TITLE:

Template switching between the leading and lagging strands at replication forks generates inverted copy number variants through hairpin-capped extrachromosomal DNA.

SHORT TITLE:

Template switching mechanism for generating inverted triplications.

AUTHORS:


* co-first authors

1 corresponding author

AFFILIATIONS:

Department of Genome Sciences, University of Washington, Box 355065, Seattle, WA 98195
ABSTRACT:

Inherited and de novo copy number variants (CNVs) are increasingly found to be correlated with human developmental and cancerous phenotypes. Several models for template switching during replication have been proposed to explain the generation of these gross chromosomal rearrangements. We proposed a model of template switching (ODIRA—origin dependent inverted repeat amplification) in which simultaneous ligation of the leading and lagging strands at diverging replication forks could generate segmental inverted triplications through an extrachromosomal inverted circular intermediate. Here, we created a genetic assay using split-ura3 cassettes to trap the proposed inverted intermediate. However, instead of recovering circular inverted intermediates, we found inverted linear chromosomal fragments ending in native telomeres—suggesting that a template switch had occurred at the centromere-proximal fork of a replication bubble. As telomeric inverted hairpin fragments can also be created through double strand breaks we tested whether replication errors or repair of double stranded DNA breaks were the most likely initiating event. The results from CRISPR/Cas9 cleavage experiments and growth in the replication inhibitor hydroxyurea indicate that it is a replication error, not a double stranded break that creates the inverted junctions. Since inverted amplicons of the SUL1 gene occur during long-term growth in sulfate-limited chemostats, we sequenced evolved populations to look for evidence of linear intermediates formed by an error in replication. All of the data are compatible with a two-step version of the ODIRA model in which sequential template switching at short inverted repeats between the leading and lagging strands at replication forks generates inverted interstitial triplications.
AUTHOR SUMMARY

Chromosomal rearrangements are a potent source of genetic variation in humans and other organisms. One specific type of rearrangement involves the increase in copies of segments of the genome. The variation in gene dosage that these rearrangements can cause has been associated with a wide range of neurological and other human disorders. A specific puzzling form of copy number increase consists of three tandem copies with the central copy in inverted orientation. How this rearrangement occurs is of great interest, yet the mechanisms responsible are only inferred by examining the sequence of final inverted products. Yeast provides a unique model system to explore the underlying molecular defects that give rise to inverted triplications. While the favored hypothesis suggests that double stranded DNA repair is the causative agent, we find that a particular form of template switching between strands at the replication fork, not a double stranded DNA break, is the initiating event. Using the awesome power of yeast genetics, we provide evidence in two different assays for this unique replication error that we call ODIRA (for Origin Dependent Inverted Repeat Amplification) and propose that it can also explain this form of copy number variant seen in human evolution and disease.
INTRODUCTION:

Copy number variation (CNV) refers to both increases and decreases in copies of genomic segments. In humans, many CNVs not only distinguish us from our close primate relatives, but some arise de novo and are associated with a range of human disorders [1-5]. One of the most common forms of CNV found in the human genome is the repetition of large genomic segments (referred to collectively as segmental duplications). Although the extra copies can be found as a direct repeat at the original locus, they may also be found at dispersed sites on the same or different chromosomes [6]. There are three major well characterized pathways that are thought to give rise to changes in copy number through distant interactions: non-allelic homologous recombination (NAHR), non-homologous end joining (NHEJ) and template switching during replication (FoSTeS and MMBIR) (Reviews: [7-12]). However, these mechanisms fail to easily and/or completely explain the formation of an unusual form of segmental duplication that involves a tandem triplication, where the central copy is inverted within an otherwise unrearranged chromosome. This form of CNV is seen in many human disorders and is likely under-reported because determining its inverted structure is technically challenging [8, 13, 14].

During long-term growth of the common laboratory strains of haploid budding yeast *Saccharomyces cerevisiae* in chemostats limiting for sulfate, we routinely recover inverted triplications of the *SUL1* locus, which encodes the primary sulfate transporter [15, 16]. The reproducibility of this outcome provides an ideal system and opportunity to investigate the mechanism that gives rise to this form of gene amplification. Although the size of the
amplified region varies, several structural features appear to be invariant (Figure 1A.1): 1) the amplified segment contains at least one origin of replication, 2) the junctions that mark the boundaries of the amplified segment occur at pre-existing short, interrupted inverted repeats, and 3) the arms of the inverted repeats used for amplification are within a hundred base pairs of each other [15-17]. We have proposed a unique template-switching model, called ODIRA (Origin-Dependent Inverted Repeat Amplification; [18]), in which the leading strands at divergent, regressed replication forks become ligated to the Okazaki fragments on the lagging strands (Figure 1A.2; “dogbone” ODIRA). Displacement and replication of this aberrant structure gives rise to an inverted, extrachromosomal, circular intermediate containing $SUL1$ and the adjacent origin of replication (Figure 1A.3-4). Subsequent integration into the chromosome at the original locus generates the triplication with an inverted center copy without disturbing the distal chromosomal sequences (Figure 1A.5).

A same amplicon can arise in a two-step ODIRA mechanism (Figure 1B.1; “hairpin” ODIRA) where the two template switches are temporally uncoupled. Ligation of leading and lagging strands at just the single fork moving toward the centromere (Cen-proximal junction; CJ) could produce an intermediate consisting of a hairpin capped linear segment extending to the telomere that could persist as an extrachromosomal palindromic linear duplex after replication (Figure 1B.2-3). While recombination with the initiating chromosome does not result in integration—but rather just shuffles the telomeric arms between the palindromic linear sequence and the original chromosome arm—cells containing the linear segment would enjoy a selective advantage in the sulfate-limited
In a subsequent cell cycle, the second replication error in the fork moving toward the telomere (Tel-proximal junction; TJ; Figure 1B.4-6) would generate a doubly inverted linear that could recombine with the \textit{SUL1} chromosome and generate the inverted triplication, improving its stability and selective advantage that leads to this clone sweeping the population.

Two other models of template switching—FoSTeS and MMBIR—have been proposed to account for complex chromosomal rearrangements involving distant interactions [9, 10]. These models involve the migration of the 3’ end of a nascent strand from a replication fork or the exposed 3’end from a double stranded break to other regions of homology elsewhere in the genome. At the new site, replication is reestablished, generating a junction between two disparate regions of the genome. To explain complex rearrangements, the model proposes that the same strand makes multiple sequential invasion/extension attempts in a single cell cycle. Inverted triplications would not require long-distance template switching since the homologous template is the opposite strand at the replication fork (Figure 1A.2; Figure 1B.1).

It is the nature of the inverted junction that inspired the ODIRA model. One well characterized example from the human literature seemed to fit perfectly with ODIRA because the triplication that occurred in the father of the female proband was a 2:1 mixture of SNPs from his two homologues [19]. An inverted dimeric circle that arose from one homologue and inserted into the other homologue during or preceding meiosis could be the explanation for the 2:1 SNP ratio. This case provided us with the stimulus to artificially
recreate such an event in yeast by asking whether we could detect the movement of the inverted dimeric circle (produced by template switching across the replication fork) to a new location in the yeast genome. We designed a split-ura3 construct to identify the movement of potential extrachromosomal intermediates from one location to another by the recreation of a functional URA3 gene from two overlapping partial ura3 fragments.

To capture potential inverted extrachromosomal intermediates, we integrated a 5' fragment of URA3 (“ura”) on chromosome II at the SUL1 locus, a region that is prone to inverted triplications in sulfate-limited cultures. On chromosome IX we integrated the overlapping partially complementary “ra3” fragment. Direct recombination between the homologous regions on chromosome II and IX would produce unstable dicentric chromosomes; however, an ODIRA event—a circular dimeric inverted ura fragment integrating into chromosome IX—would recreate a functional URA3 gene along with a partial inverted triplication on chromosome IX. Analysis of Ura+ clones from the split-URA3 strain, using pulse field gels, Southern blotting, microarrays, and sequencing across potential junctions, allowed us to distinguish the inverted amplicons from other chromosomal rearrangements. Among the hundreds of clones analyzed we failed to detect any Ura+ clones that were consistent with integration of circular intermediates into chromosome IX. Instead, ~20% of the URA3 clones showed genomic rearrangements that could be explained by a recombination between an inverted linear ura fragment from chromosome II with the ra3 fragment on chromosome IX (ODIRA hairpin model).
Creation of identical linear extrachromosomal intermediates could be explained by repair of a double stranded DNA (dsDNA) break that forms intrastrand hairpins at short, interrupted palindromes near the break. To distinguish between these two mechanisms we carried out the selection for Ura+ clones under two conditions: 1) altering the strand dynamics at replication forks by reducing the availability of dNTPs (growth in 50 and 200 mM hydroxyurea) and 2) increasing dsDNA breaks by targeting CRISPR/Cas9 centromere proximal to the SUL1 locus. Reducing dNTPs reduces the formation of inverted URA3 clones by more than 8 fold. And, instead of increasing the frequency of inverted URA3 clones, inducing dsDNA breaks centromere-proximal to SUL1 reduced the URA3 clone frequency to zero. Together these results suggest that double-stranded DNA breaks are not part of the pathway for inverted amplification.

To determine whether the hairpin ODIRA model could explain the occurrence of inverted triplication of the SUL1 locus in sulfate-limited chemostats [15-17], we sequenced 31 independent populations of cells that had been passaged for ~250 generations. By focusing on split reads that define junctions of inversions, we discovered that nearly a third of the cultures had different numbers of Cen-proximal and Tel-proximal inverted junctions, suggesting that there were inverted linear segments produced during the experiments. Moreover, the spacing of short-inverted repeats, the orientation of the inverted junctions and the positions of the junctions with respect to the replication map of this region of chromosome II provide further evidence that template switching between strands at a replication fork, not double strand DNA repair, is the mechanism for inverted gene amplification in yeast.
MATERIALS and METHODS:

Yeast strains and culture conditions: BY4741 (MATa, his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) and BY4742 (MATα, his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0) were used to construct the split-ura3 strain. The two partially overlapping regions of the URA3 gene, referred to as ura and ra3, were generated in two steps. We first selected (on -uracil plates) for URA3 insertion on chromosome II and chromosome IX, respectively into BY4741 and BY4742, by transformation with a PCR fragment derived from pRS406 with 100 bp homology arms (Table S1; SUL1_URA3_F and SUL1_URA3_R; Chr9_URA3_Chr9_F and Chr9_URA3_Chr9_R). These strains were then transformed with truncated PCR fragments of ura and ra3, similarly created from pRS406 and primers with the same homology arms (Table S1; SUL_URA3_F and ura_SUL1_R; Chr9_ra3_F and Chr9_URA3_Chr9_R). Transformants were selected on plates with 5-fluoro orotic acid (5-FOA) and confirmed by PCR/Sanger sequencing and Southern blotting of CHEF gels. BY4741 containing the ura fragment was mated to BY4742 containing the ra3 fragment to create the doubly heterozygous diploid. Sporulation and tetrads dissection resulted in a haploid spore, s2-1 (MATa, ura3Δ, his3Δ1, leu2Δ0, lys2Δ0; hereafter referred to as the “split-ura3 strain”; Figure 2A), with both partial ura3 fragments that were confirmed by CHEF gel/Southern blotting. This strain was used in all subsequent experiments involving selection for Ura+ clones. The ura insert lies between coordinates 789418 and 791405 on chromosome II on the Watson strand. The ra3 fragment lies in an intergenic region between coordinates 321188 and 321194 on chromosome IX, also on the Watson strand ~34 kb upstream of CEN9 (355629-355745). The overlap between ura and ra3 is 203 bp.
FY4 (MATa, srd1Δ0) was grown in mini-chemostats limiting for either sulfate (31 independent chemostats) or glucose (32 independent chemostats) for ~250 generations [20, 21].

Selection for Ura+ clones: Small independent colonies of the split-ura3 strain were picked from a synthetic complete plate with 5FOA and inoculated in 1 ml of complete synthetic yeast medium and grown to a saturation density of ~10⁸ cells. The 1 ml of cells was concentrated by centrifugation, plated onto a single -uracil plate, and incubated at 30°C for 3-5 days. A maximum of two colonies from each -uracil plate were restreaked on a fresh -uracil plate. Each purified clone was grown in 8 ml of -uracil liquid medium to make freezer stocks, CHEF-gel plugs and “NIB-n-grab” DNA preps (https://fangman-brewer.genetics.washington.edu/nib-n-grab.html; see below). To test for the effect of nucleotide depletion on the formation of Ura+ clones, the 1 ml of medium contained either 50 or 200 mM HU.

Contour-clamped Homogeneous Electric Field (CHEF) gel electrophoresis/Southern blotting: Agarose plugs for CHEF gel electrophoresis were generated using the method by L. Argueso (described in [22]) or by an adaptation of the method by S. Iadonato and A. Gnirke (described in [23]). Run conditions in the BioRad CHEF-DRII were 0.8% LE agarose in 0.5XTBE in 2.3 L 0.5% TBE running buffer at 14°C. Switch times were 47” to 170” at 165V for 39-62 hours. Standard conditions for Southern blotting and hybridization with
$^{32}$P-labeled PCR probes are described by Tsuchiyama et al. [23]. Hybridization intensity was determined using a BioRad Personal Molecular Imager.

CRISPR/Cas9 cutting in the vicinity of $SUL1$: The split-ura3 strain was transformed with plasmid pYCpGal that include a yeast centromere, the yeast $LEU2$ gene, the $GAL1$ promoter driving Cas9 expression, and a guide RNA cloning site. The Cas9-guide cassette was derived from plasmid pML104 (Addgene) (Figure S1). Guide RNAs expressed from this plasmid directed cutting to either position 708.260 kb (pYCpGAL-708b) or 792.883 kb (pYCpGAL-792b) on chromosome II (Table S2). Relative to $SUL1$, the sites are centromere-proximal and centromere-distal, respectively. Standard LiAc transformation was used to introduce a no-guide plasmid, the 708 plasmid or the 792 plasmid into the split-ura3 strain, selecting for transformants on -leucine, +glucose plates. Single colonies were then used to inoculate 1 mL of –leucine + glucose medium. The 1 ml of cells was concentrated and plated on -uracil, -leucine, +raffinose, +galactose to induce cutting by CRISPR/Cas9 and to select for uracil prototrophy. After restreaking colonies on -uracil plates, clones were expanded in liquid –uracil medium for freezer stocks and CHEF-gel plugs.

PCR/Sanger Sequencing: The insertions in the split-ura3 strain were confirmed by PCR and Sanger sequencing, in addition to Southern blotting. The state of CRISPR/Cas9 cutting at the 708 site or the 792 site was confirmed by PCR. Sanger Sequencing of PCR fragments was performed by GeneWiz (Azenta) or Eurofins.
aCGH (array Comparative Genome Hybridization) analysis: Genomic DNA from frozen chemostat samples from generation ~250 was isolated by the NIB-n-Grab protocol (see above), a modified version of the Smash-and-Grab protocol [24] that results in the recovery of DNA 20-50 kb in size. In this protocol, cells are broken by vortexing with glass beads in a buffer that stabilizes nuclei (NIB; [25]). aCGH was performed using Agilent 4x44k microarrays with probes spaced every 290 nt on average. Hybridization was executed as described previously [20]; however, sonication of the DNA samples was performed after *in vitro* labeling, rather than before labeling. This alteration allows inverted junctions to be identified by the gradual increase in signal—from single copy to multiple copies—at the site of inversion (see Figure S2).

Note: Interstitial inverted triplications in the human CNV literature are referred to as a “triplicated segment embedded in an inverted orientation between two duplicated sequences (DUP-TRP/INV-DUP)” [26]. They are found associated with a variety of genetic syndromes but remain an underappreciated form of CNV, primarily because the inverted nature of the amplicon junctions poses a challenge for DNA sequencing platforms [14]. While arrayCGH has largely been replaced by long read sequencing in genomic research there are inherent problems with nanopore sequencing of inverted templates [27]. We suggest that a modified protocol of aCGH can easily detect inverted boundaries of amplified regions.

Population short read sequencing: 150 bp paired-end sequence of whole genome fragments purified from population samples on the last day of the chemostat run were
prepared and analyzed as described [16]. Median genome read depth for the 31 sulfate-limited chemostats was ~125. Read depth analysis for the sulfate-limited cultures showed amplification of the *SUL1* gene and flanking sequences while none of the 32 glucose-limited chemostats had amplifications in the *SUL1* region.

**Split read analysis:** Split reads are defined as 150 bp reads that map to two non-contiguous regions of the yeast genome [16] and were manually curated on IGV (Integrative Genomics Viewer; [28]) as falling into one of five categories: 1) inverted junctions have reads divided between two sections of chromosome II on opposite DNA strands (n=92); 2) direct repeat junctions have the two segments of the sequence from the same strand of chromosome II (n=4); 3) de novo telomere additions contain part of a chromosome II sequence adjacent to a C<sub>T</sub>A/G<sub>T</sub>A telomere sequence (n=0); 4) telomere translocation junctions have the terminal *SUL1* part of chromosome II-R attached to an existing telomere (n=6); and 5) internal translocation junctions join chromosome II sequence to a repetitive element elsewhere in the genome—such as Ty elements, solo delta elements, tRNAs, and polyA/polyT or CAG stretches (unquantified). Inverted junction sequences (category 1) were catalogued (Table S3) and characterized by the size of the inverted sequence at the discontinuity, the spacing between the two inverted sequences, and orientation of the inversion event. To establish baselines for all available inverted repeats we used the EMBOSS Palindrome program (https://www.bioinformatics.nl/cgi-bin/emboss/palindrome) to find all potential inverted repeats and their spacing across the terminal ~80 kb of Chromosome II with an interrupted inverted repeat structure of ≤250 bp using the sacCer3 version of the genome.
For an inverted split-read junction to be considered significant, we required that at least two independent chromosome fragments from the same culture produce the same junction sequence. Inverted junctions represented by a single read were considered to be PCR artifacts produced during the sequencing protocol, while those represented by more than one unique read were copies of in vivo generated inversions.

Statistical Analysis: The significance of differences in the frequency of inverted URA3 amplicons in normal medium vs. medium containing hydroxyurea (HU, 50 or 200 mM) was assessed by Chi-squared analysis. The significance in inverted repeat length and orientation of inverted junctions among split reads from whole genome sequencing of population samples from sulfate-limiting chemostats was performed using the Mann-Whitney Rank Sum test.
RESULTS and DISCUSSION:

Genetic selection and identification of stabilized ODIRA-generated intermediates

One key feature of both versions of the ODIRA model relies on the presence of extrachromosomal intermediates that reintegrate into the chromosome through homology-dependent recombination to give rise to the inverted amplicons. To determine whether such intermediates exist, we designed a system in which homology for the integration event is at a new site where it generates a selectable phenotype. We constructed the haploid strain s2-1 (hereafter referred to as “the split-ura3 strain”) in which partially overlapping URA3 gene fragments, with a shared central region of identity, were integrated on two different chromosomes (Figure 2A). We replaced a section of the SUL1 gene on chromosome II with the 5’ portion of the URA3 gene (“ura”) in the same transcriptional orientation as SUL1. We inserted the overlapping 3’ portion (“ra3”) in a non-essential site on chromosome IX to the left of CEN9 with a transcriptional orientation toward CEN9 (Figure 2A). The design of our system allows for the reconstitution of the URA3 gene through at least two different mechanisms. One way involves integration of intermediates from either of the two ODIRA models. The linear or “dogbone” intermediates, carrying the “ura” sequence, can recombine with the ra3 sequence on chromosome IX to generate Ura+ prototrophs. A second way of reconstituting the URA3 gene is through a direct recombination event between the two chromosomes within the region of ra homology. This reconstitution, however, gives rise to an unstable dicentric chromosome and a reciprocal acentric chromosome that is lost in subsequent cell divisions (Figure 2B). Our
experimental set-up permits us to detect the insertion of the inverted dimeric circular intermediate onto chromosome IX as well as a recombination event with the inverted palindromic linear (Figure 2C,D). Both recombination events would be mitotically stable and their chromosome structures would be easily distinguishable by CHEF gel electrophoresis and Southern blotting from each other and from events that resulted from direct recombination (Figure 2E).

Starting with individual colonies of the split-ura3 strain grown on 5-FOA (to ensure that they were phenotypically uracil auxotrophs), we grew cultures to stationary phase in 1 ml of liquid complete medium before plating ~10^8 cells on plates lacking uracil (-ura plate). Each expanded colony generated at least one Ura+ clone. In the initial experiments we chose one or two colonies from each -uracil plate (27 unique events) and analyzed their chromosomes by CHEF gels and Southern blots (Figure 3A). If recombination between the two chromosomes generated the URA3 gene, then a single copy of the 5'SUL1 sequence would be on the new, unstable, dicentric chromosome. Because the left arm of chromosome IX has no centromere it would be lost, and cell survival would depend on the retention of an unrearranged copy of chromosome IX. Depending on the secondary rearrangements of the dicentric chromosome, essential genes from chromosome II might be lost as well. Therefore, we expect that these cells might also retain an intact copy of chromosome II (Figure 2B,E). If an inverted circular intermediate integrated into chromosome IX to generate a Ura+ clone, the only change in karyotype would be an increase in the size of chromosome IX and a 2:1 ratio of the 5'SUL1 probe relative to chromosome II (Figure 2C,E). Finally, if an inverted linear were to recombine with the ra3
sequences on chromosome IX (Figure 2D), then a new chromosome would be created that has two copies of $5'SUL1$ probe (also a 2:1 ratio of hybridization signal) and the right telomeres from chromosomes II and IX (Figure 2D,E). The reciprocal product, being acentric, would be lost and thus necessitate the retention of an unrearranged copy of chromosome IX.

Among the 27 Ura+ clones we examined in our initial experiment, five clones (18%) had a 2:1 ratio of the $5'SUL1$ sequences on the altered chromosome (Figure 3A). In each case, unrearranged copies of chromosomes II and IX remained, and the new chromosome had the centromere from chromosome IX and two copies of the $ura$ sequence from chromosome II with variable amounts of centromere proximal DNA. These observations were consistent with a hairpin linear ODIRA recombination (Figure 2D) in which the altered chromosomes were generated by recombination between a copy of chromosome IX and an inverted linear intermediate derived from chromosome II. To confirm the sequence arrangement of the new chromosomes in these five strains, we performed aCGH to determine the identity of the genetic material on the new chromosomes (an example is shown in Figure 3B). The most parsimonious assembly of the extra copies of segments from chromosomes II and IX (Figure 3C) is consistent with the new chromosome's size (Figure 3D). CHEF gel analysis and Southern hybridization followed by aCGH (Tables S4,S5) allowed us to determine the structure of the other 22 Ura+ clones, three of which are shown in Figures S3,4,5 and summarized in Figure S6. Each of the remaining 22 clones had structures that were consistent with the rearranged products of dicentrics that were produced by direct recombination at the $ra$ sequences on chromosomes II and IX (Figure
followed by centromere deletion (n=7; example in Figure S4), telomere capture either by BIR (n=6) or de novo telomere addition (n=6; example in Figure S5), or breakage that initiated a bridge-breakage-fusion (BBF) cycle (n=3; example in Figure S6). We failed to recover any cases where chromosome II had the internal inverted triplication that would result from integration of a circular, inverted intermediate. However, the recovery of five clones with inverted chromosome II sequences appended to chromosome IX were consistent with the linear hairpin ODIRA model.

After establishing the patterns of ODIRA-related events via aCGH and Southern blotting, we screened an additional 23 Ura+ clones for ones in which the 5'SUL1 probe on the new chromosome was in a ratio of 2:1 (relative to the signal on chromosome II). We recovered six additional clones for a total of 11 of 50 Ura+ clones. ArrayCGH analysis of each of these clones (Figure S3) revealed the same pattern of copy number variation: 3 copies of chromosome II-right and 2 copies of chromosome IX-right, with discrete jumps in copy number at the ura sequence on chromosome II and the ra3 sequence on chromosome IX. The junctions to the left of SUL1 on chromosome II were at variable positions and did not show a discrete jump in copy number from one and three copies (Figure S3). Rather, copy number gradually increased over an approximately 10 kb window, producing a pattern we refer to as a “waterfall” in which the copy number gradually transitioned from 1 to 3 copies.

The “waterfall” pattern of gradual copy number change is consistent with the inverted nature of the junction and the protocol we used for labeling the aCGH samples with Cy dyes.
(see methods; Figure S2). Shearing of the genomic DNA before labeling removes the gradual transition in the aCGH profiles and converts the junction to a discontinuous one (Figure S2). None of the other junctions attributed to dicentric rearrangements showed this gradual transition in copy number (see examples in Figures S4,5,6). In this final set of 50 Ura+ clones with the 2:1 SUL1 hybridization ratio and the gradual copy number transition in the aCGH profiles, all were consistent with a recombination between an inverted linear acentric fragment derived from the SUL1 region of chromosome II and the ra3 locus on chromosome IX. If integration of a circular inverted intermediate occurs, it is below level of detection in our system, or is unique to the conditions in the sulfate-limiting chemostats.

Centromere proximal double strand breaks do not generate inverted hairpin intermediates

While our results with the split-ura3 strain are consistent with linear hairpin formation via a template switching event, there is a possibility is that the hairpin is generated by a dsDNA break rather than by a replication error. If a dsDNA break occurs within the vicinity of short inverted repeats, resection of the 5’ end and foldback of the exposed 3’ end could create a hairpin-capped linear molecule (Figure 4A). Gap filling and ligation would yield a structure identical to the hairpin we are proposing that occurs via template switching at the regressed replication fork. To distinguish between these two sources of the hairpin intermediate, we conducted CRISPR/Cas9 experiments on the split-ura3 strain to ask specifically whether break repair is involved in the creation of the linear intermediates. If
hairpins generated from rare spontaneous dsDNA breaks had contributed to the formation of the Ura+ clones we had observed in the split-ura3 strain, then induction of dsDNA breaks centromere proximal to SUL1 should increase the frequency of URA3 products with an inverted structure (Figure S8).

We targeted CRISPR/Cas9 to a site centromere-proximal to SUL1 (708 kb; Figure 4B). As a control, we also targeted a site distal (792 kb) to SUL1 (Figure 4B). Cutting at this distal site should greatly increase the frequency of direct recombination events (Figure 2B) while reducing or eliminating inverted-linear events (Figure 2D). To introduce these dsDNA breaks, we used LEU2 CEN3 plasmids with Cas9 under the control of the GAL1 promoter (Figure S1), selecting for transformants on –leu plates supplemented with glucose to initially repress Cas9 expression (Figure 4B top). We grew 20 individual colonies to saturation before plating on -uracil, -leucine, +galactose/raffinose plates to induce Cas9 and to select for uracil prototrophs (Figure 4B bottom). In the absence of a guide we recovered the typical one or two Ura+ colonies per plate (Figure 4B,C left) and among the independent Ura+ colonies we recovered six that produced the 2:1 SUL1 hybridization ratio expected for linear ODIRA events (Figure 4D). As expected, cutting at the distal, 792 site resulted in a great increase in Ura+ colonies (~700-fold; Figure 4B,C right). While we only analyzed 20 of these Ura+ clones by CHEF gel hybridization, none were inverted products (Figure 4D) and all were consistent with repair of the dsDNA break by direct recombination with the rα3 site on chromosome IX. In contrast, inducing cleavage at the proximal, 708 site completely eliminated the production of Ura+ colonies (Figure 4B,C middle). As there are abundant short inverted repeats immediately distal to the 708 site
where a foldback could have occurred, the lack of Ura+ clones following CEN-proximal cleavage indicates that proximal breaks do not generate Ura+ colonies with an inverted chromosome architecture.

Reducing dNTP levels through inhibition of ribonucleotide reductase interferes with the production of inverted Ura+ clones

The asymmetrical nature of DNA replication plays an important role in the ODIRA model: a single stranded gap on the lagging strand template provides a site for annealing of the leading strand when it contains a newly synthesized short interrupted inverted repeat. Altering the dynamics at the fork could therefore alter the frequency of recovered inverted Ura+ clones. Growth of yeast in hydroxyurea, a drug that hinders the expansion of deoxynucleotide pools [29] and thereby slows replication fork progression by ~16-fold [30], causes the expansion of the single stranded region at replication forks from ~200 nt to ~300 nt [31]. If the gap at forks in HU treated cells is on the lagging strand, then we might expect to see an increase in the frequency of inverted Ura+ clones. However, if the single stranded gaps switch to the leading strand, then we would expect the opposite result. Finding a change in the frequency of inverted Ura+ clones would support a role for replication fork dynamics in their formation.

We repeated the Ura+ selection after growth of individual populations of split-ura3 strain in the presence of 50 or 200 mM HU. We screened a total of 76 clones (43 clones from 50 mM HU and 33 clones from 200 mM HU) and recovered three with a 2:1 5'SUL1 ratio and
the “waterfall” aCGH pattern (clone HU++3 in Figure S3 was the single clone recovered from growth in 200 mM HU). Given that the expected frequency of inverted Ura+ clones is 0.22 (11 of 50 colonies from the no HU control set of clones), there is a significant decrease in inverted clones when cells are depleted for nucleotides by HU incubated (Chi-squared p-value < 0.001). These results show that altering replication fork dynamics alters the frequency of inverted amplicons and supports our hypothesis that template switching at replication forks plays a role in the formation of the Ura+ inverted clones. The fact that we see a decrease in frequency of inverted Ura+ clones also suggests that HU treatment preferentially affects leading strand synthesis.

Inverted junctions are commonly recovered after long-term growth in sulfate-limited chemostat cultures

The evidence we collected with the split-ura3 strain argues that the major form of ODIRA amplification is through a linear hairpin intermediate. Therefore, we wanted to ask if the same is true in sulfate-limited chemostats. In our previous analysis of SUL1 amplicons we characterized clones obtained after ~250 generations of selective growth in low-sulfate medium using aCGH to look for copy number changes, CHEF gel analysis to look for changes in chromosome sizes and to determine the junctions of the amplified SUL1 sequences, and Southern blots of restriction digests and snap-back assays to look for inverted DNA junctions [16]. Because we analyzed only a few clones from each chemostat, we were unlikely to identify the rare, unstable intermediates or to accurately assess the range of variants that might be present in the culture. To address these limitations, in this
current work we collected culture samples from the last day of growth (~250 generations) for 31 independent sulfate-limited chemostats and performed Illumina 150 bp paired end DNA sequencing on the population samples. After mapping the reads back to the yeast genome we assessed read depth to identify amplified segments (Figure S9). As a control, we subjected the same yeast strain to 32 independent glucose-limiting chemostats—where there is no selection for SUL1 amplification—to assess the rate of false junctions created by PCR artifacts in the sequencing protocol.

Read depth analysis for the sulfate-limited cultures showed amplification of the SUL1 gene and/or flanking sequences in all 31 populations (Figure S9). The major amplification products appear to be interstitial; however in five of the cultures we detected one or more amplification events that appeared to extend through SUL1 to the telomere (Figure S9, red arrow heads and squares). As expected, none of the 32 glucose-limited chemostats had amplifications of the SUL1 region.

To examine amplification junctions we searched for split reads flanking SUL1 where one of the 150 bp reads mapped to two non-contiguous sites in the genome. There were three major types of junctions that could be confidently mapped: 1) junctions that map to two sites on the same strand of DNA indicating they are involved in and mark the sites of tandem duplications (Figure S9, orange bars; chemostats S15, S19, and S23); 2) junctions between unique sequences and telomeric sequences that indicate that a terminal fragment of chromosome II had become appended to an existing telomere (Figure S9, red squares; chemostats S11, S19, S21, S28 and S29); and 3) inverted junctions that mark the limits of...
inverted amplicons (Figure S9, yellow circles; all chemostats with the exception of S19 and S28).

The vast majority of the split reads defined inverted junctions (Figure 5A,B); however, the depth of these reads varied widely. To determine a cutoff threshold for significance, we required that a junction be detected in two unique reads (with different ends) from the population being sampled. In part, this threshold was determined by the comparison to the glucose-limited culture split reads. While we found single reads from inverted junctions in the glucose-grown samples, there were no junctions supported by two independent reads. In addition there was a similar frequency of single-read inverted junctions in the glucose- and sulfate-limited chemostats (Tables S6,S7). The second feature we used to set the two-read threshold takes into consideration the orientation of the split reads with respect to the SUL1 locus. To be associated with inverted amplification of the SUL1 region, the orientation of the split reads is important: in the productive orientation (Figure 5C) the centromere-proximal junction (CJ) is oriented with the duplicated segment extending toward SUL1 and the telomeric end of the chromosome; likewise, the productive orientation for the telomere-proximal junction (TJ) is oriented with the duplicated region extending towards SUL1 and the centromere. The same sequences could also be involved in non-productive orientations (Figure 5C) but these would not be associated with SUL1 amplification. In a scan of all inverted junctions supported by sequence reads from two or more independent reads, there is a perfect concordance between their positions and orientations with respect to SUL1 (Figure 5D). In contrast, no inverted junctions supported by two or more were found in the glucose-limited cultures and inverted junctions with a
single read were oriented randomly with respect to *SUL1* (Figure S10A). In addition, comparisons of the sizes of the repeats and their spacing (Figure S10B,C) between split reads from the sulfate-limited populations that were found once and those that were found more than once show significant differences, validating our split-read threshold.

If inverted amplification occurs through sequential hairpin formation then we would expect to find cases where the numbers of centromere-proximal junctions do not match the number of telomere-proximal junctions (Figure S11A,B). Most populations contained multiple proximal and distal junctions; only three of the populations contained a single amplification event where the two oppositely oriented inverted junctions correspond with the edges of the amplified region (Figure S9, chemostats S9, S25 and S27). For these populations a single expanded clone with a specific inverted triplication could have resulted from either insertion of an inverted circular ODIRA intermediate or by recombination with a doubly inverted linear (Figure S11C). We tabulated the distribution of inverted junctions across the 31 sulfate-limited populations (Table S3) and found roughly a third of the populations had mismatched number of Cen- and Tel-proximal junctions (CJ and TJ; 11 of 31 cultures; Figure S11D). While we may have missed some junctions due to their rarity and limited sequence coverage, these results are consistent with the hypothesis that the left and right junctions do not occur simultaneously and favor the hairpin-ODIRA model proposed in Figure 1B.

**Genomic features at the sites of inverted junctions**
All of the inversion junctions occurred at preexisting interrupted inverted repeats in the genome. We assessed the distribution, sizes, and spacing of interrupted inverted repeats (≤250 bp) across the terminal 81 kb of chromosome II (Figure 6A,B). Potential inverted repeats are uniformly and surprisingly frequent in the genome, with their frequency inversely correlated with the size of the repeats (Figure 6A,C). For the 31 sulfate-limited cultures, when we compare the repeats recovered in 89 of the 92 inverted junctions that were identified as being “used” with their distribution and properties in this region of chromosome II (Figure 6C), we find that the longer the repeat, the more likely it will participate in an inversion event (Figure 6C,D). Spacing between repeats also determines which repeats result in inversion events (Figure 6B). Those with spacing between 40-80 bp are preferred while the genome distribution is relatively constant across all possible spacing intervals (Figure 6B). Three out of the 92 inverted junctions identified—two CJ and one TJ—had spacing of inverted repeats of greater than 1 kb and had likely undergone a secondary deletion of one of the inverted arms. Since the initiating site of inversion could not be deduced, they were omitted from the above analysis. Similar secondary deletions were also detected in two of the split-ura3 inverted junctions (I1 and CH1, Figure S3), and in both an artificial dogbone construct and in a previously characterized SUL1 inverted triplication [17].

Given that optimal properties for frequency, size and spacing of repeats were uniformly distributed across the terminal 81 kb of chromosome II, it was surprising that the position of productive CJs were concentrated in the ~20 kb centromere-proximal to SUL1 (Figure 6E). The clustering of CJs in the region between 770 and 790 kb (Figure 6E) cannot be
explained by any bias in the distribution of potential sites but is likely the consequence of some other feature of this chromosomal region. One possibility we considered is the presence of a gene or genes to the left of 770 kb that confer a selective disadvantage when included on the SUL1 amplicon. However, the work of Sunshine et al. [32], who examined the fitness cost of chromosomal fragments that include variable extents of SUL1-adjacent DNA rules out this possibility: the first significant drop-off in fitness only occurs when DNA centromere proximal of 535 kb is included in the amplicon.

As our model is based on a potential replication error, we examined data from a high-resolution genome-wide replication study to understand this regional selection for inverted junctions [33]. There are six confirmed potential origins (ARSs) in this region of the genome (oriDB) but replication studies indicate that two are predominantly used late in S phase to complete replication of the SUL1 region (ARS224 and ARS228; Figure 6F top). Bidirectional replication initiation from these two origins predicts that the region between 743 and 768 kb would be replicated rightward by a fork from ARS224 and, as a result, would be unable to generate the centromere-proximal fork error needed in our model to amplify the SUL1 region (striped arrow). Two additional origins (ARS225 and an unnamed ARS, uknARS; Figure 6F bottom) contribute to a lesser extent to replication of the region. Initiation at ARS225 reverses the direction of replication in the region between 752 and 758 kb and may be responsible for the four centromere-proximal junctions recovered in that region.
The discontinuous nature of replication on the lagging strand at replication forks may also explain the preferential recovery of inverted repeat spacing in SUL1 amplicons. The size of the gap on the lagging strand template at a replication fork is dynamic as the leading strand advances. However, based on the size of Okazaki fragments in yeast (165 bp; [34]), we can extrapolate that the lagging strand gaps could range from a minimum of 0 to 165 bp (or more). We propose that the range of inverted repeat spacings found at inverted junctions (between 40 and 80 bp; Figure 6B) is a direct result of the size of Okazaki gaps on the lagging strand. If the repeats are close together (<40 bp) it is likely that the regressed fork would result in both copies of the repeat re-annealing between the parental strands, leaving the exposed repeats on the leading strand to anneal with one another forming a hairpin on the leading strand (Figure 7A). At the other extreme, if the copies of the repeat are more than 80 bp apart, the displaced copy of the inverted repeat on the leading strand would find its complement on the lagging strand template already occupied by an Okazaki fragment (Figure 7B). The ideal inverted repeat spacing results in one of the repeats being single stranded in both the leading and lagging strands and thus available for strand switching (Figure 7C). We suggest that the preferred size range of inverted repeat spacing that is found in inverted junctions, along with our results from the split-ura3 strains grown in hydroxyurea, is an additional argument for a template switch between leading and lagging strands at the core of the ODIRA models.
CONCLUSIONS:

We provide evidence that template switching at regressed forks generates inverted amplicons through linear, inverted, extrachromosomal intermediates. These data lead us to augment our original OD IRA model by providing evidence that the two template switches could occur simultaneously or in subsequent S phases. Using a split-ura3 yeast strain we obtained independent evidence that linear hairpin intermediates from the SUL1 region of the genome can arise in the absence of the sulfate-limited chemostat protocol and can be stabilized by selecting for their recombination with a second chromosome. The generation of these trapped inverted linear intermediates is reduced in the presence of hydroxyurea, a replication inhibitor, implicating replication as a mechanism for their formation. Furthermore, the competing hypothesis—that the inverted linear intermediates are formed through double stranded break repair—is refuted by our finding that inducing double strand breaks in the vicinity of SUL1 actually interferes with the formation of inverted linear intermediates rather than promoting their formation. A second line of evidence comes from split read analysis of populations of yeast grown in sulfate-limited chemostats where selective pressure often leads to cells with inverted triplications of the SUL1 locus that sweep the population. Out of the 31 sulfate-limited populations we sequenced, 11 cultures had unmatched numbers of left and right inverted junctions. The common features of the inverted junctions are consistent with the hybridization of the 3’ end of the leading strand with short complementary sequences in the lagging strand template, and the subsequent ligation with the lagging nascent strand. The distribution of centromere-proximal junctions is consistent with the known direction of replication in the
region centromere proximal to SUL1 and the preferred use of a specific size class of pre-
existing, interrupted inverted repeats for this strand transfer is consistent with a
replication based mechanism.
REFERENCES


Figure 1. Comparison of “dogbone” and “hairpin” ODIRA models for inverted triplication of the SUL1 locus. A) In our ODIRA model we propose that regressed forks (2a) provide an opportunity for a template switch between the nascent leading strand and the lagging strand template that occurs at short, interrupted inverted repeats (2b). Extension of the displaced leading strand and its ligation to an Okazaki fragment (2c) results in a covalent linkage between the leading and lagging nascent strands that can be expelled from the chromosome by an incoming fork from an adjacent origin (2d, and 3). A similar template switch at the divergent fork results in an extrachromosomal, self-complementary, single-stranded circular molecule (dogbone; 3). In the next cell cycle, the dogbone can replicate from its resident origin creating a duplex circular molecule that has two copies of the SUL1 region in inverted orientation (4). Recombination of the dimeric circular molecule into the chromosome through homology with the SUL1 region creates a triplication with the center copy in inverted orientation (5). The inversion junctions (Cen-proximal and Tel-proximal; CJ and TJ, respectively) map to the genomic, short interrupted, inverted repeats where the template switching occurred. B) The Cen-proximal and Tel-proximal inversions can occur in different cell cycles, generating isochromosomal linear molecules. After the second, telomere-proximal junction is created, the doubly inverted linear molecule can recombine with the SUL1 region creating an inverted triplication that is identical to that produced by the dogbone ODIRA model. The gray shaded panels in A and B show both strands of the DNA. Elsewhere, dsDNA is shown as a single line and the ssDNA loops by a thinner line.

Figure 2. Experimental design for detecting extrachromosomal intermediates during DNA amplification. (A) Sites on chromosomes II and IX were modified by insertion of overlapping fragments of the URA3 gene. A probe to the unique 5' portion of the SUL1 fragment used for Southern blotting is highlighted in cyan. (B) Recombination between the two marked
chromosomes can recreate an intact URA3 gene but results in the creation of an unstable dicentric Ura+ chromosome. Deletion of one of the centromeres results in stabilized chromosome. The reciprocal product is an acentric chromosome that is lost during mitosis. (C) The chromosome II ura fragment, amplified as an extrachromosomal, inverted circular molecule, can recombine with the target site (ra3) on chromosome IX to generate a functional URA3 gene. The resulting copy of chromosome IX contains a tandem inverted triplication of the ra segment. (D) A single replication error at the centromere-proximal fork generates a palindromic linear fragment. Recombination between one of the copies of the ura fragments and the target ra3 fragment on chromosome IX re-forms a functional URA3 gene and creates a translocation between the palindromic chromosome and chromosome IX. To cover the loss of essential genes on the left arm of chromosome IX, an unarranged chromosome IX is also expected. (E) Anticipated CHEF gel results for the four strains described in A-D. The ethidium bromide stained gel reveals all chromosomes—the rearranged chromosome for each case is indicated in red. Hybridization with chromosome specific probes is used to distinguish the different outcomes, including the 5'SUL1 probe shown in (A). The deleted CEN2 (B) is often retained as a large circular molecule through recombination between flanking TY repeated elements. In CHEF gels, these circular molecules are often found retained in the well.

**Figure 3. Physical characterization of chromosomes from Ura+ clones.** (A) Chromosomes from a wild type strain, the haploid ura and ra3 strains, the heterozygous diploid strain, split-ura3 strain, and five Ura+ clones were analyzed by CHEF gel electrophoresis and Southern blot hybridization. The ethidium bromide stained gel and the Southern hybridizations reveal that the Ura+ clones contain an additional single unique chromosome that hybridizes to the centromere of chromosome IX and also to the 5'SUL1 probe. The ratio of 5'SUL1 hybridization for each of the five Ura+ strains shows a ~2:1 hybridization ratio for the neochromosome relative to the native chromosome II (measured by probe signal). (B) Array comparative genome hybridization of clone
A1 vs. the parent split-ura3 strain indicates that the right telomeric segment of chromosome IX (labeled z) and the ~20 kb at the right telomere of chromosome II (as part of the fragment labeled y) are present at a copy number of ~2. A larger subtelomeric segment of chromosome II (labeled x) is present at a copy number of ~3. The sites of the ura and ra3 insertions are indicated by arrows. (C) The proposed structure of the inverted neochromosome that contains the three amplified segments x, y, and z is illustrated above the two non-rearranged chromosomes. (D) The sizes of each of the five inverted neochromosomes (red dots) deduced from CHEF gels correspond to their predicted sizes based on aCGH data. Size estimates of non-inverted Ura+ chromosomes are indicated by black dots.

Figure 4: Induction of DNA double stranded breaks by CRISPR/Cas9 cleavage at the centromere-proximal to 5'ura (SUL1) eliminates inverted amplicons. (A) The proposed hairpin intermediate could be formed by a single Cen-proximal ODIRA template-switching event or by a double-stranded DNA break (DSB). (B) CRISPR/Cas9 was used to induce a DSB either centromere-proximal to the ura locus (708 kb) or telomere-proximal to the ura locus (792 kb). Cells that had been transformed with the CRISPR/Cas9 plasmid (Figure S1) were selected for on-leucine plates. Twenty independent transformants were grown to saturation in 1 ml of liquid–leucine medium and the entire saturated cultures (approximately 10⁶ cells) were spread on –ura plates (+galactose, +rafinose) to induce Cas9 expression and to select for successful regeneration of a functional URA3 gene. (C) Cutting at 792 kb greatly increased Ura+ colony frequency while cutting at 708 resulted in no Ura+ colony recovery. (D) Cutting at 792 resulted in direct recombination events between chromosome II and IX, while cutting at 708 appeared to interfere with the ability to recreate the functional URA3 gene. The control plasmid lacking a guide RNA gene produced a similar ratio of hairpin URA3 chromosomes to recombined chromosomes (6:14) as found for the non-transformed split-ura3 haploid strain (11:50).
**Figure 5. Sequence analysis of inverted junctions from chemostat populations.** (A) An example of one split read from a population of cells grown ~250 generations in a sulfate-limited chemostat. In the 150 bp sequence the Watson strand rejoined the Crick strand at a four base pair inverted repeat (TGGC/GCCA). (B) The appearance of split reads in IGV. Each line is a single unique read that covers the same junction. The blue bars highlight the 4 bp inverted repeat, the orange segment is the 61 nucleotide-interruption that separates the two copies of the inverted repeats, and purple indicates the region that is present twice in inverted orientation in the 150 bp split read. (C) The right end of Chromosome II with two short interrupted, inverted repeats (red and blue triangles) flanking SUL1 and its adjacent origin ARS228 are schematized. The inverted repeats could potentially create inverted junctions in two orientations. For the inverted junctions to be part of a SUL1 amplification the Cen-proximal junction (CJ) must be oriented with the single stranded loop on the left and the Tel-proximal junction (TJ) must be oriented with the single stranded loop on the right. The other orientations, while involving the same interrupted inverted repeats, would not be productive for creating a SUL1 amplicon. (D) Among the 31 sulfate-limited chemostat populations we catalogued 92 junctions that were represented by at least two independent sequence reads. The inverted junctions are distributed across the terminal ~80 kb of chromosome II with a perfect split in orientations that occurs to either side of the SUL1 gene. The asterisks in (B) and (D) refer to the specific split read sequence shown in (A).

**Figure 6. Genomic features of the inverted junctions of SUL1 amplicons.** (A) Density of genomic inverted repeat (IR) sequences (from 2 to 14 bp) that lie within 250 bp of each other across the 81 kb right terminal region of chromosome II, binned in 1 kb intervals. (B) The number of interrupted inverted repeats of different sizes (IRs + spacers in bp) in the terminal 81 kb of chromosome II (purple) is compared to the sizes of interrupted inverted repeats (orange) used to
generate inverted SUL1 amplicons. The number of inverted repeats of different sizes is relatively constant across the genome, but most of the repeats that give rise to inverted amplicons range from 40-80 bp. (C) The frequencies of inverted repeat sizes (2-14 bp) in the genome (purple) vs. those found at the sites of inverted amplicons (orange). (D) The ratio of repeats present in amplicon junctions (used) relative to their abundance in the genome (potential). (E) The location of each of 89 inverted junctions lying between 730 and 813 kb of chromosome II. The colors indicate the size of the repeats at the inverted junctions and are the same as illustrated in (A). The asterisks indicate that the same sequence was observed in 2 or 3 independent cultures. (F) Replication origins and the direction of fork movement in the 81 kb at the right end of chromosome II. Top: The most active origins (ARS224 and ARS228) generate bidirectional forks moving left and right (arrows). The region between 743 and 768 kb is replicated by a rightward moving fork and therefore cannot create the Cen-junctions needed to generate SUL1 amplicons (striped arrow). Bottom: Infrequent initiation at two additional origins (ARS225 and unnamed ARS - uknARS) contribute change the direction that sections of the chromosome are replicated. The minor leftward moving fork between 750 and 760 kb may be responsible for the four inverted junctions in the region shown in (E).

**Figure 7. The size of single stranded gaps on the lagging strand at replication forks limits which interrupted inverted repeats contribute to inversion junctions.** Interrupted inverted repeats with a spacing of 40-80 kb are overrepresented in the inverted junctions of SUL1 amplicons. We propose that the size of the single stranded gap on the lagging strand is a primary contributor to this size selection. We illustrate this connection between interrupted inverted repeat size and lagging strand gap size for repeat spacing at three intervals. (A) If the repeats are in close proximity to one another (30 bp), during fork regression, the parental copies reform double stranded DNA and the displaced leading strand, containing the inverted repeats in single stranded form, are likely to self-hybridize. (B) If the repeats are at a distance that exceeds the average
Okazaki gap on the lagging strand (120 bp), then fork regression leads to re-pairing of the parental strands of the left inverted repeat, while the right repeat had already been covered by an Okazaki fragment. This situation results in the repeat at the free 3’end with no template to which it can hybridize. (C) When both copies of the inverted repeats are within the size of an Okazaki gap (60 bp) then the left repeat will reform double stranded DNA between parental strands during fork regression. However, the left most copy on the leading strand still find its template in single stranded form on the lagging strand.
SUPPORTING INFORMATION

Figure S1. Map of inducible CRISPR/Cas9 plasmid, pYcpGAL.

Figure S2. The gradual copy number change in aCGH profiles is due to inverted junctions. (A) The DNA isolated for aCGH had an average size of ~20-40 kb. To label this DNA with Cy dyes, the DNA was denatured and random primers added for DNA polymerase to synthesize labeled strands. After labeling, the DNA was sonicated to ~500 bp and hybridized to an array. (B) The same DNA as in (A) was sonicated before denaturation and labeling. This method eliminated the gradual transition from 1 to 3 copies and produced an abrupt transition at the same site. (C, D) A stylized depiction of the region from 600 to 700 kb of the native chromosome II (top bold lines) and the inverted amplicon (D). DNA to the left of 650 is present at one copy and DNA to the right of 650 is present in 3 copies. Random breaks are indicated by the interruptions. After denaturation of this DNA, formation of hairpin structures (grey hairpins) prevents annealing of the random oligos used for synthesis. Because breaks are in random places, different fragments will have different amounts of DNA excluded from the labeling reaction. (E) Quantification of the copy numbers expected across the site of the inversion junction generated from the schematic example in (C, D). Note that when the DNA was sheared to ~500 bp before labeling (B), the region up to within 250 bp of the inverted junction was now available for labeling with Cy dyes.

Figure S3. Array CGH of chromosome II derived from twelve Ura+ clones with 2:1 5’SUL1 hybridization ratios. All chromosomes were recovered after growth in normal medium with the exception of the clone labeled HU++3 that was obtained from incubation in 200 mM hydroxyurea. Notice the different features in the right and left amplification junctions. The left junction shows a gradual change in copy number from 1 to ~ 3 copies and occurs at variable sites along chromosome
II centromere proximal to SUL1. The left boundary of each inversion junction shows the characteristic “waterfall” transition between copy numbers. In contrast, the right junction is an abrupt change in copy number from ~3 to 2 copies created by recombination between ura and ra3 on chromosome IX. The aCGH profiles of chromosome IX were identical to that shown in Figure 2B left.

Figure S4. Centromere loss from a Ura+ dicentric chromosome produced by recombination between chromosomes II and IX. (A) Ethidium bromide stained gel of clone C1 (center lane) with a neochromosome that is larger than the native chromosome II. (The flanking gel lanes contained other Ura+ isolates.) Southern hybridizations indicate that the neochromosome has retained CEN9 and lost CEN2 and that there is an unaltered version chromosome II. (B) ArrayCGH confirms that most of chromosome II has been duplicated but one copy of the chromosome has lost its CEN2 sequence by recombination between the directly repeated Ty elements on either side of CEN2. The breaks in copy number on chromosome IX and the distal part of chromosome II mark the sites of the two URA3 fragments. (C) The most parsimonious organization of the duplicated parts of chromosomes II and IX produce a neochromosome that is consistent with the size estimated from the CHEF gel.

Figure S5. Breakage and telomere addition stabilizes a Ura+ dicentric chromosome. (A) Ethidium bromide stained gel of clone cH5 with a small neochromosome that retained CEN9 (center lane; flanking lanes are from other Ura+ clones). The cells retained unrearranged chromosomes II and IX. (B) ArrayCGH confirms that the relevant part of chromosome IX has been duplicated, beginning at the insertion site of ra3; however only the right half of chromosome II is duplicated, ending at the complementary region of the ura insertion. The sequence at ~420 kb could serve as a telomere seed after breakage of the dicentric chromosome. (C) The most parsimonious
organization of the duplicated parts of chromosomes II and IX produce a neochromosome that is consistent with the size estimated from the CHEF gel.

**Figure S6. Resolution of a dicentric Ura+ chromosome by a secondary breakage and recombination event (McClintock Bridge-Breakage-Fusion, BBF).** (A) Ethidium bromide stained gel of clone cH5 with a very small neochromosome that hybridizes to CEN9 and the 5’ SUL1 sequences from chromosome II. (B) ArrayCGH reveals that only a tiny fragment of chromosome II is retained on this neochromosome and that the left telomere has been replaced by a second copy of chromosome IX right. (C) The most parsimonious organization of the duplicated parts of chromosomes II and IX produce a neochromosome that is consistent with the size estimated from the CHEF gel. Sequences of homology where the second copy of the right telomeric fragment of chromosome IX could have been added are shown in their relative positions on the two native chromosomes.

**Figure S7. Summary of structural rearrangements that gave rise to a cohort of 27 independent Ura+ clones.** (A) No integration of circular inverted intermediates was observed. (B) Five instances of recombination of an inverted linear with chromosome IX were obtained. (C) The remaining 22 events were produced by direct recombination between chromosome II and IX with subsequent loss of a centromere, breakage and addition of a telomere, or secondary recombination presumably as a result of breakage during mitosis through BBF cycles.

**Figure S8. Alternative model for the generation of hairpin intermediates in SUL1 amplification.** (A) Repair of a double stranded break could expose an inverted repeat in the 3’ overhang that could be repaired to create the hairpin linear and its replicated isochromosomal fragment. If this double stranded break repair mechanism is responsible for the inverted Ura+
clones then CRISPR/Cas9 directed cutting on the Cen-proximal side of the SUL1 region should lead
to an increase in Ura+ clones overall and an increase frequency of inverted outcomes. (B) Resection
of the 5′end of a double strand breaks introduced distal to SUL1 would expose the urs homology in
single stranded form and stimulate recombination with the ra3 sequences on chromosome IX.
CRISPR/Cas9 cleavage on the Tel-proximal side of SUL1 should increase the overall frequency of
Ura+ clones that occur through recombination events between the two chromosomes.

Figure S9. Read depth analysis of the right telomeric regions of chromosome II from 31
sulfate-limited populations. 31 chemostat populations (S01-S31; ~250 generations) were
subjected to 150 bp paired end Illumina sequencing. The read depth is shown as a heat map with
higher copy numbers in darker shades of blue. The coordinates (in kb) are shown across the top X
axis and the positions of ORFS (navy boxes) and several identified genes are shown across the
bottom X axis. The position of SUL1 is marked by the white dotted lines. The split reads that mark
various types of amplification junctions are indicated by yellow circles (inverted junctions), orange
bars (direct repeat junctions), and red squares (junction with an existing telomere—terminal
translocations, indicated by the red triangles). Three of the inverted junctions (S06, S22 and S24)
ocurred in sequences to the left of 730 kb.

Figure S10. Comparison of split reads represented by a single PCR fragment to those
represented by two or more unique PCR fragments. (A) The distribution of inverted Cen- and
Tel-junctions recovered from 31 sulfate-limited chemostats (top) and 32 glucose-limited
chemostats (bottom) that were represented by single PCR fragment sequences (n=1). No specific
orientation with respect to SUL1 was observed (compare to Figure 5D for split reads with support
from two or more PCR fragments). Of note, there were no inverted junctions in the glucose-limited
chemostat populations with support from two or more PCR fragments. (B) The size of the inverted
repeats from the sulfate-limited chemostats with support from two or more PCR fragments (left) relative to those with support from one PCR fragment (right) are distinctly different. (C) The spacing between the inverted repeats from the sulfate-limited chemostats with support from two or more PCR fragments (left) relative to those with support from one PCR fragment (right) are distinctly different. These results provided the read-depth cut-off for distinguishing PCR artifacts from bone-fide *in vivo* inverted junctions.

Figure S11. Imbalance in numbers of Cen-junctions and Tel-junctions in the 31 sulfate-limited chemostat cultures. If inverted amplicons of the *SUL1* region arise through hairpin intermediates then in each population the number of Cen- and Tel-junctions would not necessarily be equal. (A and B) Mechanisms to explain unequal numbers of Cen- and Tel-junctions. Colored triangles indicate different inverted repeats along the chromosome. (C) Mechanisms to explain an equivalence between Cen- and Tel-junctions. (D) Among the 31 sulfate-limited chemostats, two had no inverted amplicons of the *SUL1* gene (one was a tandem duplication and the other was an amplification of the *SUL1* promoter region). The remaining 29 produced the total number of 92 inverted junctions. Eighteen cultures had matched numbers of Cen- and Tel-junctions. The remaining eleven had unbalanced numbers of Cen- and Tel-junctions.

Table S1. Oligonucleotides used in strain construction.

Table S2. Guide RNA and oligonucleotide sequences for CRISPR/Cas9 plasmids.

Table S3. *SUL1* amplicon junction sequences (n>1 instances).
Table S4. Chromosome II arrayCGH data.

Table S5. Chromosome IX arrayCGH data.

Table S6. Sulfate-limited chemostats junction sequences (n=1).

Table S7. Glucose-limited chemostats junctions sequences (n=1).

Table S8. Numerical values used for graphs in Figures and Supplemental Figures.
**Figure 1**

**A. Dogbone ODIRA**

1. **SUL1 region of chromosome II**

   ![Dogbone ODIRA Diagram](image)

2. **Replication initiation and fork stalling**

   ![Dogbone ODIRA Diagram](image)

3. **Displacement of self-complementary “dogbone” by passive replication**

   ![Dogbone ODIRA Diagram](image)

4. **Dogbone replication**

   ![Dogbone ODIRA Diagram](image)

5. **Recombination**

   ![Dogbone ODIRA Diagram](image)

   *Chr II with interstitial inverted triplication*

**B. Hairpin ODIRA**

1. **First replication initiation and fork stalling**

   ![Hairpin ODIRA Diagram](image)

2. **Displacement of self-complementary “hairpin” by passive replication**

   ![Hairpin ODIRA Diagram](image)

3. **Hairpin replication**

   ![Hairpin ODIRA Diagram](image)

4. **Second replication initiation and fork stalling**

   ![Hairpin ODIRA Diagram](image)

5. **Hairpin replication**

   ![Hairpin ODIRA Diagram](image)

6. **Recombination**

   *Chr II with interstitial inverted triplication*
A. Split-ura3 strain (s2-1)

B. Recombination between Chr II & IX

C. Dogbone recombination with Chr IX

D. Hairpin recombination with Chr IX

E. Hypothetical strains (EB stain)

Wells: A, B, C, D

Chr II: A, B, C, D

Chr IX: A, B, C, D

Figure 2

Hypothetical strains (EB stain)

5'SUL1 probe

CEN9 probe

CEN2 probe

DNA bands 1 and 2

Diagram A shows the split-ura3 strain (s2-1) with Chr II and Chr IX, each bearing CEN2 and CEN9, respectively, and a 5'SUL1 probe.

Diagram B illustrates recombination between Chr II and Chr IX, leading to dicentric and acentric chromosomes, with URA3 and ra3 as markers.

Diagram C depicts dogbone recombination with Chr IX, resulting in Chr II and Chr IX with URA3 and ra3, and CEN2 deletion.

Diagram D presents hairpin recombination with Chr IX, involving Chr II, Chr IX, and ra3.

Diagram E represents hypothetical strains (EB stain) with wells A, B, C, and D, illustrating DNA bands 1 and 2.
Figure 3

A. 

Wells →

Chr II →

Chr IX →

Probes:

- 5'SUL1
- CEN9
- CEN2

B. 

Copy number vs. Chromosome coordinates (kb)

Chr IX (A1)

Chr II (A1)

Copy number

0 1 2 3 4

410 kb

400

800

0 100 200 300 400

C. 

Neochromosome

Chr II

Chr IX

D. 

Size estimate from CHEF gels (kb)

Size estimate from aCGH (kb)

y = -9.8569 + 1.0072x

R² = 0.99879
Figure 4

A. ODIRA template-switching

B. No gRNA  708kb gRNA  792kb gRNA

C. Ura+ colonies per 10^8 cells

D. Number of Observed Events

guide: none  708 kb  792 kb
Figure 5

A. 150 bp sequencing read: Telomere-junction (TJ)

B. 5' 3' 5' 3' 5' 3'

C. Short interrupted inverted repeats

D. Sulfate n>1
Leading strand through both repeats; lagging strand gap = 80 bp

Different outcomes for leading strand after fork regression

<table>
<thead>
<tr>
<th>Variation in size of interrupted inverted repeats</th>
<th>Leading strand through both repeats; lagging strand gap = 80 bp</th>
<th>Different outcomes for leading strand after fork regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. 30 bp</td>
<td>leading strand replication through both repeats</td>
<td>fold back</td>
</tr>
<tr>
<td>B. 120 bp</td>
<td>80 bp ssDNA lagging strand gap</td>
<td>unpaired repeat</td>
</tr>
<tr>
<td>C. 60 bp</td>
<td></td>
<td>template switch</td>
</tr>
</tbody>
</table>
Figure S1

pYcpGAL
Figure S2

A. Sonication after labeling

B. Sonication before labeling

C. 610 kb 630 kb 650 kb 670 kb 690 kb

20 kb genomic fragments due to dsDNA breaks during DNA isolation

E. Inverted fragments (aligned with Chr II)

D. Inverted junction

Portions of hairpin fragments that are not labeled
Figure S3

Chromosome II coordinates (kb)

Copy number

0 1 2 3 4

0 1 2 3 4

0 1 2 3 4

0 1 2 3 4

P5-CR

P2+CR

HU++3

G2-A

P7+B

I1

CH1

G1

P2-A

J2

G3-A

Chromosome II coordinates (kb)
Figure S4: Dicentric $\rightarrow$ centromere deletion

A. C1

B. Chr II Tys

C. Resolved dicentric

$\Delta$ CEN2Δ

URA3

CEN9
Figure S5: Dicentric $\rightarrow$ de novo telomere addition

A. H-5

B. Chr II

5' CACCACCCA 3'

Copy number

1 2 3

100 200 300 400 500 600 700 800

C. Chr IX

ra3

Copy number

1 2 3

100 200 300 400

CEN9

Chr II

New telomere

Chr IX

URA3
Figure S6: Dicentric → BBF

A. F1

II

IX

*

CEN2 CEN9

B.

Chr II

0 1 2 3

copy number

100 200 300 400 500 600 700 800

C.

URA3

BBF chr

Chr II

Chr IX

5’GCTtaTTTGGTCTT 3’

5’AAGACCAAACacAGC 3’
Figure S7

A. Dogbone = 0/27

Chr IX

ra

URA3

Chr II

5'ura

Chr IX

3'ra3

Dogbone ODIRA

B. Hairpin = 5/27

Chr IX

URA3

Hairpin ODIRA

C. Direct recombination

URA3

BBF cycles

BBF = 3/27

RB = 3/27

CENΔ = 7/27

New telomere = 6/27

BIR = 6/27
Figure S8

A. CRISPR/Cas9 targeted proximal to SUL1

B. CRISPR/Cas9 targeted distal to SUL1

Does the frequency of Ura+ colonies change?

Does the frequency of URA+ colonies change?
Figure S10

A. Number of inverted products per 5 kb bins

B. Size of IRs (bp)

C. Spacing of interrupted IR (bp)
Figure S11

A. Cen-junction only

B. Tel-junctions > Cen-junctions

C. Cen-junctions = Tel-junctions

D. Diagram showing the number of centrigenic-junctions (CJs) and telomeric-inverted junctions (TJs)