Real-Time Evolution of a Subtelomeric Gene Family in *Candida albicans*

Matthew Z. Anderson,* Lauren J. Wigen,* Laura S. Burrack,* and Judith Berman*,†,1
*Department of Genetics, Cell Biology and Development, University of Minnesota, Twin Cities, Minneapolis, Minnesota 55455 and †Department of Microbiology and Biotechnology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv 69978, Israel
ORCID ID: 0000-0003-0469-9819 (J.B.)

ABSTRACT Subtelomeric regions of the genome are notable for high rates of sequence evolution and rapid gene turnover. Evidence of subtelomeric evolution has relied heavily on comparisons of historical evolutionary patterns to infer trends and frequencies of these events. Here, we describe evolution of the subtelomeric *TLO* gene family in *Candida albicans* during laboratory passaging for over 4000 generations. *C. albicans* is a commensal and opportunistic pathogen of humans and the *TLO* gene family encodes a subunit of the Mediator complex that regulates transcription and affects a range of virulence factors. We identified 16 distinct subtelomeric recombination events that altered the *TLO* repertoire. Ectopic recombination between subtelomeres on different chromosome ends occurred approximately once per 5000 generations and was often followed by loss of heterozygosity, resulting in the complete loss of one *TLO* gene sequence with expansion of another. In one case, recombination within *TLO* genes produced a novel *TLO* gene sequence. *TLO* copy number changes were biased, with some *TLOs* preferentially being copied to novel chromosome arms and other *TLO* genes being frequently lost. The majority of these nonreciprocal recombination events occurred either within the 3′ end of the *TLO* coding sequence or within a conserved 50-bp sequence element centromere-proximal to *TLO* coding sequence. Thus, subtelomeric recombination is a rapid mechanism of generating genotypic diversity through alterations in the number and sequence of related gene family members.

KEYWORDS *Candida*; telomere; experimental evolution; gene families; recombination

TELOMERES, the sequences at the ends of linear chromosomes, protect the chromosome from terminal degradation. Telomere-adjacent sequences, termed subtelomeres (Pryde *et al.* 1997), are often defined by repeat-rich and gene-poor heterochromatic sequence of variable length (Mefford and Trask 2002; Kupiec 2014). The repetitive nature and high sequence similarity (Louis and Haber 1990b; Gardner *et al.* 2002) of subtelomeres complicates their inclusion in whole-genome sequence assemblies (Kellis *et al.* 2003) and, as a result, subtelomere sequence and organization are often poorly characterized. Accordingly, genes within subtelomeric regions are often excluded from comparative genomics analyses, due to a lack of accurate assembly and sequence read assignment (Hahn *et al.* 2005; Wapinski *et al.* 2007).

Subtelomeres are composed of two segments with different levels of repetitiveness and divergence. Telomere-proximal sequences include short tandem repeats, whereas, telomere-distal domains encode unique genes, gene families, and repetitive elements of varying frequency (Pryde *et al.* 1997). Across eukaryotes, subtelomere variability is due, in part, to frequent acquisition of single-nucleotide polymorphisms (SNPs) and insertions/deletions (indels) (Cuomo *et al.* 2007; Dreszer *et al.* 2007; Anderson *et al.* 2008; Carreto *et al.* 2008), as well as copy number variation in subtelomeric gene families (Brown *et al.* 2010). Thus, subtelomeres are often the most variable region of the genome (Winzeler *et al.* 2003; Cuomo *et al.* 2007; Carreto *et al.* 2008; Kasuga *et al.* 2009; Brown *et al.* 2010) both in genome-wide studies of DNA sequence variation (Farman and Kim 2005; Dreszer *et al.* 2007) and in studies of gene copy number variation (Dujon *et al.* 2004; Carreto *et al.* 2008; Gresham *et al.* 2008). As such, subtelomeres provide ideal environments for the rapid evolution of genes upon

Copyright © 2015 by the Genetics Society of America
doi: 10.1534/genetics.115.177451
Manuscript received February 15, 2015; accepted for publication May 5, 2015; published Early Online May 8, 2015.
Available freely online through the author-supported open access option.
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.177451/-/DC1.
[1]Corresponding author: Department of Molecular Microbiology and Biotechnology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv, 69978 Israel. E-mail: judithberman11@gmail.com

which selection can then act. Accordingly, comparison of *Saccharomyces cerevisiae* strains that produce beer, sherry, or champagne revealed increased copy numbers of subtelomeric *MAL*, *SUC*, or *MEL* genes, respectively (Naumov *et al.* 1995; Naumova *et al.* 2005; Brown *et al.* 2010; Dunn *et al.* 2012). This highlights the ability of subtelomeric regions to facilitate the rapid amplification of genes that provide a selective advantage under different growth conditions. However, the mechanisms that give rise to these changes and how rapidly they arise have not been well studied.

Recombination operates extensively at subtelomeres in all eukaryote kingdoms, including protists, animals, and plants (Louis *et al.* 1994; Linardopoulou *et al.* 2005; Gaut *et al.* 2007; Rudd *et al.* 2007). Studies in protozoa and fungi found that subtelomeric recombination is particularly common in noncoding sequences (Corcoran *et al.* 1988; Rehmeyer *et al.* 2006; Boothroyd *et al.* 2009; Glover *et al.* 2013), although it also occurs within coding sequences (Kraemer *et al.* 2007; Fan *et al.* 2008). For example, in *S. cerevisiae*, the most common subtelomeric recombination junctions are the Y′ elements, a family of helicase homologs of varying functional integrity (Louis and Haber 1990a; Louis *et al.* 1994). In the malaria parasite *Plasmodium falciparum*, immune evasion is mediated by extensive recombination among subtelomeric *var* gene paralogs (Kraemer *et al.* 2007; Kyes *et al.* 2007). Recombination progresses through different mechanisms at subtelomeres including gene conversion (Morrison *et al.* 2009; Claessens *et al.* 2014) and break-induced replication (BIR) (Boothroyd *et al.* 2009; Hovel-Miner *et al.* 2012) (Supporting Information, Figure S1). However, the kinetics and stability of *var* switching events have been difficult to define (Bopp *et al.* 2013) because selection against deleterious genotypes likely discards many subtelomeric recombination events before they become fixed to a detectable level in natural populations.

The majority of subtelomeres include gene families that are important for adaptation to the ecological niches occupied by that organism (Celenza and Carlson 1985; Dujon *et al.* 2004; Kyes *et al.* 2007). For example, fluctuations frequently occur in *S. cerevisiae* subtelomeric metabolic gene families through selective pressures for efficient utilization of available carbon sources (Turakainen *et al.* 1993; Brown *et al.* 2010; Wenger *et al.* 2011; Dunn *et al.* 2012). Interestingly, natural *S. cerevisiae* populations produce new family member variants via recombination between previously established family members (Van Mulders *et al.* 2010; Christiaens *et al.* 2012). Consequently, recombination exploits existing paralog diversity to construct novel variant genes without requiring extensive mutation. Furthermore, gene noise and expression plasticity at subtelomeres is primarily due to stochastic transcriptional silencing (Gottschling *et al.* 1990; Fischer *et al.* 1997; Mondoux and Zakian 2007; Choi *et al.* 2008; Halliwell *et al.* 2012; Anderson *et al.* 2014) and facilitates a bet-hedging mechanism in which new variant genotypes can be tested without committing all cells in a population to a given expression pattern.

In *Candida albicans*, the *telo*mere-associated (*TLO*) genes are the only widespread subtelomeric gene family (Van Het Hoog *et al.* 2007), and the 13 subtelomeric paralogs are found at all but three chromosome arms (Chr2R, 6R, and 7L) (Van Het Hoog *et al.* 2007; Anderson *et al.* 2012). *TLO*s underwent a recent expansion from 1 copy in most *Candida* species to 14 paralogs in *C. albicans* (Butler *et al.* 2009). All *TLO* genes are oriented similarly in the subtelomere, with transcription proceeding toward the centromere (Figure 1C). Sequence homology comparisons distinguish three *TLO* clades, α, β, and γ, which also exhibit different transcriptional levels and cellular localization patterns (Van Het Hoog *et al.* 2007; Anderson *et al.* 2012). Indels distinguish the lengths of *TLO* genes in a clade-specific manner: the seven *TLOγ* clade genes are shortest (~525 bp), the five *TLOα* clade genes are longer, (675–750 bp), and the single *TLOβ* gene is the longest at 822 bp. Within a clade, *TLO*s are ~97% identical and, between clades, they are 82% identical (when excluding indels). All *TLO*s encode a conserved N-terminal Med2 domain followed by a gene-specific repetitive region of variable length and a C-terminal region that distinguishes the three clades (Anderson *et al.* 2012; Zhang *et al.* 2012). Investigation of the subtelomeric *TLO* gene family composition uncovered differences between the original SC5314 reference genome (Jones *et al.* 2004) and subsequently sequenced SC5314 isolates, including homologous chromosomes encoding two different *TLO* genes at the same position (Hirakawa *et al.* 2014) and a previously unidentified *TLO* (Anderson *et al.* 2012).

All *TLO*s encode a Med2 domain-containing protein, such that the 14 paralogs (13 subtelomeric and one internal on Chr1L) are thought to encode interchangeable components of the Mediator transcriptional regulation complex (Zhang *et al.* 2012). In *C. albicans*, *MED3*, which associates with *TLOs*, affects the stability of white–opaque switching (Zhang *et al.* 2013), which is necessary for mating and important for formation of some types of biofilms (Park *et al.* 2013; Jones *et al.* 2014). In the closely related species *C. dubliniensis*, which has only two *TLO* genes, deletion analysis revealed that these *TLOs* contribute to filamentous growth forms, biofilm formation, and sugar metabolism (Jackson *et al.* 2009).

Here, we describe the genome changes that occurred to members of the *TLO* subtelomeric gene family during the course of laboratory passaging. Sequencing of subtelomeres in passaged isolates revealed 16 changes in the *TLO* gene repertoire where one *TLO* gene had replaced another one on a different chromosome arm. These *TLO* "movement" events appear to involve a nonreciprocal recombination mechanism that occurs frequently, suggesting that *TLO* repertoires change frequently and most likely provide a biologically relevant mechanism for generating genetic diversity in *C. albicans*.

Materials and Methods

Strain construction and passaging

Yeast cells were passaged on solid rich medium (YPAD) agar plates at 30° under standard conditions. Two sets of nine

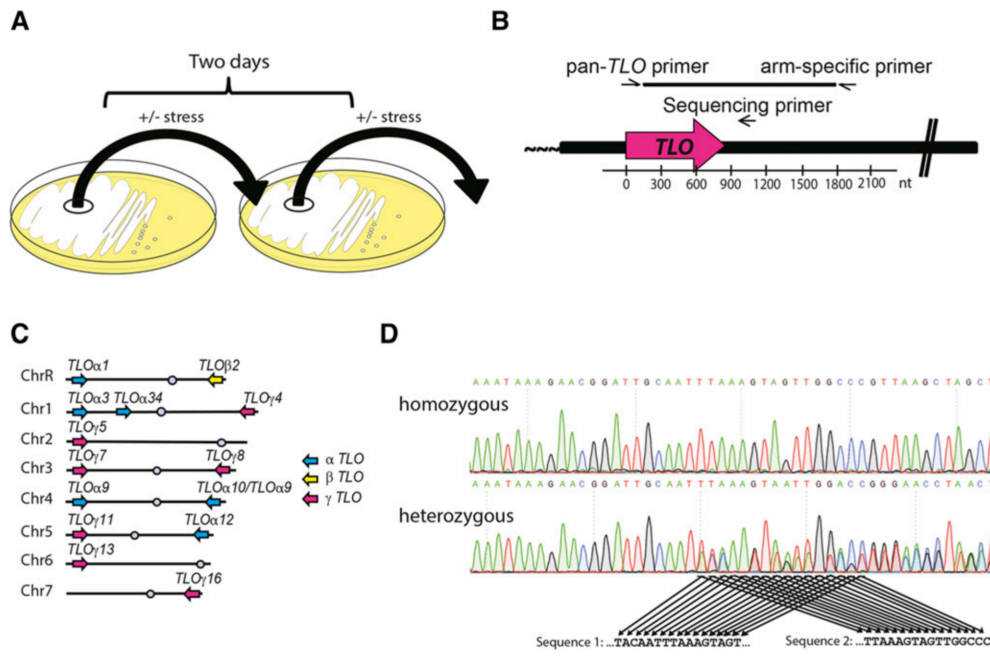


Figure 1 Evolution and sequencing of *C. albicans* subtelomeres. (A) Strains were evolved by passaging to new rich media agar plates every 2 days for 30 months using cells from the thick part of the streak to avoid bottlenecks during propagation. Passaging was performed either without stress, as described above, or with a 1hr cold shock (4°) followed by a 1hr mild heat shock (37°) immediately after being restruck every third passage. (B) *TLO* genes were amplified from each chromosome arm using primers represented as half arrows. Amplicons were sequenced from chromosome arm-specific unique sequence to the start codon of the *TLO* genes (pan-*TLO* primer). (C) Location of *TLO* genes in the SC5314 strain prior to passaging is shown. *TLO* clade is denoted by color. (D) Amplicons composed of mixed sequences as indicated in the chromatograph were separated using the chromosome arm sequence as a template to identify the second sequence. Once separated, these sequences were used to identify the second donor *TLO* gene.

strains were restruck to fresh plates every 2 days from the thick part of the streak, which minimizes bottlenecking of the population. One set, termed "stress passaged," was exposed immediately after being restruck to 1 hr at 4° and then one hour at 37° prior to incubation at 30° every three passages.

The number of generations traversed between passages was determined by cell counts before and after 2 days of growth on YPAD agar plates. Cells from two progenitor strains were struck onto YPAD plates identical to the passaged isolates. Cell counts of a streak through the struck population immediately after plating and following 2 days of growth were performed using a hemocytometer to allow for calculation of generations per passage, $P_f = P_i \times 2^N$, where P_f and P_i are the final and initial cell counts. Slightly elevated but not significant numbers of generations were found among end-point-passaged strains.

PCR amplification of gene sequences

SC5314 DNA was collected as previously described (Hoffman and Winston 1987). PCR of *TLO* genes was performed using arm-specific primers for all chromosomes together with a pan-*TLO* primer. The arm-specific primers were designed against the closest region of unique sequence centromere-proximal to the *TLO* or of similar homology on chromosome arms lacking a *TLO* gene. Attempts to sequence beyond the *TLO* gene were complicated by sequences not matching the expected chromosomal locations from the genome reference sequence.

Amplification of the core centromere sequence was performed for each of the eight *C. albicans* chromosomes. Amplification of *ENO1*, *LIP4*, *SAP1*, and *SAP7* was performed independently. Primers are listed in Table S1. All amplicons were amplified and sequenced twice independently using nested internal primers.

Analysis of TLO sequences

Sequenced amplicons were aligned to the reference *C. albicans* SC5314 genome (http://candidagenome.org/) using BLAST. Sequences were also independently aligned to the SC5314 *TLO* sequences using the multiple sequence comparison by log-expectation (MUSCLE) algorithm at http://www.ebi.ac.uk/Tools/muscle/index.html (Edgar 2004). Assignment of *TLO* encoded at specific arms was determined by *TLO*- and/or clade- specific SNPs (Table S2) and indels and comparison to the amplicon generated from the progenitor strain. Chromatographs were viewed using Sequencher (Gene Codes Corporation, Ann Arbor, MI) and Snapgene Viewer (GSL Biotech, Chicago, IL).

Mapping chromosome heterozygosity

Chromosome heterozygosity was determined using single nucleotide polymorphism–restriction fragment length polymorphism (SNP–RFLP) analysis of previously described SNP positions (Forche *et al.* 2009).

Competitive growth assay

Competitive growth assays were performed for YJB9929 evolved strains relative to an RM1000 *ENO1*–GFP strain (CAY12638) as described in Thompson *et al.* (2006). Briefly, 5×10^6 cells/ml of each evolved and reference strain were mixed and seeded into 3 ml of YPAD. The ratio of the two competitors was quantified at both the initial and final (72 hr) time points by microscopic visualization of 10 fields

of view using a Zeiss Inverted Microscope (Axio Observer Z1) fitted with an AxioCam HR. The selection coefficients were calculated as $s = [\ln (E_f/R_f) - \ln (E_i / R_i)]/T$, where E and R are the numbers of evolved and reference cells, subscripts are the final and initial populations, and T is the number of generations that reference cells have grown during the competition.

Statistical analysis

Statistics were performed using R (R Development Core Team 2013) or SPSS Statistics for Windows (IBM Corp. 2013).

Results

Sequence discrepancies between the SC5314 reference genome and SC5314 isolates suggested that the *TLO* gene family may be continuing to evolve. To determine the time scale of subtelomere evolution, we investigated *TLO* gene organization in strains that had been passaged continuously for >30 months (457 passages, ~4570 divisions) under laboratory conditions. Two duplicate sets of nine SC5314-derived lines were passaged every 2 days on solid, rich-media agar plates. Both sets (nonstressed and stressed) were propagated on YPD medium at 30°; additionally, the stressed set was exposed to a cold shock (1 hr at 4°) followed by a heat shock (1 hr at 37°) every three passages (Figure 1A). After 30 months of passaging, we PCR amplified the 16 subtelomeric regions including *TLO*s from the populations of 18 independently evolved lines, using a chromosome arm-specific primer anchored in centromere-proximal, unique sequences together with a single, pan-*TLO* primer that bound to a sequence immediately downstream of the start codon and that is conserved among all *TLO* homologs (Figure 1B). Therefore, only chromosome arms encoding *TLO*s could produce amplicons that were captured for analysis. The subtelomeric composition of the progenitor strains was identical; 13 chromosome arms encoded a *TLO* gene (Figure 1C). Only the chromosome arms encoding *TLO*s in the parental strain generated PCR products for analysis in the passaged isolates, suggesting that no *TLO* had moved to a previously unoccupied chromosome arm.

To determine the resident *TLO* on all amplified chromosome arms, we sequenced the PCR products and analyzed the sequence for homology among the *TLO* paralogs. Most subtelomeric amplicons were composed of a single nucleotide sequence, indicated by single, uniform peaks throughout the sequence trace. However, some sequence traces began as a single uniform sequence in the centromere-proximal end of the amplicon and transitioned to a heterogeneous sequence composed of double peaks as the sequence extended toward the telomere, suggesting that each homolog on that particular chromosome arm encoded distinct *TLO* sequences. The new *TLO* sequence was inferred by subtracting out the previously established *TLO* sequence for that chromosome arm from the double peaks (Figure 1D). This process was performed for the 13 chromosome arms with *TLOs* in 18 passaged strains for a total of 234 unique data points.

Timing of TLO recombination events

Fourteen major subtelomeric rearrangement events were evident following 30 months of strain passage (Figure 2A and Table 1). This was defined as any detectable alteration in the *TLO* repertoire resulting from chromosome homozygosis and/or "movement" of specific *TLO* gene family members between subtelomeres during passage.

To further refine the timing of the *TLO* recombination events, we assayed isolates at intermediate time points during passaging for either the ancestral or evolved *TLO* organization. Rearrangements occurred throughout the time course from the first month to the final month of passaging (Figure 2A). Recombination occurred randomly throughout passaging ($P = 1.0$, one-way ANOVA). Furthermore, no clear environmental perturbations or alterations in the passaging protocol coincided with these time points. Additionally, sequencing the subtelomeres at the earliest time points following recombination failed to identify co-occurrence of other subtelomeric recombination events with the exception of events at both arms of chromosome 4 (Chr4L and 4R) in YJB9929. Thus, *TLO* gene family changes occurred at random times throughout the passaging experiment.

Most studies of genome evolution compare static representations of a continuous and ongoing process. The experimentally practical time scales of *TLO* recombination allowed transitionary phases to be captured by experimental evolution. For example, in strain YJB9929, the ancestral TLOα9/*TLOα10* configuration on Chr4R had homozygosed to two copies of *TLOα10* by the end of passaging (Table 1). Interestingly, this involved two transitionary states of Chr4R prior to the final *TLOα10/TLOα10* arrangement (Figure 2B). The first transition, seen in passage 137, was loss of *TLOa9* by homozygosis of the Chr4R *TLOα10* allele; the second transition, seen between passages 168 and 184, was introduction of one copy of *TLOγ8* from Chr3R and loss of a *TLOα10* allele; and the final transition, between passage 372 and 403, was homozygosis of *TLOα10*. Identification of these different transition states during passaging suggests that additional recombination events are often missed when comparing static representatives of genome evolution.

Capture of these transitional recombination events also provided a more accurate measure of the rate of *TLO* evolution. Strains underwent ~4570 divisions during the course of the experiment based on estimates of ~10 cell divisions between passages (data not shown). Therefore, we can estimate *TLO* movement to occur in ~1:5142 cell divisions or 1:82,260 chromosome arms per division. These rates are orders of magnitude greater than either genome average per-base-pair mutation rates or insertion/deletion rates per cell division (~10^{-9}) (Lang and Murray 2008; Lynch *et al.* 2008). Thus, subtelomeric recombination provides an opportunity for high-frequency gene evolution.

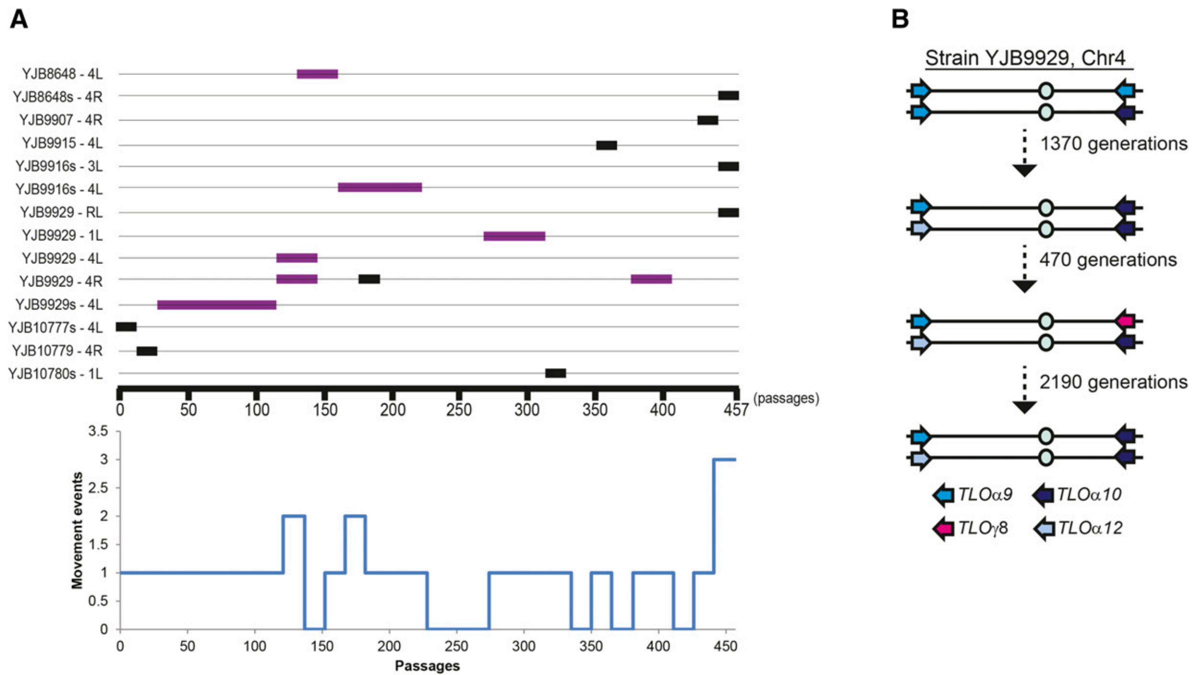


Figure 2 Evolution of subtelomeric *TLO*s during passaging. (A) The timing of identified recombination events was determined by screening archived isolates. The most narrow time frame for each event is displayed visually as a solid bar for each strain. The strain used and the chromosome arm affected by recombination are indicated on the *y*-axis. Black or purple bars delimit the time frames during which the recombination events became detectable in the different strains. Black bars indicate a 1-month interval and purple bars indicate longer intervals because isolates were archived less frequently over these intervals. Bottom: moving average for recombination events occurring at 15 passage intervals during the experiment. (B) Sequential recombination events on the right arm of chromosome 4 resulted in homozygosis of Chr4R in strain YJB9929.

Population-level strain passage requires that the detected recombination events rose to significant prominence in the passaged population. To determine the requirement and kinetics of event fixation, we sequenced the recombination site of 24 single colonies from strain YJB9929 from passage 411, immediately following recombination (*TLOα10/TLOγ8* → *TLOα10/TLOα10*). Twenty-three of 24 colonies had undergone recombination that homozygosed *TLOα10* at this time point (Figure S2), demonstrating that the detected recombination events rapidly dominate the population but are not completely fixed at the earliest time points.

Types of TLO recombination events

All of the evolved strains originated from the same genetic background with little time to accumulate unique, independent mutations prior to passaging. Thus, we assume that each strain and each *TLO* had an equal probability of recombination. Based on this assumption, there was no obvious effect of stress on either the number of strains that underwent *TLO* recombination (5 stressed *vs.* 5 unstressed) or the number of recombination events that occurred over the course of the passaging experiment (6 stressed *vs.* 10 unstressed). The distribution of recombination events among strains was random with the exception of a single strain (YJB9929) that underwent six separate events (Figure S3 and Table 1).

The distribution of *TLO*s involved in recombination events suggested specific patterns of recombination among the *TLO* genes. Recombination events were significantly clustered at certain chromosome arms, χ^2 ((N = 32) = 56.3; P = 0.0005; Monte Carlo). Interestingly, recombination was much more prevalent in α-clade *TLOs* than in *TLOs* of the β- and γ-clades.

We separated *TLO* genes that moved into two groups, the donor gene, which was copied, and the recipient gene, which was lost (Figure 3). In all cases, a copy of the donor gene remained unaltered at its native locus. *TLO* recombination occurred extensively between some *TLOs*/chromosome arms (*TLOα9*, *TLOα10*, and *TLOα12*) and was excluded from others (*TLOβ2*, *TLOγ4*, TLOγ*5*, and *TLOγ16*). Some loci were overrepresented as gained or as lost during *TLO* recombination (gained, χ^2 (N = 16) = 57.1; $P < 0.0005$; Monte Carlo); lost, χ^2 ((N = 16) = 62.0; $P < 0.0005$; Monte Carlo). For example, *TLOα12* always increased in copy number by expanding to other chromosome arms such as Chr4L, where *TLOα9* consequently reduced in copy number.

Two general types of recombination mechanisms were evident from *TLO* sequencing. The majority was ectopic recombination events; no recipient-specific sequence was seen telomere-proximal to the initial conversion tract. Thus, these were most likely due to BIR or gene conversion that extended to the end of the chromosome (Figure S1). No reciprocal genetic exchange was detected. The second set of *TLO* movements occurred due to loss of heterozygosity (LOH) of a previously heterozygous *TLO* locus following ectopic recombination for seven events (Table 1 and Table 2).

Table 1 *TLO* recombination events identified in passaged strains

Strain	Chr arm	Premovement	Postmovement	Recombination point	Recombination relative to CDS stop codon
YJB8648	4L	*TLOα9/TLOα9*	*TLOα12–α9/TLOα12–α9*	Chr4: 1612–1524	CDS 3′ end (−135, −47)
YJB8648s	4R	*TLOα9/TLOα10*	*TLOα12/TLOα12*	Chr4: 1597170–1597206	137–174 bp centromeric
YJB9907	4R	*TLOα9/TLOα10*	*TLOα12/TLOα12–α10*	Chr4: 1597488–1597612	CDS 3′ end (−265, −151)
YJB9915	4L	*TLOα9/TLOα9*	*TLOα9/TLOα12–α9*	Chr4: 1669–1519	9 nt centromeric
YJB9916s	3L	*TLOγ7/TLOγ7*	*TLOγ11/TLOγ11*	Chr3: 14321–14089	56 nt centromeric of the unspliced form
	4L	*TLOα9/TLOα9*	*TLOα12–α9/TLOα12–α9*	Chr4: 1614–1527	CDS 3′ end (−133, −45)
YJB9929	RL	*TLOα1/TLOα1*	*TLOα1/TLOγ13*	Centromeric	Centromeric
	1L	*TLOα3/TLOα3*	*TLOα3/TLOα1*	Chr1: 11547–11543	55–60 bp centromeric
	4L	*TLOα9/TLOα9*	*TLOα9/TLOα12*	Chr4: 1666–1516	6 nt 3′ to −144bp into 3′ of CDS
	4R	*TLOα9/TLOα10*	*TLOα10/TLOα10*	Centromeric	Centromeric
	4R	*TLOα10/TLOα10*	*TLOα10/TLOγ8*	Chr4: 1597146–1597167	191–170 bp centromeric
	4R	*TLOα10/TLOγ8*	*TLOα10/TLOα10*	Centromeric	Centromeric
YJB9929s	4L	*TLOα9/TLOα9*	*TLOα12–α9/TLOα12-α9*	Chr4: 1614–1527	CDS 3′ end (−133, −45)
YJB10777s	4L	*TLOα9/TLOα9*	*TLOα12–α9/TLOα12–α9*	Chr4: 1614–1527	CDS 3′ end (−133, −45)
YJB10779	4R	*TLOα9/TLOα10*	*TLOα9/TLOα9*	Chr4: 1597327–1597336	1–7 nt centromeric
YJB10780s	1L	*TLOα3/TLOα3*	*TLOα3/TLOγ11*	Chr1: 11695–11669	180 bp centromeric

The recipient *TLO* loci were completely lost in these seven events because the donor *TLO* was homozygosed at the recipient subtelomere, resulting in a net increase of two copies of the donor *TLO*. *TLOα10* was lost in two independent isolates, suggesting that it may be either particularly susceptible to invasion or more dispensable under the passaging conditions. Complete loss of the native *TLO* in the strain following homozygosis of a donor *TLO* was significantly more prevalent in stress-passaged than unstressed strains (83.3% *vs.* 14.3%, χ^2, d.f. = 1; $P = 0.024$; Fischer's exact test).

Gene loss following invasion and LOH can alter the *TLO* repertoire both at the gene and clade level. The majority of the recombination events (11/16) were between *TLOα* genes, one was between *TLOγ* genes, and four were interclade recombinations. Of the four interclade events, including the LOH events on Chr4R in YJB9929 (Figure 2B), *TLOα* and *TLOγ* were equally likely to be the donor or the recipient.

Sequence elements associated with subtelomeric recombination events

The prevalence of intraclade recombination events suggests that specific DNA regions may facilitate sequence-dependent recombination (Myers *et al.* 2008). To ask if recombination sites clustered at specific subtelomeric positions, we mapped the recombination site for each event using the junction sequence between the donor and recipient chromosome arm sequences (Figure 4A). Six of 14 recombination events occurred in the 3′ end of the *TLO* open reading frame (ORF), and two additional recombination events occurred immediately centromere-proximal of the 3′ UTR. All of these events were products of intraclade recombination, likely due to clade-specific sequence homology at the 3′ end of the *TLO* ORF. Three recombination events occurred within a highly conserved 50-bp sequence element found at all *TLO*-containing chromosome ends except for *TLOβ2* and *TLOγ16* and absent from chromosome arms lacking *TLO*s (Figure 4B and Figure S4). Previous attempts to assemble the subtelomeric regions from SC5314 whole-genome sequencing were problematic specifically due to the presence of this same sequence element that we termed the Bermuda Triangle sequence (BTS). Importantly, two of these events involved members of different *TLO* clades. Furthermore, the parental SC5314 background encodes a six nucleotide tandem duplication within the BTS that appeared at the same position that the *TLOα10* sequence on Chr4R became heterozygous for *TLOα9/TLOα10* (Figure S5, homolog 2 sequence). This suggests that recombination in the BTS produced the *TLOα9/TLOα10* heterozygous organization on Chr4R.

Sequencing of the subtelomeric region surrounding the *TLO* genes failed to identify a recombination site for 3 of the 16 gene movement events. To determine if recombination occurred centromere-proximal to the sequenced subtelomere, the heterozygosity of the altered chromosome was interrogated by SNP–RFLP using loci on each chromosome arm. In all three instances (YJB9929–RL, YJB9929–4R recombination 1, and YJB9929–4R recombination 3), heterozygosity was maintained across the SNP markers on both arms of the chromosome (Figure S6). Therefore, it is most likely that recombination occurred centromere-proximal to the subtelomere but distal to the SNP–RFLP markers.

Recombination within gene family ORFs can produce novel recombinant proteins that may have unique properties. In one case, a crossover within the 3′ ends of *TLOα9* and *TLOα12* produced a hybrid gene on Chr4L based on acquisition of a single SNP (T → C). This synonymous mutation altered the codon usage for Asp238 and the 3′ noncoding region of the gene was different (Figure 4C). Interestingly, five isolates underwent a very similar recombination event (Figure 4C, Table 1, and Table 2). Subsequently, the recombinant *TLO* was homozygosed in four of the strains, such that the *TLOα12–α9* hybrid became the only *TLO* at Chr4L (Table 1 and Table 2).

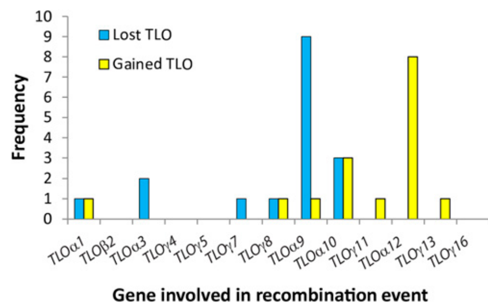


Figure 3 Inventory of *TLO* events. The *TLO* family members affected by subtelomeric recombination were separated into those lost (replaced by another paralog) and those gained (copied to other subtelomeric loci). Some *TLO*s were significantly overrepresented in the identified recombination events.

Table 2 Subtelomeric recombination mechanisms

Recombination mechanism	Movement
Crossover	YJB9907–4R
	YJB9929–4L
	YJB10780s–1L
Crossover, then LOH	YJB8648–4L
	YJB9916s–4L
	YJB9929s–4L
	YJB10777s–4L
Gene conversion	YJB9915–4L
	YJB9929–RL
	YJB9929–1L
	YJB9929–4R, 186 passages
Gene conversion, then LOH	YJB8648s–4R
	YJB9916s–3L
LOH	YJB9929–4R, 124 passages
	YJB9929–4R, 388 passages
	YJB10779–4R

The other case of inter-*TLO* recombination produced a *TLOα12–TLOα10* hybrid in strain YJB9907. In this case, a single crossover between codon 164 (265 nucleotides (nt) upstream of the stop codon) in *TLOα12* and codon 168 (151 nt upstream of the stop codon) in *TLOα10* resulted in a protein that differed from *TLOα12* by the loss of 37 amino acids in the gene-specific region of variable length and six new SNPs in the clade-specific, conserved C terminus, including two nonsynonymous codon changes (P203H and I209N) (Table S3 and Figure 4D). Thus, the recombination event produced a novel *TLOα* variant protein, *TLOα12–10*.

To ask if specific chromosome arms were biased toward specific recombination sites, we categorized recombination border locations for each subtelomere. We divided the chromosome arm into defined regions with different sequence properties represented by the *TLO* gene (~500–700 bp), the UTR immediately downstream of the *TLO* gene (UTR, ~300 bp), the BTS (50 bp), and any sequence centromere-proximal to the BTS, and inferred the frequency of recombination in each region within the passaged strains. All four regions underwent recombination with different distributions dependent on the chromosome arm (Figure 5). For example, five of the six recombination events on Chr4L occurred within the 3′ region of the *TLO* gene whereas recombination on Chr4R was more evenly divided among the four categories. This suggests that the types of subtelomeric regions prone to recombination differ depending on the chromosome arm.

TLO sequence evolution during passaging

Elevated levels of polymorphism in subtelomeres, in addition to recombination, contribute to gene evolution at chromosome ends (Cuomo *et al.* 2007; Dreszer *et al.* 2007). To investigate the potential for nucleotide polymorphisms to contribute to *TLO* gene evolution during passaging, we analyzed the *TLO* sequences for acquisition of new SNPs during passage. *TLOα* and *TLOγ* clade members accumulated 12 and 19 novel SNPs within the ORFs, respectively. In contrast, no new SNPs were identified in the single *TLOβ* family member, *TLOβ2*. The total number of polymorphisms and nonsynonymous changes in individual genes varied considerably (Figure 6A and Table S4); overall, there were similar numbers of total nonsynonymous polymorphisms within genes of the *TLOα* and the *TLOγ* clades (7 *vs.* 9, respectively). Thus, in contrast to the frequency of recombination (Figure 3), the frequency of SNP accumulation was similar between the two clades.

The *TLO* gene can be divided into three distinct regions with specific sequence properties. The N-terminal 300 nt encode the Med2 domain that provides the basis for assembly into the Mediator complex. This is followed by a short repetitive region of variable length that is gene specific and a C-terminal region of ~250–350 nt that is clade specific. In *TLOα* genes, new SNPs were primarily distributed across the clade-specific region evenly (Figure 6B); in the *TLOγ* genes, new SNPs were clustered at different points along the gene and were enriched in the Med2 domain. Therefore, there may be clade-specific selective pressures exerted on the *TLO* genes during laboratory passaging.

Previous investigation of sequence evolution among *Candida* species determined that centromeres evolve more rapidly than other loci (Padmanabhan *et al.* 2008). Thus, we compared the rate of evolution at *TLOs* relative to centromeres by interrogating an 800-bp segment of the central core of all eight *C. albicans* centromeres for SNPs that arose during passaging. Eleven SNPs accumulated across the eight centromere core regions in the nine isolates (total of 11 SNPs/12,800 bp (800 bp × 8 centromeres × 2 alleles). The distribution of mutations among the centromeres was highly skewed; a single centromere, *CEN2*, accumulated eight new mutations, *CENR* accumulated two mutations, and *CEN5* acquired a single mutation (Figure 6A and Table S5); in contrast, *TLO* genes acquired 31 novel mutations in 13 ORFs analyzed across a total of 15,970 bp (average of 614 bp × 13 telomeres × 2 alleles).

As a control for the evolution of expanded gene family members (Butler *et al.* 2009), we also analyzed 1600 bp

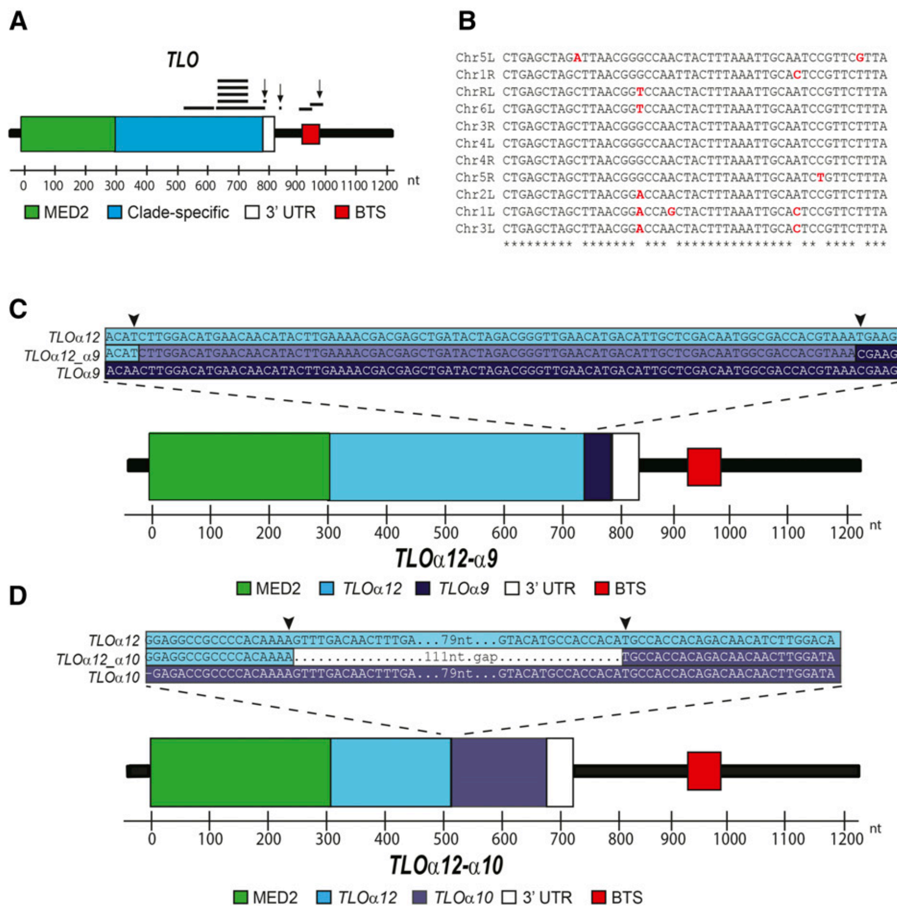


Figure 4 Recombination clusters at defined subtelomeric regions. (A) Conversion tracts were mapped onto a representation of the *C. albicans* subtelomere and enriched regions of recombination were identified in the C-terminal portion of the *TLO* coding sequence and the centromeric BTS. Arrows indicate specific sites of recombination and horizontal lines indicate a region within which a single recombination event occurred. *TLO* replacements resulting from LOH were not included in the map. (B) Alignment of the 11 subtelomeric BTS elements reveals strong sequence conservation. (C) Recombination between the *TLOα12* and *TLOα9* ORFs produced a hybrid *TLOα12–α9* sequence. Recombination occurred within a region of strong homology that is demarked by gene-specific SNPs and did not alter the Tloα12 protein sequence. The recombination region is shaded intermediate to either *TLO* gene. (D) Recombination between the *TLOα12* and *TLOα10* ORFs produced a hybrid *TLOα12–α10* sequence with a unique protein sequence. The novel Tloα12–α10 sequence is characterized by a 111-nt deletion at the site of recombination and the incorporation of nonsynonymous SNPs between the *TLO* genes.

(800 bp × 1 gene per strain × 2 alleles) within *LIP4*, a lipase within the *C. albicans* expanded *LIP* family, and two members of the expanded serine aspartyl protease (*SAP*) family, *SAP1* and *SAP7*. Sequencing detected a single SNP (C877A) within *LIP4* and two SNPs in each of the *SAP* ORFs (Figure 6A and Table S6). Additionally, the *ENO1* ORF, which encodes a metabolic enolase, did not acquire any novel SNPs following passaging of these strains (Figure 6A). Thus, the *TLO* genes as a group accumulated more mutations per locus and per base pair during the 457 passages (2.58 per *TLO and* 19.4/10 kbp) compared to the centromere core region (1.13 per centromere and 8.59/10 kbp) and other genes with multiple family members (1 per *LIP4* and 6.25/10 kbp and 2 per *SAP* and 12.5/10 kbp), indicating that telomeric genes have the potential to evolve at least as rapidly as the quickly evolving centromere regions and other expanded gene families that are acted on by host selection.

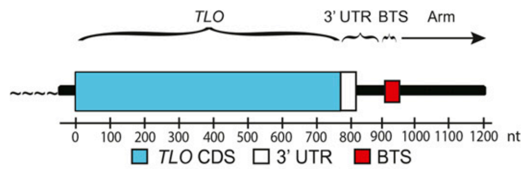
Phenotypic consequences of recombination

We next asked if the evolved strains that underwent subtelomeric recombination had acquired a fitness advantage under the passaging conditions using competitive fitness assays. We mixed equal cell concentrations of a related GFP-marked reference strain and one of several YJB9929 isolates from different passaging time points. We then allowed the cell mixtures to grow in liquid YPAD medium at 30° for 72 hr (∼30 generations). Comparison of final and initial frequencies of the two strains identified a significant increase in the competitive index of the evolved isolates after 1000 or more generations (Figure 7).

Discussion

Evolutionary histories of a wide range of species often point to subtelomeres as highly dynamic regions with rapid gene turnover (Kraemer *et al.* 2007; Linardopoulou *et al.* 2007; Anderson *et al.* 2008; Carreto *et al.* 2008; Dunn *et al.* 2012). The degree of change and the dynamics with which it occurs have been largely underappreciated, due to the complexity of repeats at subtelomeres, which obviates detailed analysis by next-generation sequencing. Here, we followed the dynamic evolution of subtelomeres over relatively short time scales in nine related isolates evolved with or without a mild temperature stress. We found that the genes in subtelomeric regions evolve rapidly, undergoing ectopic recombination events and acquiring 31 point mutations (1 per 515 bp) in the course of over ∼450 passages and ∼4500 cell divisions. Furthermore, some types of changes were recurrent and transient (*e.g.*, Chr4R changes in strain YJB9929; Figure 2 and Table 1), such that they likely would be missed by more static analyses.

Recombination resulting in the duplication of some *TLO* family genes and the loss of others was frequent, occurring



	TLO	3′ UTR	BTS	Arm
RL	0	0	0	1
1L	0	1	1	0
3L	0	1	0	0
4L	5	1	0	0
4R	1	1	2	2

Figure 5 Association of recombination site and chromosome arm. The region of recombination was defined for all recipient chromosome arms as either within the *TLO*, between the *TLO* and BTS (Mid), within the BTS, or centromeric of the BTS (Cen). Binning of the number of recombination events within each region for a given chromosome arm revealed bias in the recombination site for certain chromosome arms.

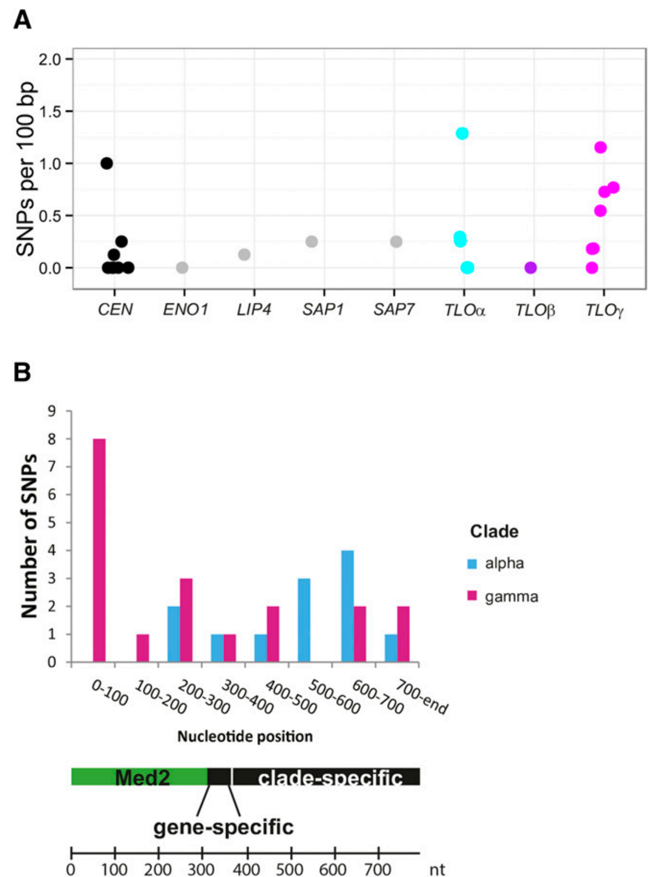


Figure 6 SNP accumulation during passaging. (A) SNPs arising during strain propagation were identified by comparison to the progenitor strain. The number of SNPs per 100 bp for each subtelomeric *TLO* (5 for *TLOα*, 1 for *TLOβ*, and 7 for *TLOγ*), eight centromere core regions (*CEN*), *ENO1*, *LIP4*, *SAP1*, and *SAP7* were plotted as individual data points for each locus. (B) Identified SNPs were grouped by clade of origin and plotted along the length of the *TLO* coding sequence. *TLOγ* clade SNPs clustered toward the Med2 domain at the 5′ end and *TLOα* SNPs arose primarily in the clade-specific region at the 3′ end of the gene.

once per ∼5000 divisions. Specific *TLO*s preferentially increased or decreased in copy number during recombination. Not surprisingly, exchanges between genes within the same clade were more frequent (75%) than interclade recombination events. Recombination was most frequent within the C-terminal clade-specific regions of *TLO* ORFs or within the highly similar BTS sequence element (88% similarity across 11 subtelomeres) and occurred via either LOH or ectopic crossover events. Point mutations accumulated across the *TLO* coding sequences and were clustered toward the N-terminal Med2 domain in *TLOγ* clade genes and toward the 3′ clade-specific region in *TLOα* clade genes (Figure 4).

The rates of subtelomeric recombination measured here are likely underestimates of recombination frequency at *C. albicans* subtelomeres. Recombination events that arise at low frequency are not captured when screening at the population level. Thus, only fixed or high-frequency recombination events could be used to calculate subtelomeric recombination rates. Furthermore, rare recombination via translocation might occur between subtelomeres and chromosome-internal positions (*i.e.*, *TLOα34*) and would not have been detected in these experiments (Fan *et al.* 2008).

Subtelomeric regions typically evolve at faster rates (Dreszer *et al.* 2007) and are the sites of more frequent gene duplication events (Ames *et al.* 2010) than other genomic regions. Accordingly, subtelomeres have been proposed to be "hotbeds" for genomic evolution (Brown *et al.* 2010). *C. albicans* subtelomeres appear to undergo a similar process of sampling evolutionary space through recombination, gene duplication, and mutation. *TLOs* are predicted to encode the same Med2 protein that is incorporated, as a monomer, into the Mediator tail and thus to regulate transcription of downstream gene sets (Ansari *et al.* 2012). Expansion of some *TLO* paralogs relative to others allows drift to operate on that gene and can result in the establishment of new interacting partners without disrupting the molecular role of the original gene. Novel *TLO* genes, produced either by mutation or recombination, result in the same molecular outcome but are likely under heavier selective pressure to be retained. We suggest that together these novel functions may modify Mediator interactions that significantly alter transcriptional responses and the phenotypes that they proscribe (Haran *et al.* 2014).

Selection appears to favor some *TLO* family genes over others. For example, two *TLO*s, *TLOγ7* and *TLOα10*, were completely lost from the genome during passaging of some strains, while *TLOα12* was gained eight times during recombination and was subsequently homozygosed in six independent isolates. Interestingly, the copy number of *TLOα9* was reduced in eight separate events, yet it was never lost: all strains encoded at least one copy of *TLOα9* at the end of the passaging experiment. Therefore, the presence of some *TLOα9* might provide a selective role in passaged strains.

Most of the observed recombination events occurred between *TLOα* clade members, which fractionate as a single stoichiometric component of the Mediator complex in

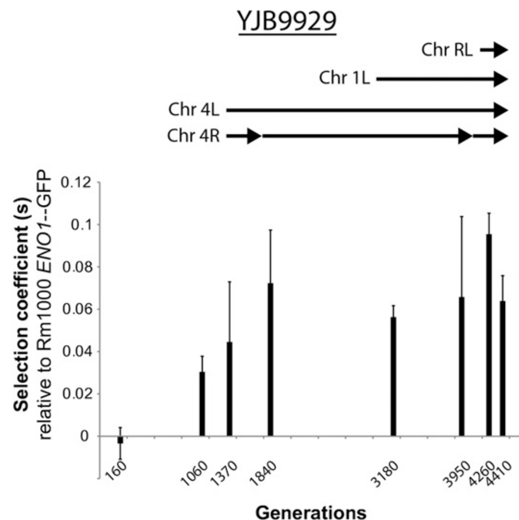


Figure 7 Increased growth rates during strain passaging. Passaged isolates from several time points associated with subtelomeric recombination were assayed for fitness using competition assays against a GFP-marked reference strain. The specific recombination events that arose in this passaged strain are illustrated for each chromosome arm by an arrow that is aligned with the time points on the *x*-axis below. As expected, relative fitness, as measured by the selection coefficient, increased with continued passaging.

C. albicans (Zhang *et al.* 2012). Thus, the recombination events observed likely alter the relative availability of different *TLOα* paralogs for incorporation into Mediator. While paralogous Mediator subunits have distinct transcriptional patterns and interacting proteins in other organisms (Tsutsui *et al.* 2008; Fukasawa *et al.* 2012), the results here suggest that some *TLO Med2 copies* are dispensable, while others are favored for *C. albicans* growth, at least under the *in vitro* passaging conditions used here.

A trend toward fixation of the recombination events identified here occurred rapidly. There was no strong evidence in the sequence traces for mixed populations composed of an ancestral and recently invaded subtelomeric population. Furthermore, targeted sequencing of individual colonies following recombination found only a single instance of the ancestral subtelomeric configuration (Figure S2). Additionally, we detected homozygosis of seven recombination events within 160 generations (1 month) of the initial invasion. Consistent with a model of constant recombination between homologous chromosome ends, most new mutations became homozygous following continued passaging (data not shown). Recombination rates are elevated near telomere across many eukaryotes (Linardopoulou *et al.* 2005; Rudd *et al.* 2007) and homozygosity may be a by-product of constant subtelomeric recombination between sister homologs. However, it is unclear if this phenomenon is a common feature in other organisms, as all previously published experiments studied tractable subtelomeric evolution in haploid organisms (Bopp *et al.* 2013).

In *S. cerevisiae*, different classes of subtelomeric genes are amplified in strains grown on different sugar sources (*e.g.*, maltose, sucrose, or melibiose) (Naumov *et al.* 1995; Dujon *et al.* 2004; Naumova *et al.* 2005; Brown *et al.* 2010). In this study *C. albicans* was grown under standard laboratory conditions (30°, rich medium) with and without occasional mild temperature stress. While it is possible that this passaging regime selected for genome evolution and *TLO* recombination, the *C. albicans* subtelomeres studied here were under selective pressures very different from, and likely less stressful than, those they encounter in their mammalian hosts. Therefore, the rates of recombination and mutation acting on *C. albicans* subtelomeres may be different in the mammalian host relative to those *in vitro*, but the relative amount of recombination that occurs at telomeres is likely higher than at other loci. Recombination frequency differences between specific subtelomeres may be influenced by the presence of recombination hotspots (Blitzblau *et al.* 2007) and sequence conservation among subsets of chromosome arms.

TLO recombination was most frequent within the C-terminal third of the *TLO* ORFs, a domain that includes clade-specific sequences. This is consistent with the higher frequency of recombination events among *TLO* paralogs with greater sequence homology (intraclade *vs.* interclade recombination). The second most frequent site for recombination between different *TLOs* was the BTS, which is 3′ (centromere-proximal) to the *TLO* ORFs and was named the Bermuda Triangle sequence because this sequence is very similar on 11 chromosomes arms and complicates the determination of the chromosome arm location of a given *TLO* gene (Figure 4B). Neither of these two sites of *TLO* recombination overlaps with recombination-prone features, such as transposable elements or tandemly repeated sequences, that can promote recombination at other loci (Rehmeyer *et al.* 2006; Boothroyd *et al.* 2009; Gemayel *et al.* 2010).

BTS recombination events should result in copying of one gene to a new chromosome arm, without affecting the *TLO* coding sequence. The benefit of moving a *TLO* to another chromosome arm is not clear, nor is it clear to what degree different subtelomeric positions affect the level of expression of a given *TLO* gene. In fact, genes expressed at different *TLO* loci appear to be affected more by Sir2p-mediated silencing and telomere-associated gene expression noise than by being located at a specific telomere (Anderson *et al.* 2014). Importantly, recombination in the BTS accounted for two of three interclade recombination events and may contribute more heavily to major alterations in the *TLO* repertoire over time compared to recombination at other points. Furthermore, the BTS is specific to *C. albicans*, which may be connected to the mechanism of *TLO* amplification in *C. albicans*. While we did not detect any amplification events or introduction of *TLOs* to chromosome arms that did not have *TLOs* or a BTS previously, we presume that such nonhomologous events are relatively rare.

In *C. albicans*, little is known about the subtelomeric recombination sites. Recombination within subtelomeric gene ORFs has been observed for Y′ genes in *S. cerevisiae* (Louis and Haber 1990a,b) and for the surface antigen ORFs in

parasitic protozoa (Jiang *et al.* 2011; Bopp *et al.* 2013). However, in protozoa, 70-bp sequences flanking subtelomeric surface antigen genes promote intergenic recombination (Boothroyd *et al.* 2009). Accordingly, we propose that the 50-bp BTS element may be analogous to the 70-bp motif that promotes homologous recombination between different subtelomeres in *Trypanosoma brucei* (Boothroyd *et al.* 2009).

A major mechanism for the construction of novel genes is recombination, as inferred from phylogenetic reconstruction of intergenic recombination events, like those that produced the *TLOα9/α12* chimeras (Kraemer *et al.* 2007; Kyes *et al.* 2007; Sander *et al.* 2014). Both of the major sites of *TLO* recombination encode physically unstable (TAA) repeats, which can promote break-induced replication after DNA replication stalling (Ohshima *et al.* 1996; Barry and Mcculloch 2001). The gain and loss of *TLO* genes and the sequence conservation between chromosome arms at the site of the crossover event strongly suggests that ectopic recombination (Haber 2000) produced most *TLO* recombination events. The absence of reciprocal exchange and the loss of nucleotide information from the recipient *TLO* suggest that break-induced replication/crossover or gene conversion is likely the major mechanisms responsible for *TLO* recombination. Consistent with this, BIR between subtelomeric Y′ elements is a common ALT mechanism in telomere replication and repair in yeasts (Lundblad and Blackburn 1993; Teng *et al.* 2000) and leads to elevated mutation rates through error-prone DNA synthesis machinery (Deem *et al.* 2011). Although close proximity to the poorly characterized chromosome end makes it difficult to differentiate between gene conversion and BIR/XO, both mechanisms would result in the same general phenomenon: the apparent movement of *TLO* genes via loss of one *TLO* gene copy and expansion of another.

C. albicans and C. dubliniensis diverged ∼20 MYA and *C. dubliniensis* has only two *TLO* genes. Thus, the *TLO* gene family expansion appears to have been a recent event in *C. albicans*. The expansion of gene families at chromosome ends is associated with a high prevalence of pseudogene fragments scattered across subtelomeric regions (Louis and Haber 1992; Gardner *et al.* 2002; Marcello and Barry 2007). *C. albicans* has only a single *TLO* pseudogene (Anderson *et al.* 2012), which appears to have been produced through insertion of a long terminal repeat retrotransposon, rather than by recombination or cumulative mutation (Anderson *et al.* 2012). The lack of extensive, degenerate *TLO* copies is consistent with the recent origin of the *TLO* expansion but could also be due to the acquisition of nonredundant functions among *TLO* paralogs. It is tempting to speculate that the bias in expansion and contraction of some family members reflects specialized roles for individual *TLO* genes in shaping the Mediator activity within the cell. This view is consistent with previous work that subtelomeric evolution facilitates adaptation to new environmental conditions in other eukaryotes, including other yeasts (Carreto *et al.* 2008; Chuma *et al.* 2011; Haran *et al.* 2014). It also suggests that analysis of *TLO* changes in strains isolated from different host niches may offer insights into the functional specialization of specific *TLO* genes.

Acknowledgements

We are grateful to Hannah Boyle for assistance in strain passaging and to Aleeza Gerstein for insight into gene bias. We thank Martin Kupiec for valuable comments on the manuscript and all members of the Berman lab for many helpful discussions and suggestions. This work was supported by the Israel Science Foundation Grant (340/13), and by the National Institute of Allergy and Infectious Diseases (AI075096-0351) to J.B. and a Research Supplement to Promote Diversity in Health-Related Research award to M.Z.A.

Literature Cited

Ames, R. M., B. M. Rash, K. E. Hentges, D. L. Robertson, D. Delneri *et al.*, 2010 Gene duplication and environmental adaptation within yeast populations. Genome Biol. Evol. 2: 591—601.

Anderson, J. A., Y. S. Song, and C. H. Langley, 2008 Molecular population genetics of Drosophila subtelomeric DNA. Genetics 178: 477—487.

Anderson, M. Z., J. A. Baller, K. Dulmage, L. Wigen, and J. Berman, 2012 The three clades of the telomere-associated TLO gene family of *Candida albicans* have different splicing, localization, and expression features. Eukaryot. Cell 11: 1268—1275.

Anderson, M. Z., A. C. Gerstein, L. Wigen, J. A. Baller, and J. Berman, 2014 Silencing is noisy: population and cell level noise in telomere-adjacent genes is dependent on telomere position and sir2. PLoS Genet. 10: e1004436.

Ansari, S. A., M. Ganapathi, J. J. Benschop, F. C. Holstege, J. T. Wade *et al.*, 2012 Distinct role of Mediator tail module in regulation of SAGA-dependent, TATA-containing genes in yeast. EMBO J. 31: 44—57.

Barry, J. D., and R. McCulloch, 2001 Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. Adv. Parasitol. 49: 1—70.

Blitzblau, H. G., G. W. Bell, J. Rodriguez, S. P. Bell, and A. Hochwagen, 2007 Mapping of meiotic single-stranded DNA reveals double-stranded-break hotspots near centromeres and telomeres. Curr. Biol. 17: 2003—2012.

Boothroyd, C. E., O. Dreesen, T. Leonova, K. I. Ly, L. M. Figueiredo *et al.*, 2009 A yeast-endonuclease-generated DNA break induces antigenic switching in *Trypanosoma brucei*. Nature 459: 278—281.

Bopp, S. E., M. J. Manary, A. T. Bright, G. L. Johnston, N. V. Dharia *et al.*, 2013 Mitotic evolution of *Plasmodium falciparum* shows a stable core genome but recombination in antigen families. PLoS Genet. 9: e1003293.

Brown, C. A., A. W. Murray, and K. J. Verstrepen, 2010 Rapid expansion and functional divergence of subtelomeric gene families in yeasts. Curr. Biol. 20: 895—903.

Butler, G., M. D. Rasmussen, M. F. Lin, M. A. Santos, S. Sakthikumar *et al.*, 2009 Evolution of pathogenicity and sexual reproduction in eight *Candida genomes*. Nature 459: 657—662.

Carreto, L., M. F. Eiriz, A. C. Gomes, P. M. Pereira, D. Schuller *et al.*, 2008 Comparative genomics of wild type yeast strains unveils important genome diversity. BMC Genomics 9: 524.

Celenza, J. L., and M. Carlson, 1985 Rearrangement of the genetic map of chromosome VII of *Saccharomyces cerevisiae*. Genetics 109: 661—664.

Choi, J. K., S. Hwang, and Y. J. Kim, 2008 Stochastic and regulatory role of chromatin silencing in genomic response to environmental changes. PLoS ONE 3: e3002.

Christiaens, J. F., S. E. Van Mulders, J. Duitama, C. A. Brown, M. G. Ghequire *et al.*, 2012 Functional divergence of gene duplicates through ectopic recombination. EMBO Rep. 13: 1145—1151.

Chuma, I., C. Isobe, Y. Hotta, K. Ibaragi, N. Futamata *et al.*, 2011 Multiple translocation of the AVR-Pita effector gene among chromosomes of the rice blast fungus *Magnaporthe oryzae* and related species. PLoS Pathog. 7: e1002147.

Claessens, A., W. L. Hamilton, M. Kekre, T. D. Otto, A. Faizullabhoy *et al.*, 2014 Generation of antigenic diversity in *Plasmodium falciparum* by structured rearrangement of Var genes during mitosis. PLoS Genet. 10: e1004812.

Corcoran, L. M., J. K. Thompson, D. Walliker, and D. J. Kemp, 1988 Homologous recombination within subtelomeric repeat sequences generates chromosome size polymorphisms in *P. falciparum*. Cell 53: 807—813.

Cuomo, C. A., U. Guldener, J. R. Xu, F. Trail, B. G. Turgeon *et al.*, 2007 The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science 317: 1400—1402.

Deem, A., A. Keszthelyi, T. Blackgrove, A. Vayl, B. Coffey *et al.*, 2011 Break-induced replication is highly inaccurate. PLoS Biol. 9: e1000594.

Dreszer, T. R., G. D. Wall, D. Haussler, and K. S. Pollard, 2007 Biased clustered substitutions in the human genome: the footprints of male-driven biased gene conversion. Genome Res. 17: 1420—1430.

Dujon, B., D. Sherman, G. Fischer, P. Durrens, S. Casaregola *et al.*, 2004 Genome evolution in yeasts. Nature 430: 35—44.

Dunn, B., C. Richter, D. J. Kvitek, T. Pugh, and G. Sherlock, 2012 Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. Genome Res. 22: 908—924.

Edgar, R. C., 2004 MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113.

Fan, C., Y. Zhang, Y. Yu, S. Rounsley, M. Long *et al.*, 2008 The subtelomere of *Oryza sativa* chromosome 3 short arm as a hot bed of new gene origination in rice. Mol. Plant 1: 839—850.

Farman, M. L., and Y. S. Kim, 2005 Telomere hypervariability in *Magnaporthe oryzae*. Mol. Plant Pathol. 6: 287—298.

Fischer, K., P. Horrocks, M. Preuss, J. Wiesner, S. Wunsch *et al.*, 1997 Expression of var genes located within polymorphic subtelomeric domains of *Plasmodium falciparum* chromosomes. Mol. Cell. Biol. 17: 3679—3686.

Forche, A., M. Steinbach, and J. Berman, 2009 Efficient and rapid identification of *Candida albicans* allelic status using SNP-RFLP. FEMS Yeast Res. 9: 1061—1069.

Fukasawa, R., T. Tsutsui, Y. Hirose, A. Tanaka, and Y. Ohkuma, 2012 Mediator CDK subunits are platforms for interactions with various chromatin regulatory complexes. J. Biochem. 152: 241—249.

Gardner, M. J., N. Hall, E. Fung, O. White, M. Berriman *et al.*, 2002 Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419: 498—511.

Gaut, B. S., S. I. Wright, C. Rizzon, J. Dvorak, and L. K. Anderson, 2007 Recombination: an underappreciated factor in the evolution of plant genomes. Nat. Rev. Genet. 8: 77—84.

Gemayel, R., M. D. Vinces, M. Legendre, and K. J. Verstrepen, 2010 Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annu. Rev. Genet. 44: 445—477.

Glover, L., S. Alsford, and D. Horn, 2013 DNA break site at fragile subtelomeres determines probability and mechanism of antigenic variation in African trypanosomes. PLoS Pathog. 9: e1003260.

Gottschling, D. E., O. M. Aparicio, B. L. Billington, and V. A. Zakian, 1990 Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. Cell 63: 751—762.

Gresham, D., M. M. Desai, C. M. Tucker, H. T. Jenq, D. A. Pai *et al.*, 2008 The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genet. 4: e1000303.

Haber, J. E., 2000 Partners and pathwaysrepairing a double-strand break. Trends Genet. 16: 259—264.

Hahn, M. W., T. De Bie, J. E. Stajich, C. Nguyen, and N. Cristianini, 2005 Estimating the tempo and mode of gene family evolution from comparative genomic data. Genome Res. 15: 1153—1160.

Halliwell, S. C., M. C. Smith, P. Muston, S. L. Holland, and S. V. Avery, 2012 Heterogeneous expression of the virulence-related adhesin Epa1 between individual cells and strains of the pathogen *Candida glabrata*. Eukaryot. Cell 11: 141—150.

Haran, J., H. Boyle, K. Hokamp, T. Yeomans, Z. Liu *et al.*, 2014 Telomeric ORFs (TLOs) in Candida spp. encode mediator subunits that regulate distinct virulence traits. PLoS Genet. 10: e1004658.

Hirakawa, M. P., D. A. Martinez, S. Sakthikumar, M. Anderson, A. Berlin *et al.*, 2014 Genetic and phenotypic intra-species variation in *Candida albicans*. Genome Res. 25: 413–25.

Hoffman, C. S., and F. Winston, 1987 A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. Gene 57: 267—272.

Hovel-Miner, G. A., C. E. Boothroyd, M. Mugnier, O. Dreesen, G. A. Cross *et al.*, 2012 Telomere length affects the frequency and mechanism of antigenic variation in *Trypanosoma brucei*. PLoS Pathog. 8: e1002900.

Corp, IBM., 2013 *IBM SPSS Statistics for Windows*. IBM Corp., Armonk, NY.

Jackson, A. P., J. A. Gamble, T. Yeomans, G. P. Moran, D. Saunders *et al.*, 2009 Comparative genomics of the fungal pathogens *Candida dubliniensis* and *Candida albicans*. Genome Res. 19: 2231—2244.

Jiang, H., N. Li, V. Gopalan, M. M. Zilversmit, S. Varma *et al.*, 2011 High recombination rates and hotspots in a *Plasmodium falciparum* genetic cross. Genome Biol. 12: R33.

Jones, Jr., S. K., M. P. Hirakawa, and R. J. Bennett, 2014 Sexual biofilm formation in *Candida tropicalis* opaque cells. Mol. Microbiol. 92: 383—398.

Jones, T., N. A. Federspiel, H. Chibana, J. Dungan, S. Kalman *et al.*, 2004 The diploid genome sequence of *Candida albicans*. Proc. Natl. Acad. Sci. USA 101: 7329—7334.

Kasuga, T., G. Mannhaupt, and N. L. Glass, 2009 Relationship between phylogenetic distribution and genomic features in *Neurospora crassa*. PLoS ONE 4: e5286.

Kellis, M., N. Patterson, M. Endrizzi, B. Birren, and E. S. Lander, 2003 Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423: 241—254.

Kraemer, S. M., S. A. Kyes, G. Aggarwal, A. L. Springer, S. O. Nelson *et al.*, 2007 Patterns of gene recombination shape var gene repertoires in *Plasmodium falciparum*: comparisons of geographically diverse isolates. BMC Genomics 8: 45.

Kupiec, M., 2014 Biology of telomeres: lessons from budding yeast. FEMS Microbiol. Rev. 38: 144—171.

Kyes, S. A., S. M. Kraemer, and J. D. Smith, 2007 Antigenic variation in *Plasmodium falciparum*: gene organization and regulation of the var multigene family. Eukaryot. Cell 6: 1511—1520.

Lang, G. I., and A. W. Murray, 2008 Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. Genetics 178: 67—82.

Linardopoulou, E. V., E. M. Williams, Y. Fan, C. Friedman, J. M. Young *et al.*, 2005 Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. Nature 437: 94—100.

Linardopoulou, E. V., S. S. Parghi, C. Friedman, G. E. Osborn, S. M. Parkhurst *et al.*, 2007 Human subtelomeric WASH genes encode a new subclass of the WASP family. PLoS Genet. 3: e237.

Louis, E. J., and J. E. Haber, 1990a Mitotic recombination among subtelomeric Y' repeats in *Saccharomyces cerevisiae*. Genetics 124: 547—559.

Louis, E. J., and J. E. Haber, 1990b The subtelomeric Y' repeat family in *Saccharomyces cerevisiae*: an experimental system for repeated sequence evolution. Genetics 124: 533—545.

Louis, E. J., and J. E. Haber, 1992 The structure and evolution of subtelomeric Y' repeats in *Saccharomyces cerevisiae*. Genetics 131: 559—574.

Louis, E. J., E. S. Naumova, A. Lee, G. Naumov, and J. E. Haber, 1994 The chromosome end in yeast: its mosaic nature and influence on recombinational dynamics. Genetics 136: 789—802.

Lundblad, V., and E. H. Blackburn, 1993 An alternative pathway for yeast telomere maintenance rescues est1- senescence. Cell 73: 347—360.

Lynch, M., W. Sung, K. Morris, N. Coffey, C. R. Landry *et al.*, 2008 A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc. Natl. Acad. Sci. USA 105: 9272—9277.

Marcello, L., and J. D. Barry, 2007 Analysis of the VSG gene silent archive in *Trypanosoma brucei* reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. Genome Res. 17: 1344—1352.

Mefford, H. C., and B. J. Trask, 2002 The complex structure and dynamic evolution of human subtelomeres. Nat. Rev. Genet. 3: 91—102.

Mondoux, M. A., and V. A. Zakian, 2007 Subtelomeric elements influence but do not determine silencing levels at *Saccharomyces cerevisiae* telomeres. Genetics 177: 2541—2546.

Morrison, L. J., L. Marcello, and R. McCulloch, 2009 Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity. Cell. Microbiol. 11: 1724—1734.

Myers, S., C. Freeman, A. Auton, P. Donnelly, and G. McVean, 2008 A common sequence motif associated with recombination hot spots and genome instability in humans. Nat. Genet. 40: 1124—1129.

Naumov, G. I., E. S. Naumova, and M. P. Korhola, 1995 Chromosomal polymorphism of MEL genes in some populations of *Saccharomyces cerevisiae*. FEMS Microbiol. Lett. 127: 41—45.

Naumova, E. S., V. Ivannikova Iu, and G. I. Naumov, 2005 Genetic differentiation of the sherry yeasts *Saccharomyces cerevisiae*. Prikl. Biokhim. Mikrobiol. 41: 656—661.

Ohshima, K., S. Kang, J. E. Larson, and R. D. Wells, 1996 TTA. TAA triplet repeats in plasmids form a non-H bonded structure. J. Biol. Chem. 271: 16784—16791.

Padmanabhan, S., J. Thakur, R. Siddharthan, and K. Sanyal, 2008 Rapid evolution of Cse4p-rich centromeric DNA sequences in closely related pathogenic yeasts, *Candida albicans* and *Candida dubliniensis*. Proc. Natl. Acad. Sci. USA 105: 19797—19802.

Park, Y. N., K. J. Daniels, C. Pujol, T. Srikantha, and D. R. Soll, 2013 *Candida albicans* forms a specialized "sexual" as well as "pathogenic" biofilm. Eukaryot. Cell 12: 1120—1131.

Pryde, F. E., H. C. Gorham, and E. J. Louis, 1997 Chromosome ends: all the same under their caps. Curr. Opin. Genet. Dev. 7: 822—828.

R Development Core Team, 2013 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rehmeyer, C., W. Li, M. Kusaba, Y. S. Kim, D. Brown *et al.*, 2006 Organization of chromosome ends in the rice blast fungus, *Magnaporthe oryzae*. Nucleic Acids Res. 34: 4685—4701.

Rudd, M. K., C. Friedman, S. S. Parghi, E. V. Linardopoulou, L. Hsu *et al.*, 2007 Elevated rates of sister chromatid exchange at chromosome ends. PLoS Genet. 3: e32.

Sander, A. F., T. Lavstsen, T. S. Rask, M. Lisby, A. Salanti *et al.*, 2014 DNA secondary structures are associated with recombination in major *Plasmodium falciparum* variable surface antigen gene families. Nucleic Acids Res. 42: 2270—2281.

Teng, S. C., J. Chang, B. McCowan, and V. A. Zakian, 2000 Telomerase-independent lengthening of yeast telomeres occurs by an abrupt Rad50p-dependent, Rif-inhibited recombinational process. Mol. Cell 6: 947—952.

Thompson, D. A., M. M. Desai, and A. W. Murray, 2006 Ploidy controls the success of mutators and nature of mutations during budding yeast evolution. Curr. Biol. 16: 1581—1590.

Tsutsui, T., H. Umemura, A. Tanaka, F. Mizuki, Y. Hirose *et al.*, 2008 Human mediator kinase subunit CDK11 plays a negative role in viral activator VP16-dependent transcriptional regulation. Genes Cells 13: 817—826.

Turakainen, H., G. Naumov, E. Naumova, and M. Korhola, 1993 Physical mapping of the MEL gene family in *Saccharomyces cerevisiae*. Curr. Genet. 24: 461—464.

van het Hoog, M., T. J. Rast, M. Martchenko, S. Grindle, D. Dignard *et al.*, 2007 Assembly of the *Candida albicans* genome into sixteen supercontigs aligned on the eight chromosomes. Genome Biol. 8: R52.

Van Mulders, S. E., M. Ghequire, L. Daenen, P. J. Verbelen, K. J. Verstrepen *et al.*, 2010 Flocculation gene variability in industrial brewer's yeast strains. Appl. Microbiol. Biotechnol. 88: 1321—1331.

Wapinski, I., A. Pfeffer, N. Friedman, and A. Regev, 2007 Natural history and evolutionary principles of gene duplication in fungi. Nature 449: 54—61.

Wenger, J. W., J. Piotrowski, S. Nagarajan, K. Chiotti, G. Sherlock *et al.*, 2011 Hunger artists: yeast adapted to carbon limitation show trade-offs under carbon sufficiency. PLoS Genet. 7: e1002202.

Winzeler, E. A., C. I. Castillo-Davis, G. Oshiro, D. Liang, D. R. Richards *et al.*, 2003 Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. Genetics 163: 79—89.

Zhang, A., K. O. Petrov, E. R. Hyun, Z. Liu, S. A. Gerber *et al.*, 2012 The Tlo proteins are stoichiometric components of *Candida albicans* mediator anchored via the Med3 subunit. Eukaryot. Cell 11: 874—884.

Zhang, A., Z. Liu, and L. C. Myers, 2013 Differential regulation of white-opaque switching by individual subunits of *Candida albicans* mediator. Eukaryot. Cell 12: 1293—1304.

Communicating editor: J. Heitman

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.177451/-/DC1

Real-Time Evolution of a Subtelomeric Gene Family in *Candida albicans*

Matthew Z. Anderson, Lauren J. Wigen, Laura S. Burrack, and Judith Berman

Copyright © 2015 by the Genetics Society of America
DOI: 10.1534/genetics.115.177451

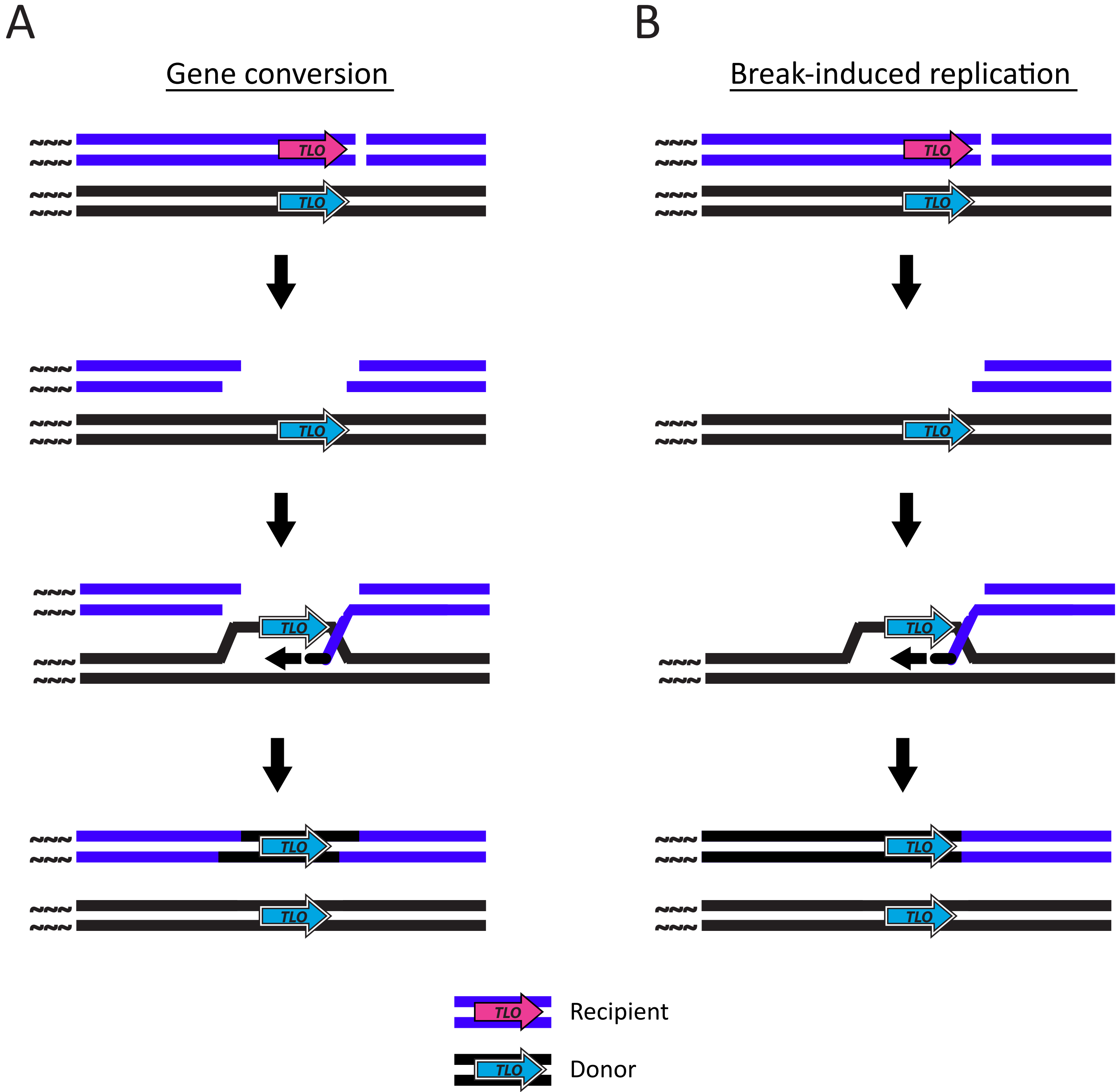


Figure S1 Mechanisms of *TLO* recombination. Subtelomeric double strand breaks in a chromosome induce repair mechanisms that may occur in different ways. During gene conversion (A), the break is repaired using an ectopic paralogous sequence found at another subtelomere. Resectioning during repair can result in loss of the native *TLO* by replacement with the *TLO* from another chromosome. The end of the broken chromosome is lost during break induced replication (B), with an ectopic chromosome end serving as a template to restore the missing chromosome end.

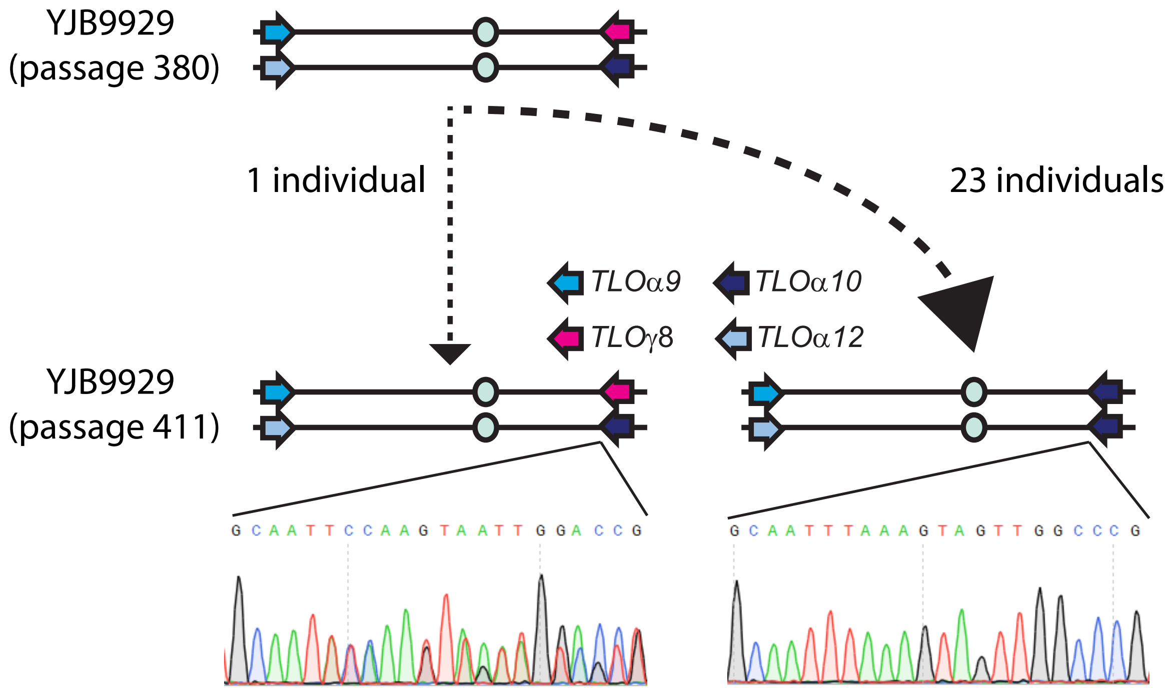


Figure S2 Kinetics of recombination events. Sequencing of the YJB9929 population identified recombination on Chr4R between passage 380 and 411. Interrogation of Chr4R by PCR and sequencing in 24 individual colonies from passage 411 revealed that 23 of the 24 colonies contained the recombination event seen in the passage 411 population and the remaining colony retained the TLO configuration seen in passage 380.

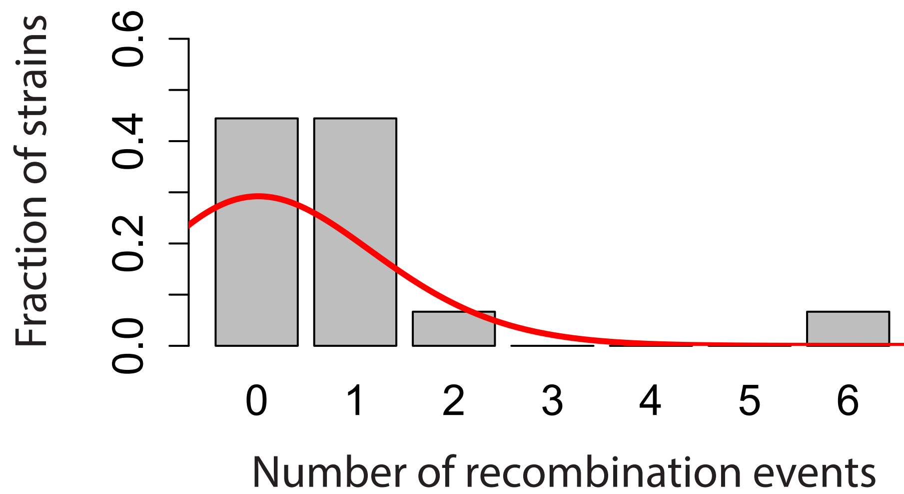


Figure S3 Distribution of recombination events. The fraction of strains encoding different numbers of recombination events was plotted (grey bars) and followed a Poisson distribution centered on the average number of recombination events per strain (0.5). One strain, YJB9929, deviated from the random distribution and encoded six separate recombination events.

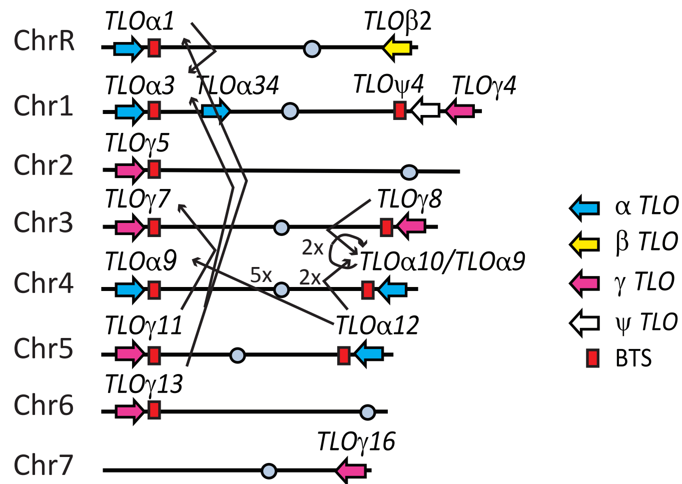


Figure S4 Organization of BTS elements in the *C. albicans* genome. A cartoon depicts the location of the thirteen subtelomeric *TLO* genes and the BTS sequence element in the parent strain for all passaged isolates. *TLO* genes and the BTS are color coded by clade as indicated.

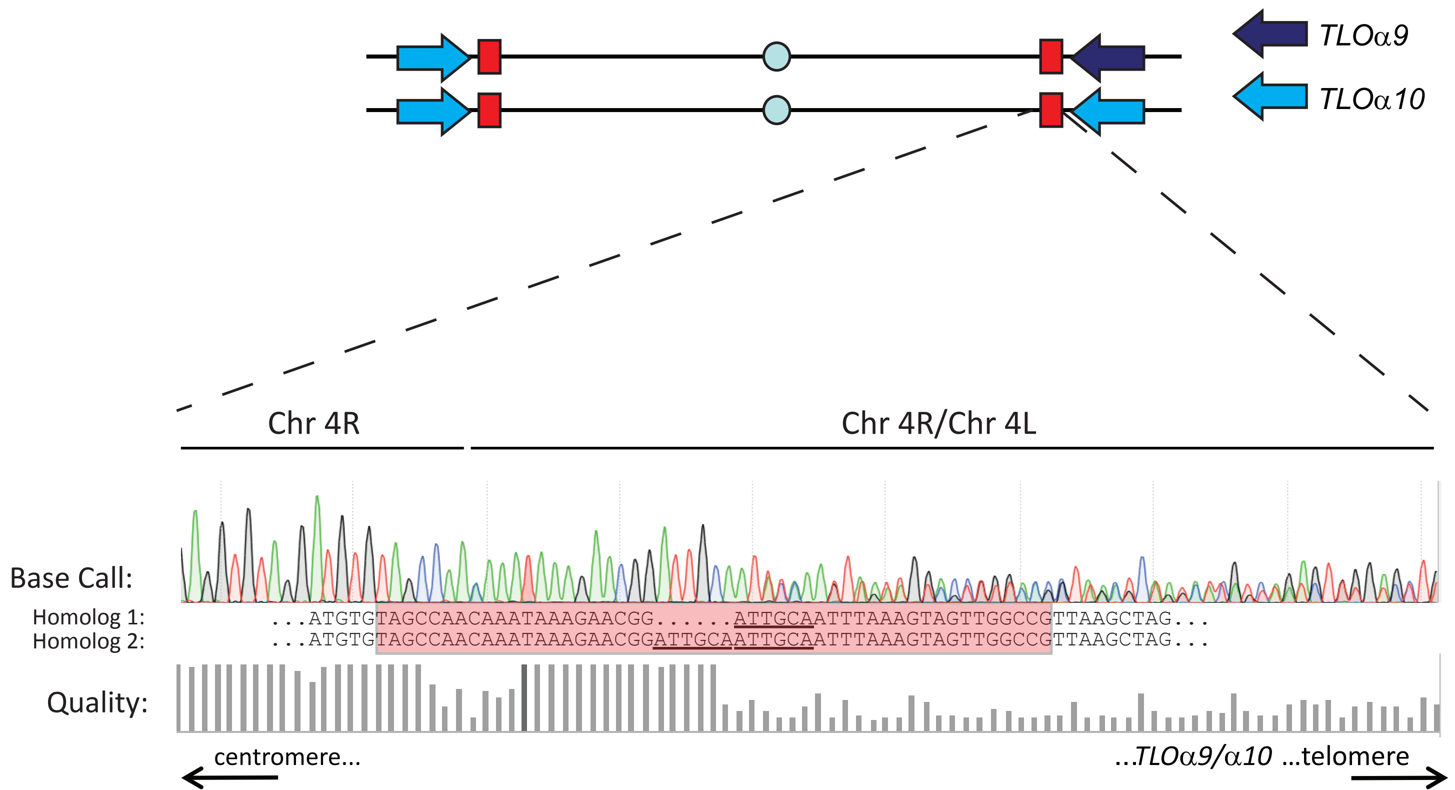


Figure S5 SC5314 encodes allelic *TLOα9* and *TLOα10* on Chr4R. The *TLO* locus on Chr4R in the parental isolates encodes a heterozygous locus with *TLOα9* on one homolog and *TLOα10* on the other homolog. Sequencing towards the telomere and across the BTS revealed two sequences coincident with a six nucleotide repeat, which is underlined for clarity.

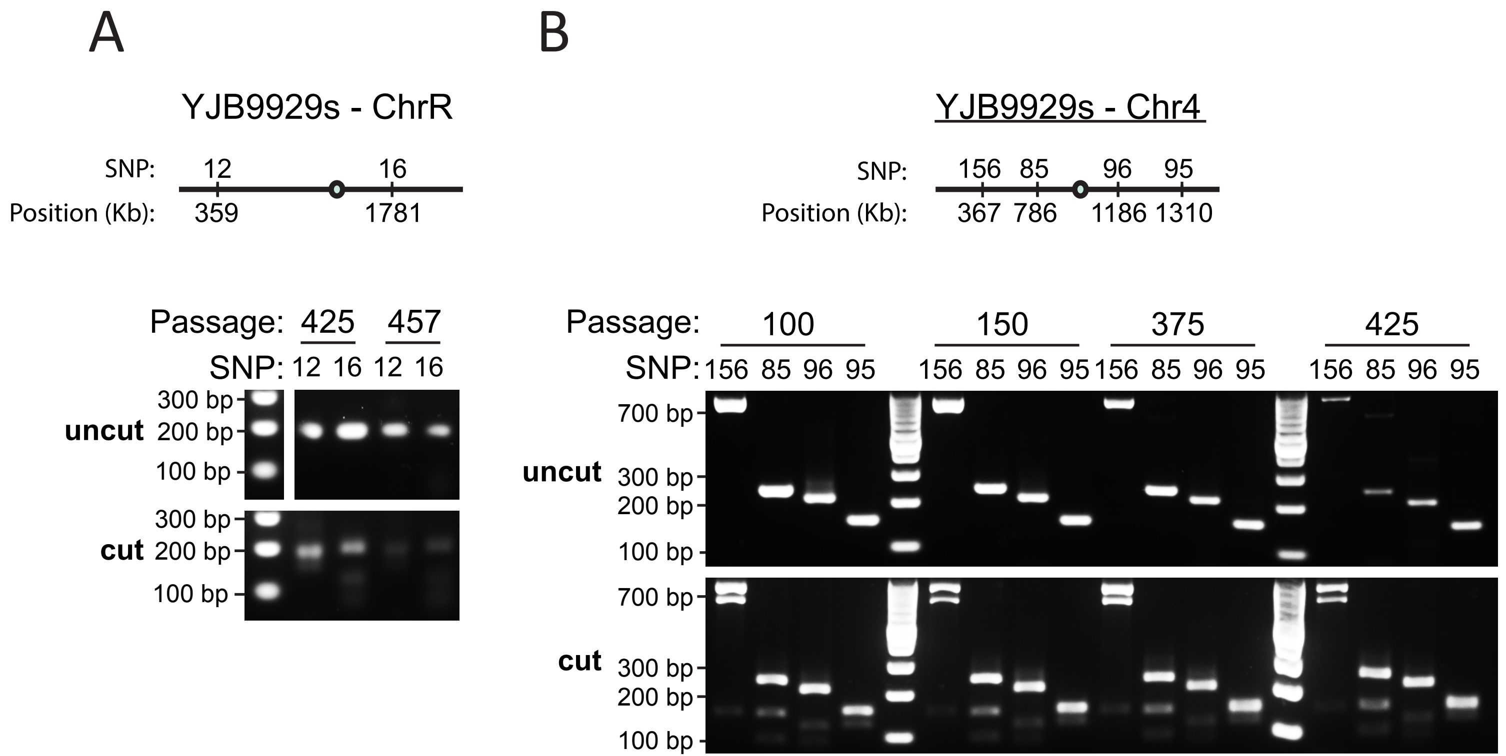


Figure S6 Chromosome heterozygosity following subtelomeric LOH. The allelic state of chromosomes with altered *TLO* configurations but lacking a defined subtelomeric recombination site were assayed by SNP-RFLP. Heterozygosity of ChrR (A) or Chr4 (B) in strain YJB9929s was assayed for both chromosome arms using previously identified SNP-RFLP positions (FORCHE et al. 2009) at the noted chromosomal positions. Both chromosomes were heterozygous for all positions tested at all time points.

Table S1 Primers used in this study

Primer	Sequence
pan-TLO amplification	ATGCCAGAAAACCTCCAAAC
RL-arm Sequencing	GGAGTACAGAAGTAGAGCAA
RR-arm Sequencing	AAAGCATCTGTAGACACGGC
1L-arm Sequencing	CGCTGGGTATCACAAGAGTG
1R-arm Sequencing	TCACCAAGACTGGGTAGAGC
2L-arm Sequencing	GGGGTATTTGGAATGGTTTG
3L-arm Sequencing	CCTATACATAGCCTTCTGCG
3R-arm Sequencing	CTGTGGTGGCAGGTAATTTT
4L-arm Sequencing	CCCACAAATATTTGTTCACC
4R-arm Sequencing	CATTATTGTGTTTGGTGGGG
5L-arm Sequencing	ATGTCTTCGACGGTATTGCC
5R-arm Sequencing	TGTTCACCTTTGACATGCTC
6L-arm Sequencing	TTTCTGGCCTCCTCTGCCTT
6R-arm Amplification	AGAGTGTGACGATGAGGATGATAGCTG
7L-arm Amplification	AGTTCTGGCTAGCTTATATTTTGCATCTCTATATCG
7R-arm Sequencing	CAGGTGTTCGCTGTCTATTC
CEN1 -L Amplification	CATCCGTATTTCACACACGC
CEN1 -R Amplification	AACACAATGGTCCCACAACG
CEN1 -R Sequencing	AGACAATCCATAGGACCCTC
CEN2 -L Amplification	GAGTGAGTTAAGCTACCCAG
CEN2 -R Amplification	TTATCACCTTTGGTTCCCTG
CEN2 -R Sequencing	CCCTCGTTGTGACGTTTTAG
CEN3 -L Amplification	GCCAACTAAGACAGCTCATG
CEN3 -R Amplification	GCGCCAAACATACAAACCAC
CEN3 -R Sequencing	TCAAACTTTCGGCTGGTAGC
CEN4 -L Amplification	GGGAGTTCAAAATATCGTCC
CEN4 -R Amplification	CGAGGCAACAATCAGAACCA
CEN4 -R Sequencing	GAACAATGGCTCACGAGTTT
CEN5 -L Amplification	CCGTCAATAGAACCGGAAGT
CEN5 -R Amplification	GACTGACATCCGTACTATCG
CEN5 -R Sequencing	GTAGCCCAGTATTAGAGATG
CEN6 -L Amplification	TGACAATGCCACTTCAAGCG
CEN6 -R Amplification	CCCTCCAATAAAGCAACCAC
CEN6 -R Sequencing	GTACTCACTAGCCCCTATTA
CEN7 -L Amplification	GGAGTTACCATCTTAGCAGC
CEN7 -R Amplification	ACAAACTCTGGATCCCACCA
CEN7 -R Sequencing	CCAGCAGCGAATATATTGGG
ENO1-L	TCAATTGTCCCATCTGGTGC
ENO1- R	TCTTTCAGATCTGGCTGGAG
LIP4 -F	GTCACTACTGTTATGGAGCC
LIP4 -R	TGAATATCAGGTCCAAGTCC
LIP4 Sequencing	CCCTTCCAAACTCGCTTCAT
SAP1 -F	CTCCAGCTAAAAGATCCCCA
SAP1 -R	GGATAAGGTTGACCGTTAGC
SAP1 Sequencing	CTCCTGTTAATGCTACTGGT
SAP7 -F	GTGCACATTTCCCTAACCGT
SAP7 -R	CCACTAGCATAAGTACCGAC
SAP7 Sequencing	GGTAATGGGATCTTTTCCGG

Table S2 SNPs differentiating intra-clade TLO members

	TLOα1	*TLOβ2*	*TLOα3*	*TLOα9*	*TLOα10*	*TLOα12*	*TLOγ4*	*TLOγ5*	*TLOγ7*	*TLOγ8*	*TLOγ11*	*TLOγ13*	*TLOγ16*
TLOα1		26	8	20	25	25							
TLOβ2			31	42	39	41							
TLOα3				19	27	25							
TLOα9					18	23							
TLOα10						22							
TLOα12													
TLOγ4								13	17	10	9	11	25
TLOγ5									14	7	7	4	22
TLOγ7										13	11	12	24
TLOγ8											3	5	19
TLOγ11												5	21
TLOγ13													20
TLOγ16													

Table S3 SNPs differentiating TLOa12 and TLOa12-a10

DNA change	Amino Acid change
C608A	P203H
T628A	I209N
C633T	-
T714C	-
T753C	-
G759A	-

Table S4 Mutations in TLO ORFs following passage

Gene	Position in ORF	Polymorphism	Type of mutation
TLOγ5	379	T → A	Non-synonymous
TLOγ7	297	A → G	Synonymous
TLOγ8	76	A → G	Non-synonymous
	96	G → A	Synonymous
	241	A → G	Non-synonymous
	333	A → T	Non-synonymous
TLOα9	444	C → T	Synonymous
	544	A → T	Non-synonymous
	565	C → T	Non-synonymous
	565	C → A	Non-synonymous
	615	C → T	Synonymous
	626	A → T	Non-synonymous
	627	C → T	Synonymous
	630	A → G	Synonymous
TLOα10	291	T → G	Synonymous
	293	T → C	Non-synonymous
TLOγ11	25	T → C	Non-synonymous
	76	A → G	Non-synonymous
	96	G → A	Non-synonymous
	111	A → T	Synonymous
	201	G → A	Synonymous
	312	G → A	Synonymous
TLOα12	304	T → G	Non-synonymous
	740	T → A	Non-synonymous
TLOγ13	24	A → T	Synonymous
	56	C → T	Non-synonymous
	531	C → T	Synonymous
TLOγ16	76	G → A	Non-synonymous
	456	A → G	Synonymous
	495	A → G	Synonymous
	522	A → T	Synonymous

Table S5 Mutations in CEN ORFs following passage

ORF	Position	Polymorphism
CENR	ChrR:1745241	G→C
	ChrR:1745422	7nt indel
CEN2	Chr2:1928980	A→G
	Chr2:1928988	T→A
	Chr2:1929024	A→C
	Chr2:1929165	G→A
	Chr2:1929181	G→A
	Chr2:1929300	G→C
	Chr2:1929343	A→C
	Chr2:1929514	T→C
CEN5	Chr5:470361	T→G

Table S6 Mutations in coding ORFs following passage

ORF	Position in ORF	Polymorphism
LIP4	877	C→A
SAP1	737	T→A
	738	C→G
SAP7	309	T→C
	1119	T→A