Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch

Raul A. Ortiz-Merino¹, Nurzhan Kuanyshev², Stephanie Braun-Galleani¹, Kevin P. Byrne¹, Danilo Porro², Paola Branduardi², Kenneth H. Wolfe¹*

¹ UCD Conway Institute, School of Medicine, University College Dublin, Dublin, Ireland, ² Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milano, Italy

* kenneth.wolfe@ucd.ie

Abstract

Many interspecies hybrids have been discovered in yeasts, but most of these hybrids are asexual and can replicate only mitotically. Whole-genome duplication has been proposed as a mechanism by which interspecies hybrids can regain fertility, restoring their ability to perform meiosis and sporulate. Here, we show that this process occurred naturally during the evolution of Zygosaccharomyces parabailii, an interspecies hybrid that was formed by mating between 2 parents that differed by 7% in genome sequence and by many interchromosomal rearrangements. Surprisingly, Z. parabailii has a full sexual cycle and is genetically haploid. It goes through mating-type switching and autodiploidization, followed by immediate sporulation. We identified the key evolutionary event that enabled Z. parabailii to regain fertility, which was breakage of 1 of the 2 homeologous copies of the mating-type (MAT) locus in the hybrid, resulting in a chromosomal rearrangement and irreparable damage to 1 MAT locus. This rearrangement was caused by HO endonuclease, which normally functions in mating-type switching. With 1 copy of MAT inactivated, the interspecies hybrid now behaves as a haploid. Our results provide the first demonstration that MAT locus damage is a naturally occurring evolutionary mechanism for whole-genome duplication and restoration of fertility to interspecies hybrids. The events that occurred in Z. parabailii strongly resemble those postulated to have caused ancient whole-genome duplication in an ancestor of Saccharomyces cerevisiae.

Author summary

It has recently been proposed that the whole-genome duplication (WGD) event that occurred during evolution of an ancestor of the yeast S. cerevisiae was the result of a hybridization between 2 parental yeast species that were significantly divergent in DNA sequence, followed by a doubling of the genome content to restore the hybrid’s ability to make viable spores. However, the molecular details of how genome doubling could occur in a hybrid were unclear because most known interspecies hybrid yeasts have no sexual cycle. We show here that Z. parabailii provides an almost exact precedent for the steps...
Mechanism of whole-genome duplication in a fertile hybrid yeast

A whole-genome duplication (WGD) occurred more than 100 million years ago in the common ancestor of 6 yeast genera in the ascomycete family Saccharomycetaceae, including *Saccharomyces* [1, 2]. Recent phylogenomic analysis has shown that the WGD was an allopolyploidization—that is, a hybridization between 2 different parental lineages [3]. One of these parental lineages was most closely related to a clade containing *Zygosaccharomyces* and *Torulaspora* (ZT), whereas the other was closer to a clade containing *Kluyveromyces*, *Lachancea*, and *Eremothecium* (KLE). The ZT and KLE clades are the 2 major groups of non-WGD species in family Saccharomycetaceae. The WGD had a profound effect on the genome, proteome, physiology, and cell biology of the yeasts that are descended from it, but the genomes of these yeasts have changed substantially in the time since the WGD occurred, with extensive chromosomal rearrangement, deletion of duplicate gene copies, and sequence divergence between ohnologs (pairs of paralogous genes produced by the WGD). These changes have made it difficult to ascertain the molecular details of how the WGD occurred. Ancient hybridizations are rare in fungi (or at least difficult to detect [4]), but numerous relatively recent hybridizations have been identified using genomics, particularly in the ascomycete genera *Saccharomyces* [5, 6], *Zygosaccharomyces* [7–9], *Candida* [10–12], and *Milleroyzyma* [13].

Marcet-Houben and Gabaldón [3] proposed 2 alternative hypotheses for the mechanism of interspecies hybridization that led to the ancient WGD in the *Saccharomyces* lineage. Hypothesis A was hybridization between diploid cells from the 2 parental species, perhaps by cell fusion. Hypothesis B was mating between haploid cells from the 2 parental species to produce an interspecies hybrid zygote, followed by genome doubling. Under both hypotheses, the product is a cell with 2 identical copies of each parental chromosome. These identical copies should be able to pair during meiosis, leading to viable spores. While there are no known examples of natural yeast hybrid species formed by diploid–diploid fusion (hypothesis A), 3 examples have been discovered in which hybrid species were apparently formed simply by mating between haploids of opposite mating types from different species (hypothesis B). These are *Candida metapsilosis* [11], *C. orthopsilosis* [10, 12], and *Zygosaccharomyces* strain ATCC42981 [8, 14]. These interspecies hybridizations occurred by mating between parents with 4%–15% nucleotide sequence divergence between their genomes. However, none of these 3 hybrids can sporulate, which could be either because the homeologous chromosomes from the 2 parents are too divergent in sequence to pair up during meiosis or because pairing occurs but evolutionary rearrangements (such as translocations) between the parental karyotypes result in DNA duplications or deficiencies after meiosis [15–18]. None of these 3 hybrids has undergone the genome-doubling step envisaged in hypothesis B.

Several groups [3, 18–20] have proposed that genome doubling could occur quite simply by means of damage to 1 copy of the *MAT* locus in the interspecies hybrid, which could cause the hybrid cell to behave as a haploid, switch mating type, and hence autodiploidize. This proposal mimics laboratory experiments carried out by Greig et al. [21] in which hybrids between

CONACyT, Mexico (grant number 440667). RAOM. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Science Foundation Ireland (grant number 13/IA/1910). KHW. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. SYSBIO Centre of Systems Biology, SysBioNet Italian Roadmap for ESFRI Research Infrastructure. PB and DP. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** BIR, break-induced replication; chr., chromosome; ICHEC, Irish Centre for High-End Computing; ITS, internal transcribed spacer; KLE, Kluyveromyces-Lachancea-Eremothecium; LOH, loss of heterozygosity; mtDNA, mitochondrial DNA; NHEJ, nonhomologous end joining; ORF, open reading frame; PacBio, Pacific Biosciences; rDNA, ribosomal DNA; RNA-Seq, RNA sequencing; WGD, whole-genome duplication; YGAP, Yeast Genome Annotation Pipeline; ZT, *Zygosaccharomyces-Torulaspora*. 

proposed to have occurred during the *S. cerevisiae* WGD. Two divergent haploid parental species, each with 8 chromosomes, mated to form a hybrid that was initially sterile but regained fertility when 1 copy of its mating-type locus became damaged by the mating-type switching apparatus. As a result of this damage, the *Z. parabailii* life cycle now consists of a 16-chromosome haploid phase and a transient 32-chromosome diploid phase. Each pair of homeologous genes behaves as 2 independent Mendelian loci during meiosis.
different species of Saccharomyces were constructed by mating. The hybrids were unable to segregate chromosomes properly and were sterile, but when 1 allele of the MAT locus was deleted, they spontaneously autodiploidized by mating-type switching and were then able to complete meiosis and produce spores with high viability. Each spore contained a full set of chromosomes from both parental species [21]. While genome doubling via MAT locus damage is an attractive hypothesis consistent with hypothesis B above [3], no examples of it occurring in nature have been described. We show here that Z. parabailii has gone through this process.

There are 12 formally described species in the genus Zygosaccharomyces [22]. The most studied of these is Z. rouxii, originally found in soy sauce and miso paste [23, 24]. Others include Z. mellis, frequently found in honey [25], and Z. sapae from balsamic vinegar [26, 27]. Species in the Z. bailii sensu lato clade (Z. bailii, Z. parabailii, and Z. pseudobailii; [28]) are of economic importance because they are exceptionally resistant to osmotic stress and low pH. Their resistance to the weak organic acids commonly used as food preservatives makes them the most frequent spoilage agent of packaged foods with high sugar content, such as fruit juices and jams, or with low pH, such as mayonnaise [29–33]. These same characteristics make Zygosaccharomyces relevant to biotechnology since high stress tolerance and rapid growth are often desirable traits in microorganisms to be used as cell factories. The strain we analyze here, Z. parabailii ATCC60483, has previously been used for production of vitamin C [34], lactic acid [35], and heterologous proteins [36].

Despite the diversity of the genus, genome sequences have been published for only 2 non-hybrid species of Zygosaccharomyces: the type strains of Z. rouxii (CBS732T; [37]) and Z. bailii (CLIB213T; [38]). The genus also includes many interspecies hybrids with approximately twice the DNA content of pure species (20 Mb instead of 10 Mb; [7, 8, 14, 39]). Mira et al. [39] sequenced the genome of Zygosaccharomyces strain ISA1307 and found that it is a hybrid between Z. bailii and an unidentified Zygosaccharomyces species. In 2013, Suh et al. [28] proposed that some strains that were historically classified as Z. bailii should be reclassified as 2 new species, Z. parabailii and Z. pseudobailii, based on phylogenetic analysis of a small number of genes. The sequences of the RPB1 and RPB2 genes that they obtained from Z. parabailii and Z. pseudobailii contained multiple ambiguous bases, consistent with a hybrid nature [39]. In the current study, we sequenced the genome of a second hybrid strain, ATCC60483. We show that ATCC60483 and ISA1307 are both Z. parabailii and are both descended from the same interspecies hybridization event. By sequencing ATCC60483 using Pacific Biosciences (PacBio) technology, we obtained near-complete sequences of every Z. parabailii chromosome, which enabled us to study aspects of chromosome evolution in this species that were not evident from the Illumina assembly of ISA1307 [39].

Results

Z. parabailii ATCC60483 genome assembly by PacBio

We first tried to sequence the Z. parabailii genome using Illumina technology, but even with high coverage, we were unable to obtain long contigs. The data indicated that the genome was a hybrid, so instead we switched to PacBio technology, which generates long sequence reads (6 kb on average in our data). Our initial assembly had 22 nuclear scaffolds, which we refined into 16 complete chromosome sequences with a cumulative size of 20.8 Mb by manually identifying overlaps between the ends of scaffolds and by tracking centromere and telomere locations. We annotated genes using the Yeast Genome Annotation Pipeline (YGAP), assisted by RNA sequencing (RNA-Seq) data to identify introns. The nuclear genome has 10,087 protein-coding genes, almost twice as many as Z. bailii CLIB213T (Table 1).
Most of the chromosome sequences extend into telomeric repeats at the ends. The consensus sequence of the telomeres is `TGTGGTGGG GGGG`, which matches exactly the sequence of the template region of the 2 homeologous *TLC1* genes for the RNA component of telomerase that are present in the genome. Chromosome sequences that do not extend into telomeres instead terminate at gene families that are amplified in subtelomeric regions or contain genes that are at chromosome ends in the inferred Ancestral (pre-WGD) gene order for yeasts [41] indicating that they are almost full length, except for 3 chromosome ends that appear to have undergone break-induced replication (BIR) and homogenization with other chromosome ends.

We identified 1 scaffold as the mitochondrial genome, which maps as a 30-kb circle containing orthologs of all *S. cerevisiae* mitochondrial genes. We also found a plasmid in the 2-micron family (5,427 bp), with 99% sequence identity to pSB2, which was first isolated [42] from the type strain of *Z. parabailii* (NBRC1047/ATCC5607).

**Z. parabailii** ATCC60483 is an interspecies hybrid, with *Z. bailii* as 1 parent

Visualization of the genome using a Circos plot [43] shows that most of the genome is duplicated, indicating a polyploid origin (Fig 1). However, although most genes have a homeolog, the chromosomes do not form simple collinear pairs. Instead, sections of each chromosome are collinear with sections of other chromosomes.

Comparison to *Z. bailii* CLIB213<sup>T</sup> shows that for each region of the *Z. bailii* genome, there are 2 corresponding regions of the *Z. parabailii* genome: 1 almost identical in sequence and 1 with approximately 93% sequence identity, which demonstrates a hybrid (allopolyploid) origin of *Z. parabailii* and suggests that *Z. bailii* was one of its parents. To analyze this relationship in detail, we estimated the parental origin of every *Z. parabailii* ATCC60483 gene based on the number of synonymous substitutions per synonymous site (*K<sub>S</sub>*) when compared to its closest *Z. bailii* homolog (Fig 2A). This analysis revealed a bimodal distribution of *K<sub>S</sub>* values in which 47.1% of the ATCC60483 genes are almost identical to CLIB213<sup>T</sup> genes (*K<sub>S</sub> ≤ 0.05) and a further 42.5% are more divergent (0.05 < *K<sub>S</sub> < 0.25).

From this relationship, we infer that *Z. parabailii* ATCC60483 is an interspecies hybrid formed by a fusion of 2 parental cells, which we refer to as Parent A (purple) and Parent B (green). Parent A was a cell with a genome essentially identical to *Z. bailii* CLIB213<sup>T</sup>. Parent B was a cell of an unidentified *Zygosaccharomyces* species with approximately 93% overall genome sequence identity to *Z. bailii* homolog (*K<sub>S</sub> = 0.16) (Fig 2A). We refer to the 2 sets of DNA in *Z. parabailii* that were derived from Parents A and B as the A-subgenome and the B-subgenome, respectively. We refer to the A- and B-copies of a gene as homeologs, and we use a suffix ("_A" or "_B") in gene names to indicate which subgenome they come from.

The genome contains ribosomal DNA (rDNA) loci inherited from each of its parents. Our assembly includes 2 complete rDNA units with 26S, 5.8S, 18S, and 5S genes. Phylogenetic

---

**Table 1. Comparison of *Z. bailii* and *Z. parabailii* genome assemblies.**

| Strain     | Species      | Genome size (Mb) | Scaffolds | Scaffold N50 (Mb) | Reference     | tRNA genes<sup>a</sup> | Protein-coding genes |
|------------|--------------|------------------|-----------|------------------|---------------|------------------------|----------------------|
| CLIB213<sup>T</sup> | *Z. bailii* | 10.2             | 27        | 0.9              | Galeote et al. [38] | 161                    | 5,084                |
| ISA1307    | *Z. parabailii* | 21.2            | 154       | 0.2              | Mira et al. [39]  | 513                    | 9,925                |
| ATCC60483  | *Z. parabailii* | 20.8            | 16        | 1.3              | This study     | 499                    | 10,087               |

<sup>a</sup>We predicted the tRNA gene content of each genome assembly using tRNAscan-SE [40].

https://doi.org/10.1371/journal.pbio.2002128.t001
analysis of their internal transcribed spacer (ITS) sequences shows that the rDNA on chromosome 11 is derived from *Z. bailii* (Parent A), whereas the rDNA on chromosome 4 is derived from Parent B and contains an ITS variant seen only in other *Z. parabailii* strains (S1 Fig). A third rDNA locus in our assembly (at 1 telomere of chromosome 15) is incomplete and does not extend into the ITS region. The rDNA unit on chromosome 4 is also telomeric, whereas the unit on chromosome 11 is located at an internal site 165 kb from the right end. None of
the genes in the interval between this rDNA and the right telomere of chromosome 11 have orthologs in *Z. bailii* CLIB213\(^T\) homologs. (B) Pie chart showing the proportions of genes classified into each category. The 2 largest categories refer to A-genes and B-genes that are in A:B pairs. “N” means genes for which no *Z. bailii* homolog was found or \(K_s\) to *Z. bailii* exceeded 0.25. “As” and “Bs” indicate other A-genes and B-genes, as analyzed in panel C. (C) Breakdown of the numbers of genes assigned to the A- or B-subgenomes that are not in A:B pairs. See S1 Data for category counts and \(K_s\) values for each gene.

https://doi.org/10.1371/journal.pbio.2002128.g002

Table 2. *Z. parabailii* ATCC60483 chromosomes and centromeres.

| Chromosome | bp     | Protein-coding genes | tRNA genes | Ancestral centromere\(^a\) | Z. rouxii centromere |
|------------|--------|-----------------------|------------|--------------------------|---------------------|
| 1          | 2,110,500 | 1,010                          | 62         | Anc_CEN5 (B)             | Zr_CEN2             |
| 2          | 2,005,801 | 1,009                          | 55         | Anc_CEN6 (A)             | Zr_CEN3             |
| 3          | 1,751,495 | 868                            | 31         | Anc_CEN4 (A)             | Zr_CEN7             |
| 4          | 1,516,135 | 718                            | 29         | Anc_CEN2 (A)             | absent\(^\text{b}\) |
| 5          | 1,443,312 | 709                            | 44         | Anc_CEN5 (A)             | Zr_CEN2             |
| 6          | 1,315,104 | 614                            | 22         | Anc_CEN7 (B)             | Zr_CEN4             |
| 7          | 1,283,838 | 638                            | 29         | Anc_CEN6 (B)             | Zr_CEN3             |
| 8          | 1,249,162 | 634                            | 28         | Anc_CEN8 (B)             | Zr_CEN6             |
| 9          | 1,240,939 | 579                            | 43         | Anc_CEN1 (B)             | Zr_CEN5             |
| 10         | 1,189,704 | 576                            | 28         | Anc_CEN3 (A)             | Zr_CEN1             |
| 11         | 1,115,933 | 522                            | 16         | Anc_CEN2 (B)             | absent\(^\text{b}\) |
| 12         | 1,091,360 | 520                            | 24         | Anc_CEN4 (B)             | Zr_CEN7             |
| 13         | 1,077,716 | 517                            | 25         | Anc_CEN3 (B)             | Zr_CEN1             |
| 14         | 1,007,293 | 494                            | 16         | Anc_CEN7 (A)             | Zr_CEN4             |
| 15         | 858,772   | 406                            | 37         | Anc_CEN1 (A)             | Zr_CEN5             |
| 16         | 571,967   | 273                            | 10         | Anc_CEN8 (A)             | Zr_CEN6             |
| mtDNA      | 29,945   | 13                             | 20         |                         |                     |
| Total (nuclear) | 20,829,031 | 10087                        | 499        |                         |                     |

mtDNA, mitochondrial DNA.

\(^a\) Synteny correspondence between *Z. parabailii* centromeres and yeast Ancestral (pre–whole genome duplication [WGD]) centromere locations [44]. A and B indicate the subgenome assignments of the *Z. parabailii* centromeres.

\(^b\) *Z. rouxii* lost Anc_CEN2 in an evolutionary fusion of 2 chromosomes [44].

https://doi.org/10.1371/journal.pbio.2002128.t002
of a telomere-to-telomere fusion between 2 chromosomes followed by loss of a centromere [44]. The missing centromere in *Z. rouxii* is Ancestral centromere Anc_CEN2, which maps to *Z. parabailii* centromeres CEN4 and CEN11, located between the genes MET14 and VPS1. The *Z. rouxii* centromere must have been lost after it diverged from the *Z. bailii/Z. parabailii* lineage. Alignment of the *Z. rouxii* MET14-VPS1 intergenic region with the *Z. parabailii* CEN4 and CEN11 regions shows that the CDE III motif of the point centromere has been deleted in *Z. rouxii* (S2 Fig).

*Z. parabailii* inherited the mitochondrial genome of its *Z. bailii* parent. A complete mitochondrial genome sequence for *Z. bailii* is not available, but we identified 55 small mitochondrial DNA (mtDNA) contigs in the CLIB213 assembly, which together account for most of the genome, and calculated an average of 96% sequence identity between these and ATCC60483 mtDNA. CLIB213 lacks 2 of the 5 mitochondrial introns that are present in ATCC60483: the omega intron of the large subunit mitochondrial rDNA and intron 2 of *COX1*. Intraspecies polymorphism for intron presence/absence and comparable levels of intra-species mtDNA sequence diversity have been reported in other yeast species [45, 46].

**Prehybridization chromosomal rearrangements in *Z. parabailii*’s parents relative to *Z. bailii***

When genes in the Circos plot are colored according to their parent of origin, it is striking that many *Z. parabailii* chromosomes are either almost completely “A” (purple) or almost completely “B” (green) (outer ring in Fig 1), even though the chromosomes do not form collinear pairs. This pattern can be seen in more detail in a dot-matrix plot between *Z. bailii* and *Z. parabailii* (Fig 3). From this plot, it is evident that most of the A-subgenome is collinear with *Z. bailii* scaffolds, whereas the B-subgenome contains many rearrangements relative to *Z. bailii*. For example, *Z. parabailii* chromosome 1 is derived almost entirely from the...
B-subgenome but maps to about 12 different regions on the Z. bailii scaffolds. In contrast, Z. parabailii chromosome 3 is derived from the A-subgenome and is collinear with a single Z. bailii scaffold.

In total, from Fig 3 we estimate that there are approximately 34 breakpoints in synteny between the Z. parabailii B-subgenome and Z. bailii but no breakpoints between the A-subgenome and Z. bailii, when posthybridization rearrangement events (described below) are excluded. This difference in the levels of rearrangement in the A- and B-subgenomes relative to Z. bailii indicates that the 2 subgenomes were not collinear at the time the hybrid was formed. Therefore, most of the rearrangements between the 2 subgenomes are rearrangements that existed between the 2 parental species prior to hybridization. The 2 parents both had 8 chromosomes, but their karyotypes were quite different. Because each event of reciprocal translocation or inversion creates 2 synteny breakpoints [47], we estimate that about 17 events of chromosomal translocation or inversion occurred between the 2 parents in the time interval between when they last shared a common ancestor and when they hybridized.

The situation in Z. parabailii (hybridization between parents differing by 17 rearrangements and 7% sequence divergence) contrasts with that in the hybrid Millerozyma sorbitophila (only 1 detectable rearrangement between the parents, despite 15% sequence divergence [13]).

Posthybridization recombination, loss of heterozygosity (LOH), and BIR

Although the Z. parabailii genome largely contains unrearranged parental chromosomes, there have been 2 major types of rearrangement after hybridization. First, posthybridization recombination between the subgenomes at homeologous sites has formed some chromosomes that are partly “A” and partly “B.” Second, a process of homogenization has occurred at some places in which 1 subgenome overwrote the other, resulting in gene pairs that are A:A or B:B. This process is commonly called loss of heterozygosity (LOH) or gene conversion. Based on their $K_s$ distances from Z. bailii, the genome contains 4,153 simple A:B homeologous gene pairs, 300 A:A pairs, and 84 B:B pairs.

To examine the genomic locations of LOH and rearrangement events in more detail, we further classified genes using a scheme that takes account of their pairing status as well as their divergence from Z. bailii. Genes were defined as “A” or “B” as before or “N” if a $K_s$ distance from Z. bailii could not be calculated (Fig 2B and 2C). We then assigned each gene to 1 of 7 categories such as “B-gene in an A:B pair” or “A-gene, unpaired” and plotted the locations of genes in each category. The resulting map of the genome (Fig 4) allows LOH and recombination events to be visualized. N-genes (black in Fig 4) are seen to be mostly located near telomeres. Several points of recombination between the A- and B-subgenomes are apparent, such as in the middle of chromosome 4. LOH tends to occur in stretches that span multiple genes. For example, on chromosome 13, LOH has formed 8 runs of consecutive A-genes in a chromosome that is otherwise “B”; these A-genes are members of A:A pairs. They were probably formed by homogenization (gene conversion without crossover), although they could also be the result of double crossovers followed by meiotic segregation of chromosomes. Patches of LOH are frequently seen adjacent to sites of recombination between the 2 subgenomes (Fig 4). Three large regions of apparently unpaired A-genes near the ends of chromosomes (1L, 5L, and 9R; light blue in Fig 4) are probably artefacts caused by BIR, which is a process that can make the ends of 2 chromosomes completely identical from an initiation point out to the telomere [48]. These regions have 2x sequence coverage in our Illumina data, and we can identify the probable locations of an identical second copy of each of them at other chromosome ends (Fig 4).
Rearrangement catalyzed by HO endonuclease and degeneration of the “B” MAT locus

The Z. parabailii genome contains 2 MAT loci (one of which is broken) and 4 HML/HMR silent loci (Fig 5). In S. cerevisiae, mating-type switching is a DNA rearrangement process that occurs in haploid cells to change the genotype of the MAT locus [49]. During switching, the active MAT locus is first cleaved by an endonuclease called HO, and its α- or α-specific DNA is removed by an exonuclease. The resulting double-strand DNA break at MAT is then repaired by copying the sequence of either the HMLα or HMRα locus. This process converts a MATα genotype to MATα, or vice versa. Repeated sequences, called Z and X, located beside MAT and the HM loci act as guides for the DNA strand exchanges that occur during this repair process. The HM loci are “silent” storage sites for the α and α sequence information because genes at these loci are not transcribed due to chromatin modification; only MAT is transcribed [49].

We infer that the parents of Z. parabailii each contained a MAT locus and 2 silent loci (HMLα and HMRα), similar to S. cerevisiae and Z. rouxii haploids [50]. Fig 5A shows that
Fig 5. (A) Organization of MAT, HML, and HMR loci in Z. parabailii ATCC60483. The genome contains 6 MAT-related regions, with 1 MAT, 1 HML, and 1 HMR locus derived from each of the A and B parents. Pink and green backgrounds indicate sequences from the A- and B-subgenomes, respectively. The MAT locus in the A-subgenome (position 294 kb on chromosome 7) is intact and expressed. The MAT locus of the B-subgenome has been broken into 2 parts by cleavage by HO endonuclease. All 6 copies of the X repeat region (654 bp) are identical in sequence, as are all 6 copies of the Z repeat region (266 bp). Gray triangles indicate the disruption of the splicing of intron 2 in MATα2 and HMLα2 of the B-subgenome. The binding sites for primers A–F used for PCR amplification are indicated by gray arrows. (B) Sequences at the MAT locus breakpoint. Red, MATα1-derived sequences. The HO cleavage site (CGCAGCA, giving a 4-nucleotide 3’ overhang) is highlighted in gray. Blue, the GDA1-YEF1 intergenic region from the equivalent region of Z. bailii CLIB213 and homologous sequences from the A-subgenome on Z. parabailii chromosomes (chrs.) 2 and 16. A 5-bp sequence (ACAAC) that became duplicated during the rearrangement is underlined. (C) Sequences of MATα2 intron 2 (lowercase) from the A- and B-subgenomes. An AG-to-AC mutation (red) at the 3’ end of the intron moved the splice site by 2 bp in the B-subgenome, causing a frameshift and premature translation termination. The splice sites in both genes were identified from RNA sequencing (RNA-Seq) data.

https://doi.org/10.1371/journal.pbio.2002128.g005
Mechanism of whole-genome duplication in a fertile hybrid yeast

Z. parabailii has a MAT locus on chromosome 7, flanked by Z and X repeats and full-length copies of the genes SLA2 and DIC1, similar to the MAT loci of many other species [50, 51]. This MAT locus is derived from Parent A. Chromosome 7 also contains HMLα and HMRα loci (derived from Parent B) near its telomeres. However, the B-subgenome’s MAT locus is broken into 2 pieces. Most of it is on chromosome 2, but its left part (the 3’ end of MATα1, the Z repeat, and the neighboring gene SLA2) is on chromosome 16 (Fig 5A). Chromosomes 2 and 16 also each contain an HMLα or HMRα locus from the A-subgenome.

Examination of the breakpoint in the B-subgenome’s MAT locus shows that the break was catalyzed by HO endonuclease, because it occurs precisely at the cleavage site for this enzyme (Fig 5B). In S. cerevisiae, HO has a long (approximately 18 bp) recognition sequence that is unique in the genome, and it cleaves DNA at a site within this sequence, leaving a 4-nucleotide 3’ overhang [52]. Although the recognition and cleavage sites of HO endonucleases in other species have not been investigated biochemically, they can be deduced because the core of the HO cleavage site (CGCAGCA) invariably forms the first nucleotides of the Z region in each species [51]. Moreover, the HO cleavage site corresponds to an amino acid sequence motif (FAQQ) in the MATα1 protein that is strongly conserved among species.

The 2 parts of the broken MAT locus are located beside the genes GDA1 and YEF1 (Fig 5A), which are neighbors in Z. bailii CLIB213T and in the Ancestral yeast genome [38, 41]. Therefore, after HO endonuclease cleaved the “B” MAT locus, the broken ends of the chromosome apparently interacted with the GDA1-YEF1 intergenic region of the A-subgenome, causing a reciprocal translocation. This site is the only syntenic breakpoint between the A-subgenome of Z. parabailii and the genome of Z. bailii (scaffold 9; Fig 3). Comparison of the DNA sequences at the site (Fig 5B) shows no microhomology between the 2 interacting sequences and that DNA repair led to duplications of a 5-bp sequence (ACAAC) from the GDA1-YEF1 intergenic region and a 2-bp sequence (CA) from MATα1, suggestive of nonhomologous end joining (NHEJ) as the repair mechanism. We hypothesize that this genomic rearrangement occurred during a failed attempt to switch mating types, which resulted in a reciprocal translocation instead of normal repair of MAT by HML or HMR.

While the B-subgenome’s MATα1 gene is clearly broken, its MATα2 gene also appears to be nonfunctional. MATα2 has 2 introns, and our RNA-Seq data show how both homeologs of this gene (ZPAR0G01480_A and ZPAR0B05090_B) are spliced. A point mutation at the 3’ end of intron 2 of the B-gene changed its AG splice acceptor site to AC, with the result that splicing now uses another AG site 2 nucleotides downstream (Fig 5C). This change results in a frameshift, truncating the B-copy of the α2 protein to 57 amino acid residues instead of 211 and presumably inactivating it.

Surprisingly, the Z. parabailii genome does not contain any MATα1 (or HMRα1) gene. This gene codes for the α1 protein, which is 1 subunit of the heterodimeric α1-α2 transcriptional repressor that is formed in diploid (a/a) cells and which acts as a sensor of diploidy by repressing transcription of haploid functions such as mating while permitting diploid functions such as meiosis [53]. The α1 gene is present in Z. rouxii and Z. sapae [27, 37, 50], but it is also absent from Z. bailii CLIB213T and must have been absent from Parent B. The Z. bailii CLIB213T MAT organization is not fully resolved [38], but it contains a MAT’ locus with α1 and α2 genes on scaffold 14 and an HMR locus with only an α2 gene on scaffold 19. Evolutionary losses of MATα1 have previously been seen in some Candida species [54, 55], but not in any species of family Saccharomycetaeae. In contrast, the gene for the other subunit of the heterodimer, MATα2, is present in all Zygosaccharomyces species and is probably maintained because it has a second role in repressing a-specific genes in this genus [56]. Solieri and colleagues have reported evidence that α1-α2 is nonfunctional in a Z. rouxii/pseudorouxii hybrid in which its 2 subunits are derived from different species [14].
**Z. parabailii** strains ATCC60483 and ISA1307 are descendants of the same interspecies hybridization event

The 2 subgenomes apparent in the Illumina scaffolds of the *Zygosaccharomyces* hybrid strain ISA1307, previously sequenced by Mira et al. [39], are both 99%–100% identical in sequence to the A- or B-subgenomes of ATCC60483. Therefore, ISA1307 is also a strain of *Z. parabailii*. Importantly, the ISA1307 genome sequence contains the same HO-catalyzed reciprocal translocation between MATα1 of the B-subgenome and the GDA1–YEF1 intergenic region of the A-subgenome (Fig 5A). Because this rearrangement is so unusual and because it did not involve recombination between repeated sequences, it is highly unlikely to have occurred twice in parallel. The rearrangement is much more likely to have occurred only once, in a common ancestor of the 2 *Z. parabailii* strains after the hybrid was formed. It cannot pre-date the hybridization because it formed junctions between the A- and B-subgenomes, which originated from different parents.

ATCC60483 and ISA1307 are independent isolates of *Z. parabailii*, both from industrial sources. ATCC60483 was isolated from citrus concentrate used for soft drink manufacturing in the Netherlands [57, 58], and ISA1307 was a contaminant in a sparkling wine factory in Portugal [39, 59–61]. We found several examples in which the 2 strains differ in their patterns of LOH, which confirms that they have had some extent of independent evolution. All 3 large regions of BIR (on chromosomes 1, 5, and 9; Fig 4) are unique to ATCC60483. ISA1307 contains A:B homeolog pairs throughout these regions, whereas ATCC60483 has only A-genes, which we infer to be in A:A pairs. Other examples of differential LOH include a 4-kb region around homologs of the *S. cerevisiae* gene YLR049C, which exists as B:B pairs in ATCC60483 but A:B pairs in ISA1307, and the gene KAR4, which is an A:B pair in ATCC60483 but only a B-gene (single contig) in ISA1307. Notably, the section of the RPB1 gene (also called RPO21) that Suh et al. [28] used for taxonomic identification of *Z. parabailii* and *Z. pseudobailii* exists as an A:B pair in ATCC60483, but only as an A-gene in the ISA1307 genome. The absence of the B-copy of RPB1 made Mira et al. [39] hesitant to conclude that ISA1307 is *Z. parabailii*.

**Z. parabailii** ATCC60483 is fertile and haploid

Both ATCC60483 and the type strain of *Z. parabailii* ATCC56075T have previously been reported to be capable of forming ascospores [28, 57, 58]. We confirmed that our stock of ATCC60483 is able to sporulate (Fig 6A and 6B). On malt extract agar plates, we observed that sporulation occurs directly in zygotes formed by conjugation between 2 cells, resulting in asci in which the 2 former parental cell bodies typically contain 2 ascospores each. Such dumbbell-shaped (conjugated) asci, indicative of sporulation immediately after mating, are characteristic of the genus *Zygosaccharomyces* [25] and have previously been described in other *Z. bailii* (sensu lato) strains [25, 62–66]. The presence of conjugating cells in a culture grown from a single strain indicates that ATCC60483 is functionally haploid (capable of mating) and that it is homothallic (capable of mating-type switching). Since the zygote proceeds immediately into sporulation without further vegetative cell divisions, the diploid state of *Z. parabailii* appears to be unstable. Although Suh et al. [28] reported that asci of the type strain of *Z. parabailii* contain 2 spores, we consistently observed that asci occur in pairs of mated cells connected by a conjugation tube (Fig 6A and 6B), indicating that 4 spores are formed per meiosis.

We dissected tetrad asci from ATCC60483, grew colonies from the spores, and then used colony PCR to determine their genotype at the intact MAT locus on chromosome 7. Among 13 tetrads analyzed, 9 showed a ratio of 2 MATα colonies to 2 MATα colonies (Fig 6C and 6D). Two tetrads showed 1:3 or 3:1 ratios, and the other two yielded both MATα and MATα PCR products from some single-spore colonies. The genotype of the ATCC60483 starting strain is
Fig 6. (A,B) Ascospore formation in *Z. parabailii* ATCC60483. White arrows show conjugation tubes in dumbbell-shaped asci. Black arrows show budding vegetative cells. Scale bars, 10 μm. Cultures were grown on 5% malt extract agar for 6–10 days at 25˚C. (C) Examples of PCR determination of *MAT* locus genotypes in tetrads. Pairs of PCR primers as shown in Fig 5A were used to amplify the *MAT* locus in colonies grown from spores after dissection of conjugated asci. PCR primer pairs AB and AE amplify the left side of the *MAT* locus, including the Z region (AB, 1,485-bp product from *MATα*; AE, 2,103-bp product from *MATα*). Primer pairs DF and DC amplify the right side of the *MAT* locus, including the X region (DF, 1,882-bp product from *MATα*; DC, 2,027-bp product from *MATα*). PCR products were sequenced to determine whether they originated from the A- or B-subgenome. (D) Summary of *MAT* genotypes in colonies grown from spores from 13 dissected tetrads. Magenta circles denote colonies with A-subgenome alleles (*MATα_A* or *MATα_A*), and green circles denote colonies with B-subgenome alleles (*MATα_B* or *MATα_B*). Half circles represent colonies that gave both *MATα* and *MATα* PCR products.

https://doi.org/10.1371/journal.pbio.2002128.g006
MATα from the A-subgenome (designated MATα_A), so the presence of MATα genotypes in colonies derived from spores made by this strain confirms that mating-type switching occurred at some point. We sequenced the PCR products and found that the A- and B-subgenome HMRα loci were both used as donors for mating-type switching: among the pure MATα colonies, 18 were MATα_A and 7 were MATα_B (Fig 6D). Quite surprisingly, 4 tetrads with 2a:2α segregation had 1 MATα_A and 1 MATα_B spore colony, which is inconsistent with simple meiotic segregation from an a/α diploid. Because all the spores contain a functional HO gene, the genotypes of these 4 tetrads (#1, #7, #19, and #20) probably result from additional switches during the early growth of some colonies. Similarly, switching during early colony growth may explain the presence of MATα_B genotypes in tetrad #11 and the colonies with mixed a+α genotypes (in tetrads #11 and #13), as well as the presence of faint PCR products corresponding to the alternative MAT genotype in some other colonies (Fig 6C). In S. cerevisiae, homothallic diploid (HO/HO MATα/MATα) strains show 2:2 segregation of MAT alleles in tetrads, but after spore germination the haploid cells can then switch mating types as often as once per cell division [67], leading to mating and colonies that contain mostly diploid cells [68]; by contrast, most (but not all) of the Z. parabailii spore-derived colonies contained a single mating type (Fig 6C and 6D).

We found that almost all the genes involved in mating and meiosis that Mira et al. [39] reported to be missing from the Z. parabailii ISA1307 genome are in fact present in both ATCC60483 and ISA1307 (S1 Table). For example, we annotated A- and B-homeologs of IME1, UME6, DON1, SPO21, SPO74, REC104, and DIG1/DIG2 as well as MATα2, MATα1, and MATα2. We also identified genes for the α-factor and a-factor pheromones whose consensus sequence, AHYRILSPGAMF, is quite different from that of other yeasts, including Z. rouxii (7/13 matches) and S. cerevisiae (4/13 matches) [2]. Z. parabailii and Z. bailii do lack most of the ZMM group of genes, involved in crossover interference during recombination [69], even though these are present in Z. rouxii (S1 Table). Interestingly, identical sets of ZMM genes have been lost in Z. bailii/Z. parabailii relative to Z. rouxii, as were lost in most Lachancea species relative to Lachancea kluyveri [70]: ZIP2, CST9 (ZIP3), SPO22 (ZIP4), MSH4, MSH5, and SPO16 are absent, as well as MLH2, which is not known to be a ZMM gene, whereas ZIP1 is retained. A similar loss of ZMM genes has occurred in Eremothecium gossypii relative to E. cumbalariae [71].

Posthybridization gene inactivations
A small number of Z. parabailii ATCC60483 genes have “disabling” mutations—frameshifts or premature stop codons that prevent translation of a normal protein product. The majority of these mutations are present in only 1 subgenome of ATCC60483 and are unique to this strain. For example, there is a 1-bp insert in the A-homeolog of the DNA repair gene MLH1 that is not present in the B-homeolog or in ISA1307 or CLIB213T. In a systematic search, we found a total of 10 A-genes and 9 B-genes that were inactivated only in strain ATCC60483 (S2 Table). In each case, the other homeolog was intact, and the mutations, discovered in the PacBio assembly, were confirmed by our Illumina contigs of the ATCC60483 genome.

We found a further 8 disabling mutations that are shared between ATCC60483 and ISA1307. One of these is the AC-to-AG splice site mutation in the B-homeolog of MATα2 described above (Fig 5C). Another is the HO endonuclease gene, whose A-homeolog contains an identical 1-bp deletion in both ATCC60483 and ISA1307, whereas the B-homeolog of HO is intact in both strains (S2 Table). It is perhaps surprising that the HO gene that degenerated is the A-homeolog, whereas the broken MAT locus is the B-homeolog, but the 2 endonucleases
are likely to have had identical site specificities because the HO cleavage site is well conserved among species. The existence of these 8 shared disabling mutations provides further support for the idea that the 2 strains of *Z. parabailii* are descended from the same hybrid ancestor, because these mutations may not be viable in the absence of the intact homeologous copies of these genes. Only one of them is present also in CLIB213 (S2 Table).

**In-frame introns and other features of the genome**

We annotated 447 introns in the *Z. parabailii* ATCC60483 genome, most of which are confirmed by our RNA-Seq data. There are 428 intron-containing genes, including 19 that have 2 introns. We did not find any examples of intron presence/absence differences between homeologs. Interestingly, we found several genes with an in-frame intron—that is, an intron that is a multiple of 3 bp long and contains no stop codons, so that both the spliced and unspliced forms of the mRNA can be translated into proteins. Genes with in-frame introns are likely to undergo alternative splicing, making 2 forms of the protein with different functions. One of these loci is *PTC7* (ZPAR0J04940_A and ZPAR0A06900_B). Both of the *Z. parabailii* homeologs contain a 69-bp in-frame intron within the open reading frame (ORF) of the gene. It has previously been shown that alternative splicing of a similar in-frame intron in *S. cerevisiae* *PTC7* leads to the translation of a mitochondrial protein isoform from the spliced mRNA and a nuclear envelope protein isoform from the unspliced mRNA and that the intronic region codes for a transmembrane domain of the protein [72]. Thus, the alternative splicing mechanism in *PTC7* is conserved between *Saccharomyces* and *Zygosaccharomyces*. We also found in-frame introns in the *Z. parabailii* orthologs of *S. cerevisiae NUP100*, *NCB2*, and *HEH2*, identically in their A- and B-homeologs. None of these genes is known to be alternatively spliced in *S. cerevisiae*. In each of these examples, there are typical splice donor, branch, and acceptor sequences within the long form of the ORF.

Programmed "+1" ribosomal frameshifting, a process whereby the ribosome skips forward by 1 nucleotide when translating an mRNA, is known to occur in 3 genes in *S. cerevisiae: OAZ1, ABP140*, and *EST3* [73], and we found that +1 frameshifting is also required to translate the *Z. parabailii* orthologs of these 3 genes, in both the A- and B-homeologs. We also found 2 new loci that apparently undergo +1 frameshifting. Translation of both homeologs of *BIR1* (ZPAR0002690_A and ZPAR0I02720_B) requires a +1 frameshift at a sequence identical to the *EST3* frameshifting site: CTT-A-GTT, where the A is the skipped nucleotide. Translation of both homeologs of *YJR112W-A* (ZPAR0002960_A, ZPAR0I02990_B) requires a +1 frameshift at a sequence identical to the *ABP140* frameshifting site: CTT-A-GGC.

In *S. cerevisiae*, the *CUP1* locus confers resistance to copper toxicity by a gene amplification mechanism. *CUP1* codes for a metallothionein, a tiny cysteine-rich copper-binding protein. The reference *S. cerevisiae* genome sequence contains 2 identical copies of *CUP1* duplicated in tandem, but under copper stress this locus can become amplified to contain up to 18 tandem copies of the gene [74, 75]. There are at least 5 different types of *CUP1* repeats in different *S. cerevisiae* strains, which must have originated independently from progenitors with a single *CUP1* gene [75, 76]. In *Z. parabailii*, we found a slightly different organization. At homeologous loci on chromosomes 2 and 7, ATCC60483 has multiple identical copies of a 1,454-bp repeating unit. Each unit contains 2 metallothionein genes, *MT-58* and *MT-47*, coding for proteins of 58 and 47 residues, respectively. There is only 56% amino acid sequence identity between *MT-58* and *MT-47* proteins. The chromosome 7 locus contains 5 copies of the repeating unit, and the chromosome 2 locus contains 2 copies, so ATCC60483 has 14 metallothionein genes in total. These loci are not syntenic with *S. cerevisiae CUP1*, but they are syntenic with metallothionein genes in *C. glabrata* and *Z. rouxii* [77, 78].
Discussion

Our results show that *Z. parabailii* is a hybrid species that was formed by fusion between two 8-chromosome parental species, one of which was *Z. bailii*. The low sequence divergence of the ATCC60483 A-subgenome from the type strain of *Z. bailii* (the modal synonymous site divergence is less than 1%; Fig 2A) and the almost complete collinearity of these genomes (Fig 3) indicate that the A-parent of *Z. parabailii* should be regarded as *Z. bailii* itself, and not merely as a species closely related to *Z. bailii*.

The unusual MAT locus structure of this hybrid raised questions about how it was formed and whether *Z. parabailii* currently has a full sexual cycle. At first glance, the MATα/MATα hybrid genotype of ATCC60483 might suggest that *Z. parabailii* could not have been formed by mating. However, this genotype could also be the result of mating-type switching. We propose that the following steps occurred (Fig 7). *Z. parabailii* was formed by mating between strains of parent A (Z. bailii) and parent B, of opposite mating types. These parental genomes already differed by about 34 chromosomal rearrangement breakpoints, so the hybrid was unable to produce viable spores by meiosis. The hybrid also had no MAT locus due to an illegitimate recombination with the GDA1-YEF1 intergenic region instead of HML or HMR. At some point after hybridization, the HO gene from the A-subgenome also degenerated by acquiring a frameshift mutation.

The breakage of the “B” MAT locus can be inferred to have been one of the first rearrangement events that occurred after the hybridization but also to have been recent. It must have been recent because the pseudogene fragments of the broken HO gene, and 4 silent HML/HMR loci have not yet accumulated any other mutations. There are no nucleotide differences in 2,298 bp between the broken MATα_B locus on chromosome 2 and HMLα_B on chromosome 7. Together, these 2 observations suggest that the interspecies mating that formed *Z. parabailii* occurred less than 10^5 generations or 1,000 years ago [79]. Such a recent origin is consistent with the very low numbers of gene inactivations that have occurred since hybridization, with the fact that most of these are not shared between the 2 sequenced *Z. parabailii* strains (only 8 of 27 inactivating mutations are shared; S2 Table), and with the retention of rDNAs from both parents. We expect that, if the *Z. parabailii* lineage survives, it will accumulate extensive inactivations and deletions of redundant duplicated genes over the next few million years, as seen in older WGDs.

The net result of the evolutionary changes to the genome is that *Z. parabailii* now has 16 chromosomes (all different in structure but containing homoeologous regions), 1 active MAT locus, 1 active HO gene, and 4 silent HML/HMR loci. A genome with this structure resembles haploid *S. cerevisiae* [1] and is potentially capable of both mating-type switching and mating. We confirmed that both of these processes occur in ATCC60483. *Z. parabailii* has a life cycle in which 16-chromosome haploids mate to produce 32-chromosome diploids (Fig 7) that sporulate immediately because the diploid state is unstable; there is no MATα_1 gene, and hence, there is no a1–α2 heterodimer. Thus, *Z. parabailii* is an allopolyplid that regained
fertility by genome doubling after interspecies mating, as a consequence of damage to 1 copy of its MAT locus.

Two previous reports that Z. parabailii strains produce only mitotic spores [31, 80] can be reinterpreted in view of the hybrid nature of the genome. Their experimental data are fully compatible with the meiotic sexual cycle we propose for Z. parabailii. Rodrigues et al. [80] made a derivative of ISA1307 in which 1 copy of ACS2 was disrupted by the G418-resistance marker APT1 and the other copy was not. After sporulation of this strain, all 80 spores they tested were G418 resistant, and all 16 spores from 4 tetrads contained both an intact copy of ACS2 and an acs2::APT1 disruption, which led Rodrigues et al. [80] to conclude that the spores were made by mitosis. However, this inheritance pattern is exactly the pattern expected if the 2 copies of ACS2 are homeologs (different Mendelian loci) rather than alleles of a single
Mendelian locus and if ISA1307 is a haploid that autodiploidized before it sporulated. Thus, their strain could be described as haploid \textit{acs2}_a::\textit{APT1 ACS2}_B, where the \textit{ACS2}_A and \textit{ACS2}_B loci have independent inheritance (they are on chromosomes 10 and 13 in our genome sequence). Similarly, Mollapour and Piper [31] disrupted 1 of the 2 copies of \textit{YME2} in strain NCYC1427 with a \textit{kanMX4} cassette and found that all the spores produced by this strain retained both an intact \textit{YME2} and \textit{yme2::kanMX4}. They concluded that the spores were vegetative, but again, the result is consistent with meiotic spore production if the 2 \textit{YME2} loci have independent inheritance (they are on chromosomes 4 and 6) and if the disruption was made in a haploid strain that autodiploidized before sporulating. The sequence data in [31] allow NCYC1427 to be identified as \textit{Z. parabailii} and not \textit{Z. bailii} as originally described. Furthermore, in both ISA1307 [80] and NCYC1427 [25, 31], spores are formed in pairs of conjugated cells, similar to Fig 6A. We conclude that ISA1307 and NCYC1427 have sexual cycles identical to the one we describe for ATCC60483.

The evolutionary steps that formed \textit{Z. parabailii} by interspecies mating, and restored its fertility by damage to one of its \textit{MAT} loci, are essentially identical to one of the mechanisms (hypothesis B) proposed for the origin of the ancient WGD in the \textit{S. cerevisiae} lineage [3, 18–20]. Our study therefore validates genome doubling after \textit{MAT} locus damage as a real evolutionary process that occurs in natural interspecies hybrids, enabling them to resume mating and meiosis. The \textit{Z. parabailii} hybridization was very recent, so any period of clonal reproduction that elapsed before fertility was restored must have been short, which is as expected because there is no selection to maintain meiosis genes during clonal growth [18, 20]. The possible role of \textit{MATa1} in the ancient WGD remains unclear. In \textit{Zygosaccharomyces}, the absence of this gene makes zygotes proceed into sporulation. In the ancient WGD, it is likely that a \textit{MATa1} gene was present in the initial zygote, in which case the zygote would have been stable until it sustained \textit{MAT} locus damage, but this is not certain because the ZT parent might have lacked \textit{MATa1}. The specific cause of damage to the \textit{MAT} locus in \textit{Z. parabailii} was incorrect DNA repair after cleavage by the mating-type switching endonuclease HO. The \textit{HO} gene is present in the ZT clade, but not in the KLE clade [51, 81], and these 2 clades were the 2 parental lineages of the interspecies hybridization that led to the ancient WGD [3]. Species that contain \textit{HO} show evolutionary evidence of repeated deletions of DNA from beside their \textit{MAT} loci, caused by accidents during mating-type switching [51]. Indeed, the disappearance of the \textit{MATa2} gene from Saccharomycetales genomes, which occurred at approximately the same time as the WGD, must have been due to some sort of mutational damage to the \textit{MAT} locus. Although \textit{HO}-mediated damage can only occur in the small clade of yeasts that contain \textit{HO}, other types of mutational damage to 1 copy of \textit{MAT} are a plausible mechanism for fertility restoration in other fungal interspecies hybrids.

\section*{Materials and methods}

\subsection*{Strain and growth media}

The strain analyzed here originally came from the collection of Thomassen & Drijver-Verblifa NV in the Netherlands [57, 58] and was called “\textit{Saccharomyces bailii} strain 242” in those studies. It was isolated from citrus concentrate being used as raw material for soft drinks. It was later deposited at the American Type Cultures Collection as ATCC60483. Suh et al. [28] identified it as \textit{Z. parabailii} by molecular methods.

\subsection*{PacBio DNA sequencing, assembly, and annotation}

ATCC60483 genomic DNA was prepared using the Blood & Cell Culture DNA Mini Kit (Qiagen), according to the manufacturer’s manual. To prevent fragmentation of the DNA,
the sample was not vortexed. The final genomic DNA amount was 15 μg as determined by Qubit Fluorometer (Thermo Scientific). PacBio sequencing was carried out by the Earlham Institute (Norwich, United Kingdom) using 8 SMRT cells, which generated 218x mean coverage for the nuclear scaffolds. We assembled the raw data using the computational facilities at the Irish Centre for High-End Computing (ICHEC), with the HGAP3 protocol of the SMRT Analysis suite version 2.3.0 [82]. We initially obtained 22 nuclear scaffolds, which we reduced to 16 chromosomes by manually identifying overlaps between scaffolds. In parallel, we also obtained 198x Illumina read coverage of the genome (Genome Analyzer IIx; University of Milano-Bicocca, Department of Clinical Medicine), which we assembled separately into contigs that were used to verify the status of rearrangement points and pseudogenes discussed in the text.

The \textit{Z. parabailii} chromosomes were annotated using an improved version of our automated YGAP [83], which uses information in the Yeast Gene Order Browser [78] and the Ancestral (pre-WGD) gene order [41] to generate a synteny-based annotation. The automated annotation was curated using transcriptome data from ATCC60483 cultures grown in a bioreactor; Illumina RNA-Seq was generated at Parco Tecnologico Padano (Italy). We made a de novo transcriptome assembly using Trinity [84] and compared the transcripts against YGAP’s gene models using PASA [85] and by manual inspection of spliced mRNA reads.

Chromosomes were numbered 1 to 16 from largest to smallest. Genes were given systematic names by YGAP such as \textit{ZPAR0D01210_B}, where ZPAR indicates the species; 0 indicates the genome sequence version; \textit{D} indicates chromosome 4; 01210 is a sequential gene number counter that increments by 10 for each protein-coding gene (genes that were added manually have numbers that end in 5 or other digits); and the suffix \textit{_B} indicates that this gene is assigned to the B-subgenome as described below. NCBI nucleotide sequence database accession numbers are CP019490–CP019505 (nuclear chromosomes), CP019506 (mitochondrial genome), and CP019507 (2-micron plasmid).

The mitochondrial genome of \textit{Z. bailii} CLIB213\textsuperscript{T} was not reported with the rest of this strain’s genome [38] and is highly fragmented in the assembly. We identified mitochondrial contigs in the original CLIB213\textsuperscript{T} assembly by BLASTN using the ATCC60483 mtDNA as a query, assembled these contigs into 55 larger contigs using the CAP3 assembler and SSPACE3 [86, 87], and calculated a weighted average nucleotide identity of 96% from nonoverlapping alignments totaling 23,197 bp.

**Gene assignments to the A- and B-subgenomes**

We assigned most genes in \textit{Z. parabailii} ATCC60483 to either the A-subgenome (highly similar to the \textit{Z. bailii} CLIB213\textsuperscript{T} genome) or the B-subgenome (derived from the other parent in the hybridization), using their levels of synonymous nucleotide sequence divergence from CLIB213\textsuperscript{T} genes. For this purpose, we used BLASTP [88] to compare every annotated protein from ATCC60483 to the CLIB213\textsuperscript{T} proteome and designated the best hit as a homolog. The corresponding ATCC60483 and CLIB213\textsuperscript{T} DNA sequence pairs were then aligned using CLUSTALW [89], and their levels of sequence divergence were calculated using the \textit{yn00} program from the PAML suite [90]. ATCC60483 genes were assigned to the A-subgenome if the level of synonymous divergence was \( K_S \geq 0.05 \) and to the B-subgenome if \( 0.05 < K_S \leq 0.25 \) and given an \textit{_A} or \textit{_B} suffix on the gene name accordingly. Genes for which \( K_S > 0.25 \) or for which no \textit{Z. bailii} homolog was identified were given the suffix \textit{_N}. To identify inactivated genes systematically, we searched the annotated A:B gene pairs for cases in which one of the homeologs was less than 90% of the length of the other, and we then examined these cases manually (S1 Table).
Note that our use of the labels “A” and “B” differs from the scheme used by Mira et al. [39] for strain ISA1307. We designated each gene (homeolog) as either “A” or “B” based on its divergence from *Z. bailii* CLIB213, with “A” always indicating the *Z. bailii*-like homeolog. Some chromosomes therefore contain mixtures of “A” and “B” genes due to posthybridization recombination or homogenization between the 2 subgenomes. In contrast, Mira et al. [39] identified homeologous pairs of scaffolds in their assembly and arbitrarily designated 1 scaffold as “A” and the other as “B” so that each scaffold is homogeneous, but there is no consistent relationship between the “A” and “B” labels and the parent-of-origin of a homeolog in their scheme.

**Tetrad dissection and MAT locus PCR amplification**

Cells were left for sporulation on malt extract (5%) agar for 5 days. A small loop of cells was washed in sterile distilled water, resuspended in a 1:20 dilution of Zymolyase 100T, and incubated for 10 min at 30°C. The Zymolyase solution was removed by centrifugation, and the pellet resuspended in distilled water (500 μl). A 10-μl drop was placed in the middle of a YPD plate, and dumbbell-shaped asci were dissected using a Singer Sporeplay dissection microscope. The YPD plate was incubated for 2 days at 30°C. Individual spore-derived colonies were used for MAT locus genotyping by colony PCR using Q5 polymerase high-fidelity 2x master mix (NEB) and annealing temperature 55°C. Sequences of PCR primers A–F are given in S3 Table. Primers E and F were designed to bind equally to the *HMR* regions of the A- and B-subgenomes. Primers A–D are specific for the A-subgenome.

**Supporting information**

**S1 Fig.** Phylogenetic tree of internal transcribed spacer (ITS) regions of rDNA. Chr4 and Chr11 are the ITS sequences from the chromosome 4 and 11 rDNA units in the *Z. parabailii* ATCC60483 genome. All other sequences are from Suh et al. [28] for strains of *Z. parabailii* (Zpar), *Z. bailii* (Zbai), and *Z. pseudobailii* (Zpse). Letters a-n are ITS variant designations [28]. The tree was constructed by PhyML in the Seaview package using default parameters.

(TIF)

**S2 Fig.** Sequence alignment of the *VPS1-MET14* intergenic regions from *Z. rouxii* and *Z. parabailii*. The *Z. parabailii* regions contain CEN4 and CEN11 whereas the *Z. rouxii* region is not a centromere. Putative CDE I and CDE III motifs are boxed.

(TIF)

**S1 Table.** *Z. parabailii* orthologs of *S. cerevisiae* genes involved in mating or meiosis.

(XLSX)

**S2 Table.** Homeolog pairs in which one gene is damaged and the other is intact, in ATCC60483.

(XLSX)

**S3 Table.** PCR primer sequences used for MAT locus genotyping.

(DOC)

**S1 Data.** Excel spreadsheet containing, in separate sheets, the underlying numerical data for Fig panels 2A, 2B and 2C.

(XLSX)
Acknowledgments

We thank John Morrissey and Francesca Doonan for encouragement and support, Simon Wong at ICHEC for help with genome assembly, Laura Dato for initial work on Illumina genome sequencing, Isabel Sá-Correia for strain ISA1307, Virginie Galeote for CLIB213 data, and Geraldine Butler for comments on the manuscript.

Author Contributions

Conceptualization: Danilo Porro, Paola Branduardi, Kenneth H. Wolfe.
Data curation: Raúl A. Ortiz-Merino, Kevin P. Byrne.
Formal analysis: Raúl A. Ortiz-Merino.
Funding acquisition: Danilo Porro, Paola Branduardi, Kenneth H. Wolfe.
Investigation: Raúl A. Ortiz-Merino, Nurzhan Kuanyshev, Stephanie Braun-Galleani, Kevin P. Byrne.
Methodology: Nurzhan Kuanyshev, Kevin P. Byrne, Paola Branduardi.
Project administration: Danilo Porro, Paola Branduardi, Kenneth H. Wolfe.
Resources: Paola Branduardi.
Software: Raúl A. Ortiz-Merino, Kevin P. Byrne.
Supervision: Danilo Porro, Paola Branduardi, Kenneth H. Wolfe.
Validation: Raúl A. Ortiz-Merino, Nurzhan Kuanyshev, Kevin P. Byrne.
Visualization: Raúl A. Ortiz-Merino, Kevin P. Byrne.
Writing – original draft: Raúl A. Ortiz-Merino, Kenneth H. Wolfe.
Writing – review & editing: Paola Branduardi, Kenneth H. Wolfe.

References

1. Wolfe KH, Shields DC. Molecular evidence for an ancient duplication of the entire yeast genome. Nature. 1997; 387: 708–713. https://doi.org/10.1038/42711 PMID: 9192896
2. Wolfe KH, Armisen D, Proux-Wera E, O’Eigeartaigh SS, Azam H, Gordon JL, et al. Clade- and species-specific features of genome evolution in the Saccharomycetaceae. FEMS Yeast Res. 2015; 15: fov035. https://doi.org/10.1093/femsyr/fov035 PMID: 26066552
3. Marcet-Houben M, Gabaldon T. Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker’s yeast lineage. PLoS Biol. 2015; 13: e1002220. https://doi.org/10.1371/journal.pbio.1002220 PMID: 26252497
4. Campbell MA, Ganley AR, Gabaldon T, Cox MP. The Case of the Missing Ancient Fungal Polyploids. Am Nat. 2016; 188: 602–614. https://doi.org/10.1086/688763 PMID: 27860510
5. Hittinger CT. Saccharomyces diversity and evolution: a budding model genus. Trends Genet. 2013; 29: 309–317. https://doi.org/10.1016/j.tig.2013.01.002 PMID: 23395329
6. Wendland J. Lager yeast comes of age. Eukaryot Cell. 2014; 13: 1256–1265. https://doi.org/10.1128/EC.00134-14 PMID: 25084862
7. James SA, Bond CJ, Stratford M, Roberts IN. Molecular evidence for the existence of natural hybrids in the genus Zygosaccharomyces. FEMS Yeast Res. 2005; 5: 747–755. https://doi.org/10.1016/j.femsyr.2005.02.006 PMID: 15851103
8. Gordon JL, Wolfe KH. Recent allopolyploid origin of Zygosaccharomyces rouxii strain ATCC 42981. Yeast. 2008; 25: 449–456. https://doi.org/10.1002/yea.1598 PMID: 18509846
9. Solieri L, Dakal TC, Croce MA, Giudici P. Unravelling genomic diversity of Zygosaccharomyces rouxii complex with a link to its life cycle. FEMS Yeast Res. 2013; 13: 245–258. https://doi.org/10.1111/1567-1364.12027 PMID: 23279556
Mechanism of whole-genome duplication in a fertile hybrid yeast

10. Pryszcz LP, Nemeth T, Gacser A, Gabaldon T. Genome comparison of Candida orthopsilosis clinical strains reveals the existence of hybrids between two distinct subspecies. Genome Biol Evol. 2014; 6: 1069–1078. https://doi.org/10.1093/gbe/evu082 PMID: 2474362

11. Pryszcz LP, Nemeth T, Saus E, Ksiezeopolska E, Hegeduova E, Nosek J, et al. The genomic aftermath of hybridization in the opportunistic pathogen Candida metapsilosis. PLoS Genet. 2015; 11: e1005626. https://doi.org/10.1371/journal.pgen.1005626 PMID: 26517373

12. Schroder MS, Martinez de San Vicente K, Prandini TH, Hammel S, Higgins DG, Bagagli E, et al. Multiple origins of the pathogenic yeast Candida orthopsilosis by separate hybridizations between two parental species. PLoS Genet. 2016; 12: e1006404. https://doi.org/10.1371/journal.pgen.1006404 PMID: 27806045

13. Leh Louis V, Despons L, Friedrich A, Martin T, Durrens P, Casarégola S, et al. Pichia sorbitophila, an interspecies yeast hybrid, reveals early steps of genome resolution after polyploidization. G3. 2012; 2: 299–311. https://doi.org/10.1534/g3.111.007045 PMID: 22384408

14. Bizzarri M, Giudici P, Cassanelli S, Solieri L. Chimeric Sex-Determining Chromosomal Regions and Dysregulation of Cell-Type Identity in a Sterile Zygosaccharomyces Allodiploid Yeast. PLoS ONE. 2016; 11: e0152558. https://doi.org/10.1371/journal.pone.0152558 PMID: 27065237

15. Hunter N, Chambers SR, Louis EJ, Borts RH. The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. EMBO J. 1996; 15: 1726–1733. PMID: 8612597

16. Delneri D, Colson I, Grammenoudi S, Roberts IN, Louis EJ, Oliver SG. Engineering evolution to study inter-species yeast hybridization, reveals early steps of genome resolution after polyploidization. G3. 2012; 2: 839–850. https://doi.org/10.1534/g3.116.021666 PMID: 16951060

17. Liti G, Barton DB, Louis EJ. Sequence diversity, reproductive isolation and species concepts in Saccharomyces. Genetics. 2006; 174: 839–850. https://doi.org/10.1534/genetics.106.061266 PMID: 16951060

18. Morales L, Dujon B. Evolutionary role of interspecies hybridization and genetic exchanges in yeasts. Microbiol Mol Biol Rev. 2012; 76: 721–739. https://doi.org/10.1128/MMBR.00022-12 PMID: 22304364

19. Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature. 2006; 440: 341–345. https://doi.org/10.1038/nature04562 PMID: 16541074

20. Wolfe KH. Origin of the yeast whole-genome duplication. PLoS Biol. 2015; 13: e1002221. https://doi.org/10.1371/journal.pbio.1002221 PMID: 26252643

21. Greig D, Borts RH, Louis EJ, Travisono M. Epistasis and hybrid sterility in Saccharomyces. Proc Biol Sci. 2002; 269: 1167–1171. https://doi.org/10.1098/rspb.2002.1989 PMID: 12061961

22. Hulin M, Wheals A. Rapid identification of Zygosaccharomyces with genus-specific primers. Int J Food Microbiol. 2014; 173: 9–13. https://doi.org/10.1016/j.ijfoodmicro.2013.12.009 PMID: 24382328

23. Ohnishi H. Osmophilic yeasts. Adv Food Res. 1963; 12: 53–94. PMID: 14280864

24. Mori H, Windisch S. Homothallism in sugar-tolerant Saccharomyces rouxii. J Ferment Technol. 1982; 60: 157–161.

25. James SA, Stratford M. Zygosaccharomyces Barker (1901). In: Kurtzman CP, Fell JW, Boekhout T, editors. The Yeasts, a Taxonomic Study (5th edition). 2. Amsterdam: Elsevier; 2011. p. 937–947.

26. Delneri D, Colson I, Grammenoudi S, Roberts IN, Louis EJ, Oliver SG. Engineering evolution to study inter-species yeast hybridization, reveals early steps of genome resolution after polyploidization. G3. 2012; 2: 839–850. https://doi.org/10.1534/g3.116.021666 PMID: 16951060

27. Liti G, Barton DB, Louis EJ. Sequence diversity, reproductive isolation and species concepts in Saccharomyces. Genetics. 2006; 174: 839–850. https://doi.org/10.1534/genetics.106.061266 PMID: 16951060

28. Pryszcz LP, Nemeth T, Gacser A, Gabaldon T. Genome comparison of Candida orthopsilosis clinical strains reveals the existence of hybrids between two distinct subspecies. Genome Biol Evol. 2014; 6: 1069–1078. https://doi.org/10.1093/gbe/evu082 PMID: 2474362

29. Ohnishi H. Osmophilic yeasts. Adv Food Res. 1963; 12: 53–94. PMID: 14280864

30. James SA, Stratford M. Zygosaccharomyces Barker (1901). In: Kurtzman CP, Fell JW, Boekhout T, editors. The Yeasts, a Taxonomic Study (5th edition). 2. Amsterdam: Elsevier; 2011. p. 937–947.

31. Solieri L, Chand Dakal T, Giudici P, Zygosaccharomyces sapae sp. nov., isolated from Italian traditional balsamic vinegar. Int J Syst Evol Microbiol. 2013; 63: 364–371. https://doi.org/10.1099/ijsem.0.043323-0 PMID: 23024146

32. Solieri L, Chand Dakal T, Giudici P, Zygosaccharomyces sapae sp. nov., isolated from Italian traditional balsamic vinegar. Int J Syst Evol Microbiol. 2013; 63: 364–371. https://doi.org/10.1099/ijsem.0.043323-0 PMID: 23024146

33. Palma M, Roque Fde C, Guerreiro JF, Mira NP, Queiroz L, Sa-Correia I. Search for genes responsible for the remarkably high acetic acid tolerance of a Zygosaccharomyces bailii-derived interspecies hybrid strain. BMC Genomics. 2015; 16: 1070. https://doi.org/10.1186/s12864-015-2278-6 PMID: 26673744
34. Sauer M, Branduardi P, Valli M, Porro D. Production of L-ascorbic acid by metabolically engineered *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii*. Appl Environ Microbiol. 2004; 70: 6086–6091. https://doi.org/10.1128/AEM.70.10.6086-6091.2004 PMID: 15466554

35. Dato L, Branduardi P, Passolunghi S, Cattaneo D, Riboldi L, Frascotti G, et al. Advances in molecular tools for the use of *Zygosaccharomyces bailii* as host for biotechnological productions and construction of the first auxotrophic mutant. Fems Yeast Research. 2010; 10: 894–908. https://doi.org/10.1111/j.1567-1364.2010.00668.x PMID: 21040442

36. Vigentini I, Brambillia L, Branduardi P, Merico A, Porro D, Compagno C. Heterologous protein production in *Zygosaccharomyces bailii*: physiological effects and fermentative strategies. FEMS Yeast Res. 2005; 5: 647–652. https://doi.org/10.1016/j.femsyr.2004.11.006 PMID: 15780664

37. Souciet JL, Dujon B, Gaillardin C, Johnston M, Baret PV, Cliften P, et al. Comparative genomics of protoploid Saccharomycetaeae. Genome Res. 2009; 19: 1696–1709. https://doi.org/10.1101/gr.091546.109 PMID: 19525356

38. Galeote V, Bigey F, Devillers H, Neveuglise C, Dequin S. Genome sequence of the food spoilage yeast *Zygosaccharomyces bailii* CLIB 213T. Genome Announc. 2013; 1; e00606–00613. https://doi.org/10.1128/genomeA.00606-13 PMID: 23969048

39. Mira NP, Münsterkötter M, Dias-Valada F, Santos J, Palma M, Roque FC, et al. The genome sequence of the highly acetic acid-tolerant *Zygosaccharomyces bailii*-derived interspecies hybrid strain ISA1307, isolated from a sparkling wine plant. DNA Res. 2014; 21: 299–313. https://doi.org/10.1093/dnares/dst056 PMID: 24453040

40. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997; 25: 955–964. PMID: 9203104

41. Gordon JL, Byrne KP, Wolfe KH. Additions, losses and rearrangements on the evolutionary route from yeast sex chromosomes by mating-type switching accidents. Proc Natl Acad Sci USA. 2011; 108: 20024–20029. https://doi.org/10.1073/pnas.1112808108 PMID: 22129960

42. Watanabe J, Uehara K, Mogi Y. Diversity of mating-type chromosome structures in the yeast *Zygosaccharomyces rouxii* caused by ectopic exchanges between MAT-like loci. PLoS ONE. 2013; 8: e62121. https://doi.org/10.1371/journal.pone.0062121 PMID: 23814024

43. Johnson AD. Molecular mechanisms of cell-type determination in budding yeast. Curr Opin Genet Dev. 1995; 5: 552–558.

44. Logue ME, Wong S, Wolfe KH, Butler G. A genome sequence survey shows that the pathogenic yeast *Candida parapsilosis* has a defective MTL1 at its mating type locus. Eukaryot Cell. 2005; 4: 1009–1017. https://doi.org/10.1128/EC.4.6.1009-1017.2005 PMID: 15947193

45. Butler G. Fungal sex and pathogenesis. Clin Microbiol Rev. 2010; 23: 140–159. https://doi.org/10.1128/CMR.00053-09 PMID: 2065328
56. Baker CR, Booth LN, Sorrells TR, Johnson AD. Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. Cell. 2012; 151: 80–95. https://doi.org/10.1016/j.cell.2012.08.018 PMID: 23021217

57. Put HMC, De Jong J, Sand FE, Van Grinsven AM. Heat resistance studies on yeast spp. causing spoilage in soft drinks. J Appl Bacteriol. 1976; 40: 135–152. PMID: 1270367

58. Put HMC, De Jong J, Sand FE, Van Grinsven AM. Heat resistance studies on yeast spp. causing spoilage in soft drinks. J Appl Bacteriol. 1976; 40: 135–152. PMID: 1270367

59. Malfeito-Ferreira M, Loureiro V, Wiium H, Aubyn S. A rapid characterization of yeast contaminants associated with sparkling wine production. Industrie dell Bevande. 1990; 19: 504–506.

60. Malfeito-Ferreira M, Loureiro-Dias MC, Loureiro V. Weak acid inhibition of fermentation by Zygosaccharomyces bailii and Saccharomyces cerevisiae. Int J Food Microbiol. 1997; 36: 145–153. PMID: 9217103

61. Malfeito-Ferreira M, Tareco M, Loureiro V. Fatty acid profiling: a feasible typing system to trace yeast contamination in wine bottling plants. Int J Food Microbiol. 1997; 38: 143–155. PMID: 9506280

62. Kudrjawzew WI. Die Systematik der Hefen. Berlin: Akademie Verlag; 1960.

63. Phaff HJ, Miller MW, Mrak EM. The life of yeasts. Harvard University Press; 1966.

64. Barnett JA, Payne RW, Yarrow D. Yeasts: Characteristics and identification. 1 ed: Cambridge University Press; 1983.

65. Barnett JA, Payne RW, Yarrow D. Yeasts: Characteristics and identification. 3 ed: Cambridge University Press; 2000.

66. Kurtzman CP, James SA. Zygosaccharomyces and related genera. In: Blackburn CdW, editor. Food spoilage microorganisms. Boca Raton: CRC Press; 2006. p. 289–305.

67. Strathern JN, Herskowitz I. Asymmetry and directionality in production of new cell types during clonal growth: the switching pattern of homothallic yeast. Cell. 1979; 17: 371–381. PMID: 378408

68. Herskowitz I. Life cycle of the budding yeast Saccharomyces cerevisiae. Microbiol Rev. 1988; 52: 536–553. PMID: 3070323

69. Shinohara M, Oh SD, Hunter N, Shinohara A. Crossover assurance and crossover interference are distinctly regulated by the ZMM proteins during yeast meiosis. Nature Genet. 2008; 40: 299–309. https://doi.org/10.1038/ng.83 PMID: 18297071

70. Vakirlis N, Sarilar V, Drillon G, Fleiss A, Agier N, Meyniel JP, et al. Reconstruction of ancestral chromosome architecture and gene repertoire reveals principles of genome evolution in a model yeast genus. Genome Res. 2016; 26: 918–932. https://doi.org/10.1101/gr.204420.116 PMID: 27247244

71. Wendland J, Walther A. Genome evolution in the Eremothecium clade of the Saccharomyces complex revealed by comparative genomics. G3: Genes, Genomes, Genetics. 2011; 1: 539–558.

72. Juneau K, Nislow C, Davis RW. Alternative splicing of PTC7 in Saccharomyces cerevisiae determines protein localization. Genetics. 2009; 183: 185–194. https://doi.org/10.1534/genetics.109.105155 PMID: 19564484

73. Farabaugh PJ, Kramer E, Vallabhaneni H, Raman A. Evolution of +1 programmed frameshifting signals and frameshift-regulating tRNAs in the order Saccharomycetales. J Mol Evol. 2006; 63: 545–561. https://doi.org/10.1007/s00239-005-0311-0 PMID: 16838213

74. Karin M, Najarian R, Haslinger A, Valenzuela P, Welch J, Fogel S. Primary structure and transcription of an amplified genetic locus: the CUP1 locus of yeast. Proc Natl Acad Sci USA. 1984; 81: 337–341. PMID: 6364141

75. Zhao Y, Strope PK, Kozmin SG, McCusker JH, Dietrich FS, Kokoska RJ, et al. Structures of naturally evolved CUP1 tandem arrays in yeast indicate that these arrays are generated by unequal nonhomologous recombination. G3 (Bethesda). 2014; 4: 2259–2269.

76. Warringer J, Zorgo E, Cubillos FA, Zia A, Gjuvsland A, Simpson JT, et al. Trait variation in yeast is defined by population history. PLoS Genet. 2011; 7: e1002111. https://doi.org/10.1371/journal.pgen.1002111 PMID: 21698134

77. Mehra RK, Garey JR, Butt TR, Gray WR, Wing DR. Candida glabrata metallothioneins. Cloning and sequence of the genes and characterization of proteins. J Biol Chem. 1989; 264: 19747–19753. PMID: 2584191

78. Byrne KP, Wolfe KH. The Yeast Gene Order Browser: combining curated homology and synteny context reveals gene fate in polyploid species. Genome Res. 2005; 15: 1456–1461. https://doi.org/10.1101/gr.3672305 PMID: 16169922

79. Rolland T, Dujon B. Yeasty clocks: dating genomic changes in yeasts. C R Biol. 2011; 334: 620–628. https://doi.org/10.1016/j.crvi.2011.05.010 PMID: 21819943
80. Rodrigues F, Ludovico P, Sousa MJ, Steensma HY, Corte-Real M, Leao C. The spoilage yeast Zygosaccharomyces bailii forms mitotic spores: a screening method for haploidization. Appl Environ Microbiol. 2003; 69: 649–653. https://doi.org/10.1128/AEM.69.1.649-653.2003 PMID: 12514054

81. Butler G, Kenny C, Fagan A, Kurischko C, Gaillardin C, Wolfe KH. Evolution of the MAT locus and its Ho endonuclease in yeast species. Proc Natl Acad Sci USA. 2004; 101: 1632–1637. https://doi.org/10.1073/pnas.0304170101 PMID: 14745027

82. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nature Methods. 2013; 10: 563–569. https://doi.org/10.1038/nmeth.2474 PMID: 23644548

83. Proux-Wéra E, Armisén D, Byrne KP, Wolfe KH. A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. BMC Bioinformatics. 2012; 13: 237. https://doi.org/10.1186/1471-2105-13-237 PMID: 22984983

84. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology. 2011; 29: 644–U130. https://doi.org/10.1038/nbt.1883 PMID: 21572440

85. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Hannick LI, et al. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Research. 2003; 31: 5654–5666. https://doi.org/10.1093/nar/gkg770 PMID: 14500829

86. Huang X, Madan A. CAP3: A DNA sequence assembly program. Genome Res. 1999; 9: 868–877. PMID: 10508846

87. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011; 27: 578–579. https://doi.org/10.1093/bioinformatics/btq683 PMID: 21149342

88. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215: 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2 PMID: 2231712

89. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007; 23: 2947–2948. https://doi.org/10.1093/bioinformatics/btm404 PMID: 17846036

90. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007; 24: 1586–1591. https://doi.org/10.1093/molbev/msm088 PMID: 17483113