



Mating-type switching by chromosomal inversion in methylotrophic yeasts suggests an origin for the three-locus *Saccharomyces cerevisiae* system

Sara J. Hanson, Kevin P. Byrne, and Kenneth H. Wolfe¹

UCD Conway Institute, School of Medicine and Medical Science, University College Dublin, Dublin 4, Ireland

Edited by Jasper Rine, University of California, Berkeley, CA, and approved October 7, 2014 (received for review August 19, 2014)

Saccharomyces cerevisiae has a complex system for switching the mating type of haploid cells, requiring the genome to have three mating-type (*MAT*)-like loci and a mechanism for silencing two of them. How this system originated is unknown, because the three-locus system is present throughout the family Saccharomycetaceae, whereas species in the sister *Candida* clade have only one locus and do not switch. Here we show that yeasts in a third clade, the methylotrophs, have a simpler two-locus switching system based on reversible inversion of a section of chromosome with *MATa* genes at one end and *MATalpha* genes at the other end. In *Hansenula polymorpha* the 19-kb invertible region lies beside a centromere so that, depending on the orientation, either *MATa* or *MATalpha* is silenced by centromeric chromatin. In *Pichia pastoris*, the orientation of a 138-kb invertible region puts either *MATa* or *MATalpha* beside a telomere and represses transcription of *MATa2* or *MATalpha2*. Both species are homothallic, and inversion of their *MAT* regions can be induced by crossing two strains of the same mating type. The three-locus system of *S. cerevisiae*, which uses a nonconservative mechanism to replace DNA at *MAT*, likely evolved from a conservative two-locus system that swapped genes between expression and nonexpression sites by inversion. The increasing complexity of the switching apparatus, with three loci, donor bias, and cell lineage tracking, can be explained by continuous selection to increase sporulation ability in young colonies. Our results provide an evolutionary context for the diversity of switching and silencing mechanisms.

Hansenula polymorpha | *Pichia pastoris* | yeast genetics | comparative genomics

Mating-type switching in yeasts is a highly regulated process that converts a haploid cell of one mating type into a haploid of the opposite type (1–5). Switching involves the complete deactivation of one set of regulatory genes and activation of an alternative set, but, unlike most regulatory changes, the switch is achieved by physically replacing the DNA at an expression site that is shared by both types of cell. In *Saccharomyces cerevisiae* the switching system uses a menagerie of molecular components (3) including three mating-type (*MAT*)-like loci (the expressed *MAT* locus and the silent loci *HML* and *HMR*); an endonuclease (HO) that creates a double-strand break at the *MAT* locus, which then is repaired using *HMLalpha* or *HMRalpha* as a donor; a mechanism (Sir1 and Sir2/3/4 proteins) for repressing transcription and HO cleavage at the silent loci; two triplicated sequences (the Z and X regions) that guide repair of the dsDNA break; a donor-bias mechanism (the recombination enhancer, RE) to ensure that switching happens in the correct direction; and a cell lineage-tracking mechanism (Ash1 mRNA localization) to ensure that switching occurs only in particular cells. Most of these components have no function other than facilitating switching.

Given its complexity, it is surprising that switching seems to have evolved at least twice (5–10). Only unicellular fungi (yeasts) switch mating type, and these fungi have evolved from multicellular filamentous fungal ancestors on at least five separate lineages (11). Switching has arisen in two of these lineages represented by

S. cerevisiae (subphylum Saccharomycotina) and *Schizosaccharomyces pombe* (Taphrinomycotina). A large clade of nonswitching filamentous ascomycetes, the Pezizomycotina, is related more closely to *S. cerevisiae* than to *S. pombe* (Fig. 1). Despite their apparently independent origins, the switching systems of *S. cerevisiae* (3) and *S. pombe* (4, 5) have many parallels, including the presence of three *MAT*-like loci (two of which are silenced) with triplicated guide sequences, a mechanism for specifically inducing a DNA break at *MAT*, donor bias (27), and cell-lineage tracking. However, the details of switching and silencing are quite different in the two species (3, 4).

Switching is an error-prone process (28, 29), so it must confer a benefit to yeast species or it would not have been maintained by natural selection. However, the nature of this evolutionary benefit is debated (2, 6, 8, 30). We and others have proposed that it is related to the control of spore germination in uncertain environments (29, 30). The goal of switching is not to restore diploidy, because some species that switch grow primarily as haploids, but rather is to maximize the ability of a young colony to make new spores if nutrient levels fall. Yeasts are dispersed to new habitats when they are eaten and excreted by insects (31, 32), and ascospores are structures that assist yeast cells to survive passage through the insect digestive tract (33). Although many tetrads (sets of four haploid spores formed by meiosis of a diploid) remain intact, digestion by the insect removes the ascus wall and causes some spores to become isolated (33). If an isolated spore germinates, it has no way of making new spores unless it can find a mating partner of the opposite mating type [the “lonely spore

Significance

Saccharomyces cerevisiae undergoes a programmed DNA rearrangement to switch between mating types a and alpha. The origins of this complex and multifaceted process, which requires three copies of the mating-type (*MAT*) locus (with two silenced), have remained unknown. In this study we present a mechanism for mating-type switching in methylotrophic yeasts that shares a common origin with the well-characterized system in *S. cerevisiae* but has simpler components. This system requires only two copies of the *MAT* locus, with one copy transcriptionally repressed by proximity to centromeric or telomeric chromatin. Switching between the mating types occurs by recombination between inverted-repeat sequences flanking the *MAT* loci. This system suggests an ancestral mechanism for mating-type switching in yeasts.

Author contributions: S.J.H. and K.H.W. designed research; S.J.H., K.P.B., and K.H.W. performed research; S.J.H., K.P.B., and K.H.W. analyzed data; and S.J.H. and K.H.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹To whom correspondence may be addressed. Email: kenneth.wolfe@ucd.ie.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1416014111/-DCSupplemental.

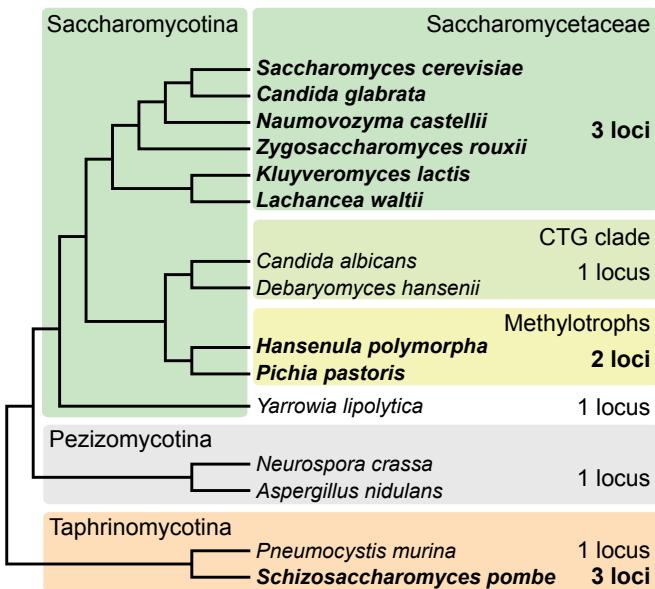


Fig. 1. Cladogram showing numbers of *MAT*-like loci. Species named in bold have been shown experimentally to switch mating types (3, 4, 12–18). Saccharomycotina, Pezizomycotina, and Taphrinomycotina are subphyla within phylum Ascomycota (19). Saccharomycetaceae, the CTG clade, and the methylophilous clade are clades within the Saccharomycotina. The tree topology is based on refs. 20–26.

scenario” (2, 7)]. Switching provides a partner, allowing the cells to become diploid and able to make new spores. In *S. cerevisiae*, new spores can be made just two cell divisions after spore germination (1). Switching enables isolated spores to germinate in very poor environments, replicate for a few cell cycles as permitted by the environment, and then sporulate (29). In contrast, species that cannot switch would need to be more cautious about germinating. Over time, species that can switch are predicted to have a growth advantage over species that cannot switch, because they can risk germinating in poorer environments and so germinate earlier in improving environments. The formation of spores by asexual means does not appear to be an option for most ascomycete yeasts because ascosporulation is linked to meiosis, unlike molds, which can make asexual spores in conidia (19).

In the family Saccharomycetaceae, switching has been reported experimentally in five other genera as well as in *Saccharomyces* (Fig. 1). Genome sequences show that almost all species in this family have three *MAT*-like loci and are likely to switch mating types by a mechanism similar to that of *S. cerevisiae* (8, 29, 34, 35), although some species use an alternative mechanism instead of HO endonuclease to cut the *MAT* locus (15). Family Saccharomycetaceae is one clade within the subphylum Saccharomycotina, which also includes a large clade of *Candida* and related species (called the “CTG clade” because of its unusual genetic code). Species in the CTG clade have only one *MAT*-like locus and do not switch (36, 37). More recently, genomes have been sequenced from a third clade within the Saccharomycotina which includes the methanol-assimilating species *Hansenula polymorpha* (20) and *Pichia pastoris* (21, 38) and relatives such as *Dekkera bruxellensis* (22, 23) and *Kuraishia capsulata* (24). For convenience we refer to this clade as the “methylophilous.” Phylogenomic analyses consistently have shown that the methylophilous are monophyletic and are a sister of the CTG clade, with the Saccharomycetaceae outside the grouping (Fig. 1). *H. polymorpha* and *P. pastoris* are regarded as homothallic both on classical grounds [observation of sporulation in colonies grown from single spores or observation of mating between a mother cell and a bud

(19)] and genetic grounds [any strain can mate with any other strain (17, 18, 39–41)]. Genetic evidence for mating-type switching has been reported in a few methylotrophs, including *H. polymorpha* (17, 40) and *P. pastoris* (18), but the molecular details have not been investigated.

Results

An Inversion Polymorphism Correlates with *MAT* Gene Expression in *H. polymorpha*. *H. polymorpha* grows primarily as a haploid and is homothallic (17, 39, 40). The genomes of two strains, NCYC495 and DL-1, have been sequenced. In a taxonomic revision (42) these strains were reclassified as different species, *Ogataea polymorpha* (NCYC495) and *Ogataea parapolymorpha* (DL-1); the genomes are about 10% divergent in sequence. We found that they are completely collinear in gene order (no inversions or translocations) throughout the genome except for one 19-kb inversion. The region inverted between the strains has homologs of *MATA1* and *MATA2* genes at one end and homologs of *MATalpha1* and *MATalpha2* at the other end, separated by seven other genes whose known functions are unrelated to mating (Fig. 2). These are the only *MAT*-like loci in the genome. The inverted region is located between two copies of a 2-kb inverted repeat (IR) sequence, derived from the 3' end of the gene *SLA2* and part of *MATA1*. PCR amplification across the possible inversion endpoints in genomic DNAs from the two strains confirmed that they differ in orientation and that the orientation of a third strain, CBS4732, is the same as that of DL-1 (Fig. 3A). Transcription of the *MATalpha* or *MATA* genes depends on the orientation of the 19-kb region (Fig. 3B). In NCYC495, *MATA1* and *a2* are expressed, and *MATalpha1* and *alpha2* are repressed, so we designated this orientation as the “*a* orientation.” The opposite occurs in strains DL-1 and CBS4732, which have the alpha orientation. All four *MAT* genes are expressed in diploids. The seven other genes in the 19-kb inverted region show expression in both orientations (*SI Appendix*, Fig. S1).

Gene Expression at the Silent Locus in *H. polymorpha* Is Repressed by a Centromere. The region immediately to the left of the silenced *MAT* genes and the leftmost copy of the 2-kb IR in *H. polymorpha* (as drawn in Fig. 2) has several characteristics that indicate it could be a regional centromere (20, 43). It coincides with a trough of G+C content (*SI Appendix*, Fig. S2), and in both NCYC495 and DL-1 the region contains clusters of retrotransposons and LTRs from the Ty5 family, which tends to target centromeres (43). (The sequences and locations of the retro-elements differ in the two strains).

We confirmed that the region beside the *MAT* is a centromere by two methods. First, we overexpressed a tagged version of the centromeric histone Cse4 (CenH3) and used ChIP to measure binding to the region in two haploid strains of opposite orientations derived from an NCYC495 background. The Cse4 signal is high in the putative centromeric region to the left of the IR in both strains. It extends onto the *MATA1* and *a2* genes only in the alpha-orientation strain (Fig. 3C) and onto the *MATalpha1* and *alpha2* genes only in the *a*-orientation strain (Fig. 3D). Second, we reanalyzed Hi-C (chromatin conformation capture) data from a synthetic metagenomics experiment that included *H. polymorpha* NCYC495 (44) to extract signals resulting from the 3D clustering of its centromeres in the nucleus (45). This analysis revealed centromeric signals from each *H. polymorpha* chromosome (*SI Appendix*, Fig. S2) and placed the peak signal from chromosome 3 at position 897 kb, which is 3 kb from the leftmost IR and 5 kb from the *MATalpha* genes. The centromeric region is transcriptionally silent (*SI Appendix*, Fig. S2).

Homothallic Mating in *H. polymorpha* Occurs by Inverting the *MAT* Region. *H. polymorpha* is homothallic, so any strain can mate with any other strain. If the orientation of the *MAT* region determines

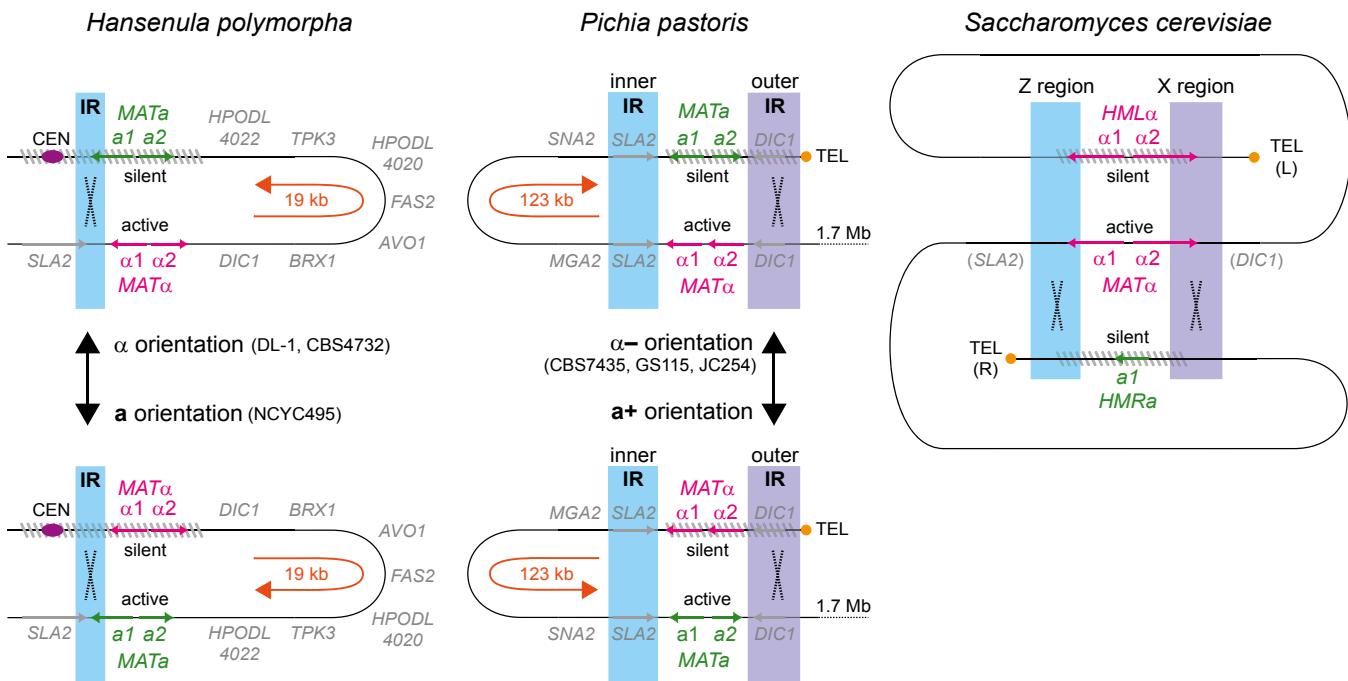


Fig. 2. MAT locus organization and mating-type switching in *H. polymorpha*, *P. pastoris*, and *S. cerevisiae*. Blue and purple shading show the IRs. Orange arrows show the orientation of the invertible regions. Two other arrangements of the *P. pastoris* region ($\alpha+$ and $\alpha-$), resulting from exchanges in the inner IR that do not change the mating type, are not shown (Fig. 5). In the *S. cerevisiae* diagram *SLA2* and *DIC1* indicate the genes that flanked the *MAT* locus in the ancestor of family Saccharomycetaceae, although they no longer are at this position in *S. cerevisiae* because of chromosome erosion (29).

mating type in this species, then mating between two strains with the same orientation should be possible only if one of them first switches mating type by inverting the region. We crossed pairs of NCYC495-derived strains (*O. polymorpha*) in the alpha orientation or the **a** orientation by growing them together in mating (nitrogen starvation) medium and using genetic markers to select for diploids (Fig. 4*A*). In each case, the diploids contained chromosomes with both orientations. This result indicates that one parental chromosome underwent inversion before, during, or after mating. Growth of haploid strains with the alpha orientation or **a** orientation in liquid minimal medium without nitrogen resulted in the appearance of both *MAT* region orientations within 12 h (Fig. 4*B*). Because the short timeframe (*H. polymorpha* typically requires >24 h for mating) and agitation of the cultures on a shaker preventing prolonged contact between cells made mating events in these cultures unlikely, these data suggest that inversion of the *MAT* region occurs before the formation of diploids. The appearance of both orientations was suppressed by the addition of ammonium sulfate (Fig. 4*B*), indicating that inversion of the *MAT* region is induced by the same nitrogen-starvation conditions that induce mating. We suggest that inversion in a haploid cell leads to a change of expressed mating type, leading to mating. PCR analysis of 153 random spores formed by sporulation of a diploid strain that was made by crossing two alpha-orientation haploids showed that 84 (55%) were in the **a** orientation and 69 (45%) were in the alpha orientation, a frequency that is not significantly different from 50% ($P = 0.26$ by two-tailed binomial test).

***P. pastoris* also Has Two MAT-Like Loci at the Ends of an Invertible Region.** *P. pastoris* is a related homothallic methylotroph that is predominantly haploid (18, 41). The taxonomy of *P. pastoris* strains also has been revised recently, but all the strains we discuss here are from a single species that is properly called “*Komagataella phaffii*” (46). The genome sequence of strain CBS7435 (38) contains two *MAT*-like loci. One locus is close to the telomere of chromosome 4 and contains *MAT_{a1}* and *a₂*, whereas the other locus is 138 kb away

and contains *MAT_{a1}* and *alpha₂*. These loci are flanked on one side by a 3-kb IR and on the other side by a 6-kb IR, which we refer to as the “outer” and “inner” IRs, respectively (Fig. 2). The reported genome sequence of *P. pastoris* strain GS115 (21), which was derived from strain CBS7435 by mutagenesis, terminates in the telomere-proximal copy of the inner IR and so does not include the second *MAT*-like locus, but we have confirmed by PCR that it is present in GS115 and is similar to the structure in CBS7435.

The organization of *P. pastoris* chromosome 4 suggested that the *MAT*-like locus close to the telomere might be silenced by telomere position effect (TPE) (47), making the telomere-distal *MAT*-like locus the active one, and that haploids could switch mating type by recombining the outer IRs to invert the whole 138-kb region. Indeed, haploid strains of *P. pastoris* differ as to whether the telomere-distal locus is *MAT_{a1}* (designated as the alpha orientation) or *MAT_a* (designated as the **a** orientation). As in *H. polymorpha*, crossing two strains in the same orientation results in diploids that are heterozygous for orientation (Fig. 5*A*). Surprisingly, RT-PCR analysis indicates that haploid strains of both orientations transcribe all four *MAT* genes, as do diploids (SI Appendix, Fig. S3), and public RNA sequencing data from strain GS115 also show expression of all four genes (48). However, by quantitative RT-PCR (qRT-PCR) we find that haploid expression of *MAT_{a1}* and *MAT_{a2}* differs in the two orientations, with an alpha-orientation strain having 15-fold higher expression of *MAT_{a1}* and fourfold lower expression of *MAT_{a2}* than an **a**-orientation strain (Table 1 and SI Appendix, Table S1). For genes *a₂*, *alpha₁*, and *alpha₂*, a diploid strain has a level of expression that is intermediate between the two haploids. *MAT_{a1}* is highly expressed in both haploid orientations, at a level only slightly lower than the control gene *ACT1* in our experiments, and is three- to fourfold lower in diploids. Southern blot analysis shows that a haploid *P. pastoris* culture grown in rich medium does not contain a mixture of chromosomal orientations, so the expression of both

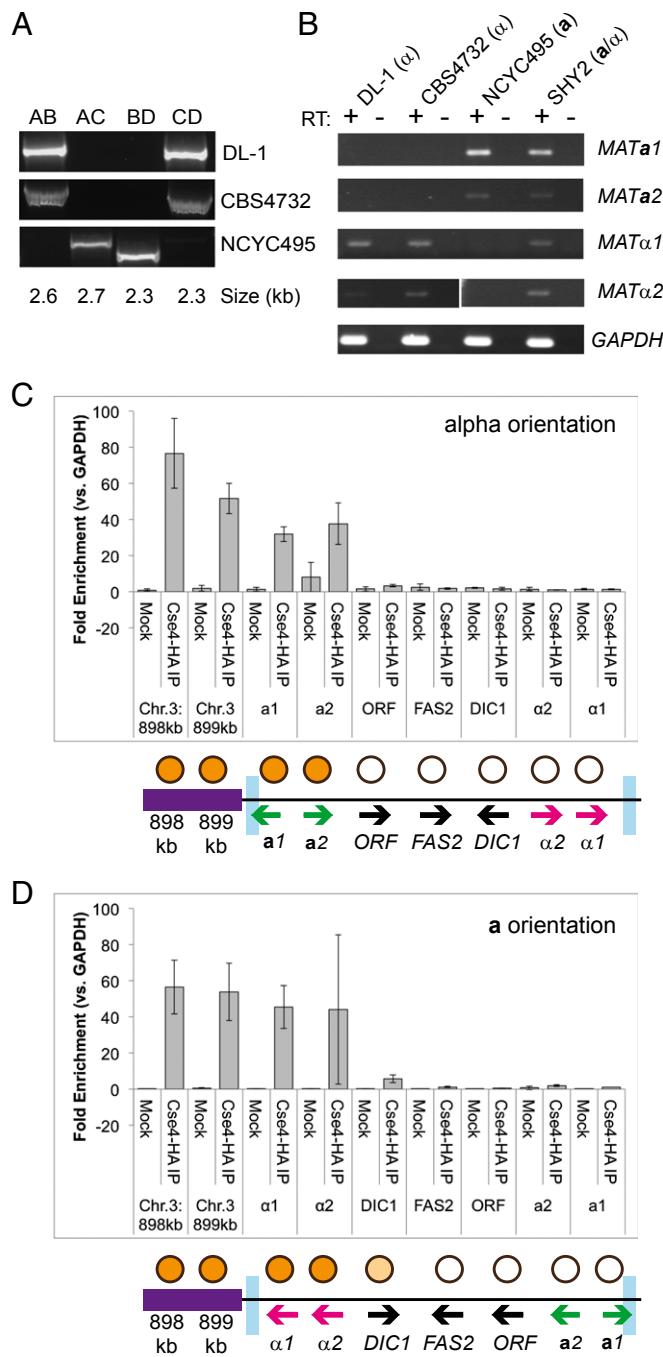


Fig. 3. Structure and expression of the *MAT* locus in *H. polymorpha* strains. (A) Orientation of the 19-kb region differs among *H. polymorpha* haploid strains. PCR products were amplified with primer pairs flanking IRs as shown schematically in Fig. 4A. Products from the AB and CD primer pairs indicate the alpha orientation, and products from the AC and BD pairs indicate the a orientation. (B) Expression of *H. polymorpha* *MAT* genes depends on orientation of the 19-kb region. Amplified cDNA was generated from haploid strains in both orientations and from a diploid strain (SHY2, produced by crossing two NCYC495-derived haploid strains with opposite orientations). “+ RT” indicates the addition of reverse transcriptase to cDNA synthesis reaction; “– RT” indicates no reverse transcriptase was added. (C and D) ChIP indicates the presence of Cse4 nucleosomes at the silenced *MAT* loci in haploids. The graphs present the fold enrichment of the target sequence relative to a control sequence (GAPDH) outside the 19-kb invertible region. The schematics below represent the relative positions of loci examined by ChIP. Positions highlighted in purple indicate putative centromeric regions. Blue boxes represent the IRs. Orange circles indicate the binding of tagged

sets of *MAT* genes probably is caused by leaky silencing rather than by chromosomal rearrangement (*SI Appendix*, Fig. S4).

Four Isomers of *P. pastoris* Chromosome 4. Because inversions could occur by recombination between the inner IRs, the outer IRs, or both, there are four possible isomeric structures for *P. pastoris* chromosome 4. We name these structures “alpha+,” “alpha–,” “a+,” and “a–” (Fig. 5B). The prefixes “alpha” and “a” indicate the mating type (the *MAT* genotype at the telomere-distal position), and the suffixes “+” and “–” indicate the orientation of the unique 123-kb region between the inner IRs relative to the telomere. PCR amplification of genomic DNAs with orientation-specific primer pairs showed that homothallic mating causes orientation alpha– to switch to a+ (Fig. 5A). Switching therefore occurs by recombination in the outer IRs, which is equivalent to swapping the locations of the telomere and the remaining 1.7 Mb of the chromosome. Telomeric repeats (TGCTGGA_n) begin immediately beside the outer IR with no DNA between them (*SI Appendix*, Fig. S5).

Inversion across the inner IRs would not change the mating type but would make the two copies of chromosome 4 in a diploid collinear except at the *MAT* loci themselves. We guessed that such inversions might occur during meiosis, because otherwise any meiotic recombination within the 123-kb unique region would produce inviable chromosomes (*SI Appendix*, Fig. S6), but we found no evidence by random spore analysis to support this idea. However, all four possible chromosome structures are found among natural isolates of *P. pastoris* (Fig. 5B), indicating that inversion across the inner IR must occur at some rate.

Discussion

Our results show that homothallic methylotrophic yeast species switch mating types using a flip-flop inversion mechanism. In the late 1970s flip-flop inversion was discussed as a possible model for mating-type switching in several yeast species (49–54) and was found to be the mechanism of flagellar-phase variation in *Salmonella* (55). In a classic study in 1977, Hicks and Herskowitz (51) rejected the two-locus flip-flop model for *S. cerevisiae* in favor of the three-locus cassette model, which was proven soon after (56). In 1981, Tolstorukov and colleagues showed that the cassette model did not explain the pattern of switching in the methylotroph *Pichia methanolica* and proposed instead that this organism has a flip-flop system (52, 53). Our results now validate this proposal for methylotrophs. It remains to be determined whether the mechanism of inducible inversion involves a site-specific recombinase (such as Hin in phase variation or Flp in the yeast 2-μm plasmid) or uses the general recombination machinery. The methylotroph genomes do not have homologs of *HO* endonuclease.

Their conserved syntetic location shows clearly that the *MAT* loci of methylotrophs are orthologous to those of other yeasts in subphylum Saccharomycotina (Fig. 2). The gene *SLA2*, which is present in the *H. polymorpha* IR and the *P. pastoris* inner IR, also forms the Z region in many Saccharomycetaceae family yeasts (29) and is located beside the single *MAT* locus of many Pezizomycotina species (6). On the other side *DIC1*, which is located in the outer IR of *P. pastoris* and beside *MATalpha* in the 19-kb invertible region of *H. polymorpha*, was ancestrally beside the *MAT* locus in family Saccharomycetaceae and forms the X region in some species (29). Therefore, the *MAT* regions of methylotrophs and Saccharomycetaceae share a common ancestor.

Did the common ancestor of methylotrophs and Saccharomycetaceae switch mating types, and if so did it use a two-locus

Cse4 to the tested regions. The ORF is the gene *HPODL4022*. C shows strain SH4330 (alpha orientation), and D shows strain NCYC495 (a orientation).

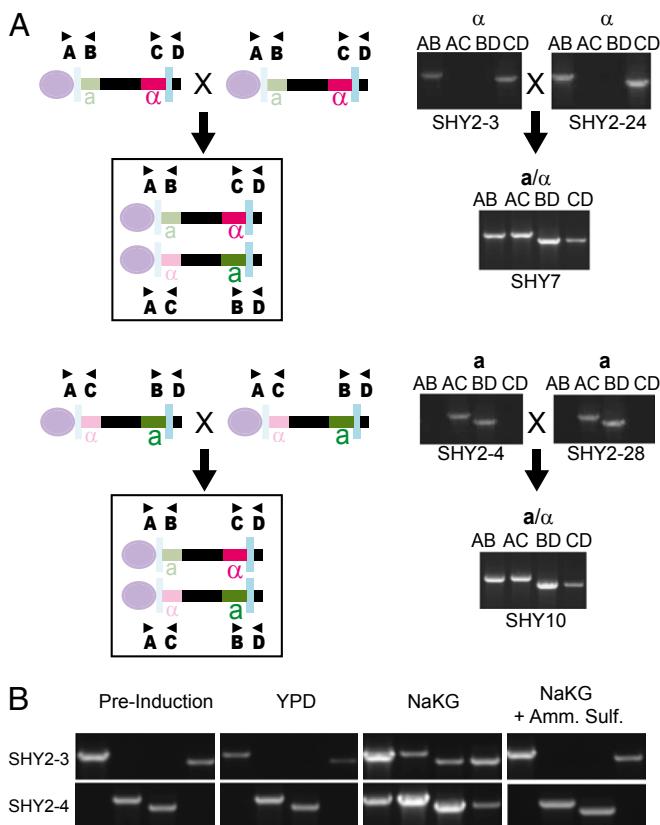


Fig. 4. Mating-type switching by inversion in *H. polymorpha*. (A) Crossing two strains in the same orientation results in diploids that are heterozygous for orientation. Haplid *MAT*α (SHY2-3 and SHY2-24) and *MAT*a (SHY2-4 and SHY2-28) strains with complementary auxotrophic mutations were obtained by random spore analysis of a sporulated diploid (SHY2) and were mated to produce diploids (SHY7 from α × α and SHY10 from a × a). (Left) Schematics indicate the *MAT* locus organization in the strains used in each cross and the resulting diploids. Purple ovals represent the centromere. The *MAT* genes closer to the centromere are silenced (pale colors). (Right) Gel images show amplification across the IR regions with the primer combinations indicated. (B) *MAT* locus inversion is induced by nitrogen limitation. Gel images show *MAT* locus organization for haploid *MAT*α (SHY2-3) and *MAT*a (SHY2-4) strains after growth for 12 h in YPD, minimal medium without nitrogen (NaKG), and NaKG supplemented with 40 mM ammonium sulfate as a nitrogen source.

or a three-locus system? Parsimony suggests that it did switch, because a single loss (in the CTG clade) is more likely than two independent gains in Saccharomycetaceae and methylotrophs. We propose that the ancestor had a simple two-locus system with one IR, similar to *H. polymorpha*. Such a system can be formed spontaneously by DNA introgression between the *MAT* idiomorphs in a heterothallic fungus (*SI Appendix*, Fig. S7), as appears to have occurred very recently in *Sclerotinia sclerotiorum* (Pezizomyctina), which inverts part of its *MAT* locus in every meiotic generation (57).

Further evidence for a two-locus ancestor comes from computer simulations (Fig. 6 and *SI Appendix*), which show that a simple two-locus system can evolve into a fully featured three-locus system as the result of continuous natural selection for a single trait—the ability of tiny colonies (4–16 cells) to produce the highest possible number of spores if nutrients run out. Selection for this trait would be expected if it was the reason that switching originated (29). Our simulations measured allele frequency and the potential yield of spores in colonies expanding from 1 to 128 haploid cells under different switching systems (Fig. 6).

Taking the eight-cell stage as an example, potential spore yield from the colony would be maximized if it had four α cells and four a cells and so could make four diploids and hence 16 spores. However, no switching system can achieve a 4:4 ratio consistently in every eight-cell colony. In an *H. polymorpha*-like system, if inversion of the *MAT* region occurred in 10% of cells, simulations show that on average only five spores would be produced per colony instead of 16, so the potential spore yield is only 31% of the theoretical maximum (Hpol_10 in Fig. 6). As shown in Fig. 6, the maximum spore yield from eight-cell colonies switching by the *H. polymorpha* system is 62% (10 spores), which occurs when the *MAT* region inverts in 50% of cells (Hpol_50). In comparison, the highly efficient systems of *S. cerevisiae* (Scer_90) and *S. pombe*

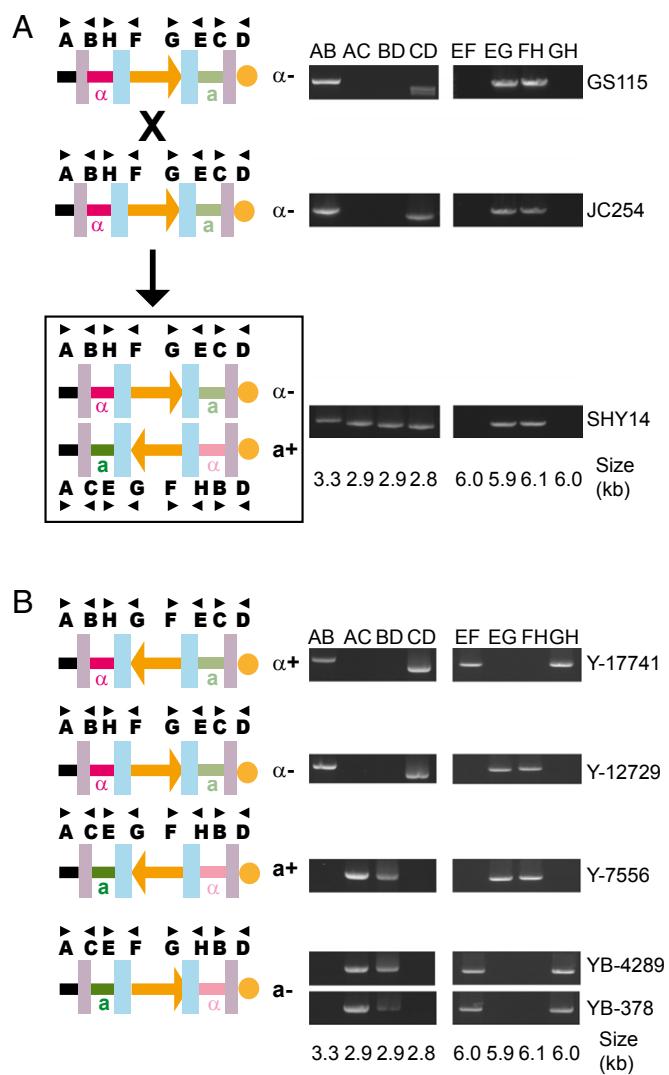


Fig. 5. Mating-type switching by inversion in *P. pastoris*. (A) Crossing two strains in the same orientation results in diploids that are heterozygous for orientation. (Left) The schematic indicates chromosome 4 organization in the haploid parents and their diploid progeny. The orange arrow shows the orientation of the 123-kb region between the inner IRs. The *MAT* locus closer to the telomere (orange circle) is repressed (pale colors). Purple and blue boxes represent the outer and inner IRs, respectively. (Right) Gel images show amplification across the IRs using primer combinations as indicated. (B) All four possible structures of chromosome 4 are found among natural isolates of *P. pastoris*. The isolates are from tree fluxes and were obtained from the US Department of Agriculture Agricultural Research Service collection as *Komagataella phaffii* (*SI Appendix*, Table S2).

Table 1. Fold change in expression of *P. pastoris* MAT loci determined by qRT-PCR

Gene	Ratio GS190(a)/GS115(α)*	Ratio GS115(α)/GS190(a)*	Ratio GS190(a)/SHY15(α/α)*	Ratio GS115(α)/SHY15(α/α)*
a1	0.75 (0.57–0.97)	—	3.36 (2.57–4.41)	4.50 (3.86–5.24)
a2	4.38 (3.41–5.62)	—	2.83 (2.20–3.63)	0.65 (0.47–0.89)
α1	—	1.82 (1.19–2.77)	0.66 (0.43–1.00)	1.20 (1.03–1.39)
α2	—	15.03 (13.18–17.15)	0.11 (0.06–0.20)	1.61 (1.41–1.84)

$$n = 2^{-\Delta\Delta C_t}$$

*Values in parentheses show \pm SD; see *SI Appendix, Table S1* for details of calculations.

(Spom_90) achieve about 90% of theoretical maximum spore yield (>14 spores) at the eight-cell stage.

The simulations show that continued selection for increased spore yield can lead to a two-locus system becoming more complex. If the invertible region between the IRs in a two-locus system is large, meiotic recombination within this region becomes more probable and will decrease potential spore yield (*SI Appendix, Fig. S6*), but this decrease can be overcome by adding a second IR (compare the yield in Ppas_50/10 with that of Ppas_50/0 in Fig. 6, which shows the effect of allowing inversion at the inner IR in *P. pastoris*). Duplicating one set of MAT genes will convert a two-locus system into a three-locus system and form Z and X regions from the two previous IRs (Fig. 2). This step will either increase potential spore yield or leave it unchanged if the two-locus system already was at the maximum productivity of

switching (the yields from Scer_50 and Hpol_50 are identical and higher than Hpol_10; Fig. 6). Once three loci are in place, the system can change from a conservative single Holliday junction mechanism to a nonconservative synthesis-dependent strand-annealing mechanism (3), which introduces the potential for biased choice of donor and increased efficiency (*SI Appendix*). When donor bias is possible, cell-lineage tracking becomes advantageous, because preventing switching in some cells while activating it in others can maximize potential spore yield in the important early generations of the colony (in Fig. 6, compare spore yields at the eight-cell stage in Scer_90 vs. Scer_50 and in Spom_90 vs. Spom_50, i.e., strong donor bias vs. unbiased switching in these species). Thus, an *S. cerevisiae*-like system with three MAT-like loci, Z and X guide sequences, HO, RE, and Ash1 can evolve step by step, starting from a simple *H. polymorpha*-like system, by continuous selection for one trait: increased spore yield from young colonies. Although haploidy and diploidy may each be beneficial in particular growth environments (58), and the ability to transition between them may be advantageous, selection for efficient spore production alone is sufficient to explain the origin and subsequent elaboration of the mating-type switching system.

What silencing mechanism was used in the ancestral system? It seems likely that the primordial system used either centromeric or telomeric heterochromatin to silence the nonexpressed MAT genes. However, all Saccharomycotina species, including *H. polymorpha* and *P. pastoris*, have lost the proteins (Clr4 and Swi6/HP1) that make the histone H3K9me modification that is characteristic of silent heterochromatin in most eukaryotes, including *Schizosaccharomyces* and *Pezizomycotina* (59). Instead, *H. polymorpha* may use Cse4-containing nucleosomes to silence its centromere-proximal MAT genes, similar to the mechanism proposed for neocentromeres in *Candida albicans* (60). It is less clear how expression differences between the MAT genes in *P. pastoris* haploids are achieved. TPE is a good candidate, but we see differential expression of only two of the four MAT genes, a result that is unexpected under TPE. TPE is a stochastic process that silences expression of telomere-linked genes in a proportion of cells in a population (47), possibly explaining why we see diminution rather than complete silencing of expression. It is unclear also how differential expression of only MATalpha2 and MATalpha2 could determine cell type under the current paradigm for the function of these genes (36).

Our results push back the origin of mating-type switching in the lineage leading to *S. cerevisiae* to before the common ancestor of methylotrophs and Saccharomycetaceae and hence almost to the base of subphylum Saccharomycotina. Consequently, they imply that the ability to switch has been lost in the CTG clade which includes the pathogen *C. albicans*. Genomic data indicate that, despite its benefits, switching also has been lost several other times including in the methylotrophs *Kuraishia capsulata* and *Dekkera bruxellensis* (22–24), and in the Saccharomycetaceae *Lachancea kluyveri* and *Kazachstanica africana* (29, 61). Null alleles of HO are frequent among natural populations of *S. cerevisiae* (62). Losses of switching may simply reflect life in environments that do not require frequent spore formation for survival or dispersal. Studying the mechanisms of switching and

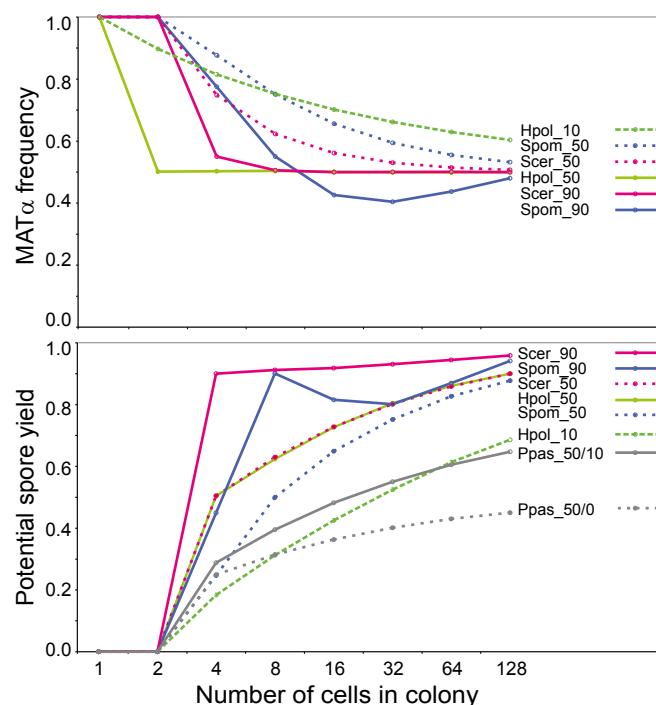


Fig. 6. Simulations of changes in allele frequency and potential spore yield under different switching systems during growth from 1–128 haploid cells in a colony started by a germinating MATalpha spore. (Upper) *S. cerevisiae* (Scer) and *S. pombe* (Spom) systems were simulated at switching productivities of 90% (strong donor bias) and 50% (no bias), and *H. polymorpha* (Hpol) systems were simulated at productivities of 50% (all cells switch randomly) and 10% (fewer cells switch randomly). The productivity of switching is the proportion of cells that successfully change mating type when permitted to do so. Allele frequency curves for *P. pastoris* (not plotted) are identical to those for *H. polymorpha*. (Lower) The gray lines show the effect of allowing inversion in the inner IR of *P. pastoris* at a frequency of 10% (Ppas_50/10) compared with no inversion (Ppas_50/0), when the productivity of switching is 50%. Graphs are averages from 10,000 replicate colonies. Details of simulations are given in *SI Appendix*.

silencing in deeper lineages of Saccharomycotina predating and postdating the loss of H3K9me heterochromatin will help us learn more about the origins of these systems and the evolutionary pressures that created them.

Materials and Methods

Yeast Strains and Culture Conditions. Strains of *H. polymorpha* and *P. pastoris* used in this study are listed in *SI Appendix, Table S2*. Cultures were grown under standard rich-medium conditions. The medium used for the induction of mating in *H. polymorpha* was 2% malt extract and in *P. pastoris* was minimal medium without nitrogen (NaKG: 0.5% sodium acetate, 1% potassium chloride, 1% glucose). Isolation of spores from sporulating diploid populations was performed by random spore analysis. After growth on mating/sporulation medium for 4 d, diploids were suspended in diethyl ether for ~1 h before plating on rich medium. Individual colonies were screened for *MAT* locus orientation by colony PCR (see below). To induce inversion of the *MAT* locus, overnight cultures of *H. polymorpha* grown in yeast extract/peptone/dextrose (YPD) were diluted to OD₆₀₀ ~0.5 in YPD, NaKG, or NaKG with 40 mM ammonium sulfate and were grown in an orbital shaker for 12 h at 30 °C with agitation at ~200 cycles per min.

DNA Extraction and *MAT* Locus PCR. DNA was harvested from stationary-phase cultures by homogenization with glass beads followed by phenol-chloroform extraction and ethanol precipitation. Primers used for *MAT* locus amplification are listed in *SI Appendix, Table S3*. Amplification was done using GoTaq (Promega) or Phusion Taq (New England Biolabs) polymerase for 25–35 cycles with an annealing temperature of 55–60 °C. Colony PCR was performed by heating cells in 10 µL water at 95 °C for 5 min followed by addition of PCR reagents.

RNA Extraction and RT-PCR. RNA was extracted from log-phase cultures by either hot acid phenol-chloroform extraction or homogenization with glass beads followed by phenol-chloroform extraction and ethanol precipitation. Following DNase I (Invitrogen) treatment, cDNA was synthesized using SuperScript III Reverse Transcriptase (Invitrogen). Primers used for amplification of *MAT* gene cDNA are listed in *SI Appendix, Table S3*. For nonquantitative RT-PCR, amplification was performed using GoTaq (Promega) polymerase for 30 cycles with an annealing temperature of 55 °C. qRT-PCR was performed on an Agilent Mx3005P system using Sensifast Lo-ROX SYBR Green mix (Bioline) according to the manufacturer's instructions.

Cse4 ChIP. A synthetic construct containing the *H. polymorpha* *CSE4* gene tagged at an internal site with 3HA and expressed from the *S. cerevisiae* *TEF1* promoter was designed (*SI Appendix, Fig. S8*) and made by Integrated DNA Technologies. The construct was inserted into a panARS vector (63) containing a KanMX marker by restriction digestion and ligation and was transformed into *H. polymorpha* by electroporation as previously described (64). Log-phase cells were crosslinked using formaldehyde (1% final concentration) for 20 min followed by the addition of glycine to stop the reaction. Cells were harvested and lysed using glass beads followed by chromatin fragmentation by sonication with a Bioruptor Standard (Diagenode). Immunoprecipitation of chromatin fragments was performed with EZview Red Anti-HA Affinity Gel (Sigma-Aldrich) or mouse IgG1 (Cell Signaling Technology) as isotype control. After washes, bound DNA was eluted with HA peptide (Sigma-Aldrich), crosslinks were reversed, and the samples were purified by phenol-chloroform extraction. qPCR was performed on an Agilent Mx3005P system using Sensifast Lo-Rox SYBR Green mix (Bioline) according to the manufacturer's instructions. Primers used for amplification of centromeric and *MAT* region DNA sequences are listed in *SI Appendix, Table S3*.

- Strathern JN, Herskowitz I (1979) Asymmetry and directionality in production of new cell types during clonal growth: The switching pattern of homothallic yeast. *Cell* 17(2): 371–381.
- Herskowitz I (1988) Life cycle of the budding yeast *Saccharomyces cerevisiae*. *Microbiol Rev* 52(4):536–553.
- Haber JE (2012) Mating-type genes and *MAT* switching in *Saccharomyces cerevisiae*. *Genetics* 191(1):33–64.
- Klar AJ (2007) Lessons learned from studies of fission yeast mating-type switching and silencing. *Annu Rev Genet* 41:213–236.
- Egel R (2005) Fission yeast mating-type switching: Programmed damage and repair. *DNA Repair (Amst)* 4(5):525–536.
- Lin X, Heitman J (2007) Mechanisms of homothallism in fungi and transitions between heterothallism and homothallism. *Sex in Fungi*, eds Heitman J, Kronstad JW, Taylor JW, Casselton LA (American Society for Microbiology, Washington, DC), pp 35–57.
- Lee SC, Ni M, Li W, Shertz C, Heitman J (2010) The evolution of sex: A perspective from the fungal kingdom. *Microbiol Mol Biol Rev* 74(2):298–340.
- Dujon B (2010) Yeast evolutionary genomics. *Nat Rev Genet* 11(7):512–524.
- Rusche LN, Rine J (2010) Switching the mechanism of mating type switching: A domesticated transposase supplants a domesticated homing endonuclease. *Genes Dev* 24(1):10–14.
- Knop M (2011) Yeast cell morphology and sexual reproduction—a short overview and some considerations. *C R Biol* 334(8–9):599–606.
- Nagy LG, et al. (2014) Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. *Nat Commun* 5:4471.
- Edskes HK, Wickner RB (2013) The [URE3] prion in *Candida*. *Eukaryot Cell* 12(4):551–558.
- Naumov GI, Naumova ES, Marinoni G, Piskur J (1998) Genetic analysis of the yeast *Saccharomyces castellii*, *S. exigua*, and *S. martiniae*. *Russ J Genet* 34(4):457–460.
- Watanabe J, Uehara K, Mogi Y (2013) Diversity of mating-type chromosome structures in the yeast *Zygosaccharomyces rouxii* caused by ectopic exchanges between *MAT*-like loci. *PLoS ONE* 8(4):e62121.

15. Barsoum E, Martinez P, Aström SU (2010) Alpha3, a transposable element that promotes host sexual reproduction. *Genes Dev* 24(1):33–44.
16. Di Rienzi SC, et al. (2011) Genetic, genomic, and molecular tools for studying the protoploid yeast, *L. waltii*. *Yeast* 28(2):137–151.
17. Gleeson MA, Sudbery PE (1988) Genetic analysis in the methylotrophic yeast *Hansenula polymorpha*. *Yeast* 4(4):293–303.
18. Tolstorukov II, Cregg JM (2007) *Pichia Protocols*. Classical Genetics. Methods in Molecular Biology, ed Cregg JM (Humana, Totowa, NJ), 2nd Ed, Vol 389, pp 189–201.
19. Kurtzman CP, Fell JW, Boekhout T, eds (2011) *The Yeasts, a Taxonomic Study* (Elsevier, Amsterdam).
20. Ravin NV, et al. (2013) Genome sequence and analysis of methylotrophic yeast *Hansenula polymorpha* DL1. *BMC Genomics* 14:837.
21. De Schutter K, et al. (2009) Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nat Biotechnol* 27(6):561–566.
22. Curtin CD, Borneman AR, Chambers PJ, Pretorius IS (2012) De-novo assembly and analysis of the heterozygous triploid genome of the wine spoilage yeast *Dekkera bruxellensis* AWRI1499. *PLoS ONE* 7(3):e33840.
23. Piškur J, et al. (2012) The genome of wine yeast *Dekkera bruxellensis* provides a tool to explore its food-related properties. *Int J Food Microbiol* 157(2):202–209.
24. Morales L, et al. (2013) Complete DNA sequence of *Kuraishia capsulata* illustrates novel genomic features among budding yeasts (Saccharomycotina). *Genome Biol Evol* 5(12):2524–2539.
25. Kunze G, et al. (2014) The complete genome of *Blastobotrys (Arxula) adeninivorans* LS3 – a yeast of biotechnological interest. *Biotechnol Biofuels* 7:66.
26. James TY, et al. (2006) Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* 443(7113):818–822.
27. Jakočiūnas T, Holm LR, Verhein-Hansen J, Trusina A, Thon G (2013) Two portable recombination enhancers direct donor choice in fission yeast heterochromatin. *PLoS Genet* 9(10):e1003762.
28. Hicks WM, Kim M, Haber JE (2010) Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. *Science* 329(5987):82–85.
29. Gordon JL, et al. (2011) Evolutionary erosion of yeast sex chromosomes by mating-type switching accidents. *Proc Natl Acad Sci USA* 108(50):20024–20029.
30. Knop M (2006) Evolution of the hemiascomycete yeasts: On life styles and the importance of inbreeding. *BioEssays* 28(7):696–708.
31. Reuter M, Bell G, Greig D (2007) Increased outbreeding in yeast in response to dispersal by an insect vector. *Curr Biol* 17(3):R81–R83.
32. Stefanini I, et al. (2012) Role of social wasps in *Saccharomyces cerevisiae* ecology and evolution. *Proc Natl Acad Sci USA* 109(33):13398–13403.
33. Coluccio AE, Rodriguez RK, Kernan MJ, Neiman AM (2008) The yeast spore wall enables spores to survive passage through the digestive tract of *Drosophila*. *PLoS ONE* 3(8):e2873.
34. Butler G, et al. (2004) Evolution of the MAT locus and its Ho endonuclease in yeast species. *Proc Natl Acad Sci USA* 101(6):1632–1637.
35. Gabaldón T, et al. (2013) Comparative genomics of emerging pathogens in the *Candida glabrata* clade. *BMC Genomics* 14:623.
36. Bennett RJ, Johnson AD (2005) Mating in *Candida albicans* and the search for a sexual cycle. *Annu Rev Microbiol* 59:233–255.
37. Butler G (2010) Fungal sex and pathogenesis. *Clin Microbiol Rev* 23(1):140–159.
38. Kübler A, et al. (2011) High-quality genome sequence of *Pichia pastoris* CBS7435. *J Biotechnol* 154(4):312–320.
39. Sudbery PE, Gleeson MAG (1989) Genetic manipulation of the methylotrophic yeasts. *Molecular and Cell Biology of Yeasts*, eds Walton EF, Yarranton GT (Blackie, Glasgow), pp 304–329.
40. Lahtchev K (2002) *Basic Genetics of Hansenula polymorpha*. *Hansenula polymorpha: Biology and Applications*, ed Gellissen G (Wiley-VCH, Weinheim), pp 8–20.
41. Sreekrishna K, Kropp KE (1996) *Pichia pastoris*. *Nonconventional Yeasts in Biotechnology: A Handbook*, ed Wolf K (Springer, Berlin), pp 203–253.
42. Kurtzman CP (2011) A new methanol assimilating yeast, *Ogataea parapolymorpha*, the ascosporic state of *Candida parapolymorpha*. *Antonie van Leeuwenhoek* 100(3):455–462.
43. Lynch DB, Logue ME, Butler G, Wolfe KH (2010) Chromosomal G + C content evolution in yeasts: Systematic interspecies differences, and GC-poor troughs at centromeres. *Genome Biol Evol* 2:572–583.
44. Burton JN, Liachko I, Dunham MJ, Shendure J (2014) Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps. *G3 (Bethesda)* 4(7):1339–1346.
45. Marie-Nelly H, et al. (2014) Filling annotation gaps in yeast genomes using genome-wide contact maps. *Bioinformatics* 30(15):2105–2113.
46. Kurtzman CP (2009) Biotechnological strains of *Komagataella (Pichia) pastoris* are *Komagataella phaffii* as determined from multigene sequence analysis. *J Ind Microbiol Biotechnol* 36(11):1435–1438.
47. Gottschling DE, Aparicio OM, Billington BL, Zakian VA (1990) Position effect at *S. cerevisiae* telomeres: Reversible repression of Pol II transcription. *Cell* 63(4):751–762.
48. Liang S, et al. (2012) Comprehensive structural annotation of *Pichia pastoris* transcriptome and the response to various carbon sources using deep paired-end RNA sequencing. *BMC Genomics* 13:738.
49. Egel R (1977) "Flip-flop" control and transposition of mating-type genes in fission yeast. *DNA Insertion Elements, Plasmids and Episomes*, eds Bukhari AI, Shapiro JA, Adhya SL (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), pp 447–455.
50. Hicks JB, Strathern JN, Herskowitz I (1977) The cassette model of mating-type interconversion. *DNA Insertion Elements, Plasmids and Episomes*, eds Bukhari AI, Shapiro JA, Adhya SL (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), pp 457–462.
51. Hicks JB, Herskowitz I (1977) Interconversion of yeast mating types. II. Restoration of mating ability to sterile mutants in homothallic and heterothallic strains. *Genetics* 85(3):373–393.
52. Tolstorukov II, Benevolensky SV, Efremov BD (1981) Mechanism of mating and self-diploidization in the haploid yeast *Pichia pinus*. VI. Functional organization of the mating type locus. *Sov Genet* 17(6):685–690.
53. Tolstorukov II, Benevolensky SV, Efremov BD (1982) Genetic control of cell type and complex organization of the mating type locus in the yeast *Pichia pinus*. *Curr Genet* 5(2):137–142.
54. Oshima Y (1993) Homothallism, mating-type switching, and the controlling element model. *Saccharomyces cerevisiae. The Early Days of Yeast Genetics*, eds Hall MN, Linder P (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), pp 291–304.
55. Silverman M, Simon M (1980) Phase variation: Genetic analysis of switching mutants. *Cell* 19(4):845–854.
56. Hicks J, Strathern JN, Klar AJ (1979) Transposable mating type genes in *Saccharomyces cerevisiae*. *Nature* 282(5738):478–483.
57. Chitrampalam P, Inderbitzin P, Maruthachalam K, Wu BM, Subbarao KV (2013) The *Sclerotinia sclerotiorum* mating type locus (MAT) contains a 3.6-kb region that is inverted in every meiotic generation. *PLoS ONE* 8(2):e56895.
58. Zörgö E, et al. (2013) Ancient evolutionary trade-offs between yeast ploidy states. *PLoS Genet* 9(3):e1003388.
59. Hickman MA, Froyd CA, Rusche LN (2011) Reinventing heterochromatin in budding yeasts: Sir2 and the origin recognition complex take center stage. *Eukaryot Cell* 10(9):1183–1192.
60. Ketel C, et al. (2009) Neocentromeres form efficiently at multiple possible loci in *Candida albicans*. *PLoS Genet* 5(3):e1000400.
61. Souciet JL, et al.; Génolevures Consortium (2009) Comparative genomics of protoploid Saccharomycetaceae. *Genome Res* 19(10):1696–1709.
62. Katz Ezov T, et al. (2010) Heterothallism in *Saccharomyces cerevisiae* isolates from nature: Effect of HO locus on the mode of reproduction. *Mol Ecol* 19(1):121–131.
63. Liachko I, Dunham MJ (2014) An autonomously replicating sequence for use in a wide range of budding yeasts. *FEMS Yeast Res* 14(2):364–367.
64. Faber KN, Haima P, Harder W, Veenhuis M, Ab G (1994) Highly-efficient electro-transformation of the yeast *Hansenula polymorpha*. *Curr Genet* 25(4):305–310.
65. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
66. Dobin A, et al. (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1):15–21.
67. Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH (2009) JBrowse: A next-generation genome browser. *Genome Res* 19(9):1630–1638.
68. Neuvéglise C, Feldmann H, Bon E, Gaillardin C, Casaregola S (2002) Genomic evolution of the long terminal repeat retrotransposons in hemiascomycetous yeasts. *Genome Res* 12(6):930–943.

Hanson et al; Supporting Information

Table of contents

Section	Page
Simulation of potential spore yield in colonies with different switching systems	2
Figure legends	4
References	7

Simulation of potential spore yield in colonies with different switching systems

S. cerevisiae and *S. pombe* switch mating types by destroying and replacing the active *MAT* locus, whereas the methylotrophs make a reciprocal swap between the active and silent loci. More specifically, *S. cerevisiae* and *S. pombe* use synthesis-dependent strand annealing (SDSA) to copy information unidirectionally from a silent locus into the broken *MAT* locus. SDSA involves two strand-exchange events, one in each of the homologous guide regions that flank the non-homologous *Yalpha* and *Ya* regions, and is resolved without crossover (1). These species achieve highly productive switching through biased choice of the appropriate donor; about 80-90% of the cells that are expected to switch do switch correctly to the opposite mating type (2, 3). In contrast, switching in the methylotrophs must preserve the outgoing *MAT* sequence for future re-use and probably proceeds via a single Holliday junction in the IR. This difference makes it unlikely that the methylotrophs can achieve the same levels of productive switching as *S. cerevisiae* and *S. pombe*, because the Holliday junction is expected to be resolved randomly into crossover (inverted) and non-crossover molecules (4), leading to a maximum of 50% successful switching from each attempt.

We used computer simulation (Fig. 6; Fig. S9) to study how efficiently different switching systems enable a small colony to make new spores if the environment demands it. Cells were followed in 10,000 replicate colonies, each starting from a single *MATalpha* spore and going through 7 mitotic divisions to the 128-cell stage. At each cell division, probabilistic rules for switching were applied corresponding to each known system (Fig. S9). All switching systems cause the frequency of the founder *MATalpha* allele to decline towards 50% (Fig. 6). In the *S. cerevisiae* and *S. pombe* systems, changing the productivity of switching from 50% (no donor bias) to 90% (strong donor bias) causes the frequency of *MATalpha* to decline faster. In the methylotroph system, the founder allele frequency falls quickly if switching is 50% productive, and more slowly if it is only 10% productive.

The ability of a simulated colony to make new spores was measured as the ‘potential spore yield’, defined as S/S_{max} where S is the number of spores that would be made if all pairs of cells in the colony with opposite mating types mated and sporulated, and S_{max} is the number that would be made if mating type did not matter and any two cells could mate. Potential spore yield is reduced by unequal allele frequencies, leaving some cells without a partner (5). In all systems except *P. pastoris*, it is simply twice the frequency of the rarer *MAT* allele (Fig. 6). Donor bias greatly increases potential spore yield relative to unbiased switching, especially at the 4-32 cell stages in *S. cerevisiae* and the 8-16 cell stages in *S. pombe*. Yield in *H. polymorpha* is low if it switches with 10% productivity, but at 50% productivity its yield is identical to that of *S. cerevisiae* in the absence of donor bias. As argued above we do not think that the productivity of switching in *H. polymorpha* can exceed 50%, but even if it does, potential spore yield declines again (the curves for 10% and 90% productivity are identical).

We simulated *P. pastoris* under the assumption that meiotic recombination can occur in the 123-kb unique region between the inner IRs, and that the genetic length of this region is 50 centimorgans, which leads to 2 inviable spores per meiosis. Thus potential spore yield in the *P. pastoris* model is low relative to *H. polymorpha*, for which we assumed no meiotic recombination in the 19-kb region. However, if inversion at the inner IRs of *P. pastoris* is permitted, yield increases (Fig 6).

Supplementary Figure Legends

Fig. S1. RT-PCR showing expression of genes between *MAT* loci within the 19-kb invertible region in *H. polymorpha*. Amplified cDNA was generated from haploid strains of **a** or alpha orientation. SH4331 is derived from NCYC495 background. +/- RT indicates addition of reverse transcriptase to the cDNA synthesis reaction.

Fig. S2. Localization of *H. polymorpha* centromeres by reanalysis of Hi-C data (6).
(A) Hi-C chromosomal contact matrix. Dots show single (green) and multiple (yellow) interactions, in 10 kb windows. Signals at consistent positions in each inter-chromosomal comparison derive from physical interactions between the centromeres of different chromosomes.
(B) Local interaction intensity on NCYC495 chromosome 3 shows a peak at 897 kb. LTRs of the Ty5-like retrotransposon Tpa5 are marked by orange circles. The red box shows the region magnified in D.
(C) G+C content variation (7) on chromosome 3.
(D) Expressed sequence tag coverage (blue) of the region between 875-975 kb on chromosome 3, showing that centromere is transcriptionally silent. The graphic below shows a complete Tpa5 retrotransposon (brown) in the centromeric region, the two IRs (gray) at 900 and 921 kb, *MATa* genes (green), *MATalpha* genes (purple), and other annotated genes (orange). The EST data was generated by the Joint Genome Institute (NCBI SRR346565) and mapped onto a version of the NCYC495 genome that was converted bioinformatically to the alpha orientation because the *MATalpha* genes are more strongly expressed in the EST data.

Fig. S3. RT-PCR showing expression of all four *MAT* genes in *P. pastoris* haploids with both orientations of the 138-kb region, and in diploids. Amplified cDNA was generated from haploid and diploid strains representing all four possible *MAT* locus organizations (alpha+, alpha-, **a+**, **a-**). +/- RT indicates addition of reverse transcriptase to cDNA synthesis reaction.

Fig. S4. Southern blot showing *MAT* locus orientation in *P. pastoris* strains GS115 and GS190. Schematics show location of restriction sites and probes (dark purple bars) used in the blot to determine orientation of (A) the outer inverted repeats, shown as blue rectangles, and (B) the inner inverted repeats shown as purple rectangles. Magenta and green regions represent *MATalpha* and *MATa* regions, respectively. Letters A and G correspond to primer locations used in PCRs in Figure 5. The orange circle the telomere. DNA was extracted from haploid strains grown in YPD or minimal (NaKG) medium for 0 or 4 hours. Hybridization reactions were performed with random-primed digoxigenin-labeled probes (Roche) generated according to manufacturer's instructions.

Fig. S5. Scale map of *MAT* locus regions in *H. polymorpha* and *P. pastoris*. Inverted repeat regions are highlighted in purple and blue. Numbers below the *MAT* genes are the percentage amino acid sequence identities to their *K. lactis* orthologs in a ClustalW alignment.

Fig. S6. Consequences of meiotic recombination between oppositely-oriented *MAT* regions in *P. pastoris*. The upper panel represents a diploid genome entering meiosis I with replicated copies of chromosome 4 in alpha- and **a+** orientations. If a meiotic recombination occurs within the 123-kb invertible region (orange arrow), chromosomes with large duplications and deficiencies will be produced, probably making two of the four spores inviable. If the centromere is located in the 1.7 Mb region, the recombinant chromosomes will also be dicentric and acentric. A general conclusion is that the larger the invertible region, the higher the probability of recombination within it, and hence the higher the expected number of inviable spores. If the two ends of the invertible region are so far apart that they are genetically unlinked, the fraction of recombinant chromosomes is expected to be 50%, and the average number of viable spores per meiosis will be 2 (as simulated in model Ppas_50/0 in Fig. 6). Alternatively (for example), if the two ends of the invertible segment are only 10 centimorgans apart the average number of viable spores per meiosis will be 3.6.

Fig. S7. Model for the origin of a flip-flop inversion system by introgression between *MAT* idiomorphs in a heterothallic species, based on the model of Chitrampalam *et al.* (8) for *Sclerotinia sclerotiorum*. The three chromosome rearrangement events shown could all involve recombination at short sequence repeats.

Fig. S8. Sequence of *H. polymorpha* tagged CSE4 construct.

Fig. S9. Models of cell lineages during mating-type switching in different yeast species, and computational framework for simulations. Panels A-C show DNA replication and cell division events from a single haploid cell (spore) to the 8-cell stage in 3 switching systems. Red arrows indicate steps in which switching is permitted, and black asterisks indicate cells that are competent to switch. The frequency of productive switching, f , is the probability that a cell that attempts to switch succeeds in changing its mating type.

(A) *S. cerevisiae* model using the nomenclature of Herskowitz (5). Cells are mating type alpha (green/magenta DNA strands) or **a** (blue DNA strands). Only mother cells can switch, and switched cells appear in pairs because switching occurs prior to S phase. If switching is fully productive, 2/4 grandchildren and 4/8 great-grandchildren of the original cell will show a switched mating type. In practice, switching is about 90% productive ($f = 0.9$). We also simulated *S. cerevisiae* switching without donor bias ($f = 0.5$).

(B) *S. pombe* strand-segregation model using the nomenclature of Klar (2). Cells are mating type P or M, and are either switchable (s) or unswitchable (u). Green and magenta indicate the Watson and Crick DNA strands of a P allele at the *mat1* locus, and blue indicates an M allele. Arrowheads on DNA strands indicate new synthesis. An origin of replication is located to the right of the *mat1* locus as drawn, so the Watson strand is always synthesized as a lagging strand. A switchable cell is one that contains two ribonucleotides (green asterisk) at a specific site in the Watson strand, left behind by incomplete removal of an RNA primer during lagging strand synthesis. When the cell attempts to use this imprinted strand as a template for DNA synthesis, replication fails and a double-strand break is formed, leading to mating type switching. If switching is fully productive, 1/4 grandchildren and 4/8 great-grandchildren of the original cell will show a switched mating type. In practice, switching is about 90% productive ($f = 0.9$). We also simulated *S. pombe* switching without donor bias ($f = 0.5$).

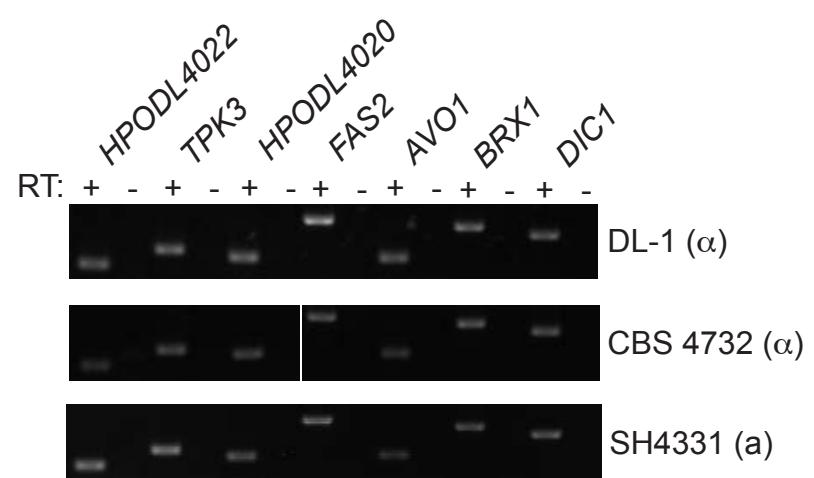
(C) Model for *H. polymorpha* and other species that use a 2-locus inversion system. Cells are mating type alpha or **a**. All cells can switch. Switched cells appear in pairs because switching occurs prior to S phase. Switching is maximally 50% productive ($f = 0.5$) because no more than 50% of Holliday junctions in the IR can be resolved as crossovers. Unlike the *S. pombe* and *S. cerevisiae* models, which are essentially deterministic because of their high productivity, the *H. polymorpha* model is probabilistic so different 8-cell colonies will contain different numbers of alpha and **a** cells. The example shown is one of many possible outcomes, depending on whether each cell that is competent to switch actually switches. We simulated *H. polymorpha* with all cells attempting to switch ($f = 0.5$), and with 20% of cells attempting to switch ($f = 0.1$). The model for *P. pastoris* is identical to *H. polymorpha* except that we tracked orientation of the 123-kb unique region and modeled inversion by recombination in the inner IR at either 0% or 10% probability (Fig. 5).

(D) Generic framework for computer simulations of switching. Cells have a mating type (1 or 0) and are switchable (s) or unswitchable (u). P_0 is the probability of switching mating type (from 1 to 0 or vice versa) in a parental cell prior to DNA replication, which will cause switched cells to appear in pairs. P_1 and P_2 are probabilities of a switch occurring in progeny 1 or progeny 2 after DNA replication. The probabilities P_0 , P_1 , and P_2 are either zero or f , depending on the species being modeled, as shown in the Table, where f is the expected average frequency of productive switching in cells that are competent to switch. There are no unswitchable cells in the *H. polymorpha* model. In *S. cerevisiae*, progeny 1 and 2 correspond to mother and daughter cells, respectively. Genotypes under this generic framework are shown at the bottom of panels A-C. Our simulations started with a single cell in state 1u (or 1s for *H. polymorpha*) and followed cell divisions for 7 generations with P_0 , P_1 and P_2 as specified, for various values of f (0.1, 0.5, 0.9), in 10,000 replicate colonies for each set of parameters.

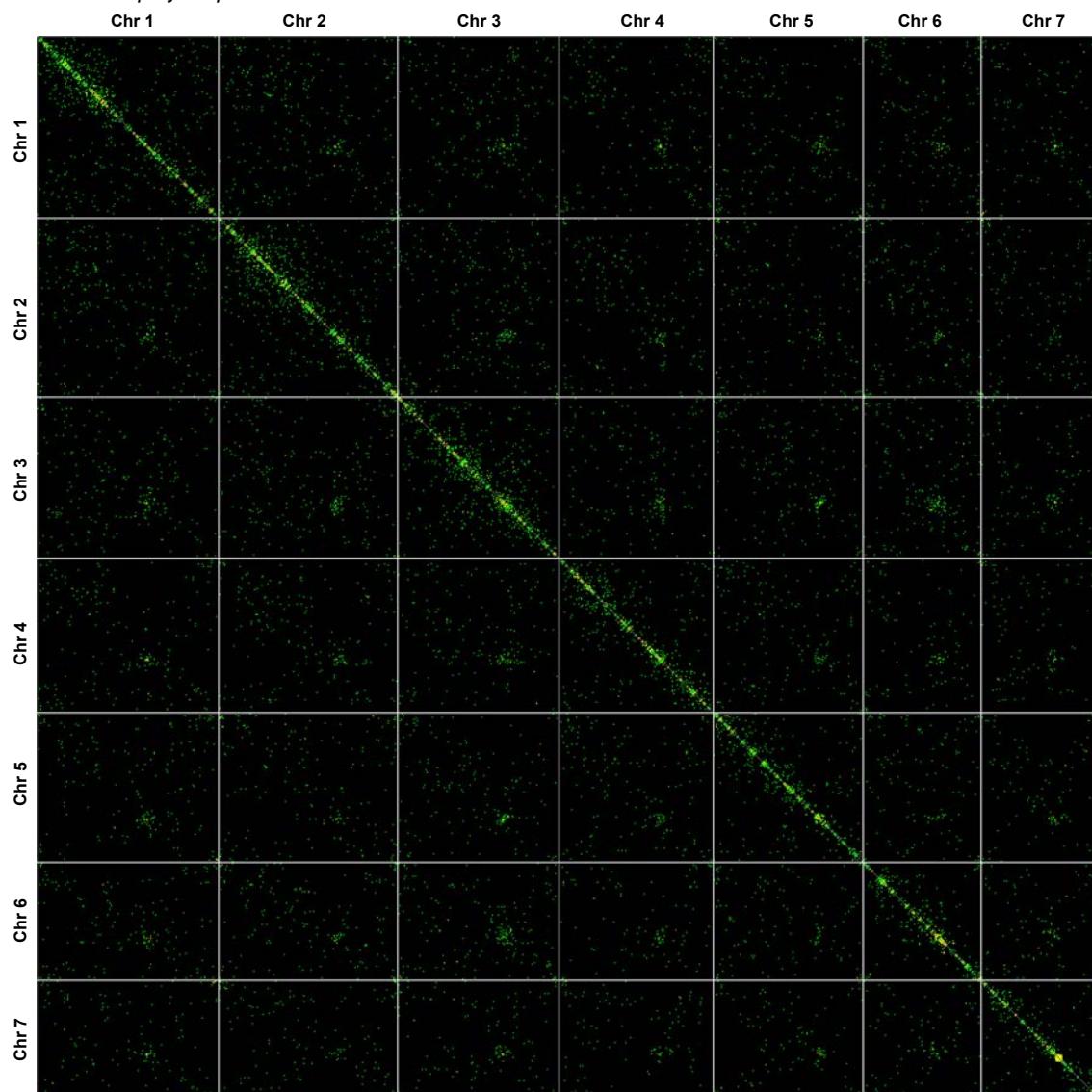
References for SI Appendix.

1. Haber JE, Ira G, Malkova A, Sugawara N (2004) Repairing a double-strand chromosome break by homologous recombination: revisiting Robin Holliday's model. *Philos Trans R Soc Lond B Biol Sci* 359: 79-86.
2. Klar AJ (2007) Lessons learned from studies of fission yeast mating-type switching and silencing. *Annu Rev Genet* 41: 213-236.
3. Strathern JN, Herskowitz I (1979) Asymmetry and directionality in production of new cell types during clonal growth: the switching pattern of homothallic yeast. *Cell* 17: 371-381.
4. Klein HL, Symington LS (2012) Sgs1--the maestro of recombination. *Cell* 149: 257-259.
5. Herskowitz I (1988) Life cycle of the budding yeast *Saccharomyces cerevisiae*. *Microbiol Rev* 52: 536-553.
6. Burton JN, Liachko I, Dunham MJ, Shendure J (2014) Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps. *G3 (Bethesda)* 4: 1339-1346.
7. Lynch DB, Logue ME, Butler G, Wolfe KH (2010) Chromosomal G + C content evolution in yeasts: systematic interspecies differences, and GC-poor troughs at centromeres. *Genome Biol Evol* 2: 572-583.
8. Chitrampalam P, Inderbitzin P, Maruthachalam K, Wu BM, Subbarao KV (2013) The *Sclerotinia sclerotiorum* mating type locus (*MAT*) contains a 3.6-kb region that is inverted in every meiotic generation. *PLoS One* 8: e56895.

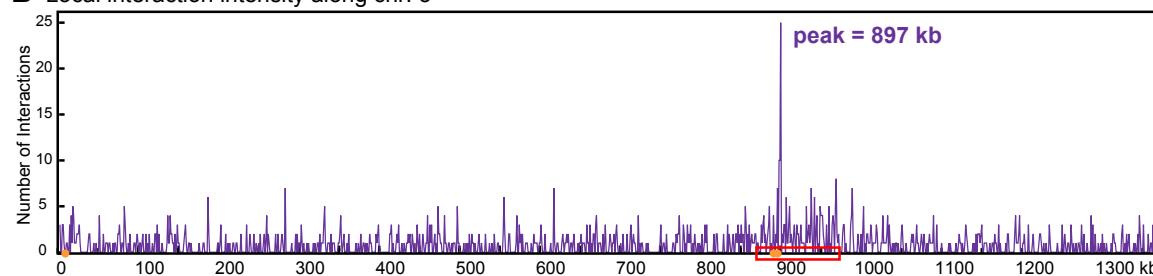
Hanson et al; Fig. S1



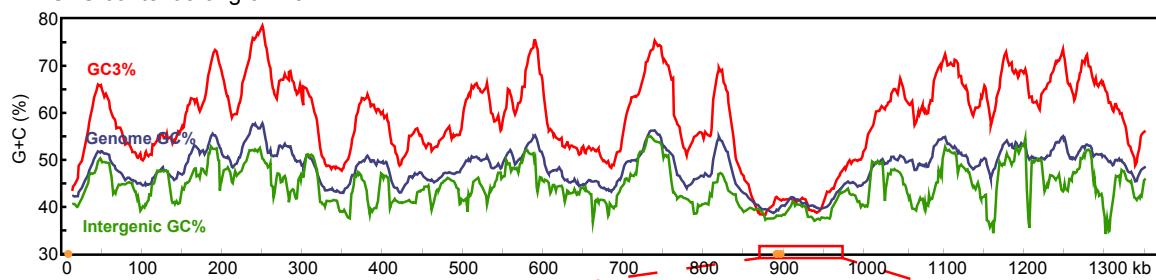
A *Hansenula polymorpha* NCYC495 Hi-C interaction matrix



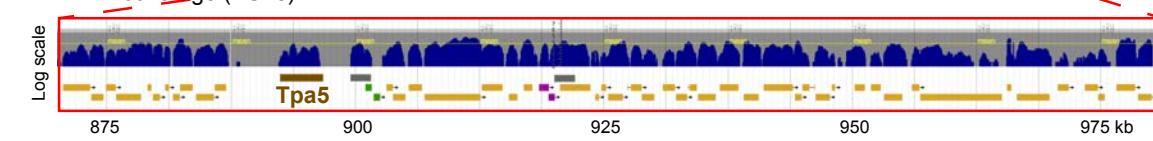
B Local interaction intensity along chr. 3



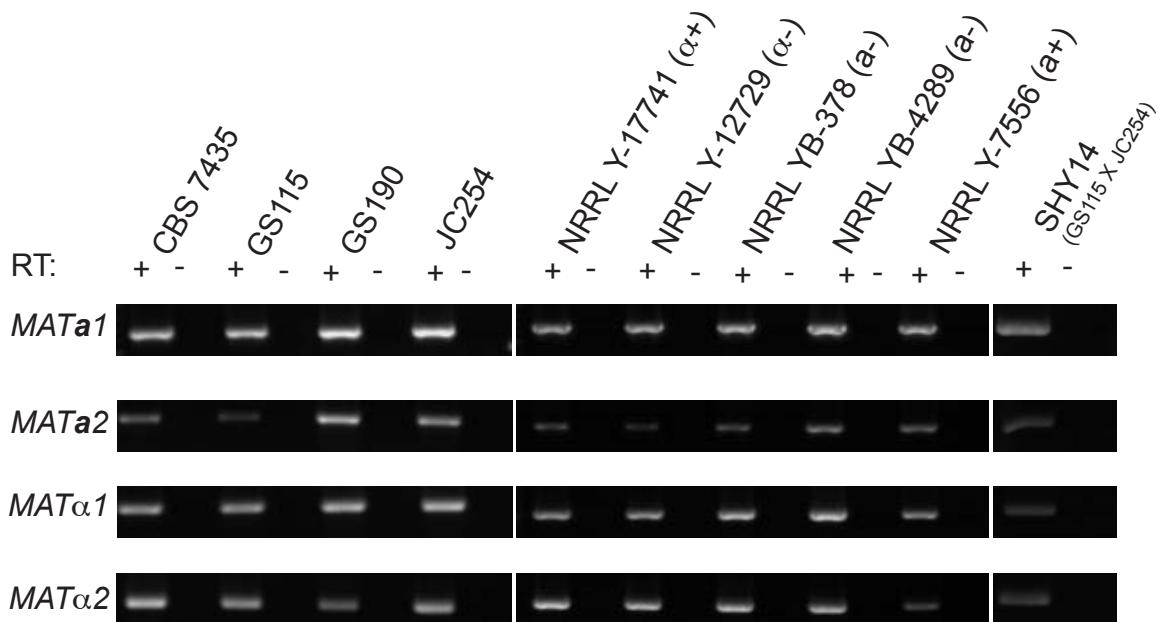
C G+C content along chr. 3



D mRNA coverage (ESTs)

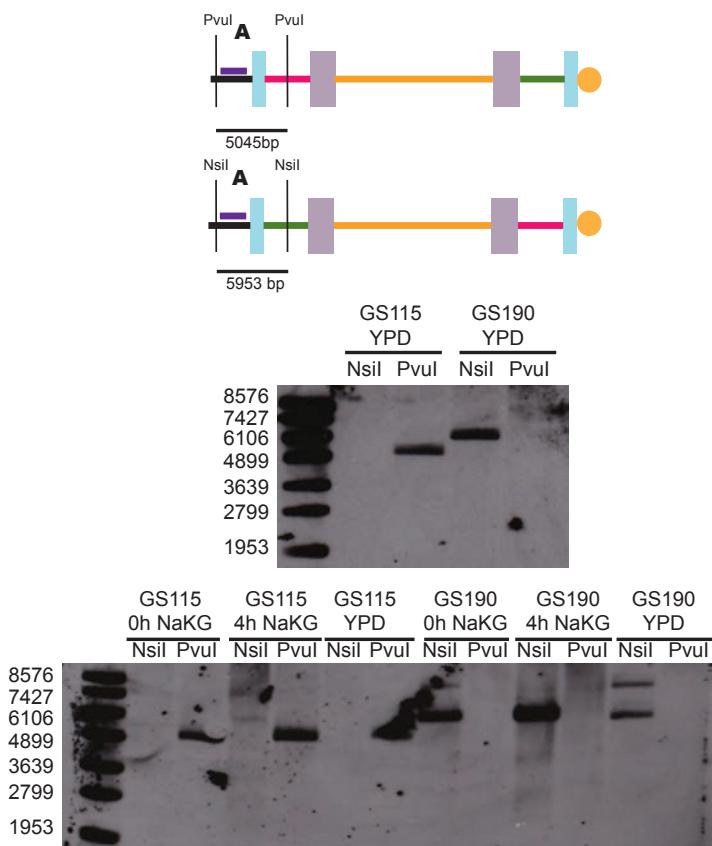


Hanson et al; Fig. S3

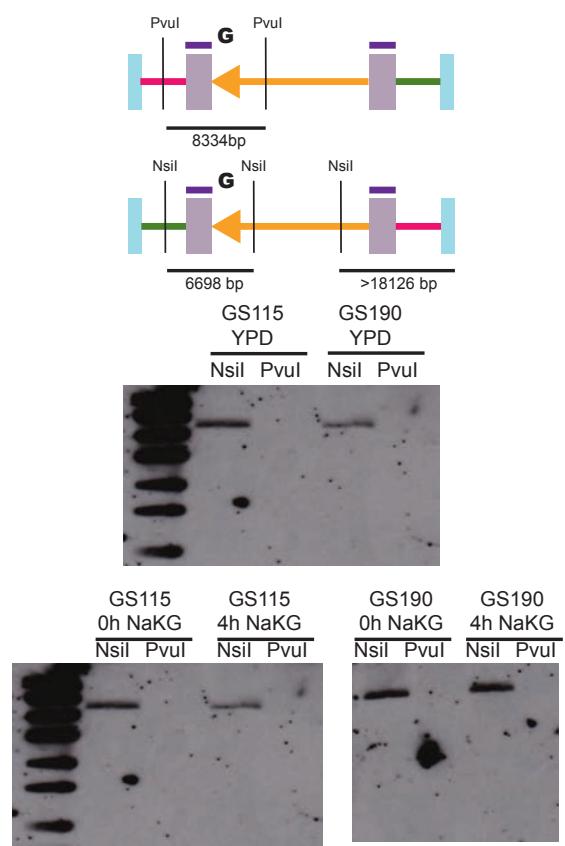


Hanson et al; Fig. S4

A

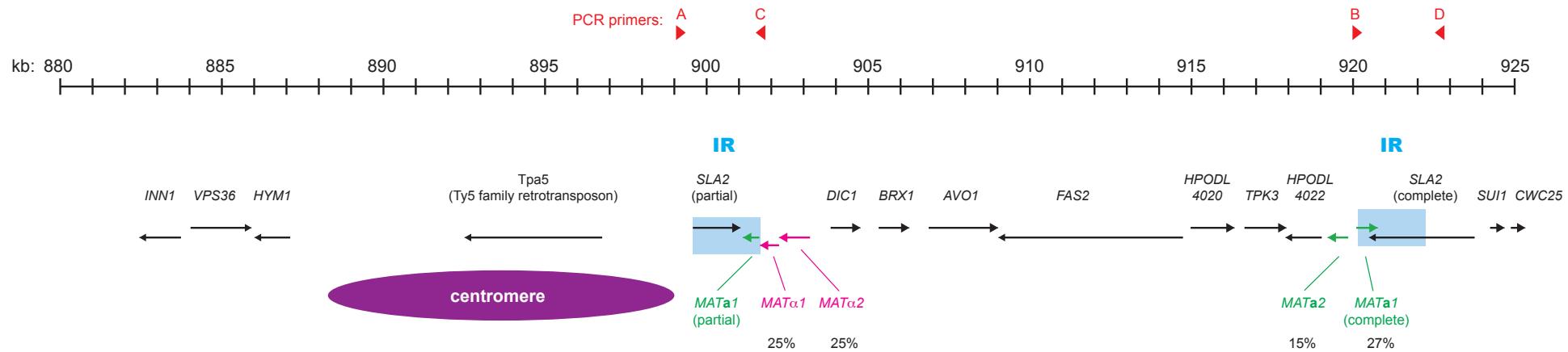


B

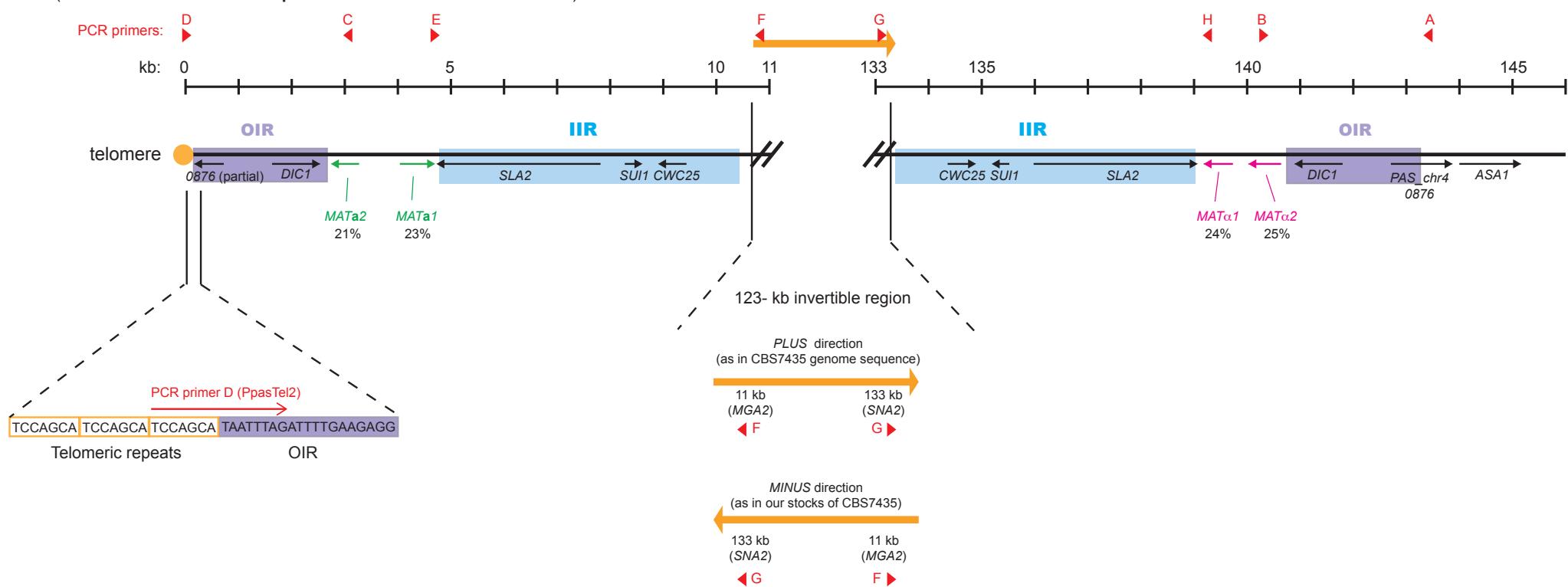


Hanson et al; Fig. S5

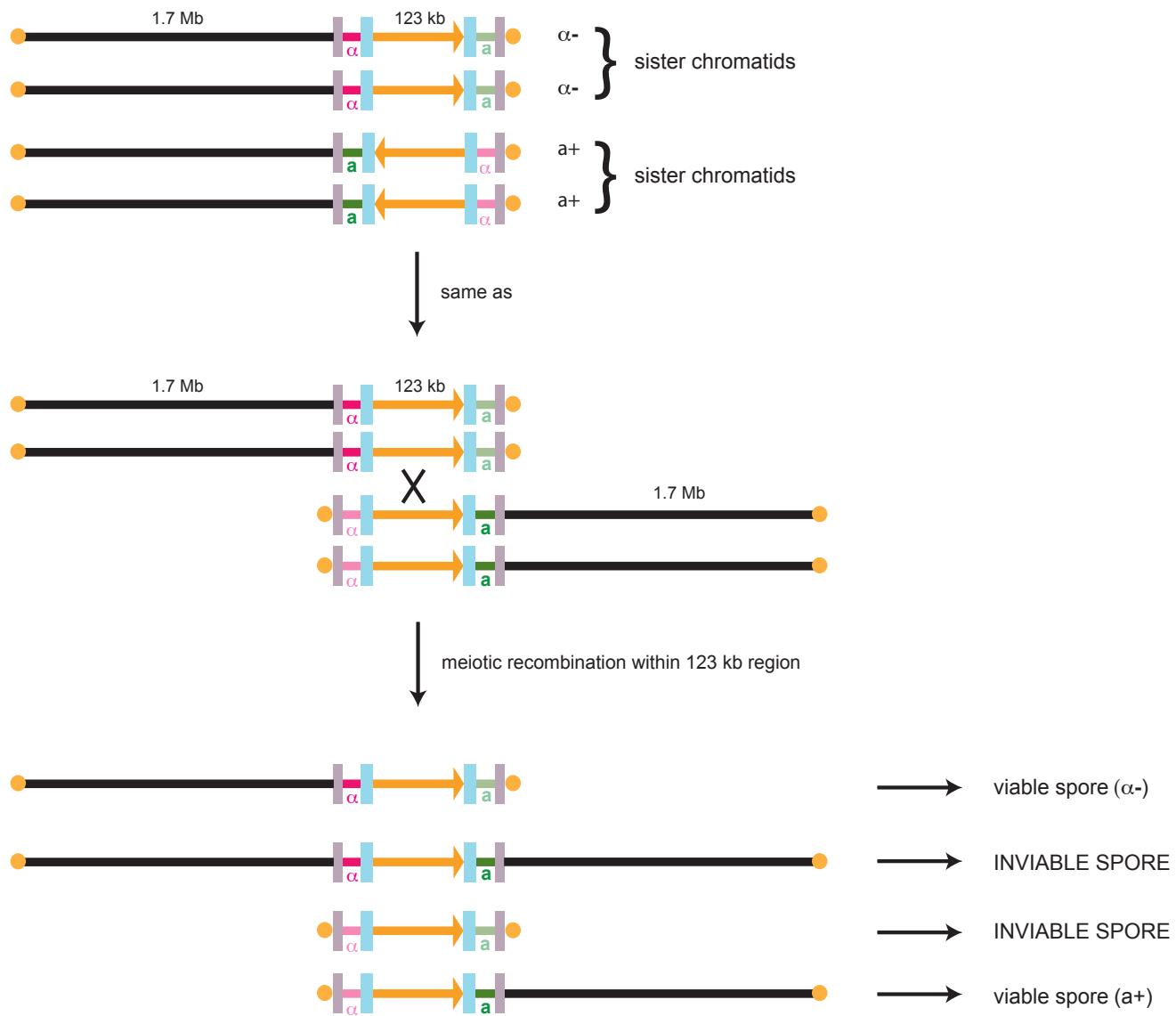
Hansenula polymorpha (NCYC495 chromosome 3 in **a** orientation; <http://genome.jgi.doe.gov/Hanpo2/Hanpo2.home.html>)



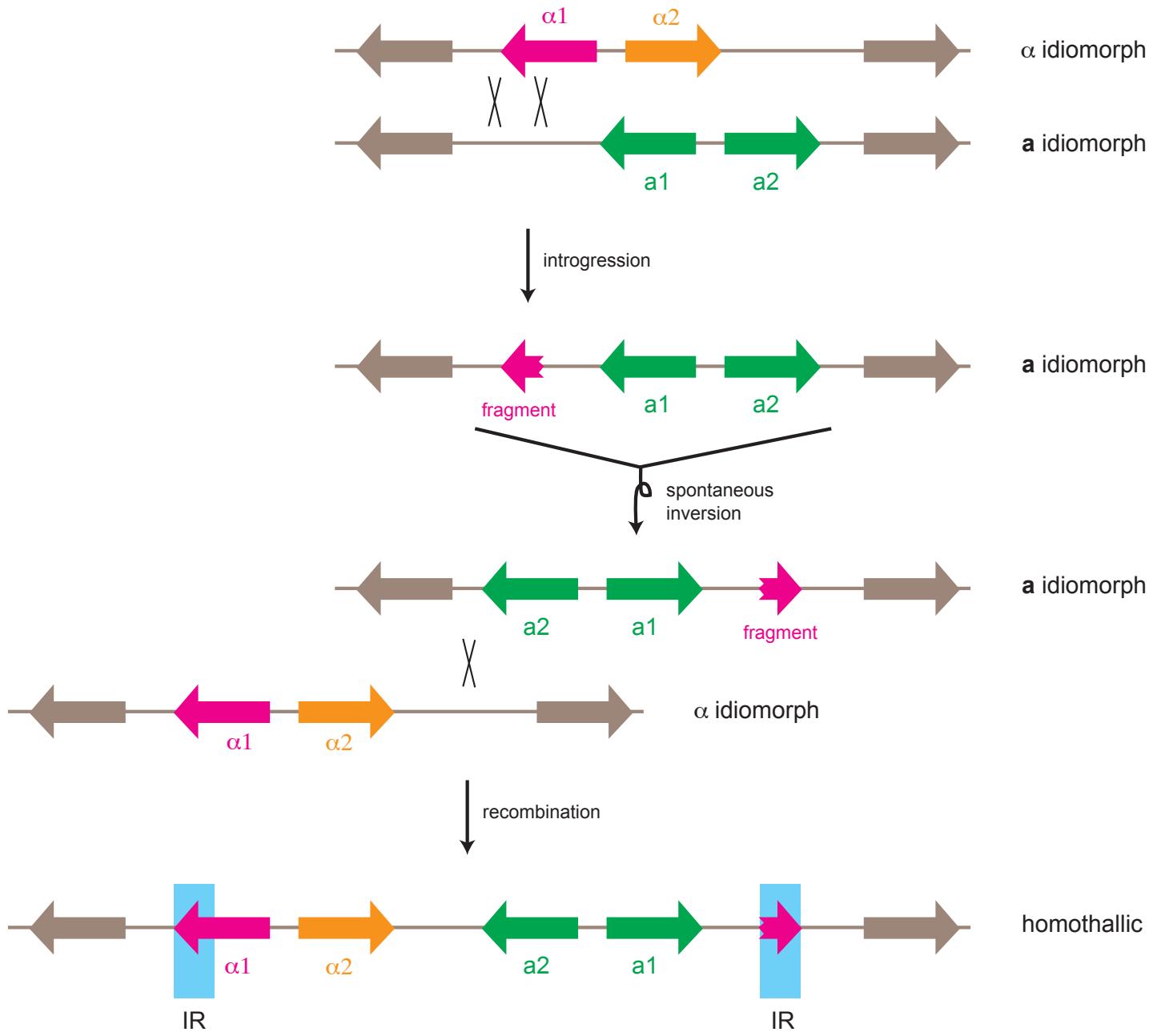
Pichia pastoris (CBS7435 chromosome 4 in **alpha+** orientation GenBank FR839631.1)



Hanson et al; Fig. S6



Hanson et al; Fig. S7



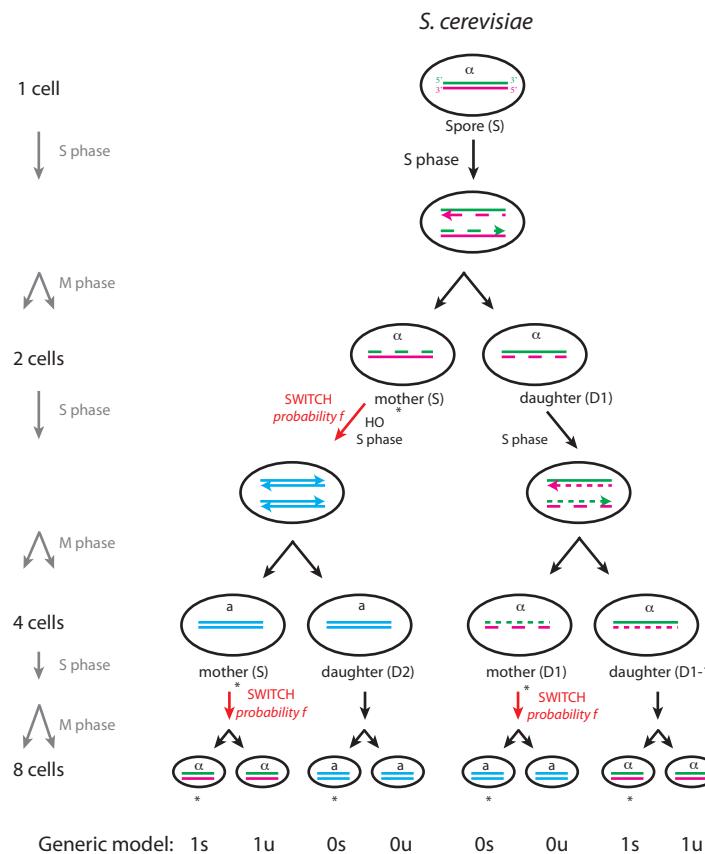
Hanson et al; Fig. S8

ccccacacccatagctcaaattgtttactcccttttactcttcaggatttctcgactccgcgcacgcggaccgttccacttccaaac
acccaaggcacagcataactaaattttcccttttcgtcaaaaaaggcaataaaaaattttatcacgttctttcttgcgaaaagaaaaaaaaga
gaccgcctcgitttttttcgtcaaaaaaggcaataaaaaattttatcacgttctttcttgcgaaaagaaaaaaaaga
acctccattgtatatttaagttaataaaacggctcaatttcaagtttcagtttcattttctgttcatataactttttacttttgcgatgt
gaaagaaaagcatagcaatctaagggcgatggccaggtaataaggcacaagccgtgtcgacaacgcggaaaagac
gcccaagatggcgtccagcacaagctcaagaatttgcgcaggccagttcaaaaaacaccgttcaaaatccccatcagatgttc
ctgactatgcggctatccgtatgcgtccggactatgcaggatccatcatgcgttccagattacgttagctctgagacgc
caaaggctcaaggccggctcgacgttagctggctcgagaaaccagccaggcgaccggacagaaaatcaaaacgcgg
aaaaaaaaagattcaagccggactgtggcgtcgagatcagaaggccagaagtccacagagctgttatcaggaaact
acccttgctagactcgtcagagaaaatcgtccaggacgcggatccatcatgcgttccagattacgttagctctgagacgc
gttgcaggaaagccgtcgaggcgatctgtcatcttttgcgatccacgcggacagcggactgtccatcatgcgttccagattacgttagctctg
tgcagaaaagacatccatctgtccaggagactgcgggggatgtaatcgtactgcgttccatcatgcgttccagattacgttagctctg
tcatttgtatagttttatattgttagttgttcatattaaatcaaattgttagcgttccatcatgcgttccagattacgttagctctg
aagttaagtgcgcagaaagtaatatcatgcgtcaatgtatgtgaatgcgttccatcatgcgttccagattacgttagctctg
catccatcatgcgtqaaaac

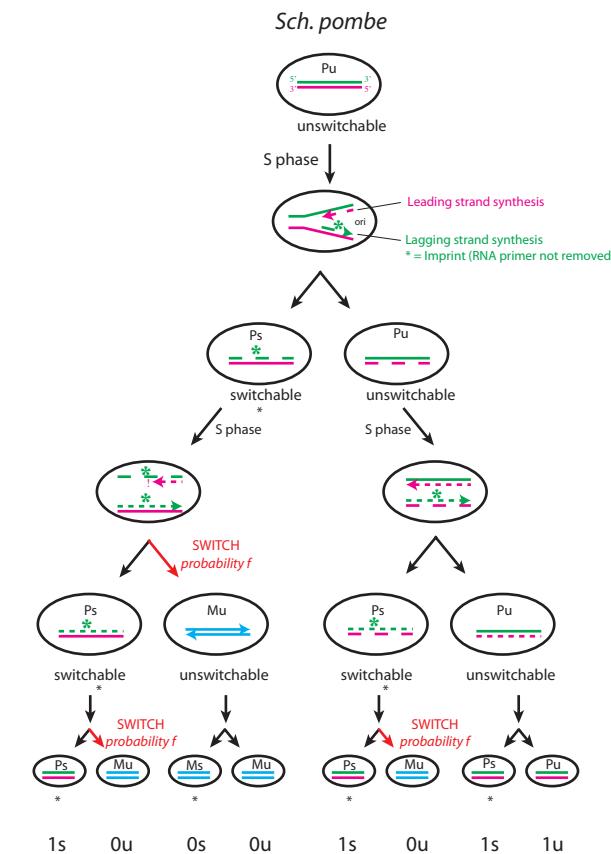
Saccharomyces cerevisiae TEF1 promoter
Hansenula polymorpha CSE4
3xHA tag
Ashbya gossypii TEF1 terminator

Hanson et al; Fig. S9

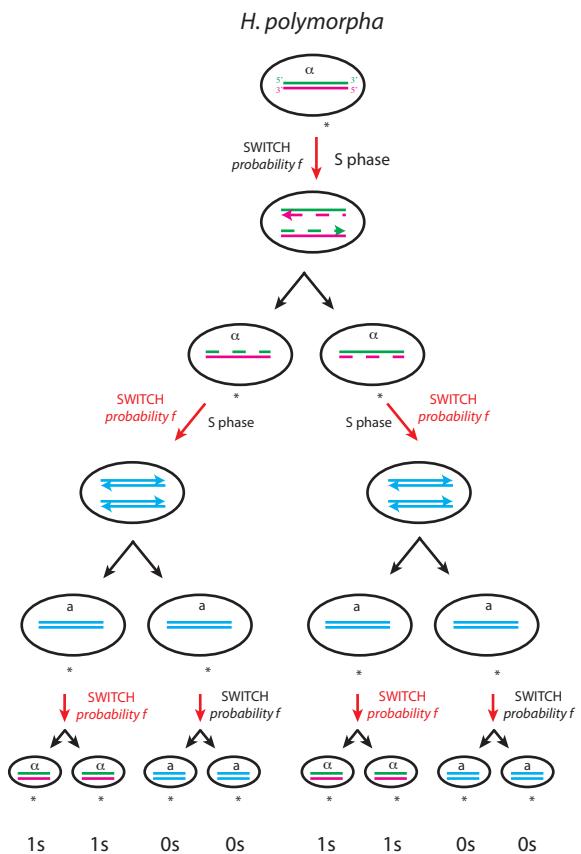
A



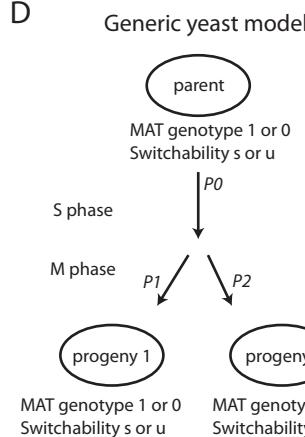
B



C



D



		<i>S. cerevisiae</i> parameters	<i>Sch. pombe</i> parameters	<i>H. polymorpha</i> parameters
		Switchability of parent	Switchability of parent	Switchability of parent
Probability of switching	<i>P</i> 0	s 0	0 0	f N/A
	<i>P</i> 1	0 0	0 0	0 N/A
	<i>P</i> 2	0 0	f 0	0 N/A
Switchability of progeny	progeny 1	s	s	s
	progeny 2	u	u	s

Table S1. qRT-PCR of MAT loci in *P. pastoris*.

Gene	Strain	ACT1 Average C _T	a1 Average C _T	ΔC _T (a1-ACT1)	ΔΔC _T (GS190 vs. GS115)	a1 _N relative to GS115	ΔΔC _T (GS115 vs. SHY15)	ΔΔC _T (GS190 vs. SHY15)	a1 _N relative to SHY15
a1	GS115 (α)	22.40 ± 0.18	22.89 ± 0.14	0.49 ± 0.22	0.00 ± 0.22	1.0 (0.85-1.16)	-2.17 ± 0.22	-----	4.50 (3.86-5.24)
	GS190 (α)	22.99 ± 0.18	23.99 ± 0.34	0.91 ± 0.39	0.42 ± 0.39	0.75 (0.57-0.97)	-----	-1.75 ± 0.39	3.36 (2.57-4.41)
	SHY15 (α/α)	22.38 ± 0.57	25.04 ± 0.73	2.66 ± 0.93	-----	-----	0.00 ± 0.93	0.00 ± 0.93	1.0 (0.52-1.91)
	Strain	ACT1 Average C _T	a2 Average C _T	ΔC _T (a2-ACT1)	ΔΔC _T (GS190 vs. GS115)	a2 _N relative to GS115	ΔΔC _T (GS115 vs. SHY15)	ΔΔC _T (GS190 vs. SHY15)	a2 _N relative to SHY15
a2	GS115 (α)	22.40 ± 0.18	28.41 ± 0.42	6.01 ± 0.46	0.00 ± 0.46	1.0 (0.73-1.38)	0.63 ± 0.46	-----	0.65 (0.47-0.89)
	GS190 (α)	22.99 ± 0.18	26.87 ± 0.31	3.88 ± 0.36	-2.13 ± 0.36	4.38 (3.41-5.62)	-----	-1.50 ± 0.36	2.83 (2.20-3.63)
	SHY15 (α/α)	22.38 ± 0.57	27.76 ± 0.19	5.38 ± 0.60	-----	-----	0.00 ± 0.60	0.00 ± 0.60	1.0 (0.66-1.52)
	Strain	ACT1 Average C _T	$\alpha 1$ Average C _T	ΔC _T ($\alpha 1$ -ACT1)	ΔΔC _T (GS115 vs. GS190)	$\alpha 1_N$ relative to GS190	ΔΔC _T (GS115 vs. SHY15)	ΔΔC _T (GS190 vs. SHY15)	$\alpha 1_N$ relative to SHY15
$\alpha 1$	GS115 (α)	22.40 ± 0.18	28.01 ± 0.14	5.60 ± 0.22	-0.86 ± 0.22	1.82 (1.19-2.77)	-0.26 ± 0.22	-----	1.20 (1.03-1.39)
	GS190 (α)	22.99 ± 0.18	29.46 ± 0.58	6.47 ± 0.61	0.00 ± 0.61	1.0 (0.85-1.16)	-----	0.61 ± 0.61	0.66 (0.43-1.00)
	SHY15 (α/α)	22.38 ± 0.57	28.25 ± 1.25	5.86 ± 1.38	-----	-----	0.00 ± 1.38	0.00 ± 1.38	1.0 (0.38-2.6)
	Strain	ACT1 Average C _T	$\alpha 2$ Average C _T	ΔC _T ($\alpha 2$ -ACT1)	ΔΔC _T (GS115 vs. GS190)	$\alpha 2_N$ relative to GS190	ΔΔC _T (GS115 vs. SHY15)	ΔΔC _T (GS190 vs. SHY15)	$\alpha 2_N$ relative to SHY15
a2	GS115 (α)	22.40 ± 0.18	26.92 ± 0.08	4.51 ± 0.19	-3.91 ± 0.19	15.03 (13.18-17.15)	-0.69 ± 0.19	-----	1.61 (1.41-1.84)
	GS190 (α)	22.99 ± 0.18	31.42 ± 0.86	8.43 ± 0.88	0.00 ± 0.88	1.0 (0.54-1.84)	-----	3.23 ± 0.88	0.11 (0.06-0.20)
	SHY15 (α/α)	22.38 ± 0.57	27.59 ± 0.89	5.20 ± 1.06	-----	-----	0.00 ± 1.06	0.00 ± 1.06	1.0 (0.48-2.08)

N = 2^{-ΔΔCT}; Values given + standard deviation

Table S2. Strains and plasmids used in this study.

Species	Strain name	Background	MAT locus designation (this study)	Genotype	Source
Hansenula polymorpha	CBS4732	CBS4732	MATalpha	<i>ade3 ura3</i>	Kantcho Lahtchev, Bulgarian Academy of Sciences, Sofia, Bulgaria
Hansenula polymorpha	DL-1	DL-1	MATalpha	<i>leu2 ura3</i>	Kantcho Lahtchev, Bulgarian Academy of Sciences, Sofia, Bulgaria
Hansenula polymorpha	NCYC495	NCYC495	MAT α	<i>ade11 met6</i>	Kantcho Lahtchev, Bulgarian Academy of Sciences, Sofia, Bulgaria
Hansenula polymorpha	SH4330	NCYC495	MATalpha	<i>ura3</i>	Satoshi Harashima, Osaka University, Osaka, Japan
Hansenula polymorpha	SH4331	NCYC495	MATalpha	<i>ade11</i>	Satoshi Harashima, Osaka University, Osaka, Japan
Hansenula polymorpha	ura3	NCYC495	MATalpha	<i>ura3</i>	Ida van der Klei, University of Groningen, Groningen, Netherlands
Hansenula polymorpha	leu1.1	NCYC495	MATalpha	<i>leu1</i>	Ida van der Klei, University of Groningen, Groningen, Netherlands
Hansenula polymorpha	ade11.1	NCYC495	MAT α	<i>ade11</i>	Ida van der Klei, University of Groningen, Groningen, Netherlands
Hansenula polymorpha	SHY2	NCYC495 X SH4330 (NCYC495)	MAT α/α	<i>ade11/ADE11 met6/MET6 URA3/ura3</i>	This study
Hansenula polymorpha	SHY2-3	NCYC495 X SH4330 (NCYC495)	MATalpha	<i>ura3</i>	This study
Hansenula polymorpha	SHY2-4	NCYC495 X SH4330 (NCYC495)	MAT α	<i>ura3</i>	This study
Hansenula polymorpha	SHY2-24	NCYC495 X SH4330 (NCYC495)	MATalpha	<i>met6</i>	This study
Hansenula polymorpha	SHY2-28	NCYC495 X SH4330 (NCYC495)	MAT α	<i>met6</i>	This study
Hansenula polymorpha	SHY7	SHY2-3 X SHY2-24 (NCYC495)	MAT α/α	<i>MET6/met6 ura3/URA3</i>	This study
Hansenula polymorpha	SHY10	SHY2-4 X SHY2-28 (NCYC495)	MAT α/α	<i>MET6/met6 ura3/URA3</i>	This study
Hansenula polymorpha	SHY17	NCYC495	MAT α	<i>ade11 met6 [pSH02]</i>	This study
Hansenula polymorpha	SHY32	NCYC495	MATalpha	<i>ura3 [pSH02]</i>	This study
Pichia pastoris	CBS7435	CBS7435	MATalpha-	WT	Spanish Type Culture Collection (CECT)
Pichia pastoris	GS115	CBS7435	MATalpha-	<i>his4</i>	Kristof De Schutter, Ghent University, Ghent, Belgium
Pichia pastoris	GS190	CBS7435	MAT $\alpha+$	<i>arg4</i>	James Cregg, Keck Graduate Institute of Applied Life Science, Claremont, CA
Pichia pastoris	JC254	CBS7435	MATalpha-	<i>ura3</i>	James Cregg, Keck Graduate Institute of Applied Life Science, Claremont, CA
Pichia pastoris	SHY14	GS115 X JC254 (CBS7435)	MAT $\alpha+/\alpha$	<i>HIS4/his4 ura3/URA3</i>	This study
Pichia pastoris	NRRL YB-378	NRRL YB-378	MAT α	WT	USDA Agricultural Research Service Culture Collection (from slime flux of elm tree, <i>Ulnus americana</i>)
Pichia pastoris	NRRL YB-4289	NRRL YB-4289	MAT α -	WT	USDA Agricultural Research Service Culture Collection (from black oak tree <i>Quercus kelloggii</i>)
Pichia pastoris	NRRL Y-7556	NRRL Y-7556	MAT $\alpha+$	WT	USDA Agricultural Research Service Culture Collection (from black oak tree <i>Quercus kelloggii</i>)
Pichia pastoris	NRRL Y-12729	NRRL Y-12729	MATalpha-	WT	USDA Agricultural Research Service Culture Collection (unknown habitat)
Pichia pastoris	NRRL Y-17741	NRRL Y-17741	MATalpha+	WT	USDA Agricultural Research Service Culture Collection (from sap flux of Emory oak tree, <i>Quercus emoryi</i>)
Plasmid	pSH02	pRS406	N/A	panARS - KanMX - pScTEF-3HA-HpCSE4	This study

Table S3. Primers used in this study.

Primer name	sequence (5'-3')	Label in figure
MAT locus PCR		
HpolMATa2	CCACTCATGGGAAATGATCCG	A
HpolMATb1	GAGTCATGGGTCTGGTTG	B
HpolMATb2	CTGCATGATATGACTACCAGCC	B
HpolMATc1	CTCAGATGATCCCACCACTAGG	C
HpolMATd1	CTGCGTCAGCTCAGGAATC	D
PpasAlpha2_1	GAGAGTTTCTTGGAGGAGC	B
PpasA2_1	GGCATAACCACGCAGGATATC	C
PpasMFS_1	CCCAGGTAAGTCAAAGCTGC	A
PpasMFS_2	GTTTGAGATTGAGGCCAGCATAG	A
PpasTel_1	TCCAGCATCCAGCAGTCAGCA	D
PpasTel_2	TCCAGCATCCAGCATAATTAGA	D
PpInt_E1	CTTCAAGCTCAGTCCATCC	E
PpInt_F1	CGTGGTGCTATCAGCTAATGTGCC	F
PpInt_G1	GTGCGACTCTCCACTAGAGC	G
PpInt_H1	CCTGGAATGAGCTACTCAGC	H
RT-PCR		
Hpola1-2	CACTCTTGTAGACGTCCAGTTC	
HpolMATb2	CTGCATGATATGACTACCAGCC	
Hpola2-1	ATCAGTGAAGGGCGTAGCAAAC	
Hpola2-2	GGGGCCAAGTGGATTTAGGTC	
Hpolalpha2-1	GGTTAGTGTGCTCCCTTCTTC	
Hpolalpha2-2	GAACGTCAAATACGAGCAGCCAG	
Hpolalpha1-1	GCATTCCGTAAATACATCCTCGAC	
HpolMATc2	CCATTCAAAGAGCAAGGAACCG	
HpGAPDHF1	GGAAGAATTGGTAGACTGGTGTG	
HpGAPDHR1	GTTCCCTTGAACTTCCGTGTG	
PpA1F1	CCAAGCTCTCACTTATCGACTGC	
PpA1R1	CGAGATGACTGACGGGTCA	
PpA2F3	CGAGTACCTGCAAATCGAGG	
PpA2R1	GGTTACGGTCTCCTGCG	
PpAlpha1F3	ACTGGTAAAATCAATACAGCGA	
PpAlpha1R1	CCATGCGACAGATAATAGTCTCG	
PpAlpha2F2	GCAAGGGATATCTCATAATGCG	
Ppalpha2_2	CACGTCGTTCTGAGCATTCTC	
PpGAPDHF1	CCCACAAGGCTTACAAGGGT	
PpGAPDHR1	ACACCGGTGGACTCAATGAC	
PpTUB1F3	CCAACGGTCCATACCCA	
PpTUB1R1	CACACTTGACCATTGGTTGCTG	
PpACT1F1	ACACAGTGTCCCACCGTC	
PpACT1R1	ACCGTGCTCGATTGGGTATC	
ChIP qPCR		
Hp_RegB2F	CTTGGAAAACCTTACTGCTGGTAG	899 kb
Hp_RegB1R	TGCCCCGTGAGAATGGGATC	899 kb
Hp_Tpa5ds_3F	GGCCAACCGAACCACTTCTATTTC	898 kb
Hp_Tpa5ds_2R	TTCGTCCCAATTCCAGGATAAC	898 kb
Hp_alpha1-2	CCAACATACGTCGCCCTTAC	
Hp_alpha1-3	GCCATGAAACCATTCAAACGC	
Hpolalpha2-2	GAACGTCAAATACGAGCAGCCAG	
Hp_alpha2-3	CGGTAGTGAACCTTCCATTGAAG	
HpolMATb2	CTGCATGATATGACTACCAGCC	
Hp_a1-4	TCCAAGATCCTCACGCCCTTC	
Hp_a2-2	GGGGCCAACGGATTTAGGTC	
Hp_a2-1	ATCAGTGAAGGGCGTAGCAAAC	
HpDIC1_1	GCATGCATTGGATGGCTTAC	
HpDIC1_2	CATGAATCCCAAGCCTTCTCTAG	
Hp_FAS2_1	GGCGCCTTACACAGATGAC	
Hp_FAS2_3	GCATCTCAGCTGTTCTGTGCG	
Hp_ORF_1	GGCATTGTTGTTAGTTACACCC	
Hp_ORF_2	CCATATGAAAGTCGATGCCGTC	
Hp_GAPDHF1	GGAAGAATTGGTAGACTGGTGTG	
Hp_GAPDHR1	GTTCCCTTGAACTTCCGTGTG	