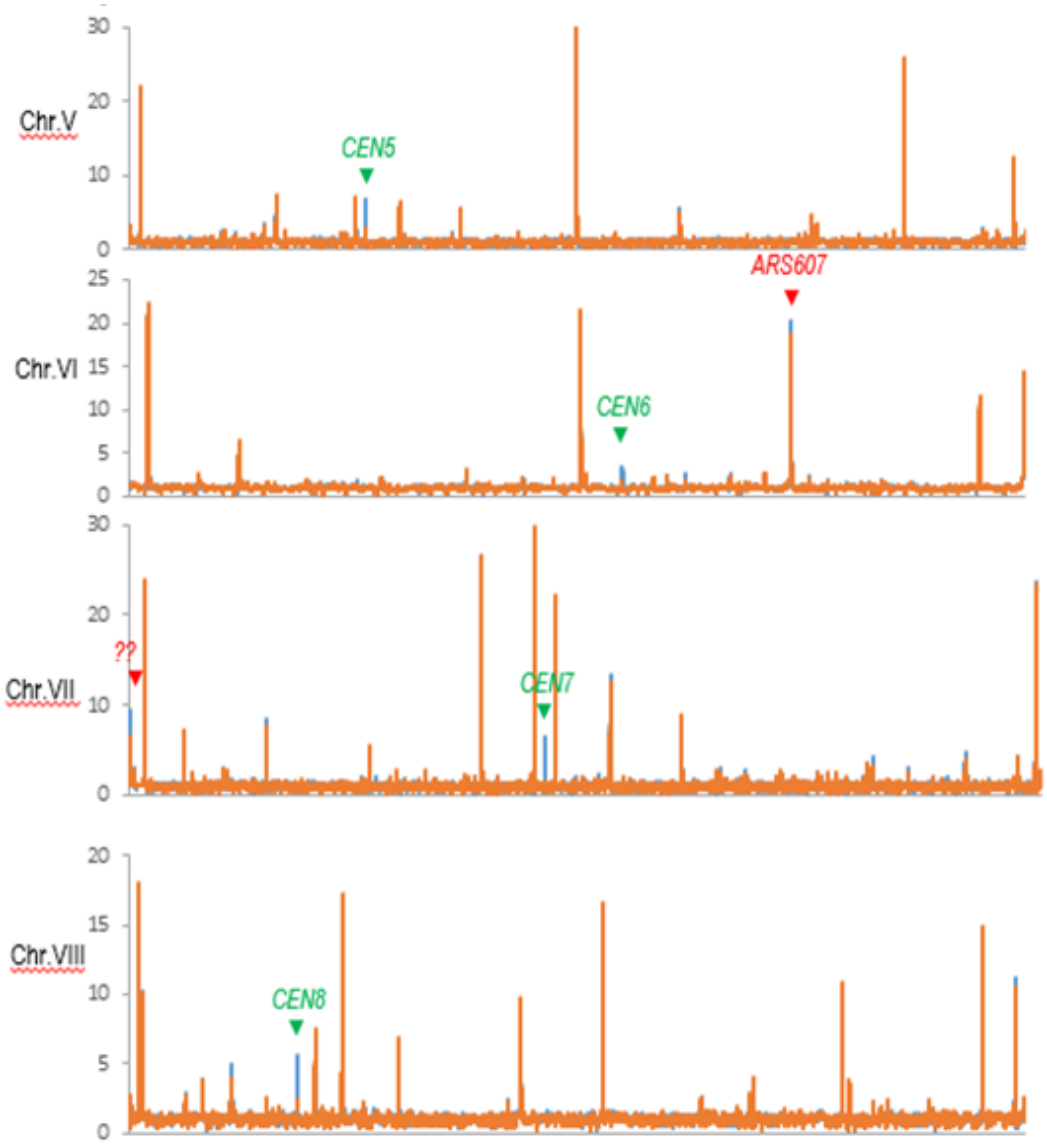

Fig3.5.B. The ChIP enrichment across chromosomes V-VIII of *S. cerevisiae*. The ChIP signal in the WT strain is depicted in orange and the chimeric strain in blue. Overall, signals in the majority of the genome are very similar. There are not any ARS loci where the chimeric strain has significantly decreased the enrichment signals suggesting that the chimeric ORC4 construct does not interfere with the binding of the WT *S. cerevisiae* complex. Additionally there is not any significant signal enrichment in the chimeric strain that does not show some enrichment in the WT *S. cerevisiae* strain. Locations of increased enrichment in the chimeric strain correspond to locations where there is also some enrichment in the WT.

Figure 3.5 C

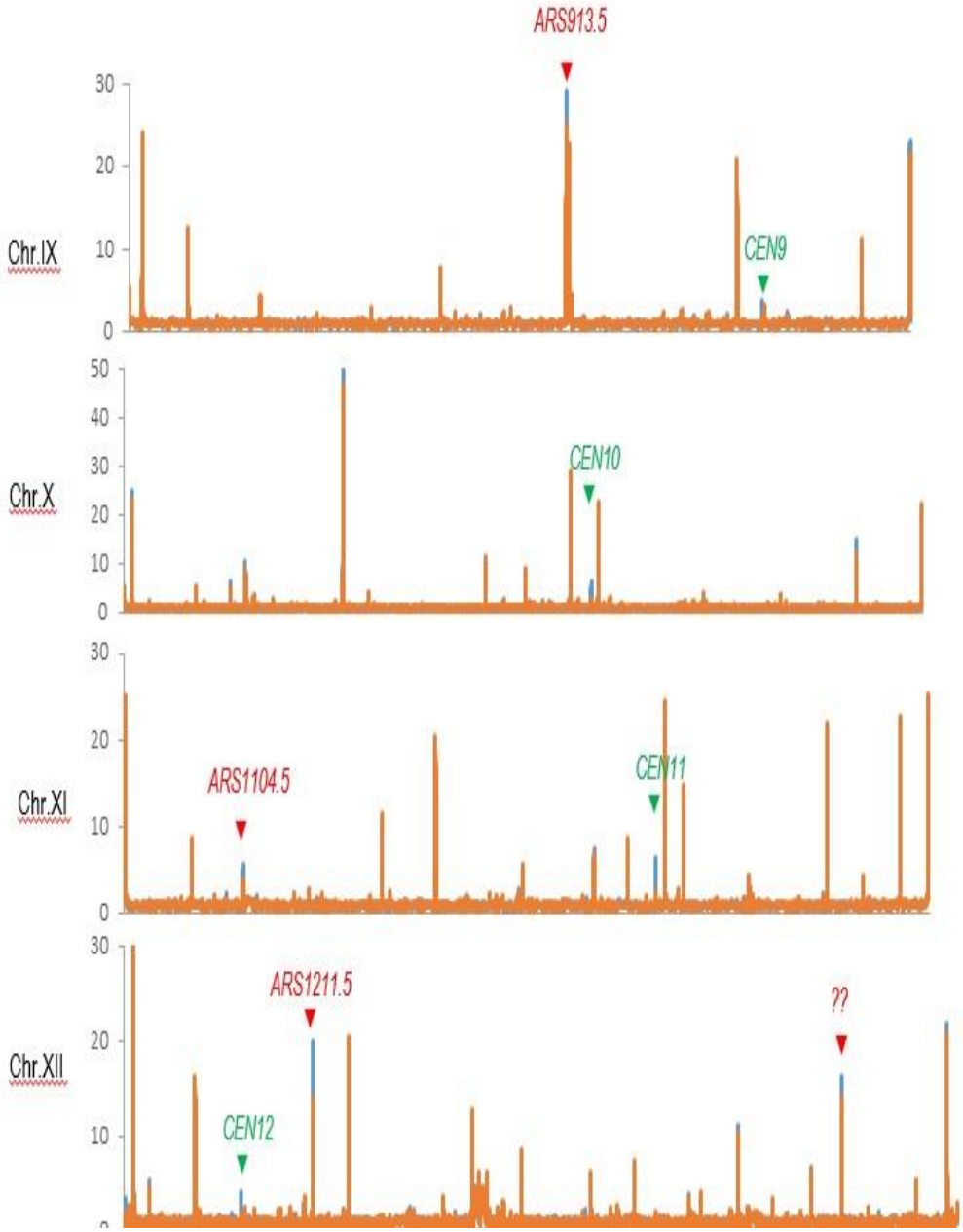

Fig3.5.C. The ChIP enrichment across chromosomes IX-XII of *S. cerevisiae*. The ChIP signal in the WT strain is depicted in orange and the chimeric strain in blue. Overall, signals in the majority of the genome are very similar. There are not any ARS loci where the chimeric strain has significantly decreased the enrichment signals suggesting that the chimeric ORC4 construct does not interfere with the binding of the WT *S. cerevisiae* complex. Additionally there is not any significant signal enrichment in the chimeric strain that does not show some enrichment in the WT *S. cerevisiae* strain. Locations of increased enrichment in the chimeric strain correspond to locations where there is also some enrichment in the WT.

Figure 3.5 D

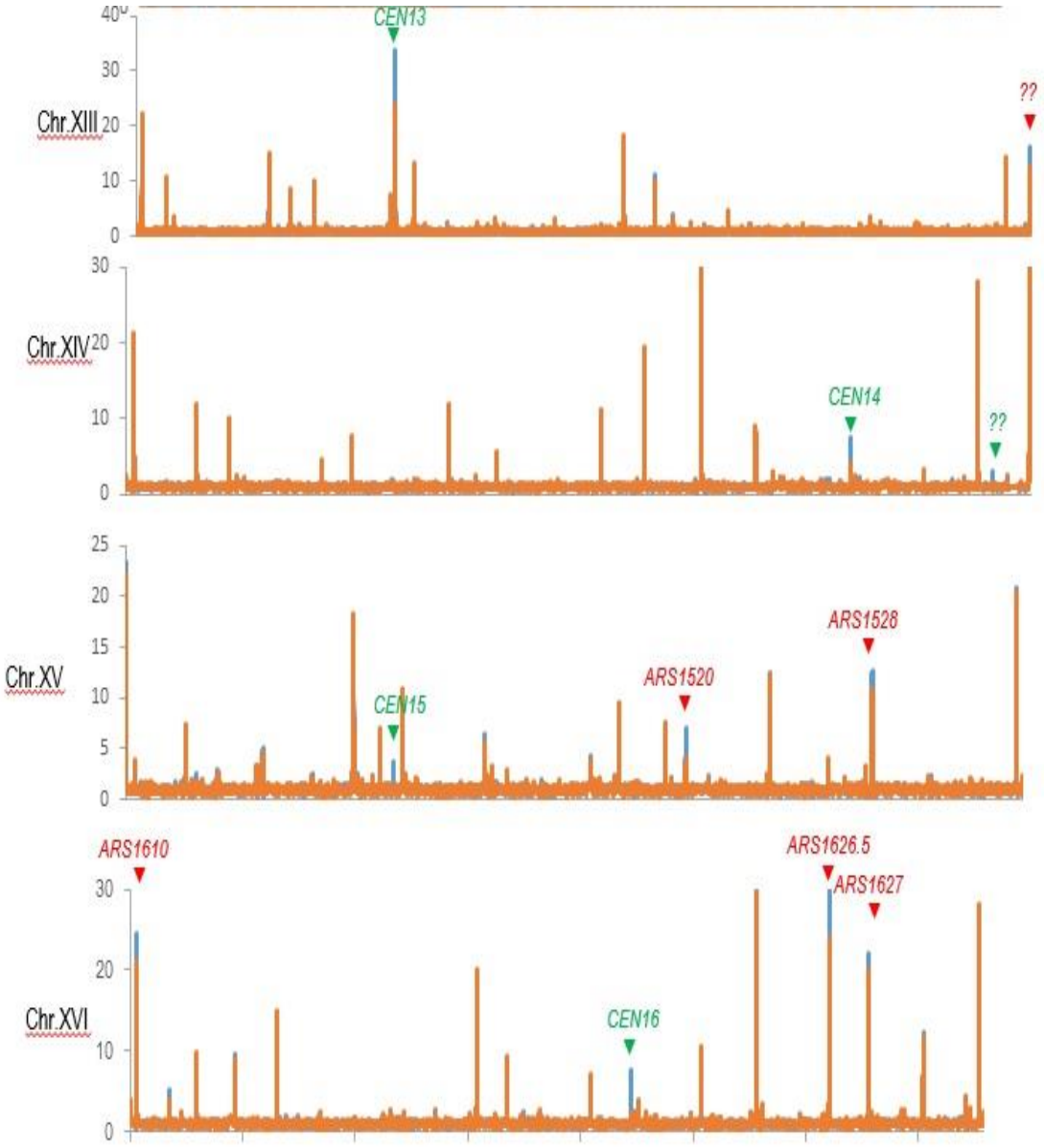

Fig3.5.D. The ChIP enrichment across chromosomes XIII-XVI of *S. cerevisiae*. The ChIP signal in the WT strain is depicted in orange and the chimeric strain in blue. Overall, signals in the majority of the genome are very similar. There are not any ARS loci where the chimeric strain has significantly decreased the enrichment signals suggesting that the chimeric ORC4 construct does not interfere with the binding of the WT *S. cerevisiae* complex. Additionally there is not any significant signal enrichment in the chimeric strain that does not show some enrichment in the WT *S. cerevisiae* strain. Locations of increased enrichment in the chimeric strain correspond to locations where there is also some enrichment in the WT.

ChIP-seq (Lefrancois, 2010) was performed using a diploid strain containing only wild-type *ORC4* as the control and a heterozygous diploid strain containing a wildtype *S. cerevisiae ORC4* and a chimeric *ORC4* that is a fusion of the N-terminal 207 aa of the *S. cerevisiae* Orc4 with the C-terminal 304 aa of the *K. lactis* Orc4. In the heterozygous chimeric strain, DNA bound to both the wild-type and the chimeric complexes were analyzed. Comparative analyses of these two strains provided important information on the differences in the binding patterns of the wild type and the chimeric ORC.

ChIP-seq analysis showed that the chimeric complex did not result in a general disruption of the binding of the ORC complex. The majority of origins showed very similar levels of enrichment when normalized against the input background in both the wild-type *S. cerevisiae* and chimeric strain. Peaks associated with known *S. cerevisiae* origins are within ten percent enrichment of each other. The similarity in the binding pattern suggests that the chimeric Orc4 does not cause a general defect or interference with the binding of the endogenous ORC complex (Fig 3.5 and Fig 3.7.A). The similarity in the binding signal at the majority of origins provides confidence for a comparative analysis of the small number of about a dozen dissimilar sites observed genome-wide.

Perhaps the most surprising result was that we did not observe any enrichment in ORC signals that is unique to the chimeric strain. However, there were a number of sites with increased signal enrichment in the chimeric strain and all of them fall at known *Sc*ARSs or in the centromeric region (Fig 3.7).

We first focus our attention on the *HMR-E* locus which was used for our visual assay for the chimeric Orc4 function. The *HMR-E* locus showed a 1.5 fold increase in signal enrichment in the chimeric strain compared to the wild-type strain (Fig 3.6). In addition, the peak of signal enrichment at the *HMR-E* locus is slightly shifted, corresponding to the slight shift in the centers of the *S. cerevisiae* and *K. lactis* ACSs. In contrast, the *HMR-I* locus and the loci flanking the *HML* did not show any noticeable differences in enrichment or peak position, suggesting that the chimeric Orc4 did not result in a general disruption of binding at the sites involved in ORC mediated silencing, but rather that the effect on binding was specific to the *HMR-E* locus.

Figure 3.6.

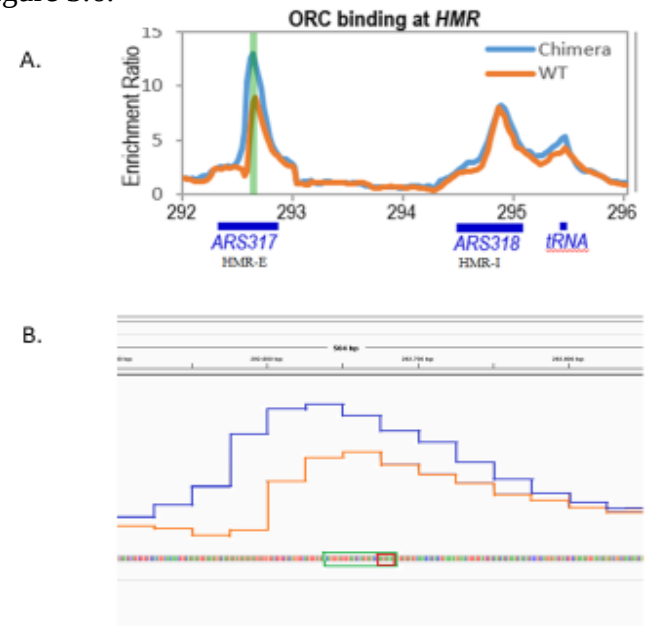

Fig3.6.A. ChIP-Seq analysis of the HMR locus. ChIP-Seq was performed on a wild-type strain and a strain heterozygous for a chimeric ORC4. ORC6 was tagged with 6HA. ChIP data suggests very similar levels of binding for both strains at the *HMR-I* (ARS318) locus but increased binding at the *HMR-E* (ARS317) element in the chimeric strain, corresponding to the *K. lactis* ACS at this locus. B. A closer examination of *HMR-E ARS*. The *Sc* ACS is marked in the sequence below in red and the *Kl* ACS is marked in green. The enrichment in the chimeric strain is offset as compared to the enrichment in the WT strain roughly corresponding to the offset of the *Kl* ACS from the *Sc* ACS.

Analysis of the sequences at these four loci demonstrated that the *HMR-E* locus is the only site that contains overlapping *S. cerevisiae* and *K. lactis* ACSs consistent with the loss of silencing being due to the presence of the *K. lactis* ACS and the binding of the chimeric complex to this ACS.

In addition to the increased binding at the *HMR-E* locus, the chimeric Orc4 resulted in increased binding at a few select origins (Fig 3.7 B and D). This increased binding was not a general trend as most origins displayed very similar levels of enrichment in both the wild-type and chimeric strains (Fig3.5 and 3.7.A). The increases at these select sites were similar in magnitude to that observed at the *HMR-E* locus on the order of 20% to 200% greater enrichment in the chimeric than the wild-type strain. Examination of the sequences at these origins showed that they also contain a good match to the *K. lactis* ACS in the vicinity of the *S. cerevisiae* ACS.

While the chimeric Orc4 complex recognizes and binds *K. lactis* ACSs, this binding specificity alone may not be sufficient for establishing a functional replication origin. In *S. cerevisiae* there are over 14,000 *S. cerevisiae* ACS matches in the genome but only a small subset of 300-400 actually serve as replication origins (Breier et al, 2004, Nieduszynski et al, 2007). Other factors such as the openness of the local chromatin state as well as the presence of actively transcribed genes influence the accessibility of ORC to their cognate ACS (Eaton et al, 2010). Analysis of the *S. cerevisiae* genome for strong matches to the *K. lactis* ACS (at a stringent cutoff threshold of 10) is 167, of these matches, only 30 fall within intergenic regions. ChIP-seq analysis shows a >20% enrichment of ORC binding in the chimeric *ORC4* strain only at 11 sites all of which are also known *Sc*ARSs. In addition, every centromere showed ORC binding in the wild type strain and enhanced signals in the chimeric strain. The small number of binding sites for the chimeric ORC complex may explain the inability of the chimeric *ORC4* to

substitute for the endogenous *ORC4* because of insufficient initiation events for the complete replication of the entire yeast genome.

ChIP Seq Results Demonstrate Presence of Increased ORC Binding at Centromeres in the Chimeric Strain

Figure 3.7.

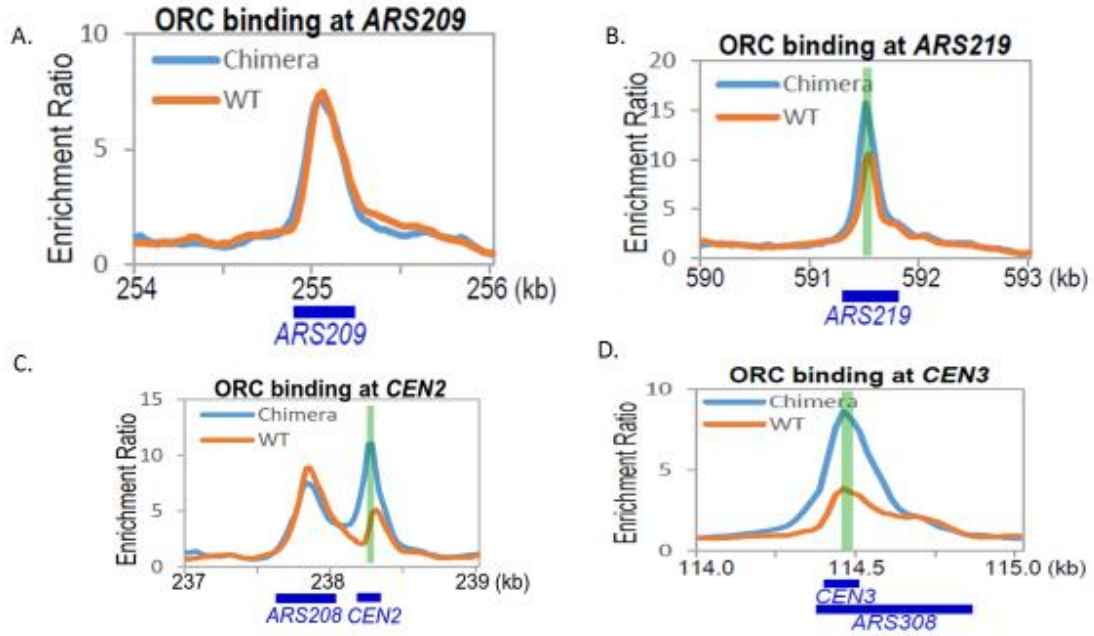

Fig3.7.A. Example of ORC signals at an *S. cerevisiae* ARS. ARS209 is an example of a typical *S. cerevisiae* ARS. The level of enrichment is near identical in both the WT *S. cerevisiae* strain and the ORC4 chimeric strain. This type of signals is observed at the vast majority of *S. cerevisiae* ARSs. B. Examples of ChIP-seq enrichment at an *S. cerevisiae* ARS which also contains a match to the *K. lactis* ACS. The vertical green line indicates the positions of the *K. lactis* ACS. The chimeric strain shows a significant increase in enrichment over the WT *S. cerevisiae* at this ARS, suggestion that there is increased binding of ORC complexes at this loci due to the affinity of the chimeric complex for the *K. lactis* ACS. C. Increased enrichment of ORC signal at the CEN2 loci. The nearby ARS (ARS208) does not show any significant difference in enrichment. D. Increased enrichment of ORC signal at the CEN3 loci in the chimeric ORC4 strain is typical of all centromeres. A *S. cerevisiae* ARS (ARS308) is also present in CEN3. The chimeric strain shows a dramatic increase in enrichment possibly due to the presence of a strong match to the *K. lactis* ACS at this locus as highlighted by the vertical green line.

One interesting piece of information gleaned from the ChIP-Seq data was the binding of the ORC complex to the centromeres. The association of the ORC complex to centromeres has been observed in other species but not observed *in vivo* in *S. cerevisiae* (Prasanth et al, 2002). All

centromeres showed a significant enrichment of the ORC complex association, despite the absence of *S. cerevisiae* or *K. lactis* ACSs at these loci.

Interestingly, the tested chimeric strain displayed an even greater ORC enrichment at the centromeres (Fig3.7.C and D). While some centromeres did contain a possible match to the *K. lactis* ACS, this was not true of all centromeres. This observation suggests that the recruitment of the ORC complex to the centromeres may be dependent on factors other than DNA sequence affinity or the ACS derived for *K. lactis* ARSs represents only a subset of the chimeric ORC recognition sequences. It is also possible that the centromeres fit a consensus sequence to which the *K. lactis* ORC complex would bind but not necessarily initiate replication. The initial *K. lactis* ARS screen was dependent upon the replication of the plasmids (Liachko et al, 2010). Therefore it would only find sequences to which the *K. lactis* ORC both binds and initiates replication, but would not find sequences to which the ORC complex binds but does not initiate replication. The sequences found at the centromeres in *S. cerevisiae* could be such sites. It has been found that several ORC subunits in human cells associate with the centromeres and other heterochromatin (Prasanth et al, 2004). Further examination revealed that depletion of multiple or even a single ORC subunit leads to gross defects in spindle formation and chromosome condensation both of which can result in genetic instability (Prasanth et al, 2004, Prasanth et al, 2010). While this effect has not been directly observed in *S. cerevisiae* or *K. lactis* the importance of the association of ORC at the centromeres in humans suggests that it may be conserved.

Chimeric Orc4 strains replicate plasmids containing a *K. lactis* ACS

Based on the ChIP-seq results, the chimeric ORC has some specificity for the *K. lactis* ACS. Since the C-terminal domain of Orc4 comes from *Kl*Orc4, this result suggests that the C-

terminal portion of the Orc4 alone determines the binding specificity of the whole ORC complex. If so, the chimeric ORC should be able to bind native *K. lactis* ARSs and this could be confirmed by plasmid maintenance assays. To investigate the ability of the chimeric Orc4 subunit to direct the establishment of replication origins and initiate DNA replication at *K. lactis* ACSs, I used a plasmid-based assay developed by Ivan Liachko (Liachko et al. 2010) to determine the ARS consensus sequence of *K. lactis*. Basically, Dr. Liachko had constructed a

Figure 3.8.

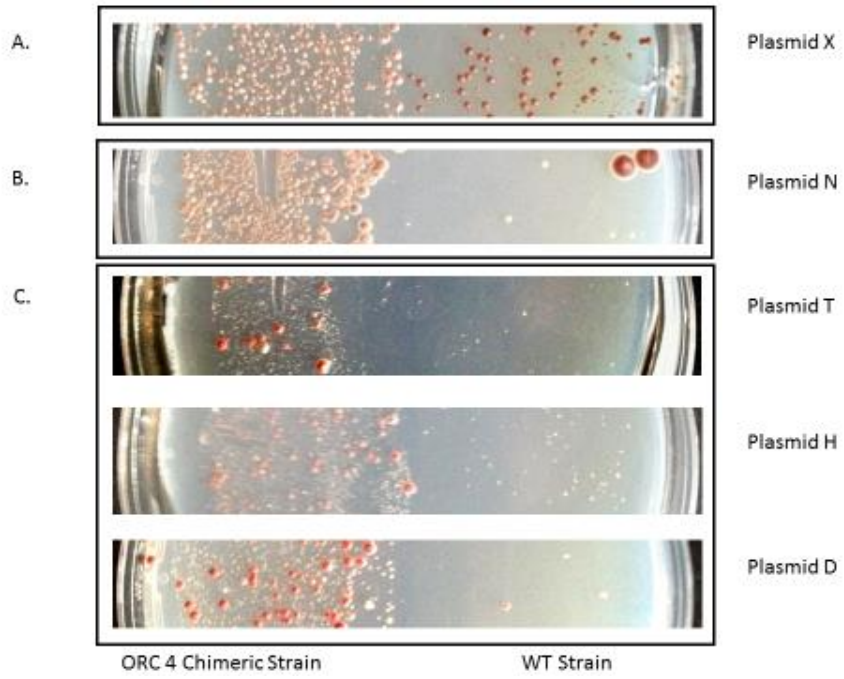

Fig3.8. Yeast transformation assay for autonomously replicating plasmids containing *K. lactis* ARSs on selective media. Strains containing the chimeric ORC4 used for the ChiP-seq are shown on the left, wild-type cells on the right. A. A plasmid containing both *Sc* and a *Kl* ARS acivities. This plasmid can be replicated by both the Orc4 207 chimeric strain and the WT *Sc* strain. B. A plasmid containing a strong *Kl* ARS and a very weak *Sc* ARS activity. The chimeric Orc4 207 strain is able to replicate such plasmids effectively, while the transformation frequency is greatly reduced in the WT strain. C. Plasmids which contain only a *Kl* ACS. The ORC4 207 chimeric strain is able to replicate such plasmids effectively, while the WT strain only forms abortive colonies. Other Orc4 chimeric strains containing the *Kl* AAA+ domain are also able to replicate such plasmids.

library of the entire repertoire of *K. lactis* ARSs that have been confirmed to be functional in *K. lactis*. The plasmids contained several auxotrophic markers and a *CEN5* sequence, but contained no *S. cerevisiae* ACS. *K. lactis* ARSs were inserted into these plasmids to determine which

sequences could function as origins in *K. lactis*. Most of these *K. lactis* ARSs could only function in *K. lactis* but a subset (17%) of *this K.* lactis ARS library were able to replicate in *S. cerevisiae* because of the serendipitous occurrence of a *S. cerevisiae* ACS nearby. Using Figure 3.9.

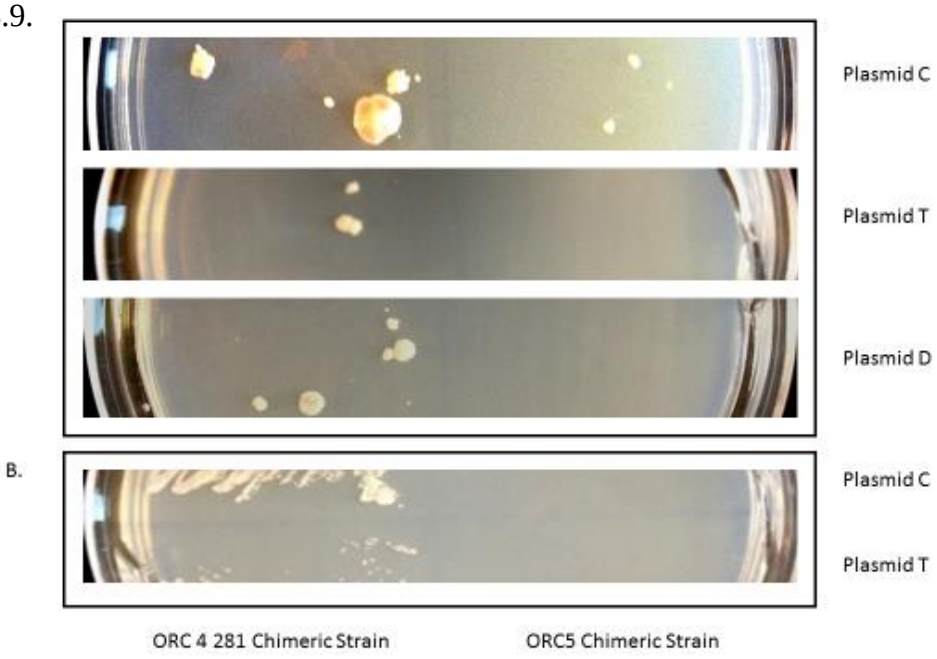

Figure3.9.A. Yeast DNA transformation assay for autonomously replicating plasmids containing *K. lactis* ARSs in strains containing the chimeric ORC4 WHD from *K. lactis* (Orc4 281) (Left) and the Orc5 chimeric strain (Right). Plasmids containing class C *K. lactis* ACS are used. The Orc4 281 Chimeric strain is able to maintain these plasmids while the ORC5 chimeric strain is not able to. The serrated colony morphology observed in the Orc4 281 chimeric strain suggests that it does not maintain the plasmids containing a *Kl* ACS as well as other Orc4 chimeric strains. B. A restreaking of transformed colonies from the transformations of the Orc4-281 chimeras and ORC5 chimeras. The Orc4-281 chimera is able to maintain the plasmids while the ORC5 chimera is not. However, the Orc4 281 chimeric appears to thin out in the restreak fairly rapidly suggesting it does not maintain plasmids containing a *Kl* ACs well.

representative subsets of this library, we were able to verify the specificity of the chimeric ORC.

Three classes of *K. lactis* ARS plasmids were tested with the wild-type *S. cerevisiae* and chimeric ORC strains: plasmids which replicate in both *S. cerevisiae* and *K. lactis* cells (Class 1), plasmids which replicate poorly in *S. cerevisiae* but strongly in *K. lactis* (Class 2), and plasmids which replicate only in *K. lactis* (Class 3). Cells of Wild-type *S. cerevisiae* were transformed in parallel with the cells of heterozygous strains containing chimeric Orc4. Plasmids that are able

to replicate in both species replicated in both the wild-type and heterozygous chimera strain (Fig3.8.A). Plasmids that contain a poor *S. cerevisiae* ACS but a robust *K. lactis* ACS replicated poorly in *S. cerevisiae* showing only a few small colonies but replicated robustly in the chimeric strain (Fig3.8.B). Finally plasmids which do not replicate in *S. cerevisiae* show no growth or only abortive colonies in a wild-type *S. cerevisiae* strain however there is robust growth in the Orc4-207 chimeric strain (Fig3.8.C). Strains containing the chimeric Orc5 constructs did not show the versatility of replicating plasmids containing *K. lactis* ACSs, suggesting that this ability to establish the replicative complex at *K. lactis* ACSs is unique to the Orc4 chimera (Fig3.9). The ORC4-281 chimeric strain shows unusual colony morphology suggesting that it does not maintain the *Kl* ACS plasmids as well as the Orc4-207 chimeric strain. As the Orc4-207 chimera contains a portion of the *K. lactis* AAA+ domain that contains the ISM and the ORC4-281 chimera does not (Fig3.1D), these results suggest that the WHD domain is necessary for recognition of the *K. lactis* ACS but the AAA+ plays a role in stabilizing the binding of the ORC complex to the ACS. The fact that only the *K. lactis* portion of the AAA+ domain but not the *S. cer* portion cooperates with the *K. lactis* WHD in stabilizing ORC binding to the *K. lactis* ACS suggests that it may also have sequence specificity for the *K. lactis* ACS or it only works in combination with the *K. lactis* WHD. Overall, our results showed that the C-terminal region of the *K. lactis* Orc4, especially the WHD, is the primary determinant of the sequence specificity of the ORC complex.

To further examine the effect of the AAA+ domain on plasmid stability, plasmid maintenance assays were performed comparing the ORC4-207 and Orc4-281 chimeric strains. Both Orc4-207 and Orc4-281 strains maintained class 1 and class 2 plasmids at roughly equal efficiency, approximately 77% to 79% in the case of class 1 plasmids and 44% to 50% in the

case of class 2 plasmids. However the Orc4-281 strain maintained class 3 plasmids at a significantly lower rate than the Orc4-207 strain as shown in figure 3.10 where class 3 plasmids are maintained at 82% in Orc-207 strains but only at 19% in Orc-281 strains. Class 3 plasmids are the only class in which a significant difference ($p = 0.019$) can be seen between the two chimeric strains. This result suggests that the while the *K. lactis* WHD is sufficient to initiate replication at *K. lactis* ARSs the *K. lactis* AAA+ domain is necessary for efficient initiation function at *K. lactis* ARSs. The similar levels of stability in class 1 and class 2 plasmids between

Figure 3.10.

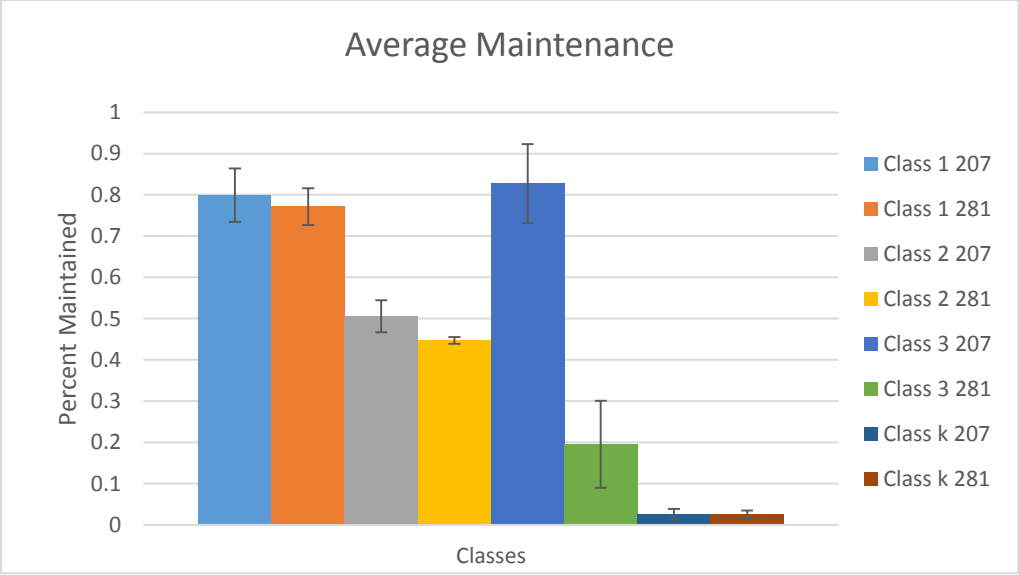

Fig3.10. Plasmid maintenance assay. The Orc4 207 and Orc4 281 strains were transformed with each class of plasmid, selected, and then allowed to grow for 10 rounds of cell division in nonselective media. The strains were then plated on YPD, incubated at 30C and then replica plated onto media selective for the plasmid. These plates were then scored and the percent of plasmid containing cells were calculated. Class 1 plasmids were maintained at 79% and 77% in Orc4-207 and Orc4-281 respectively, while Class 2 were maintained at 50% and 44% respectively. Class 3 plasmids containing only a *K. lactis* ACS was maintained at 82% in the Orc4-207 strain but only at 19% in the Orc4-281 strain. Finally in the class k plasmid which contains overlapping *K. lactis* and *S. cerevisiae* ACSs there are very low levels of maintenance. All classes were performed with a minimum of 3 replicates.

the two strains suggest that the *S. cerevisiae* ACSs are being utilized at a similar level in both strains, providing a baseline of stability. This stabilizing effect of the *S. cerevisiae* ACS is also suggested by the increasing deviation in stability of the class 2 and especially the class 3

plasmids observed between the two strains, suggestion an increasing importance of the *K. lactis* ARSs in maintaining the plasmids.

One interesting effect can be seen in a class 2 plasmid (class k) in which the *S. cerevisiae* and *K. lactis* ARSs overlap. In this case the plasmid loss rate is much higher in both the Orc4-207 and Orc4-281 strains, with the plasmid being maintained at only a few percent in both strains. This would seem to suggest that overlapping ARSs lead to a loss of stability of the pre-replicative complex perhaps due to the competition for binding of both types of ORC.

K. lactis ORC shows an elevated rate of amino acid changes

The ACS sequence of *K. lactis* and *S. cerevisiae* have diverged greatly and the C-terminal end of the Orc4 subunit is primarily responsible for this change in sequence recognition. A cross-species comparison of the Orc4 primary sequences would yield important information about the relationship between the protein structure and the DNA contact residues. There are several regions that may contribute to the DNA binding specificity. An alignment of the Orc4 C-terminal half of *S. cerevisiae*, *K. lactis* and several related species showed that some regions are highly conserved while others diverged significantly (Fig 3.11). The highly diverged regions are candidates responsible for determining sequence specificity. The ACS of *S. bayanus* is nearly identical to that of *S. cerevisiae* and both species can faithfully replicate the ARS plasmids of the other species. Therefore amino acid differences between these two species are not likely to be important. On the other hand, amino acid sequences that are conserved between *S. cerevisiae* and *S. bayanus* but are diverged from *K. lactis* are likely to be more important.

Figure 3.11.

Figure 3.11. A. An alignment of the C terminal region of Orc4 in various yeast species and human. Highlighted in the blue box is a region where *S. cerevisiae* contains 18 more amino acids than *K. lactis*. This region encompasses a small part of a loop and part of a helix in the winged helix domain and therefore likely to be responsible for determining the DNA binding specificity of the protein. Interestingly most of this region is also missing in *S. pombe* and human. Highlighted in the purple box is another candidate region where *S. cerevisiae* and *S. bayanus* are identical but there are many changes between *S. cerevisiae* and *K. lactis.*

Comparing these alignments, a few interesting features become obvious. The first is a region of 18 amino acids in the WH domain that are missing in *K. lactis* but present though not conserved in the other yeast species. Smaller gaps in this region are found in other yeast species however none of the examined species have gaps in this region that are as large as that found in *K. lactis*. Several other regions show differences in the conserved sequence between *S. cerevisiae* and *S. bayanus* but not *K. lactis* (Fig3.11). However it is not possible to ascertain how each of these differences affects the specificity of binding of the ORC complex without further analysis. It is interesting to note that the human and *S. pombe* ORC, which shows dramatically different binding specificities to the *Sc* ORC, also contain large gaps in the same region. Another related question is how the changes in the specificity changed throughout various species. Would the

ancestral state more closely resemble *S. cerevisiae* or *K. lactis* or some intermediate form that closely resembles neither species? To examine this question the Ka (nonsynonymous substitutions) of each of the ORC subunits (Orc1 – Orc5) of a particular species relative to that of *K. lactis* and *S. cerevisiae* were compared to determine the rate of divergence of each subunit. To compare the pairwise changes a ratio was calculated of the Ka for each species against *K. lactis* over the Ka for each species against *S. cerevisiae* which can be expressed as (Ka when compared to *K. lactis*) / (Ka when compared to *S. cerevisiae*). In this manner a ratio of 1 would suggest that both *S. cerevisiae* and *K. lactis* subunit have had the same number of changes as compared to the given species. A ratio of less than 1 would imply that *S. cerevisiae* has undergone a greater number of amino acid changes compared to the given species than *K. lactis,* while a ratio of greater than 1 would suggest a greater number of changes have occurred in the *K. lactis* amino acid sequence. Several pieces of interesting information can be gleaned from this analysis. Compared to the other subunits, Orc1 is most similar among the different species in the level of changes (Fig3.12). However, the actual Ka of Orc1 shows that, it has the highest absolute number of changes compared to the other subunits. This result suggests that the Orc1 has more flexibility in its sequence variability, perhaps due to its larger size.

The general trend found in the Ka for all subunits reveals unequal evolutionary rates of ORC among the yeast species. All but one Ka ratio is greater than 1, suggesting that the amino acid sequence in *K. lactis* ORC has undergone more changes than *S. cerevisiae* relative to the other species. The only exception to this pattern is found in Orc1 of *E. gossypii* which is much more closely related to *K. lactis* than to *S. cerevisiae*. In all other subunits the ratio consistently being greater than one suggests a higher degree of similarity with *S. cerevisiae,* even with species that are more closely related to *K. lactis* (Fig3.12). This result would seem to suggest that

K. lactis has undergone more changes and perhaps a more rapid evolution in its ORC and ACS than other species.

Figure 3.12.

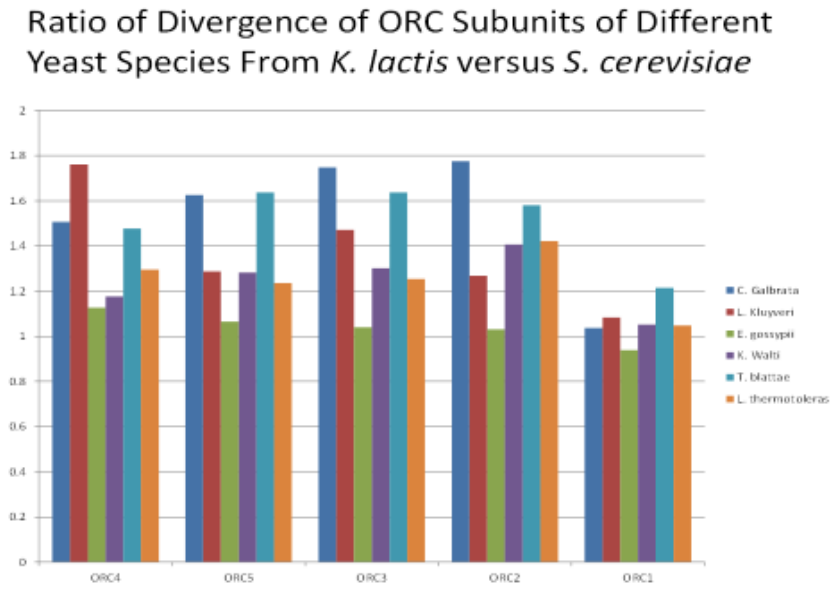

Figure 3.12. A ratio of the changes of amino acids between various species presented as a ratio of the changes between a give species and *K. lactis* and the changes between a given species and *S. cerevisiae.* A ratio equal to one indicates that an equal number of changes have occurred. A ratio of greater than one indicates a greater number of changes have taken place between that species and *K. lactis* than that species and *S. cerevisiae*. From this data it can be seen that most species show a high degree of changes relative to *K. lactis* than to *S. cerevisiae* even in species that are more closely related to *K. lactis.*

Presence of an *ORC4* paralog *RIF2* in a subset of post Whole Genome Duplication species

In examining the evolution of yeast species in the post whole genome we find that most duplicated genes were lost with only one functional copy retained. This pattern is true for the majority of the ORC subunits with two exceptions, *ORC1* and *ORC4*. *ORC1* has a paralog, *SIR3*, which is an important component of the silencing apparatus in *S. cerevisiae* and is required for the spreading of silenced chromatin. *ORC4* has a paralog, *RIF2* which is involved in telomere silencing and maintaining telomere length. Neither *SIR3* nor *RIF2* are essential unlike their ORC subunit paralogs, but have acquired new related functions (Bell SP 2002, Byrne and Wolfe 2005).

Figure 3.13.

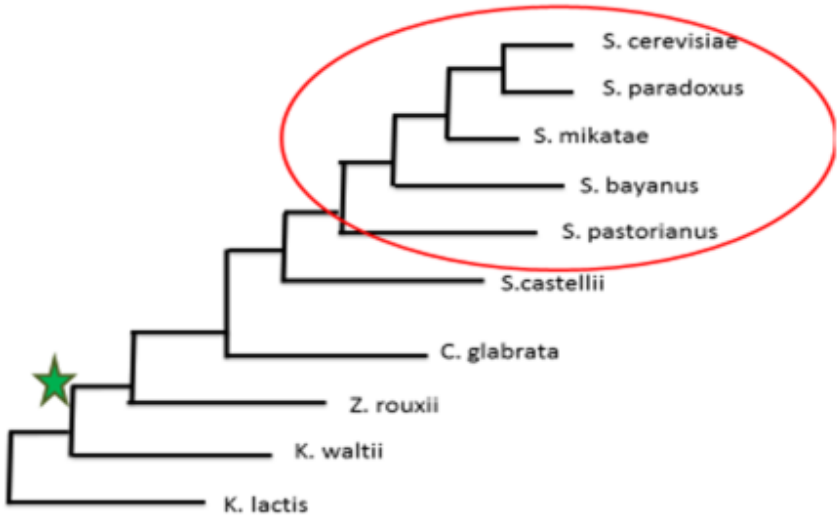

Figure 3.13. A partial phylogenetic tree of yeast species. The group highlighted in the red circle is the only subset of the tree that contains both ORC4 and its paralog RIF2, which presumably resulted from the whole genome duplication event (Marked with the green star). This observation would seem to suggest that the divergence of the functions of ORC4 and RIF2 occurred by the time the sensu stricto group had diverged from the other yeast species.

Altering the ORC specificity during evolution would mean a total change in the landscape for replication initiation events, both in location and frequency. How do evolving

species survive under such turmoil? One hypothesis is that the evolving species retain the ability to initiate at both old and newly evolving ORC bound sites by retaining two copies of *ORC4*, the old version and the evolving version during the transitional state. To examine the importance of the duplication of *ORC4* and how it may have affected ORC evolution, I, in collaboration with Xuepeng Sun, examined post whole-genome duplication species for the presence of a *RIF2* gene and the syntenic position of *ORC4* and *RIF2* in these species. Immediately following the whole genome duplication both copies of *ORC4* would be functional. If one *ORC4* copy evolved to recognize an alternative ACS while the other copy of *ORC4* could recognize the ancestral ACS. The presence of both copies would allow for more freedom in the ability of the ACS landscape to evolve. If one of the *ORC4* has changed to recognize a new ACS and this new ACS gradually populates a sufficient number of locations to allow for complete replication of the genome, then there could be a total switch to the use of the new ACS without any harmful consequence. If this switch occurs the copy that recognizes the ancestral ACS would become dispensable and would be free to acquire a new function. In this scenario, the ancestral copy would be the one to become dispensable. The synteny of the duplicated genes should provide a clue as to whether ORC4 co-evolved with the ACSs among the yeast species via this mechanism or pathway. In other words, one should find the functional *ORC4* alternating in the synteny group of the *ORC4* or *RIF2* gene rather than fixed in one syntenic location among yeast species that emerged after WGD.

Upon examination of the presence and synteny of *RIF2* in post WGD species we made a few interesting observations. *RIF2* appears to be found only in species closely related to *S. cerevisiae* as illustrated in figure 3.13, suggesting that the paralog of *ORC*4 was acquired later and that *ORC4* was duplicated by means other than through WGD. The strong similarity of the

ACSs among the sensu stricto species (based on *S. cerevisiae* and *S. bayanus*) does not offer insight into the mystery of the co-evolution of ORC and ACS.

This observation would also suggest that the function of RIF2 is maintained in the other species' ORC4 subunit as it likely was in the ancestral state (It is also possible that other proteins involved in telomere silencing have taken up the functions that RIF2 has in *S. cerevisiae*). Additionally this would suggest that RIF2 and ORC4 functions had become fixed by the time the sensu stricto group had diverged from the other species. In all post whole genome duplication groups it appears that all *ORC4* are in the same synteny group suggesting that the *ORC4* function was fixed early on after the whole genome duplication.

Summary

DNA replication is a highly regulated process that begins with the binding of the ORC complex to many replication origins. In yeast, these origins are defined by specific sequences, however it was not clear how the ORC complex determined this sequence specificity. In order to examine this we chose to work with *S. cerevisiae* and *K. lactis* which have distinct ACSs only recognized by their cognate ORCs. We were able to determine that the WHD of ORC4 is largely responsible for the sequence specificity of the ORC complex, with the AAA+ domain also playing a role in the ability of the complex to bind to an ARS.

We were able to identify differences in binding between a WT *S. cerevisiae* strain and a strain with a chimeric ORC4 subunit by using a ChIP-Seq analysis. This analysis found increased presence of the ORC complex in the chimeric strain at regions that correspond to the *K. lactis* ACS, most significantly at the HMR region, perhaps explaining the loss of silencing at this region, as well as increased signal at the centromeres.

We found that ORC4 chimeras containing the WHD as well as the AAA+ of *K. lactis* maintained plasmids more stably than ORC4 chimeras containing only the WHD of *K. lactis*. This result suggests that while the WHD the ORC4 of *K. lactis* is sufficient to initiate replication at *K. lactis* ARSs, and the AAA+ plays a role in either the strength of recognition or the stability of binding.

Analysis of the amino acids sequences of *S. cerevisiae*, *K. lactis*, and other yeast species reveals that *K. lactis* has undergone a large number of changes relative to other yeast species. This result suggests that *K. lactis* ORC has undergone a rapid evolution. The polyploid nature of yeast provides an explanation of how such large changes could occur in an essential and highly regulated system.

Works Cited:

Gaudier, M. et al. “Structural Basis of DNA Replication Origin Recognition by an ORC Protein” Science (2007) Vol 317: 1213-6

Yang Zhang. “I-TASSER server for protein 3D structure prediction BMC Bioinformatics”, Vol 9, 40 (2008)

Ambrish Roy, Alper Kuckural, Yang Zhang, “I-TASSER: a unified platform for automated protein structure and function prediction” Nature Protocols, (2010) Vol 5, 725-738

Lee, D.G. and Bell, S.P. “Architecture of the Yeast Origin Recognition Complex Bound to Origins of DNA Replication” Molecular Cell Biology (1997) (12):7159-68

Gupta, A., P. Mehra, and S. K. Dhar. "Plasmodium Falciparum Origin Recognition Complex Subunit 5: Functional Characterization and Role in DNA Replication Foci Formation." Molecular microbiology 69.3 (2008): 646-65.

Duncker, B. P., I. N. Chesnokov, and B. J. McConkey. "The Origin Recognition Complex Protein Family." Genome biology 10.3 (2009): 214

Dueber, E.C. et al. “Molecular Determinants of Origin Discrimination by Orc1 Initiators in Archaea” Nucleic Acids Research (2011) 39(9) 3621-31

Fox, C.A., Loo, S., Dillin, A., and Rine, J. The origin recognition complex has essential

functions in transcriptional silencing and chromosomal replication. Genes & development (1995) 9, 911-924.

Sussel, L., Vannier, D., Shore, D. “Epigenetic Switching of transcriptional states: cis- and trans-acting factors affecting establishment of silencing at the HMR locus in Saccharomyces cerevisiae” Mol Cell Biol (1993) Jul; 13(7):3919-28

Chen, Z., et al. "The Architecture of the DNA Replication Origin Recognition Complex in Saccharomyces Cerevisiae." Proceedings of the National Academy of Sciences of the United States of America 105.30 (2008): 10326-31

Lefrancois, P., Zheng, W., Snyder, M. “ChIP-Seq using high-throughput DNA sequencing for genome-wide identification of transcription factor binding sites” Methods Enzymol 2010;470:77-104. doi: 10.1016/S0076-6879(10)70004-5

Breier, A., S. Chatterji, and N. Cozzarelli. "Prediction of Saccharomyces Cerevisiae Replication Origins." Genome biology 5.4 (2004): r22

Nieduszynski, C. A., et al. "OriDB: A DNA Replication Origin Database." Nucleic acids research 35.Database issue (2007): D40-6

Eaton, M. L., et al. "Conserved Nucleosome Positioning Defines Replication Origins." Genes & development 24.8 (2010): 748-53

Prasanth, S. G. et al. “Orc6 Involved in DNA Replication, Chromosome Segregation, and Cytokinesis” Science (2002) Vol. 297 no. 5583 1026-1031

Prasanth, S.G. et al “Human Orc2 localizes to centrosomes, centromeres, and heterochromatin during chromosome inheritance” EMBO J. (2004); 23(13): 2651–2663

Prasanth, S.G. et al. “Human origins recognition complex is essential for HP1 binding to chromatin and heterochromatin organization” PNAS (2010) 15093-15098

Malik, H.S., Henikoff, S. “Major evolutionary transitions in centromere complexity” Cell 2009 Sep 18;138(6):1067-82.

Liachko, I., et al. "A Comprehensive Genome-Wide Map of Autonomously Replicating Sequences in a Naive Genome." PLoS genetics 6.5 (2010): e1000946

Bell, S.P. (2002). The origin recognition complex: from simple origins to complex functions. Genes Dev 16, 659-672.

Bryne, K.P. and Wolfe, K.H. “The Yeast Gene Order Browser: Combining Curated Homology and Syntenic Context Reveals Gene Fate in Polyploid Species” Genome Research 15(10):1456-61

Albertin, W and Marullo, P. “Polyploidy in fungi: evolution after whole-genome duplication” Proceedings. Biological sciences/ The Royal Scociety (2012) Jul 7; 279(1738:2497-509)

Chapter 4: Discussion:

The Wing Helix and +AAA domains of ORC

The lack of crystal structures for the *S. cerevisiae* ORC subunits poses some difficulty for studying the ORC complex. The construction of possible three-dimensional models for the relevant ORC subunits proved to be helpful in studying the binding of the ORC complex to its cognate ARSs. While crosslinking and EM microscope studies provide some insight into the association of the complex to the ARS DNA, they tell us little about the actual structure of the subunits. By comparing the predicted structures of the ORC subunits to the known archeal crystal structure in a process called threading, we are in a better position to project how the ORC interacts with DNA and specifies its binding affinity. The different behaviors of the chimeras containing only the WHD and those containing the WHD and part of the AAA+ domain allow us to determine which regions are important for the recognition of the *K. lactis* ARS and which for the stability of the bound complex. When we trace these domains to the predicted 3-dimensional structure we find that they correspond strongly to the regions that make contact with DNA in the archeal structure (Guadier et al, 2007). Furthermore mutational analysis of archeal ORC determined that while the WHD does confer specificity of binding, the AAA+ domain does confer some specificity as well, although the majority of contact with the DNA is made by the WHD (Dueber et al, 2011). While it is tempting to try to narrow down the exact residues that are important from this data by substituting individual residues and look for altered specificity or weakened DNA binding, it is beyond the scope of this research to test every relevant substitution for altered functions. Alignment of the *K. lactis* or chimeric Orc4 structure to the archeal ORC structure does not offer too much of an advantage in predicting likely residues. While the predicted 3-D structures are largely similar to the archeal structure, alignments of the amino acid

sequences show very little conservation. Further work by solving the NMR structure of the WH and AAA+ domains bound to DNA would be much more informative as to which residues actually contacts the ARS DNA as well as measuring their binding affinity. [1]

Why certain chimeras fail to substitute for the endogenous Orc protein

The failure of the ORC4 and ORC5 chimeras to substitute for their *S. cerevisiae* counterparts are likely to be due to different reasons. The most trivial explanation is that the chimera fails to form a complex with the other five *S. cerevisiae* ORC subunits. It is also possible a chimeric ORC complex is formed but other factors are affecting the ability of the ORC complex to initiate DNA synthesis. For example, the chimeric subunit may directly affect the interaction of the ORC complex with the rest of the replication machinery, or it may alter the conformation of the complex as a whole, which results in a loss of interactions with the rest of the replication machinery. The latter possibility is more likely as Orc4 and Orc5 are not known to interact with components of the replication or silencing machinery.

Our results show that Orc4 changes the binding specificity of the ORC complex as a whole. This conclusion is supported by the ORC4 chimeric strains having a silencing defect where a *Sc* and a *Kl* ARS overlap, the increased binding seen at Kl ARSs in the ChIP-SEQ data in the chimeric strain as compared to the wild type, and most importantly the ability of the chimeric strain to replicate plasmids containing only a *K. lactis* ACS. This last set of data show that not only does the complex containing the chimeric ORC4 subunit direct the binding of the

[1] After submission of this thesis, the crystal structure of the Drosophila ORC was solved (Bleichert et al. 2015). The AAA+ domains of Orc1, 2, 3, 5, 4 interact to form a quasi-spiral arrangement of DNA binding elements around an approximately 20Å wide central channel. Unless the *K. lactis* AAA+ domain contributes to a more stable chimeric ORC, the crystal structure supports the notion that the C-terminal proximal portion of the Orc4 AAA+ domain contributes to the sequence specific binding of ORC.

ORC complex to *KL* ARSs but it is able to successfully recruit the replication machinery to these loci. We cannot rule out that the ORC5 chimeric subunit also alters the binding specificity of the ORC complex in a way similar to the ORC4 chimeric except the newly acquired specificity directs the chimeric ORC to sites other than the *Sc* and *Kl* ARS at loci that we are not aware of. We do not believe that the ORC5 chimera has the specificity for *K. lactis* ARSs as it did not exhibit a silencing defect in the ORC5 chimeric strain, nor showed the ability to replicate plasmids containing *K. lactis* ARSs. In fact, we do not know whether the ORC5 chimera forms a functional hexameric complex that binds DNA and if such a complex is formed, whether it could form part of the replisome. We have decided not to follow this line of investigation because it was clear that the chimeric Orc4 alone was sufficient to convert the specificity of the *Sc*ORC to that of the *Kl*ORC.

The dominant silencing defect of the ORC4 chimera

If the chimeric Orc4 subunit is able to recruit the entire ORC to *K. lactis* ARSs then it is not immediately clear as to why it exerts a silencing defect. If the chimeric ORC is able to recruit the replication machinery, then it seems likely that it should also be able to recruit the silencing machinery to the HMR locus. Additionally the loss of silencing takes place in a dominant manner when the WT copy of ORC4 should be sufficient to establish silencing. I can offer alternative models to explain this odd effect. (Figure 4.1).

Model 1 (Fig 4.1B): The chimeric ORC complex binds more tightly to the HMR E locus, preventing the WT complex from binding, but it fails to interact with the silencing machinery. We note that the *S. cerevisiae* ACS at the HMR-E is a very strong match but so is the *K. lactis* ACS, but the ORC complex that has a stronger affinity would be dominant over the other. That the chimeric ORC may have a stronger affinity of the two is supported by the increase in ChIP

signal seen at the HMR-E locus. In this scenario, the chimeric ORC would have to interfere with the interaction between the silencing machinery and the endogenous ORC complex. So far there is no evidence of a direct interaction between Orc4 and any of the silencing machinery. However it is possible that the chimera results in a conformational change that destroys the interaction of the ORC complex with the silencing proteins. However this same conformational change must not affect the interaction of the ORC complex with the replication machinery as the chimeric ORC is able to recruit the replication machinery to the *Kl*ARS.

Model 2 (Fig4.1C): The chimeric ORC complex binds more tightly to the HMR E locus, preventing the WT ORC from binding, and the difference in position, due to the slightly offset *K. lactis* ACS with respect to the *S. cerevisiae* ACS, prevents the proper establishment of silencing. Similarly, as the ORC complex is not symmetric, the chimeric ORC complex may bind in a reverse orientation. The stronger binding of the chimeric ORC complex is supported in the increased ChIP signal observed at the HMR-E locus. The position of the ORC complex has been shown to play a role in how effectively it carries out its functions with even a single base pair deletion being sufficient to disrupt binding in certain cases (Bolon and Bielinsky, 2006). Additionally, if the chimeric ORC complex is oriented in a reverse orientation compared to the wild-type, this may disrupt silencing as reversing the E element disrupts silencing at HMR (Zou, Y. et al 2006). However more investigation would be needed to determine if this is what results in the loss of silencing.

Model 3 (Fig4.1D): The competition between the chimeric ORC and the WT ORC prevents either from establishing a strong presence at the overlapping ACS and thus prevents the establishment of the silencing machinery. In this scenario, both the chimeric complex and the WT complex have similar affinities for the HMR-E locus such that neither complex can

Figure 4.1.

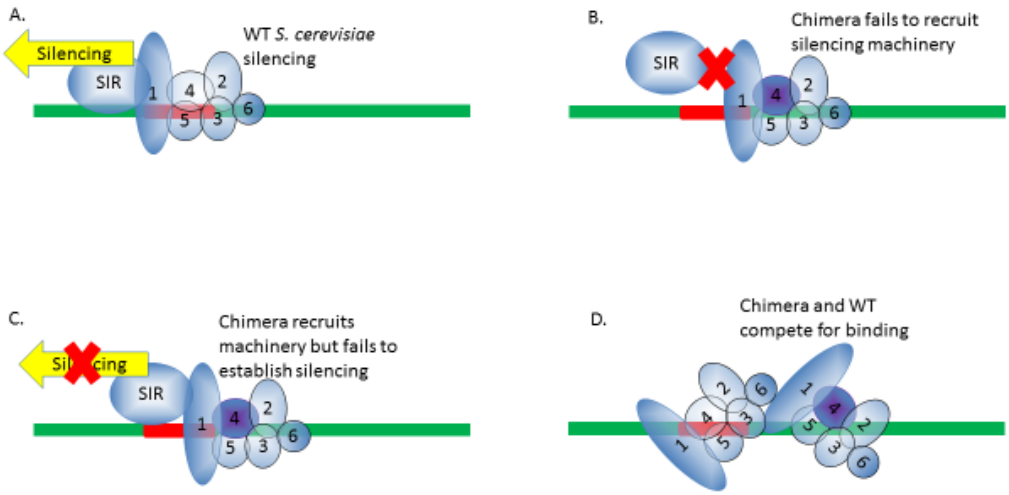

Fig4.1. Possible models for the loss of silencing seen at the HMR-E locus. The dominant loss of silencing seen in the heterozygous strain containing both the WT *S. cerevisiae* and the chimeric ORC4 subunit can possibly be explained in several different models. The ORC complex is shown in transparent blue with the numbers designating each subunit. The darker purple ORC 4 subunit represents the chimeric ORC 4 subunit while the lighter colored ORC4 subunit is the WT *S. cerevisiae* subunit. The *S. cerevisiae* ACS is marked in red and the *K. lactis* ACS is denoted in green. A.) Silencing occurring in a WT *S. cerevisiae* cell. The WT ORC complex binds to the *S. cerevisiae* ACS and recruits the silencing machinery to the HMR-E locus. The silencing then spreads to the rest of the HMR locus. B.) In this model the ORC complex containing the chimeric subunit binds more strongly than the wild-type ORC complex and excludes the WT ORC complex from binding. However the chimeric ORC4 changes the overall conformation of the ORC complex and therefore fails to interact with the silencing machinery and therefore does not establish silencing. C.) In this model the chimeric complex binds strongly excluding the WT complex from binding. In this model the chimeric ORC complex interacts with the silencing machinery, however silencing is not established properly. In this model the chimeric ORC complex does not bind in the same location as the WT ORC complex and therefore the silencing is not properly established. D.) In this model the chimeric and WT. ORC complexes compete for binding at the HMR-E locus. In this model, neither version of the complex can bind strongly, and therefore the silencing machinery is not properly recruited to the HMR locus.

completely exclude the other complex resulting in neither complex being able to bind stably enough to recruit the silencing machinery. This sort of interaction could also be the result of both complexes attempting to bind, but due to the fact that the two ARSs are offset, each complex prevents the other from assuming the necessary conformation to recruit the silencing machinery. If so, the elevated signal in the ChIP-SEQ data seems incongruous with this hypothesis as one might expect this transient occupancy to reduce the ChIP signal. It is possible the weak

association is sufficient for crosslinking of the complex to the DNA, and that the presence of additional tagged Orc6, at this loci, despite the weaker binding of the ORC complex, leads to higher levels of signal in the ChIP-SEQ. We have also considered that the increase in signal is due to the unraveling of DNA in this region due to the loss of the establishment of silencing making it more amenable to chromatin IP. A higher proportion of DNA would be recovered in this case leading to a higher signal. This scenario of competitive binding is supported by the apparent lower maintenance of plasmids which contain overlapping *K. lactis* and *S. cerevisiae* ARSs in the chimeric Orc4 strains, whereas plasmids that contain only a *K. lactis* ARS or plasmids that contain both a *K. lactis* ARS and a *S. cerevisiae* ARS that do not overlap show much higher levels of plasmid maintenance. This model of competitive exclusion between the two types of complexes at this locus is tempting as it explains both of these occurrences with a single mechanism.

ORC binding at centromeres

The enrichment of ORC at the centromeres in the chimeric strain brings up some interesting possibilities. While there is some enrichment at the centromeres in the WT strain, there is a consistent relative increase in enrichment at the centromeres in the chimeric strain. The data from the WT strain suggests that the presence of ORC at the centromeres in *S. cerevisiae* is a normal facet of the ORC complexes interactions. There is evidence that some of the ORC complex subunits, Orc6 in particular, (Prasanth et al. 2002) play a role in the kinetochores in higher eukaryotes, however this has not been shown in *S. cerevisiae*. While the ORC complex is involved in silencing at the mating type loci and the telomeres, the mechanisms governing the silencing of the centromeres do not utilize the ORC complex. Additionally, the recruitment of Orc6 is not dependent upon the presence of a *S. cerevisiae* ACS. Similarly the increased

presence of ORC at the centromeres in the chimeric strain cannot be attributed to a *K. lactis* ACS. While a subset of the centromeres do have a possible *K. lactis* ACS nearby this is not true for all of the centromeres. Furthermore the peaks do not necessarily center around the *K. lactis* ACSs, but rather around the centromeres. Additionally, the sizes of the peaks are not proportionate to the strength of the *K. lactis* ACS as would be expected if the ACSs were recruiting the ORC. There are a few possible explanations for these centromeric peaks. One possibility is that there is a loss of silencing at the centromeres which results in the DNA being more easily ChIPed leading to a higher signal. This loss of silencing could be due to competitive association of the chimeric ORC complex at the centromere, similar to how silencing is lost at the HMR-E region. However, the reason for this loss of silencing is not clear, as silencing at centromeres is not ORC dependent in *S. cerevisiae* Another possibility is that the chimeric ORC binds to the CEN sequences found at all centromeres, but in a manner that does not recruit the replication machinery. If it is the case that the *K. lactis* Orc4 also has a binding affinity for centromeric sequences, but does not initiate replication, then these sequences would not have been found in the initial screen of *K. lactis* ACSs. While this remains unknown, if the *K. lactis* ORC associates in such a manner, it may inhibit the establishment of silencing at the centromeres by preventing the normal association of the silencing machinery.

Evolution of ORC specificity for differing ARS sequences

It is evident that the sequence specificity of the ORC complexes in *S. cerevisiae* and *K. lactis* are very different, and that each will only initiate replication at their own ACS. While we have been able to determine that the C-terminus of the Orc4 subunit is the primary determinant of the ORC specificity, it is not clear how the difference in recognition evolved. I will explore the possible circumstances of how this evolution may have occurred.

Through comparisons of the amino acid sequence of the Orc4 of multiple yeast species, we identified two key regions which contain significant divergence between *S. cerevisiae* and *K. lactis*, but are largely conserved between *S. cerevisiae* and *S. bayanus* which share the same ACS (Liachko et al, 2010). We hope that further structural study will determine which specific residues are important for the interaction with the DNA at the ACS, similarly to how these residues were identified for archeal ORC proteins (Guadier et al 2007).

Even without knowing which residues are important, the ability of the chimeric ORC4 subunits to initiate replication is informative as to the structure of the ORC complex. This result tells us that the WHD and AAA+ domains can be altered quite drastically without interfering with protein-protein interactions with the other subunits. Additionally, while the chimeric Orc4 cannot substitute for the endogenous Orc4 it does not prevent the *S. cerevisiae* ORC complex from functioning in a diploid. This lack of interference is interesting as it suggests that in a diploid or polyploid cell, one copy of the *ORC4* gene would be free to evolve to bind to other distinct sequences while the cell would be maintained by other conserved copies. As most yeast species found in the wild are polyploid, this seems quite likely (Albertin W and Marullo P., 2012). Given this opportunity to evolve and the fact that the C-terminus of the Orc4 protein can be changed quite drastically without interfering with its association with the rest of the ORC complex, this would be a likely manner for the differing ORC sequence specificities to evolve.

While this hypothesis provides a possible mechanism by which the *K. lactis* ORC could evolve to recognize a different ACS, it is not clear why *K. lactis* has diverged so significantly from the other yeast species for which the ACSs are known. Our analysis of the amino acid changes between *S. cerevisiae, K. lactis,* and several other yeast species, shows that the ORC complex has changed significantly in *K. lactis* compared to other species. Interestingly however,

this rapid evolution is specific to the ORC complex, as other genes involved in replication do not show a similar pattern (Richard et al, 2005). These two results would suggest that the *K. lactis* ORC complex has evolved faster, but the replication machinery as a whole has not, and that these changes are likely related to the change in specificity of the ORC complex. It seems possible that once the ORC complex was able to initiate replication at the *K. lactis* ACS, selection pressures may have resulted in a number of changes to the complex as a whole to achieve a greater level of stability.

Figure 4.2.

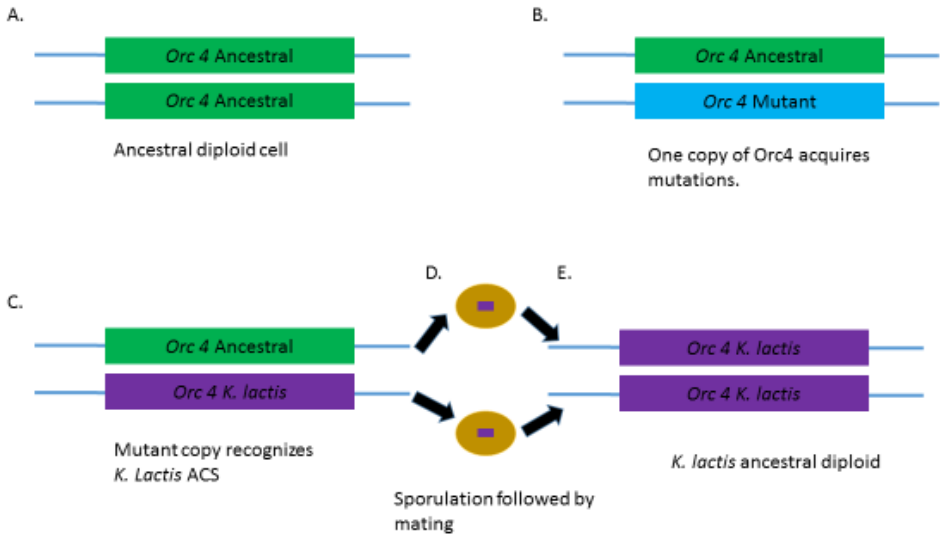

Fig4.2. The genomes of yeast in the wild are generally polyploid. In this figure the genome will be shown as diploid for simplicity.

A. The genome of the ancestral yeast species prior to the divergence of *K. lactis*. All copies of the Orc4 gene are the WT state that recognize the ARS of the ancestral species. B. Mutations can arise which result in a copy of the *ORC4* gene which no longer results in a protein that recognizes the ancestral ACS. The multiple copies of *ORC4* allow for maintenance of the genome as long as one copy remains functional. C. The mutant copy (or copies), no longer under selection, can mutate further until one copy recognizes the *K. lactis* ACS. Both copies may be functional, provided enough copies of the *K. lactis* ACS are present in the genome. D. The cell containing both copies sporulates and haploid cells containing the *K. lactis ORC4* mate. E. A diploid cell in which both copies of Orc4 recognize the *K. lactis* ACS resulting in a cell that is the ancestor to *K. lactis*.

Works Cited

Gaudier, M. et al. “Structural Basis of DNA Replication Origin Recognition by an ORC Protein” Science (2007) Vol 317: 1213-6

Dueber, E.C. et al. “Molecular Determinants of Origin Discrimination by Orc1 Initiators in Archaea” Nucleic Acids Research (2011) 39(9) 3621-31

Bolon, Y.T. and Bielinsky, A.K. “The Spatial Arrangement of ORC Binding Modules Determines the Functionality of Replication Origins in Budding Yeast” Nucleic Acids Research (2006);34(18):5069-80

Zou, Y. et al. “Asymmetric Positioning of Nucleosomes and Directional Establishment of Transcriptionally Silent Chromatin by *Saccharomyces cerevisiae* Silencers” Molecular and Cellular Biology ORC. (2006) p. 7806-7819

Prasanth, S. G. et al. “Orc6 Involved in DNA Replication, Chromosome Segregation, and Cytokinesis” Science (2002) Vol. 297 no. 5583 1026-1031

Liachko, I., et al. "A Comprehensive Genome-Wide Map of Autonomously Replicating Sequences in a Naive Genome." PLoS genetics 6.5 (2010): e1000946

Albertin, W and Marullo, P. “Polyploidy in fungi: evolution after whole-genome duplication” Proceedings. Biological sciences/ The Royal Scociety (2012) Jul 7; 279(1738:2497-509)

Richard, G. F. et al. 2005 “Comparative Genomics of Hemiascomycete Yeasts: Genes Involved in DNA Replication, Repair, and Recombination Mol Biol Evol (2005) 22 (4):1011-1023

Chapter 5: Perspectives:

We have shown that ORC4 is the subunit that determines the binding specificity of the ORC complex. Furthermore we have shown that the winged helix domain is the primary determinant of this specificity, however the AAA+ domain is necessary for efficient binding and maintaining function. While these functional studies have been extremely informative, there are further structural and biochemical studies that will help with further understanding of the manner in which the ORC complex interacts with the DNA at origins.

Biochemical analysis of the DNA binding of purified ORC

Further steps can be taken through a biochemical approach to study the DNA binding specificities of the WT *S. cerevisiae* ORC complex, WT *K. lactis* ORC complex, as well as *S. cerevisiae* ORC complexes containing ORC4 chimeras. The ORC complex can be purified from a strain that overexpresses all six ORC subunits. The ORC4 subunits in each could be tagged for purification, perhaps by using a TEV-3FLAG tag. The purified complex could then be treated with TEV protease, removing the 3FLAG tag, in order to prevent any possible interference with the complex's interaction with DNA. The ORC complex could then be fractionated by glycerol gradient sedimentation through which fractions containing the whole ORC complex can be recovered for ORC-DNA binding assays.

The binding specificity of the *S. cerevisiae*, *K. lactis*, ORC4-207, and ORC4-281 ORC complexes could be examined by use of gel mobility-shift assays. Additionally *in vitro* pull down assays for the loading of ORC onto DNA using DNA coupled to streptavidin magnetic beads could also prove informative. Each of these four ORC complexes could be tested on DNA

which contains matches to either the *S. cerevisiae* ACS, the *K. lactis* ACS, or both. Nonspecific competitor DNA could be used to eliminate any nonspecific interactions of the ORC complex with DNA.

Structural analysis of the WHD and AAA+ domains in DNA binding

The ORC4 WHD appears to be largely responsible for the binding specificity of the ORC complex, which likely means that it is directly interacting with the DNA at the ACS. The crystal structure of the *Drosophila* ORC has just been published after this thesis was completed. It is believed that Drosophila ORC has little DNA binding specificity and may have selectivity for negatively supercoiled DNA topology. It may be interesting to compare the 3-D structure of Drosophila ORC, the *S. cerevisiae* ORC and the chimeric ORC in future studies to gain more insight about the DNA binding specificity of some of the fungal ORC. NMR structural determination is an attractive method for determining the structure and dynamics of the WHD and AAA+ domains as they make contact with the DNA. Previous NMR studies on other DNA replication proteins have success with high resolution structures of the WHD domain of Mcm6 interacting with CDT1 (Wei et al., 2010). Determining the structure *S. cerevisiae* and *K. lactis* ORC4 as they interact with their ARS sequences would allow us to determine which residues contact which base in the DNA to establish DNA binding specificity and stability of the ORC-DNA complex.

ChIP Comparisons of ORC4-207 and ORC4-281 Strains

While the ChIP experiment carried out highlights important differences between the ORC4-207 and the WT strains, especially as seen at the HMR-E locus and the centromeres, the

experiment utilized a tag on the ORC6 subunit and therefore did not directly probe for the presence of the chimeric subunit. By directly examining the binding of the ORC4 chimera we would gain a better understanding of where the chimera binds and why it is unable to substitute for the *S. cerevisiae* ORC4. Furthermore, by comparing the ORC4-207 and ORC4-281 chimeras in a genome wide context we can gain a greater understanding of how the AAA+ domain affects the binding of ORC to DNA. If there is a difference in the locations to which each chimera binds it would be very informative as to how much of the specificity is conferred by the AAA+ domain or if the complex binds to only the best matches to the *K. lactis* ACS without the *K. lactis* AAA+ domain to provide the necessary stability.

Significance Beyond Yeast

While the mechanics that determine the binding specificity of ORC in *S. cerevisiae* and *K. lactis* are likely similar in other simple eukaryotes and archea; it is difficult to say how these findings inform us to the mechanics of metazoans. Metazoans lack the sequence specificity found in *S. cerevisiae* and there is some evidence that metazoans use additional proteins to recruit ORC to their origins (Moriyama et al 2012, Dhar et al, 2001). However, it has not been ruled out that DNA binding specificity plays a minor role in ORC recruitment or a small subset of origins are selected based sequence specificities in metazoans. Our studies have allowed us a better understanding of how the ORC complex binds to DNA and additional structural studies will further elucidate exactly how and which parts of the complex are important for this binding. Even if the residues themselves are not well conserved in more complex metazoans, the overall structure will still likely be informative.

We generated a total genome map of the binding of the ORC complex. In addition to a better understanding of global ORC binding, we found several novel features such as the binding of the ORC complex to centromeres. These results may be informative in the role that ORC binding plays in the establishment of centromeres in a general context. Additionally, knowing how two ORC complexes compete for binding may provide insight into how metazoan ORCs have evolved to bind at their origins.

Works Cited:

Zhun Wei, et al. “Characterization and Structure Determination of the Cdt1 Binding Domain of Human Minichromosome Maintenance (Mcm) 6” The Journal of Biological Chemistry (2010) Vol. 281, No 17, pp 12469-12473

Moriyama, K., et al. “Epstein-Barr nuclear antigen 1 (EBNA1)-dependent recruitment of origin recognition complex (Orc) on oriP of Epstein-Barr virus with purified proteins: stimulation by Cdc6 through its direct interaction with EBNA1” J Biol Chem 2012 Jul 6;287(28):23977-94

Dhar, S.K., et al.” Replication from oriP of Epstein-Barr virus requires human ORC and is inhibited by geminin” Cell (2001) Aug 10;106(3):287-96.