Global deceleration of gene evolution following recent genome hybridizations in fungi

Sira Sriswasdi,1 Masako Takashima,2 Ri-ichiroh Manabe,3 Moriya Ohkuma,2 Takashi Sugita,4 and Wataru Iwasaki1,5,6

1Department of Biological Sciences, Graduate School of Science, University of Tokyo, Bunkyo-ku, Tokyo 113-0032, Japan; 2Japan Collection of Microorganisms, RIKEN BioResource Center, Tsukuba, Ibaraki 305-0074, Japan; 3Division of Genomic Technologies, RIKEN Center for Life Science Technologies, Yokohama, Kanagawa 230-0045, Japan; 4Department of Microbiology, Meiji Pharmaceutical University, Kiyose, Tokyo 204-8588, Japan; 5Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba 277-8568, Japan; 6Atmosphere and Ocean Research Institute, University of Tokyo, Kashiwa, Chiba 277-8564, Japan

Polyploidization events such as whole-genome duplication and inter-species hybridization are major evolutionary forces that shape genomes. Although long-term effects of polyploidization have been well-characterized, early molecular evolutionary consequences of polyploidization remain largely unexplored. Here, we report the discovery of two recent and independent genome hybridizations within a single clade of a fungal genus, *Trichosporon*. Comparative genomic analyses revealed that redundant genes are experiencing decelerations, not accelerations, of evolutionary rates. We identified a relationship between gene conversion and decelerated evolution suggesting that gene conversion may improve the genome stability of young hybrids by restricting gene functional divergences. Furthermore, we detected large-scale gene losses from transcriptional and translational machineries that indicate a global compensatory mechanism against increased gene dosages. Overall, our findings illustrate counteracting mechanisms during an early phase of post-genome hybridization and fill a critical gap in existing theories on genome evolution.

[Supplemental material is available for this article.]
Results

Recent and independent genome hybridizations in Trichosporon yeasts

We sequenced the genomes of Trichosporon coremiiforme, T. faecale, T. inkin, and T. ovoides and downloaded that of T. asahii strain CBS2479 (Supplemental Table S1; Yang et al. 2012). The genome sizes of T. coremiiforme and T. ovoides (40–42 Mb) relative to those of closely related T. asahii, T. faecale, and T. inkin (20–25 Mb) suggested that these two species experienced recent genome expansions. Although some basidiomycetous fungi are able to enter dikaryotic states, no dikaryons were detected via fluorescent microscopy (Supplemental Fig. S1). Gene orthology analysis revealed that ~70% of ortholog groups in T. coremiiforme (4472 of 6260 groups) and T. ovoides (3932 of 5657 groups) each consists of two homeologs (paralogs resulting from polyploidization). Most importantly, we were able to reconstruct extensive syntenic alignments between 74% of T. coremiiforme genes and 89% of T. asahii genes and between 79% of T. ovoides genes and 98% of T. inkin genes (Supplemental Fig. S2), which altogether constituted convincing evidence that the larger genomes resulted from polyploidizations.

To determine the type of the two polyploidization events (self-duplication or inter-species hybridization), we reconstructed the two subgenomes in each of T. coremiiforme and T. ovoides and characterized their phylogenetic histories relative to other Trichosporon species. For simplicity, we named the two subgenomes A and B, with subgenome A representing the one that is evolutionarily closer to the reference nonpolyploid genomes (T. asahii for T. coremiiforme and T. inkin for T. ovoides). For T. ovoides, a homeolog copy whose sequence is more identical to that of the ortholog in T. inkin was assigned to subgenome A and the other homeolog copy was assigned to subgenome B. This produced an assignment that is >93% consistent with the fact that T. inkin was assigned to subgenome A and the other homeolog copy was assigned to subgenome B. This produced an assignment that is >93% consistent with the fact that T. inkin was assigned to subgenome A and the other homeolog copy was assigned to subgenome B. This produced an assignment that is >93% consistent with the fact that T. inkin was assigned to subgenome A and the other homeolog copy was assigned to subgenome B.

Losses of genes belonging to transcription and translation machineries

The gene-loss patterns after the two hybridization events are largely independent at the individual gene level. Regardless of whether a homeolog pair is retained as two-copy in T. coremiiforme, its gene-loss rate is consistently ~30% in T. ovoides (Table 1), although these species are closely related and each gene-loss event might convey similar changes in fitness potentials. When mapped onto an S. cerevisiae protein complex data set (Pu et al. 2009), of 192 complexes, 27 and 44 complexes experienced multiple gene losses in T. coremiiforme and T. ovoides, respectively, with 14 complexes in common.

A considerable number of transcription and translation machineries are found among these complexes (Supplemental Tables S2, S3), and they generally experience higher loss rates of their component genes than other protein complexes. In T. coremiiforme, the overall gene-loss rate among protein complexes involved with transcription and translation (43 of 192 complexes) is 33% and is significantly higher than that among other complexes (21%, permutation test P-value < 1 × 10^-5). A modest but similar trend was observed in T. ovoides (28%) for complexes involved with transcription and translation and 26% for other complexes, permutation test P-value = 0.301. Cytoplasmic ribosomal proteins, in particular, lost a copy of 27 of 55 genes in T. coremiiforme and 16 of 59 genes in T. ovoides (Supplemental Table S4). If stoichiometry constraint—a pressure to maintain dosage balance between interacting proteins—is the driving force behind the elevated gene-loss rates in protein complexes involved with transcription and translation, gene-loss events will be concentrated among interaction partners of single-copy genes, where stoichiometry is disrupted. However, permutation tests on the protein–protein interaction networks involving these complexes revealed that this is not the case (Methods). Also, there is no difference in the levels of background sequence conservation (represented by the much higher retention rates of two-copy homeologs in T. coremiiforme and T. ovoides (>70%), compared to <15% in S. cerevisiae).
sequence identity level between nonhybrid orthologs in *T. asahii* and *T. faecale*) between ribosomal proteins that remain two-copy and those that have lost a copy (Mann-Whitney *U* test, *P*-value = 0.7478). Therefore, extensive losses of transcriptional and translational machineries may instead indicate a mechanism to alleviate unfavorable effects of global gene and protein overexpression caused by increased gene dosages.

### Different levels of subgenome divergences and dominances in the two hybrids

The existence of two independent hybridization events among closely related *Trichosporon* fungi gives us a unique opportunity to compare and contrast evolutionary consequences of hybrid genomes in these lineages. As evident from the reconstructed phylogenetic relationship, evolutionary distance between *T. ovoides* subgenomes are 91.2% identical (Fig. 2A). On the other hand, amino acid sequence divergence in *T. coremiiforme* subgenomes is greater than that between those of *T. coremiiforme* (Fig. 1A). At amino acid sequence level, *T. ovoides* subgenomes are 84.7% identical, whereas *T. ovoides* subgenome A and *T. inkin*’s genome are 91.2% identical (Fig. 2A). On the other hand, amino acid sequence identity levels between the *T. coremiiforme*’s subgenomes and *T. asahii*’s genome are largely the same at 92%–93% (Fig. 2A), indicating a degree of resemblance between the genome hybridization in *T. coremiiforme* and a WGD. Such differences in subgenome divergences in *T. coremiiforme* and *T. ovoides* can have a profound influence on the gene-loss patterns in these hybrids, particularly because more diverged subgenomes likely produce more incompatible protein complexes and interactions. Indeed, we identified a strong gene-loss bias in *T. ovoides* (70% of gene-loss events occurred on subgenome B, binomial test *P*-value = 1.9 × 10−44) in contrast to a balanced gene-loss distribution in *T. coremiiforme* (40% of gene-loss events occurred on subgenome B) (Fig. 2B). These results were derived exclusively from gene-loss events that occurred inside doubly conserved synteny structures so as to minimize the impact of genome sequencing, assembly, and subgenome assignment qualities; the higher number of gene-loss events for *T. ovoides* directly resulted from higher coverage of the doubly conserved synteny structures on its genome. Losses of genes belonging to cytoplasmic ribosomal large and small subunits are also more biased in *T. ovoides*, with 10 of 13 or 77% loss on subgenome B (binomial test *P*-value = 0.0461), than in *T. coremiiforme*, with 10 of 16 or 62.5% loss on subgenome A (binomial test *P*-value = 0.2272) (Supplemental Table S4). In terms of evolutionary rates, as represented by the ratio of nonsynonymous to synonymous substitution rates (ds/ds), *T. ovoides* subgenomes exhibit larger and more significant divergences than those of *T. coremiiforme*. The median ds/ds differences between homeologs from each subgenome is 7.4% in *T. ovoides* (paired Wilcoxon signed-rank test, *P*-value = 3.3 × 10−6) and 2.7% in *T. coremiiforme* (paired Wilcoxon signed-rank test, *P*-value = 0.01). These findings illustrate the larger discrepancy of post-hybridization evolutionary pressures on *T. ovoides*’ subgenomes than on those of *T. coremiiforme*.

### Global deceleration of evolutionary rates following genome hybridization

With the genome sequences of three closely related nonhybrid *Trichosporon* species, we quantitatively probed the impact of recent hybridization on gene evolutionary rates. Since genome hybridization is expected to relax gene evolutionary constraint, the evolutionary rates of redundant genes in the two hybrids would be higher than those of their counterparts in nonhybrid species. We defined the background evolutionary rate of each *T. coremiiforme* gene as the weighted average ds/ds of its orthologs in *T. asahii* and *T. faecale* (with the synonymous substitution rates as weights) and the background evolutionary rate of each *T. ovoides* gene as the ds/ds of its ortholog in *T. inkin* (Fig. 3A). We then calculated the fold differences between the ds/ds of genes in the two hybrids and the corresponding background ds/ds and classified these genes as either evolutionarily accelerated, neutral, or decelerated at a twofold threshold. This revealed that decelerated homeolog pairs—those whose evolutionary rates decreased after hybridization—significantly outnumbered accelerated homeolog pairs (binomial test *P*-value = 0.0052 in *T. coremiiforme* and 2.24 × 10−5 in *T. ovoides*) (Fig. 3B). These observations were stable upon changing the fold-change threshold for calling genes as accelerated, neutral, or decelerated. Furthermore, we identified significant overlaps between evolutionarily decelerated homeolog pairs in the two hybrids, which indicate that deceleration of evolutionary rates may be a systematic response in early post-hybridization genomes (Fig. 3C). The absence of any significant enrichment for acceleration or deceleration of evolutionary rates among single-copy genes suggests that the involved mechanisms operate exclusively on redundant homeologs (Supplemental Fig. S3).

In general, decelerated gene evolution can be due to strengthened negative selection. If this is the case, redundant genes that are more evolutionarily decelerated might be enriched for those that have specific functions or high numbers of protein interaction partners. However, we did not observe any significant relationship between the decelerated genes and gene ontology annotations.

---

**Table 1. Gene losses are largely independent between *T. coremiiforme* and *T. ovoides***

|                | Two-copy | *T. ovoides* | Total |
|----------------|----------|--------------|-------|
|                | Two-copy | Single-copy  |       |
| *T. coremiiforme* | 2784a    | 1224 (30.5%) | 4008  |
|                | 872      | 346 (28.4%)  | 1218  |
|                | 3656     | 1570 (30.0%) | 5226  |

*aNumbers shown here are numbers of gene ortholog groups (e.g., 2784 two-copy ortholog groups contain 5568 genes).

*bPercentages represent the fraction of the total number of genes in each row.

---

**Figure 2. Considerable subgenome divergence in *T. ovoides*. (A)** Distribution of amino acid sequence identity levels between homeologous genes from the two subgenomes of *T. coremiiforme* or *T. ovoides* (top) and between orthologous genes from *T. coremiiforme* and *T. asahii* or between those from *T. ovoides* and *T. inkin* (bottom). (B) Distribution of gene losses in *T. coremiiforme* and *T. ovoides* on each of their subgenomes. Only gene losses occurring inside doubly conserved synteny are included.
Convergent changes in evolutionary rates among homeolog pairs

Another curious observation was the lack of homeolog pairs with divergent changes in evolutionary rates—i.e., the pair whose one gene becomes evolutionarily accelerated, whereas the other becomes decelerated (Fig. 3B). We observed unique and consistent enrichments of homeologous genes with convergent directions of change in evolutionary rates (2.6–2.8 fold enrichments, Fisher’s exact test, P-values <4.4 × 10^{-10} in all cases) (Fig. 4A; Methods) that are in contrast with the classical evolutionary scenario in which one homeolog copy remains under evolutionary pressure to maintain its function, whereas the other becomes less constrained. Quantitatively, homeologous genes in each hybrid exhibit significant positive correlations in their post-hybridization evolutionary rate changes (Pearson’s correlation coefficient = 0.4880 in T. coremiiforme and 0.6267 in T. ovoides, permutation test P-values <0.0001) (Fig. 4B). Because these analyses were based on the changes in evolutionary rate following genome hybridization, no correlation should be expected if each homeologous gene evolves independently. Importantly, the lack of such strong positive correlations between nonhybrid T. asahii and T. faecale orthologs (Pearson’s correlation coefficient = 0.1937) clearly indicates that the background evolutionary closeness between T. ovoides and T. coremiiforme’s subgenomes cannot account for these findings.

Finally, we turned our attention to homeolog pair in which both gene copies became evolutionarily accelerated. Levels of amino acid sequence conservation of accelerated homeolog pairs are significantly higher than those of others with a median of 96%

or numbers of protein interaction partners (S. cerevisiae gene annotation and protein–protein interaction data sets were used) (Methods). An alternative explanation could be that gene conversion—a mechanism capable of preserving and homogenizing homeolog sequences and functions (Takuno et al. 2008; McGrath et al. 2014)—removed nonsynonymous mutations from these genes. Because gene conversions create long identical sequence tracts between homeologs, we utilized these signatures to probe for evidence of gene conversions among hybrid homeologs (Methods). In T. ovoides, homeologs that had likely undergone gene conversions exhibited a significant degree of decelerated evolutionary rates compared to those that had not (Mann-Whitney U test, P-value = 3 × 10^{-4}) (Supplemental Figs. S4, S5). Furthermore, we observed a threefold enrichment in the number of T. ovoides homeolog pairs that have lower sequence divergence between themselves than to their orthologous counterparts in T. inkin when only decelerated homeologs were considered (5.5% among all gene pairs versus 17.7% among decelerated homeologs) (Supplemental Fig. S4). Given that T. ovoides’ subgenome A is evolutionarily much closer to T. inkin than to subgenome B (Fig. 1A), strengthened negative selection pressure alone cannot explain such enrichment. In T. coremiiforme, we detected a weaker relationship between gene conversions and decelerated evolutionary rates. This may be due to higher intrinsic levels of sequence identity between T. coremiiforme homeologs masking the evidence of gene conversion.

Figure 3. Deceleration of evolutionary rates following genome hybridization. (A) Diagram illustrating how background d_s/d_0 is defined. Background d_s/d_0 are calculated as the weighted average of d_s/d_0 on the phylogenetic tree branches leading to nonpolyploid species (blue branches, weighted by branch d_0). d_0 ratios are then calculated as the ratio of d_s/d_0 of the branches leading to hybrid species (red branches) over the background d_s/d_0. (B) The number of homeologous gene pairs with varying patterns of acceleration or deceleration in evolutionary rates. The threshold for calling genes as evolutionarily accelerated or decelerated is twofold in d_s/d_0 ratio. Arrows and labels indicate the directions of enrichment and the binomial test P-values (under the null hypothesis that a gene pair is equally likely to become evolutionarily accelerated or decelerated). (C) Overlap of evolutionarily accelerated and decelerated homeologous gene pairs between the two hybridization events. The “accelerated” group consists of accelerated/accelerated and accelerated/neutral pairs. The “decelerated” group consists of decelerated/decelerated and decelerated/neutral pairs. The number of gene pairs in each species and the overlap size are shown. Fisher’s exact test and hypergeometric test P-values are indicated below each comparison.

Figure 4. Convergent changes in evolutionary rates among hybrid homeologs. (A) Bar plots showing the enrichment of homeologous gene pairs with varying patterns of changes in evolutionary rates. The number of expected gene pairs in each class is based on the global frequency of accelerated/neutral/decelerated genes and the assumption that each gene of a homeologous pair independently became accelerated/neutral/decelerated. Only gene pairs with convergent changes (both accelerated or both decelerated) are enriched. Numbers followed by multiplication signs indicate the fold differences in gene pairs between observed and expected. (B) Scatterplots showing the correlation between d_s/d_0 ratios of homeologous genes. Numbers in the parentheses indicate Pearson’s and Spearman’s correlation coefficients in that order. Control data set was created from the d_s/d_0 ratio between T. asahii, T. faecale, and T. coremiiforme (average between subgenome A and B) as indicated. This represents the background level of correlation in evolutionary rate changes across closely related genomes but outside the context of hybridization.
average sequence identity among T. asahii, T. faecale, and T. coremiiforme, and a median of 93% between T. asahii and T. ovoides (Mann-Whitney U test, P-value = 6.3 × 10^{-10} for T. coremiiforme and 2.7 × 10^{-15} for T. ovoides) (Fig. 5A). Also, analyses of evolutionary rates show that accelerated homeologs are not evolving faster than nonaccelerated homeologs in the hybrid genomes, but that the key differences lie in their background evolutionary rates (Fig. 5B). A plausible explanation might be that conserved genes with low background evolutionary rates can tolerate larger increases in evolutionary rates (i.e., become evolutionarily accelerated) while maintaining functional integrity.

**Discussion**

Here, we identified and characterized two recent and independent genome hybridization events in the *Trichosporon* genus of Basidiomycota fungi, in which different levels of subgenome divergence impact the degrees of subgenome dominance (Fig. 2). Frequent gene losses among transcriptional and translational machineries were observed in both hybrids (Supplemental Tables S2, S3). Although stoichiometric constraint could also explain high gene-loss rates among protein complex subunits, the lack of enrichment of gene losses between direct protein–protein interaction partners makes this hypothesis unlikely. Instead, because polyploidization introduces a drastic increase in gene dosage that has to be resolved (Gout and Lynch 2015), these gene losses may occur under a common evolutionary pressure to rapidly neutralize global mRNA and protein overexpression. Because previous studies of WGD reported that slowly evolving genes such as ribosomal proteins tend to be retained in the long term (Davis and Petrov 2004; Brunet et al. 2006; McGrath et al. 2014), the opposite tendencies found here may reflect either differences between the early and late effects of polyploidization or those between WGD and genome hybridization. We assume that the former is more likely because T. coremiiforme is almost indistinguishable from a recent WGD, and the effect of increased genes dosages would be especially severe during the early phase of post-polyploidization. The uniparental gene losses of ribosomal subunit genes (Supplemental Fig. S4) are consistent with a previous study of a young Saccharomycetaceae hybrid (Louis et al. 2012), in which uniparental losses of ribosomal RNA genes were detected. These observations may indicate a mechanism similar to nucleolar dominance, in which a hybrid genome epigenetically silences one parental set of ribosomal RNA genes. Furthermore, in addition to dosage reduction, early losses of specific ribosomal subunit paralogs can alter gene expression regulation and localization (Komili et al. 2007) and profoundly shape subsequent evolutionary landscapes.

In both *Trichosporon* hybrids, redundant homeologs exhibit consistent patterns of decelerated and correlated evolutionary rates (Figs. 3, 4), which indicates that homeologous genes in these hybrids share a significant degree of evolutionary homogeneity. Furthermore, a positive correlation between gene conversion and decelerated evolution (Supplemental Fig. S4) likely suggests a role of gene conversion in maintaining the functional integrity of redundant genes (Katju et al. 2008; Marais et al. 2010; Fawcett and Innan 2011). Although gene conversion often affects essential, highly constrained genes, we did not detect any significant difference in the levels of sequence conservation (represented by the sequence identity level between nonhybrid orthologs in T. asahii and T. faecale) between genes affected by gene conversion and those that were unaffected. This may be because these hybrid genomes are too young for the differences to show or because gene conversion in young hybrid genomes also plays a role in homogenizing incompatible genes. Although accelerated, divergent evolution of redundant homeologous genes following polyploidization may be advantageous for the resulting lineages, the same may not hold true for young hybrid genomes that are still unstable. Homogenization of homeologs would extend their lifetime until the genomes become stabilized (Takuno et al. 2008), consequently delaying subfunctionalization and neofunctionalization of redundant genes. Although some homeolog pairs became evolutionarily accelerated, they are restricted to highly conserved genes, and their evolutionary rates resemble those of other homeologs (Fig. 5). This suggests that these homeologs are actually under a similar degree of selective pressure as the genome averages and are not more evolutionarily relaxed.

In summary, we have illustrated key aspects of an early phase of post-hybridization genome evolution. Although it should be noted that a part of these mechanisms may be specific to the *Trichosporon* genus and there may be biases caused by imperfect genome sequences, assemblies, and gene ontology data sets, consistent findings were yielded for both T. coremiiforme and T. ovoides. Preservation of gene functions by deceleration and homogenization of homeolog evolutionary rates, and a global reduction of gene and protein expression by extensive gene losses from transcriptional and translation machineries, would be essential in stabilizing recently hybridized *Trichosporon* genomes. Accelerated and divergent evolution of homeologs—well-characterized

---

**Figure 5.** Accelerated evolution of highly conserved genes. (A) Distributions of average amino acid sequence identities taken over all pairwise comparison of genes from species surrounding each hybrid species, i.e., T. coremiiforme–T. asahii–T. faecale gene group or T. ovoides–T. asahii gene group. The comparisons are between gene ortholog groups that became evolutionarily accelerated following a hybridization event (gene groups that are present as two-copy in a hybrid species and both gene copies exhibit acceleration in evolutionary rates) and those that did not (other two-copy genes). Genes with accelerated evolutionary rates are enriched for highly conserved genes. (B) Box plots comparing dS/dD values of hybrid two-copy genes to their corresponding background. The whiskers cover ~99.3% of the distribution. Data for the evolutionarily accelerated gene group and the rest of the genome are shown. Mann-Whitney U test P-values between the background groups are indicated: (left) T. coremiiforme with the weighted average of T. asahii and T. faecale genes as background; (right) T. ovoides genes with T. asahii genes as background.
hallmarks of post-polyploidization evolution—likely become dominant after the adverse effects of recent polyploidization have been resolved. Our findings not only establish *Trichosporon* yeasts as a model platform for interrogating early consequences of genome hybridization, but also highlight its multistage nature (Aury et al. 2006; Conant 2014), filling in a critical gap in the current understanding of eukaryotic genome evolution.

**Methods**

**Genome sequencing and assembly**

*T. coremiiforme* JCM 2938, *T. ovoides* JCM 9940, *T. faecale* JCM 2941, and *T. inkin* JCM 9195 strains were provided by the Japan Collection of Microorganisms at RIKEN BioResource Center. Culture and DNA extraction from freeze-dried cell masses were performed as described (Raeder and Broda 1985; Takashima and Nakase 2000). Genomic DNA was purified using Genomic-tip 100 G (Qiagen) according to the manufacturer’s instructions. Paired-end DNA libraries with insert size of 200–300 bp were derived from 1 µg genomic DNA followed by fragmentation with an S2 ultrasonicator (Covaris) and sequencing with Illumina HiSeq 2500 (Illumina). Mate pair libraries with insert size of ~3 kb were prepared with Nextera Mate Pair Sample Prep Kit (Illumina) according to the manufacturer’s protocol with some modifications (Park et al. 2013) and sequenced using HiSeq 2500 to generate 151-bp paired-end reads. For *T. faecale*, additional mate pair libraries with insert sizes of ~5 and 8 kb were prepared and sequenced using MiSeq to produce 301-bp paired-end reads. All mate pair reads were processed with NextClip v0.8 (Leggett et al. 2014) to trim adapter sequences. Estimated sequencing depths range from 114× for *T. coremiiforme* to 419× for *T. faecale*. ALLPATHS-LG versions 52155 (for *T. faecale* and *T. inkin*) and 52488 (for *T. coremiiforme* and *T. ovoides*) (Gnerre et al. 2011) were used to assemble the reads into scaffolds with default parameters. Library generation, sequencing, and assembly were performed at the Genome Network Analysis Support Facility, RIKEN CLST (Yokohama, Japan). Assembly quality was assessed using CEGMA version 2.5 (Parra et al. 2009) and REAPR version 1.0.18 (Supplemental Table S1; Hunt et al. 2013). CEGMA was run with default parameters. REAPR analyses were performed using identical sets of reads that were used by ALLPATHS-LG. The smallmap and pipeline modules were run with default parameters. The perfectFromBam module was run with the following parameters: min insert = 100, max insert = 500, repetitive max qual = 3, perfect min qual = 4, and perfect min alignment score = 150. For *T. asahii*, the published genome for strain CBS 2479 (Yang et al. 2012) was used. The genome sequence data was also used in a separate study that focuses on fungal systematics (Takashima et al. 2015).

**Gene annotation**

Protein-coding genes on each draft genome were predicted using GeneMark-ES version 2 (Ter-Hovhannisyan et al. 2008) with default settings. The hidden Markov model for GeneMark was first trained on *T. asahii* CBS 2479 genome and then applied to other genomes. Genes of all sizes were predicted, but only those that translate to at least 100 amino acids in length were retained for further analyses. Performance of GeneMark was evaluated by two methods. First, we compared our prediction of *T. asahii* CBS 2479 genes to those included with the published genome. Overall, similar numbers of genes were predicted (7996 here versus 8068 published) with 7429 BLASTP matches with E-values <1 × 10−5 and 7000 matches with >80% amino acid sequence identity. Second, we compared the list of genes predicted by GeneMark to those made by AUGUSTUS version 2.5.5 (Stanke et al. 2008). The built-in gene structure model in AUGUSTUS that was trained with data from *Cryptococcus neoformans*, a Basidiomycota fungus in the same order as Trichosporonaceae, was used. Across *T. faecale*, *T. coremiiforme*, *T. inkin*, and *T. ovoides*, both methods generated similar lists of genes with 0.8%–8.8% differences in the numbers of genes predicted, 90%–92% BLASTP matches with >80% sequence identity, and 95%–97% BLASTP matches with E-values <1 × 10−5. Additionally, GeneMark produced slightly higher numbers of reciprocal best hits between *Trichosporon* and *C. neoformans* genes than AUGUSTUS did (54.68% of *C. neoformans* genes for AUGUSTUS and 55.85% for GeneMark).

**Ortholog group assignment**

Ortholog relationship assignments between *Trichosporon* species were first performed pairwise at the amino acid sequence level using inParanoid version 4.1 (Remm et al. 2001). Then, the results were merged into all-species ortholog groups using MultiParanoid (Alexeyenko et al. 2006). Ortholog groups containing one gene from *T. asahii*, at most one gene from *T. inkin* and *T. faecale*, and at most two genes from *T. coremiiforme* and *T. ovoides* were selected for further analyses. At this stage, a total of 7637 ortholog groups were obtained. Also, inParanoid was used to map orthologous relationships between genes from *T. asahii*, *Saccharomyces cerevisiae*, *Candida glabrata*, and *Lachancea waltii*. Protein sequences for *C. neoformans*, *S. cerevisiae*, *C. glabrata*, and *L. waltii* were downloaded from the National Center for Biotechnology Information (NCBI), the Saccharomyces Genome Database, the Candida Genome Database, and the Yeast Gene Order Browser database (http://ygob.ucd.ie/), respectively.

**Synteny structure reconstruction**

Synteny relationships between *Trichosporon* species were reconstructed by the aid of Gene Order Browser (GOB; source code provided by the author) (Byrne and Wolfe 2005). The list of predicted genes’ ortholog group assignments were used as inputs for GOB. Gene order in *T. asahii* genome was used as reference for aligning *T. faecale* and *T. coremiiforme* genes, and gene order in *T. inkin* genome was used as reference for aligning *T. ovoides* genes. A syntenic segment was defined as a contiguous region of conserved gene orders on *T. coremiiforme* or *T. ovoides* genome with no larger than 20-kb gaps in between consecutive genes. An overlap between two syntenic segments was defined as doubly conserved syntenic segment (Kellis et al. 2004). Extensive synteny structures were identified for both *T. asahii*–*T. coremiiforme* and *T. inkin*–*T. ovoides* pairs, each involving >88% coverage of nonpolyploid and >73% of post-polyploid genomes (Supplemental Fig. S2). Because the 20-kb gap threshold corresponds roughly to the length of six to seven consecutive genes (based on *Trichosporon* genome statistics) and may cause underdetection of synteny breaks, we also reconstructed synteny structure using a tighter 10-kb gap threshold. As evident in Supplemental Figure S2, the genome coverages of synteny structure decreased by ~1% in hybrid genomes and ~5% in nonhybrid genomes. Hence, the impact of synteny break misidentification is minimal at the 20-kb gap threshold.

**Subgenome assignment**

Two subgenome reconstruction methods were considered. The first method is a synteny-based method that assumes that each assembled scaffold belongs to a subgenome. Doubly conserved synteny structures are used to define pairs of homeologous scaffolds, and subsequent phylogenetic analyses are used to assign each of
the paired scaffolds into subgenomes. The second method is a reference genome–based method that relies on nonhybrid genomes as reference points (T. asahii for T. coremiiforme and T. inkin for T. ovoides). Each gene of a homeolog pair is assigned to one of the two subgenomes based on its sequence similarity to the nonhybrid ortholog.

For T. coremiiforme, the reference genome–based method produced an interspersed pattern of subgenome assignments, and so we investigated the level of sequence similarity between T. coremiiforme and T. asahii. Whole-genome alignments were performed using LAST version 658 (Kielbasa et al. 2011) with T. asahii genomes as database and T. coremiiforme as query. Rareness limit for initial matches (-m parameter, default = 10) was set at 100 to increase sensitivity. E-score threshold of 1 × 10^−10 was applied during post-processing of LAST’s output. Only alignments that map one region from T. asahii genome to exactly two regions of lengths at least 1000 bases on different scaffolds in T. coremiiforme were considered. This revealed that two subgenomes of T. coremiiforme are almost equidistant from T. asahii with a median difference in sequence identities of only 0.97%—a likely reason behind the interspersed subgenome assignment pattern made by the reference genome–based method.

Hence, for T. coremiiforme, only the synteny-based method could be used. To ensure high confidence, only doubly conserved synteny containing at least 10 pairs of genes were considered in the reconstruction process. This resulted in 33 homeologous scaffold pairs involving 34 distinct scaffolds being identified. Next, overlapping scaffold pairs were merged—for example, if scaffold A is homeologous to scaffolds B, C, and D, and scaffold B is homeologous to scaffolds E and F, then we can infer that scaffolds A, E, and F belong to one subgenome, whereas scaffolds B, C, and D belong to the other. However, this merging process did not connect all 33 scaffold pairs into a unique subgenome assignment. Although we could infer that (1) scaffolds 0, 6, 7, 8, 12, 13, and 17 of T. coremiiforme are homeologous to scaffolds 2, 9, 14, 18, 19, 20, 21, 23, 24, 28, 33, and 36, and (2) scaffolds 1, 22, 26, 27, and 29 of T. coremiiforme are homeologous to scaffolds 3, 10, and 15, we could not tell which of the first two subgenomes corresponds to which of the latter two subgenomes. Subsequent phylogenetic analyses were required to ultimately determine that scaffolds 0, 6, 7, 8, 12, 13, and 17 are on the same subgenome as scaffolds 3, 10, and 15.

For T. ovoides, the median difference in sequence identities among each of its two subgenomes to T. inkin is 7%, and both reconstruction methods were applicable. The synteny-based method produced a subgenome assignment covering 28 distinct scaffolds, and the reference genome–based method produced an assignment covering 46 distinct scaffolds, including all 28 covered by the former. In 25 of 28 scaffolds covered by both methods, the two assignments agreed on 93% of the genes (5466 of 5875 genes). In the other three scaffolds, strong evidence for chromosomal recombinations was observed. In scaffold S, the reference genome–based method assigned 162 genes to one subgenome and 423 genes to the other (a 27% to 73% ratio). However, when gene order was taken into account, we found that 93% of the first 143 genes on this scaffold were assigned to one subgenome and 93% of the other 442 genes were assigned to the other subgenome. Similar results were found for scaffold 6, in which 95% of its first 296 genes were assigned to one subgenome and 94% of the remaining 237 genes were assigned to the other; in scaffold 13, 98% of its first 120 genes were assigned to one subgenome, and 95% of the remaining 111 genes were assigned to the other. Because the reference genome–based method produced a subgenome assignment with higher scaffold coverage, a good agreement with synteny-based reconstruction, and capability to capture chromosomal recombination, we selected T. ovoides subgenome assignments produced by this method for subsequent analyses.

**Gene-loss analysis**

To probe the distribution of gene losses between subgenomes of T. coremiiforme or T. ovoides with high confidence, only gene losses that occurred within doubly conserved syntenic segments were counted. This consideration minimized the impact of genome misassembly and collapse of repetitive regions as those errors tend to disrupt doubly conserved syntenic structure. Based on subgenome assignments, the numbers of gene losses on syntenic segments corresponding to the same subgenome were then added together. This covered 607 gene losses in T. coremiiforme and 1214 in T. ovoides (Fig. 2B). The difference in gene counts is due to higher coverage of syntenic structure and subgenome reconstruction in T. ovoides. For each single-copy gene that does not belong to a doubly conserved syntenic segment, we assign its subgenome using the majority subgenome assignment of neighboring genes. In T. coremiiforme, because each scaffold was assigned to a subgenome as a whole, all single-copy genes on a scaffold were assigned to the same subgenome. In T. ovoides, we assigned a single-copy gene to the majority subgenome assignment of up to 20 homeologous genes surrounding it (up to 10 upstream genes and 10 downstream genes). If fewer than 10 genes are present on either side, only those genes would be included. In case of a tie or when there are no nearby homeologous genes (e.g., in region with heavy gene losses), that single-copy gene was labeled as of unknown origin.

**Phylogenetic reconstruction**

Initial phylogenetic analyses to determine the tree topology were performed in MEGA version 6 build# 6140226 (Tamura et al. 2013) using both Maximum Likelihood (ML) and Maximum Parsimony (MP) methods and PhyML version 3.1 (Guindon and Gascuel 2003; Guindon 2010). Amino acid sequences of 2623 gene ortholog groups that are present in all five Trichosporon species and remain as two-copy in both T. coremiiforme and T. ovoides were aligned using MUSCLE (Edgar 2004), and the results were mapped to their nucleotide sequences. Aligned sequences belonging to the same species or subgenomes (in the case of T. coremiiforme and T. ovoides) were concatenated together. Each isolated subgenome pair in T. coremiiforme (see “Subgenome assignment”) was analyzed separately at this stage. The Jukes-Cantor substitution model and the Subtree-Pruning-Regrafting (SPR) method were used. Bootstrap count was set to 100 in MEGA (as the software crashed at 1000 bootstrap value) and 1000 in PhyML. Designating the subgenomes of T. coremiiforme and T. ovoides, which are evolutionarily closer to T. asahii and T. inkin, respectively, as the subgenome A and the other as B, both MEGA and PhyML produced the following tree topology with 100% bootstrap support: ((T. ovoides B, (T. ovoides A, T. inkin)), (T. faecale (T. coremi., B, (T. coremi., A, T. asahii))). This tree topology was used in all subsequent analyses. Neither changing the nucleotide substitution model (in both MEGA and PhyML) nor switching between codon and nucleotide model (in MEGA) affected the topology and bootstrap support of the resulting phylogenetic trees.

**Synonymous and nonsynonymous substitution rate calculations**

PAML version 4.8 (Yang 2007) was used to calculate phylogenetic trees along with branch-specific substitution rates. First, we tested the molecular clock hypothesis of equal evolutionary rates for the Trichosporon tree using the likelihood ratio tests on the ML trees estimated with and without molecular clock assumption. The test statistic is 114.50 with a degree of freedom of 7–2 = 5, and so the
molecular clock assumption was rejected (P-value <1 × 10⁻⁷). In all cases hereafter, PAML was used without the clock assumption (clock = 0). The corresponding section of the Trichosporon phylogenetic tree topology characterized above, (((T. ovoides B, (T. ovoides A, T. asahii)), (T. faecele (T. coremii. B, (T. coremii. A, T. asahii))))), was used as input for each ortholog group. For example, for an ortholog group containing two homeologous genes in T. coremiforme and one gene each in T. asahii, T. faecale, and T. asahii, the tree topology (T. asahii, (T. faecele (T. coremii. B, (T. coremii. A, T. asahii)))) was specified. We restricted our analyses to 4951 high-confidence ortholog groups which are present in all three nonhybrid Trichosporon species. Among these, 3768 ortholog groups are present as two-copy in T. coremiforme subgenomes as $d_s = 0.6753$, between T. ovoides subgenomes as $d_s = 1.3763$, between S. cerevisiae and G. glabrata as $d_s = 8.2356$, and between S. cerevisiae-G. glabrata and L. waltii as $d_s = 68.0845$. Hence, the timing of hybridization events in Trichosporon are about at least six to 12 times younger than the whole-genome duplication event in Saccharomyces, and could be up to 50 to 100 times younger.

Evolutionary rate analyses $d_s/d_k$ values were used to represent the evolutionary rates of genes. In order to probe the changes in $d_s/d_k$, following each genome hybridization event, $d_s/d_k$ of the phylogenetic tree branches leading to nonpolyploid species were used to define the “background evolutionary rates.” For T. coremiforme, the background evolutionary rates were calculated as the weighted-average of $d_s/d_k$ of the branches leading to T. faecale and T. asahii (with $d_s$ as weights, this is equivalent to the ratio of average $d_s$ over average $d_k$ of the two branches). For T. ovoides, the $d_s/d_k$ of the branch leading to T. asahii was directly used as its background. Changes in evolutionary rates were then defined as the log (base 2) ratio of $d_s/d_k$ of the phylogenetic tree branches leading to T. coremiforme or T. ovoides over their corresponding background $d_s/d_k$. A twofold threshold was used to classify genes as evolutionarily “decelerated,” “accelerated,” or “neutral.”

To evaluate the degree of correlation of evolutionary rates between the two subgenomes of each Trichosporon hybrid, two analyses were performed. First, based on the frequency of “decelerated,” “accelerated,” and “neutral” genes in each species, an expected distribution of homeolog pairs with each of the six possible combinations—decelerated/decelerated, decelerated/accelerated, decelerated/neural, accelerated/accelerated, accelerated/neural, and neutral/neutral—was computed. Enrichment of each group was determined as the fold differences between observed and expected homeolog pair counts. Fisher’s exact tests for the enrichments of decelerated/decelerated, or accelerated/accelerated, homeolog pairs were performed on 2 × 2 tables containing the frequencies of decelerated and nondecelerated genes, or accelerated/nonaccelerated genes, on subgenome A and B, respectively. Second, the correlation coefficients of changes in $d_s/d_k$ between homeologous genes were compared to coefficients derived from sets of 10,000 randomized gene pairing (permutation test), and the coefficient was calculated using the $d_s/d_k$ ratios of T. faecale and T. asahii over T. coremiforme (Fig. 4B). The latter data set represents a baseline correlation of changes in evolutionary rates across closely related genomes outside the context of genome hybridization.

Gene conversion detection We used GENECONV v1.81a (Sawyer 1989) to detect gene conversion between two-copy genes. Sequence alignment of each ortholog group consisting of a reference gene in nonpolyploid species (T. asahii or T. inkin) and two genes in post-hybrid species (T. coremiforme or T. ovoides) was performed using MUSCLE and used as input for GENECONV. Permutation test parameters for GENECONV were set at 10,000 permutations and a maximum
An automated text extraction is not available for this document.
