Functional Divergence in a Multi-gene Family Is a Key Evolutionary Innovation for Anaerobic Growth in *Saccharomyces cerevisiae*

David J. Krause and Chris Todd Hittinger *

Laboratory of Genetics, Wisconsin Energy Institute, DOE Great Lakes Bioenergy Research Center, Center for Genomic Science Innovation, J. F. Crow Institute for the Study of Evolution, University of Wisconsin-Madison, Madison, WI

***Correponding author:** Email: cthittinger@wisc.edu.

Associate editor: Dr. Jianzhi Zhang

Abstract

The amplification and diversification of genes into large multi-gene families often mark key evolutionary innovations, but this process often creates genetic redundancy that hinders functional investigations. When the model budding yeast *Saccharomyces cerevisiae* transitions to anaerobic growth conditions, the cell massively induces the expression of seven serine/threonine-rich anaerobically-induced cell wall mannoproteins (anCWMPs): *TIP1, TIR1, TIR2, TIR3, TIR4, DAN1,* and *DAN4*. Here, we show that these genes likely derive evolutionarily from a single ancestral anCWMP locus, which was duplicated and translocated to new genomic contexts several times both prior to and following the budding yeast whole genome duplication (WGD) event. Based on synteny and their phylogeny, we separate the anCWMPs into four gene subfamilies. To resolve prior inconclusive genetic investigations of these genes, we constructed a set of combinatorial deletion mutants to determine their contributions toward anaerobic growth in *S. cerevisiae*. We found that two genes, *TIR1* and *TIR3*, were together necessary and sufficient for the anCWMP contribution to anaerobic growth. Overexpressing either gene alone was insufficient for anaerobic growth, implying that they encode non-overlapping functional roles in the cell during anaerobic growth. We infer from the phylogeny of the anCWMP genes that these two important genes derive from an ancient duplication that predates the WGD event, whereas the *TIR1* subfamily experienced gene family amplification after the WGD event. Taken together, the genetic and molecular evidence suggests that one key anCWMP gene duplication event, several auxiliary gene duplication events, and functional divergence underpin the evolution of anaerobic growth in budding yeasts.

Key words: **gene duplication, anaerobic growth, birth-and-death, multi-gene family, fermentation.**

Introduction

Gene duplication is a common evolutionary process. The pioneering work of Susumo Ohno formalized several distinct mechanisms for gene duplication, as well as theorized many potential evolutionary advantages for the retention of duplicated genes (Ohno, 1970). Gene duplication events initially create redundancy, which can resolve itself in several ways, and for which myriad models have been proposed to explain the mechanisms of duplication and the fates of the gene duplicates (Innan and Kondrashov 2010). A deeper understanding of the prevalence and conditions that favor these mechanisms requires rigorous testing of their respective hypotheses in model organisms. The birth-and-death model is a well-studied example in which multi-gene families undergo amplifications, followed by functional divergence or pseudogenization and gene loss (Nei et al. 1997). For example, this model has been applied to animal olfactory receptors (ORs), which have frequently undergone duplications followed by pseudogenization or functional differentiation (Niimura and Nei 2003; Hughes et al. 2018). Identifying functional differentiation in OR genes is complicated by difficulties in assessing specific OR gene function and the effect on sensory perception, although mapping genetic variation to sensory perception differences can provide some information of functional divergence (Trimmer et al. 2019). This model has also been applied to genes in vertebrate immune systems, including major histocompatibility complex (MHC) and immunoglobulin genes, which have undergone many duplication and pseudogenization events (Nei and Rooney 2005).

The model budding yeast *Saccharomyces cerevisiae* and its close relatives are a powerful model system for studying the fates of gene duplicates due to their genetic tractability and a whole genome duplication (WGD) event that occurred in its ancestors circa 100 Mya. This WGD was a result of allopolyploidization between ancestors of the *Kluyveromyces/Lachancea/Eremothecium* (KLE) and *Zygosaccharomyces/Torulaspora* (ZT) yeast lineages (Wolfe and Shields 1997; Marcet-Houben and Gabaldón 2015). Most gene duplicates were ultimately lost, but among those that have been retained, several cases of sub-functionalization and neofunctionalization have been described (Cliften et al. 2006; Hickman and

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

Rusche 2007; Hittinger and Carroll 2007; Dean et al. 2008; Kuang et al. 2016). Gene family expansions beyond single duplications are common for gene families found in the subtelomeres of yeast chromosomes, such as the *MAL*tose utilization genes, seri*PAU*perin genes, and *FLO*cculation genes (Luo and van Vuuren 2009; Van Mulders et al. 2009; Brown et al. 2010). Studies of the aryl-alcohol dehydrogenase family and hexose transporter family in yeast have shown the power of *S. cerevisiae* as a system for studying multi-gene families, especially those with a high degree of redundancy, by successive individual gene deletions (Delneri et al. 1999; Wieczorke et al. 1999).

The WGD event also approximately coincides with the rise of anaerobic growth in this lineage of budding yeasts. While this trait is likely underpinned by several adaptations (Hagman et al. 2013; Thompson et al. 2013), one important adaptation is sterol transport (Snoek and Steensma 2006, 2007). Sterol biosynthesis requires molecular oxygen, so sterols are not produced during anaerobic growth and are instead transported into the cell via the anaerobically induced sterol transporters Pdr11 and Aus1 (Wilcox et al. 2002; Papay et al. 2020). Several studies have suggested that anaerobically induced cell wall mannoproteins (anCWMPs) also play a role in this process (Alimardani et al. 2004; Inukai et al. 2015). Cell wall mannoproteins constitute nearly half of the fungal cell wall, along with β-glucans and chitin (Lipke and Ovalle 1998). *S. cerevisiae* contains two major types of cell wall mannoproteins that are induced during anaerobic growth: seven *DAN*/*TIR*/*TIP* genes, hereafter referred to as their encoded anCWMPs, and 24 seri*PAU*perin genes, which share some sequence similarity with the anCWMPs, but lack their serine/threonine-rich region. The inference of the roles of the anCMWPs in anaerobic growth has been complicated by conflicting and inconclusive results. One study found that *tir3Δ*, *tir1Δ*, and *tir4Δ* single-mutant strains failed to grow anaerobically (Abramova et al. 2001), while another study found that a *tir1Δ* strain and a triple-mutant *tir1Δ tir2Δ tip1Δ* strain grew anaerobically without defect (Donzeau et al. 1996). Further complicating matters, three separate deletion library screens failed to identify any anCWMP deletion as having a detectable effect on anaerobic growth (Reiner et al. 2006; Snoek and Steensma 2006; Galardini et al. 2019). In *Candida glabrata*, the *TIR3* homolog is required for anaerobic growth, but a *C. glabrata tir3Δ* mutant strain was not complemented by the *S. cerevisiae* homolog (Inukai et al. 2015). These conflicting results preclude the assignment of a definitive function for the anCWMP genes. Previous phylogenomic analyses found that the *DAN*/*TIR*/*TIP* family encoding anCWMPs underwent several gene duplication and translocation events after the WGD event in the lineage leading to *S. cerevisiae*, but the limited number of genomes available at the time hindered synteny analyses (Gordon et al. 2009). Thus, the family of genes encoding anCWMPs is ripe for both functional and evolutionary investigation.

Here, we test a set of combinatorial anCWMP gene deletion mutants for growth in anaerobic conditions and find that only two genes are required for anaerobic growth, *TIR1* and *TIR3*. Together, these two genes are also sufficient for the anCWMP contribution to anaerobic growth. However, neither gene alone supports anaerobic growth, even when overexpressed, implying they encode distinct functional roles. We construct a phylogenetic tree of the all the anCWMP homologs in all published budding yeast genome sequences and show that this gene family is conserved across many taxa but underwent several gene duplication and translocation events around the time of the WGD. We infer that the *TIR1* and *TIR3* genes diverged prior to the WGD event in the family Saccharomycetaceae, which includes *S. cerevisiae*, and we propose hypotheses for the mechanism and timing of functional divergence and its relationship to subsequent gene duplication events and the evolution of anaerobic growth.

Results and Discussion

Contributions of the Multi-gene anCWMP Family to Anaerobic Growth

The anCWMP genes encode proteins that localize to the cell wall and have an extensively glycosylated serine/threonine-rich domain. Despite genetic studies that have suggested some anCWMP genes may be involved in facilitating sterol transport (Alimardani et al. 2004; Inukai et al. 2015), a key feature of anaerobic growth, the genes still lack a clear biochemical or genetic function. We sought to definitively resolve conflicting prior genetic studies by constructing combinatorial deletion mutants in *S. cerevisiae*. We focused here on the seven anCWMP genes of the *DAN*/*TIR*/*TIP* gene family and did not investigate the *PAU* genes, which share some sequence similarity with the anCWMP genes but lack a serine/threonine-rich region. We expect the *PAU* genes to be especially difficult to study due to the large number in the *S. cerevisiae* genome (24). Since *DAN1* and *DAN4*, as well as *TIR2* and *TIR4*, are adjacent in the genome, we deleted these two pairs together, resulting in five separate anCWMP loci deletions and 32 total combinations. We measured the growth of each resulting strain under anaerobic conditions (Fig. 1; supplementary fig. S1, Supplementary Material online).

Consistent with some previous studies, we found that *TIR3* was required for anaerobic growth (Abramova et al. 2001). No other anCWMP gene was essential for anaerobic growth. By limiting further analysis to strains containing *TIR3*, we found that all *TIR3*-containing backgrounds that failed to grow anaerobically were missing *TIR1*. The only strains lacking *TIR1* that could grow anaerobically contained *TIR3*, *TIR2*/*TIR4*, and either *TIP1* or *DAN1*/*DAN4*. Further, the presence of *TIR1* and *TIR3* was sufficient to yield anaerobic growth in the absence of the other five anCWMP genes. We infer that *TIR1* contributes to anaerobic growth, but this contribution may be partially redundant with other anCWMP genes. The unknown anCWMP requirements for anaerobic growth have hampered attempts to fully engineer sterol transport into yeasts that naturally lack sterol uptake or into *S. cerevisiae*

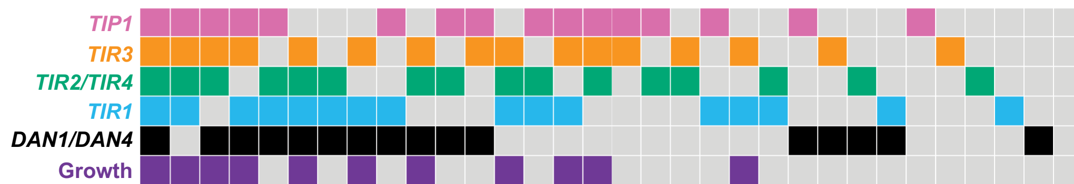

FIG. 1. Growth of combinatorial mutant strains under anaerobic conditions. All 32 combinatorial mutant strains are represented by a single column, with presence of the anCWMP gene(s) depicted as a colored box and absence depicted by a gray box. The bottom row is colored if the strain grew anaerobically to a final OD of greater than 0.5. Representative growth curves are shown in supplementary fig. S1, Supplementary Material online.

under aerobic conditions (Alimardani et al. 2004). We hypothesize that *TIR1* and *TIR3* encode the minimal anCWMP functions necessary for anaerobic growth and sterol uptake.

Why then does *S. cerevisiae* contain seven different anCWMP genes if two are sufficient for anaerobic growth? The positive dosage model of gene duplication describes a situation in which gene duplicates are retained because the resulting increased expression of the gene pair provides a fitness benefit (Kondrashov et al. 2002). We reasoned that the presence of seven anCWMP genes in the *S. cerevisiae* genome might be the result of benefits from increased gene expression of this set of genes with potentially overlapping functions. We tested this hypothesis by expressing each anCWMP gene individually on a high-copy number plasmid in the seven-gene deletion mutant. No single anCWMP gene conferred anaerobic growth to this seven-gene deletion mutant (Fig. 2A). Given that no single locus conferred anaerobic growth even when its dosage was artificially increased via a high-copy plasmid, we reject the positive dosage model as a sufficient explanation for the maintenance of the seven anCWMP genes.

Since *TIR3* is required but not sufficient for anaerobic growth, either at its native locus or when overexpressed, we performed the same overexpression experiment in a strain containing only *TIR3* (i.e., the other six genes were deleted). As expected, overexpression of *TIR1* conferred anaerobic growth in the *TIR3*-containing background. We also found that overexpression of *TIR2* conferred growth, even though a *TIR3*-containing strain containing the *TIR2/TIR4* native locus did not grow (Fig. 2B; supplementary fig. S2, Supplementary Material online). This result implies that, when overexpressed, *TIR2* may complement the function of *TIR1*, but its native expression level is not high enough to normally do so. To test this hypothesis, we analyzed published RNA-sequencing data from *S. cerevisiae* under anaerobic conditions and found that *TIR2* gene expression was 50-fold lower than *TIR1* (Myers et al. 2019). This result is consistent with the combinatorial mutant experiment, in which *tir1Δ* mutant backgrounds that grew always contained *TIR3*, *TIR2/TIR4*, and another anCWMP gene. We also overexpressed some anCWMP genes in a strain containing only *TIR1* (i.e., the other six genes were deleted). In this case, overexpression of *TIR3* conferred anaerobic growth, while *TIP1* and *TIR2* did not (Fig. 2B; supplementary fig. S2, Supplementary Material online). We conclude that *TIR3* and *TIR1* encode distinct functions that are both required for anaerobic growth, but the *TIR1* function is partially redundant with other anCWMP genes.

The anCWMP Genes Comprise at least Four Subfamilies

To investigate the evolution of this multi-gene family, we searched 332 publicly available budding yeast genomes (Shen et al. 2018) for homologs of the *S. cerevisiae* anCWMP genes, and we found homologs in most species. We also found homologs of the closely related gene *AFB1* in approximately 40% of species, and this gene family clustered outside the anCWMP genes with 100% bootstrap support (Fig. 3A). Due to the low complexity of the serine/threonine-rich region present in the middle of these gene sequences, only the short, structured N-terminal portion and the short glycosylphosphatidylinositol (GPI)-anchoring C-terminal portion were used for alignment and phylogenetic analyses (Fig. 3B). The seven members of the anCWMP gene family in *S. cerevisiae*, as well as all the anCWMP genes from the family Saccharomycetaceae, formed a single clade within the larger gene tree (Fig. 3A). This result implies that a gene family expansion occurred within the Saccharomycetaceae, so we will focus here on that clade. Even within the family Saccharomycetaceae, most non-WGD species do not contain any anCWMP genes, but those that do contain them harbor between one and four genes in a single genomic neighborhood. This gene neighborhood is generally most similar to the *TIR1* gene neighborhood of *S. cerevisiae*. In the post-WGD lineage, all species except *Vanderwaltozyma polyspora* contain anCWMP genes, which number as few as three in *Candida (Nakaseomyces) castellii* and as many as 17 in *Kazachstania unispora*.

The phylogeny of the anCWMP genes revealed several distinct clades, which we designate here as four subfamilies of the anCWMP gene family based on their related *S. cerevsiae* homologs: 1) *TIR3*; 2) *TIR1*, which includes *TIR1*, *TIR2*, and *TIR4*; 3) *TIP1*; and 4) *DAN1*, which includes *DAN1* and *DAN4* (Fig. 3, Fig. 4). The subfamilies either have strong bootstrap support, shared synteny patterns, or both. We discuss each subfamily in more detail below, as well as those homologs that do not neatly fit into the subfamilies that we currently recognize.

Genes related to *TIR3* formed a clade with 68% bootstrap support, and this subfamily was the most widely

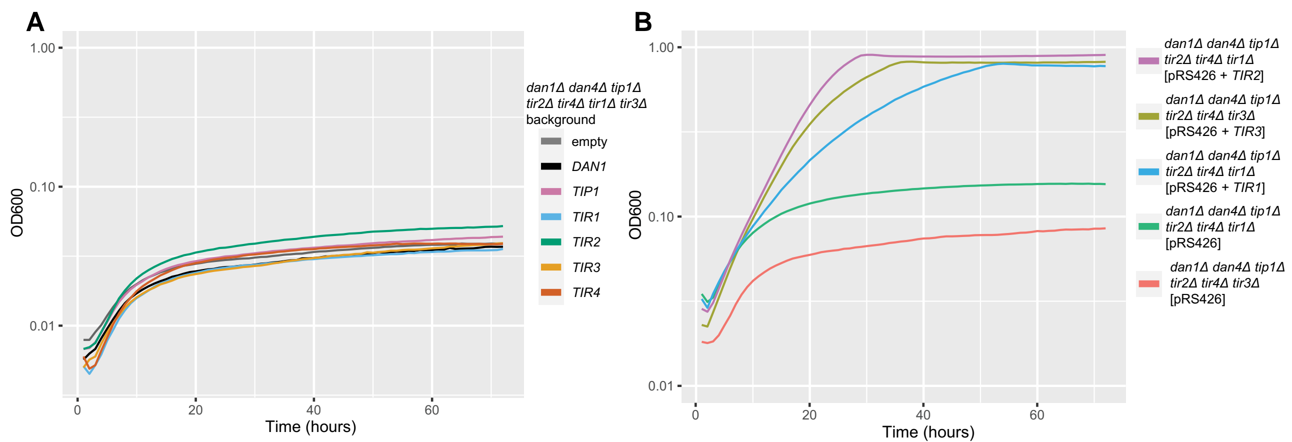

Fig. 2. Growth of strains carrying anCWMP genes on a 2μ high-copy vector with log-scaled Y-axes. (*A*) Representative growth curves for the strain lacking all genomic anCWMP genes but carrying one anCWMP gene on the high-copy vector. (*B*) Representative growth curves for backgrounds lacking all genomic anCWMP genes except either *TIR3* or *TIR1*, but carrying one anCWMP gene on the high-copy vector. All three strains that grew are shown here, but all tested strains are shown in supplementary fig. S2, Supplementary Material online.

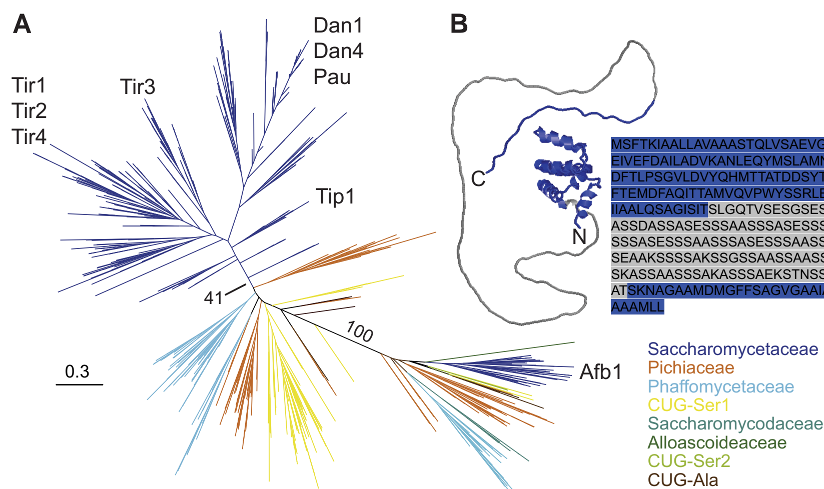

Fig. 3. (*A*) Maximum likelihood phylogeny of the anCWMP genes and *AFB1* genes from the budding yeast subphylum Saccharomycotina using amino acids. Yeast major clades are colored according to Shen et al. 2018. Key bootstrap values are shown at the origin of Saccharomycetaceae anCWMP genes and the clustering of anCWMP genes from *AFB1*. (*B*) Predicted protein structure from alpha-fold along with the amino acid sequence of *TIR3* from *S. cerevisiae* (Jumper et al. 2021; Varadi et al. 2022). Colored regions depict those used for alignment and phylogenetic analyses: the N-terminal structured region and the C-terminal GPI-anchor signal sequence, which is ultimately cleaved. The grey region depicts the serine/threonine-rich region, which was removed from protein sequences to facilitate alignments.

conserved; indeed, all post-WGD species in the dataset and several non-WGD species contain a putative *TIR3* ortholog. The members of this subfamily were generally syntenic with *S. cerevisiae TIR3*, except for the homologs in *Tetrapisispora/Yueomyces* and non-WGD species, which were found in the putative ancestral locus of all anCWMP genes. This result implies that the structure of the *TIR3* locus in *S. cerevisiae* and its relatives resulted from a translocation event after divergence from the *Tetrapisispora/Yueomyces* lineage. The presence of a *TIR3* ortholog in both ZT clade members and KLE members implies that this gene was present in both parents of taxa descended from the WGD event. Because most post-WGD species contain only one copy of *TIR3* with a few exceptions resulting from recent duplications, one copy of the *TIR3* gene was likely lost sometime after the WGD event.

Genes related to *TIR1*, *TIR2*, and *TIR4* formed a clade with 51% bootstrap support, and all post-WGD species, except *Kazachstania transvaalensis* and *Tetrapisispora/Yueomyces*, contain at least one homolog. These genes generally shared synteny with the *TIR1*, *TIR2*, or *TIR4* genes of *S. cerevisiae*, but there was no clear phylogenetic distinction between genes sharing synteny with *TIR1* and those sharing synteny with the *TIR2/TIR4* locus. Instead, gene members of this subfamily were generally more closely related to other members within a particular species or small clade of species than they were to homologs in other

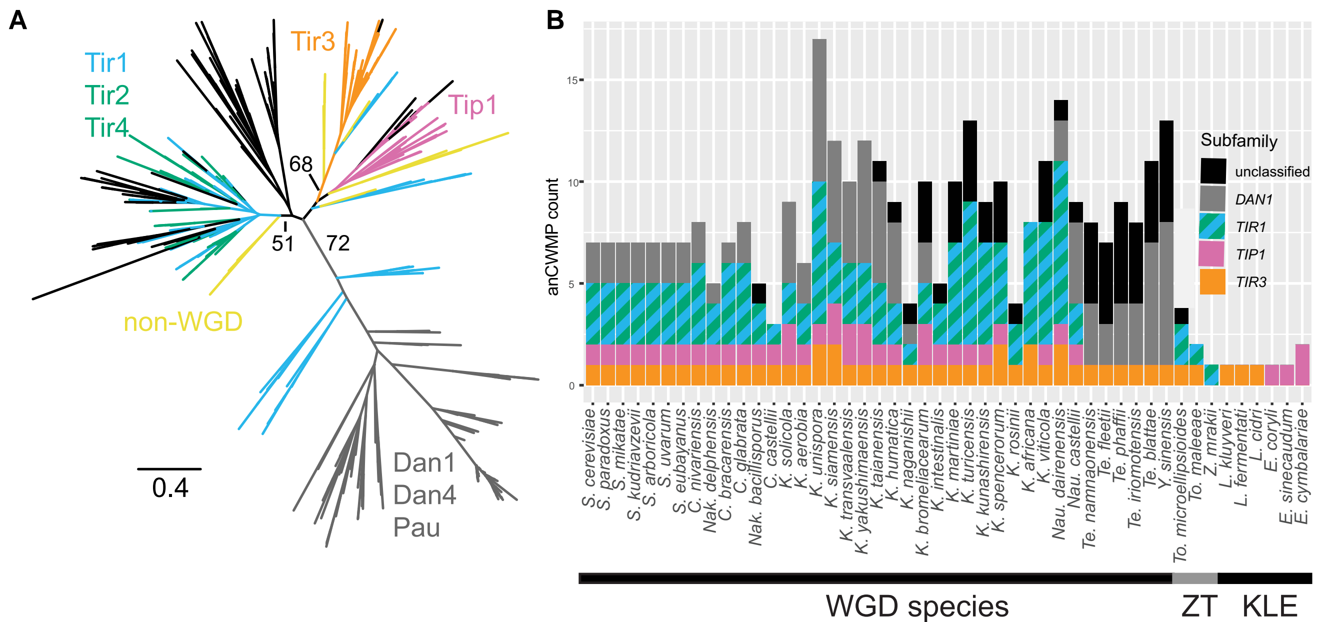

Fig. 4. (*A*) Maximum likelihood phylogeny of the anCWMP genes in Saccharomycetaceae using amino acid sequences. Branch values shown are the result of 100 bootstrap replicates. Branches are colored based on synteny with the *S. cerevisiae* homolog. Yellow branches indicate genes from non-WGD species. (*B*) Total number of each inferred anCWMP subfamily in each species of Saccharomycetaceae containing homologs. Subfamilies were defined as clades with bootstrap support values greater than 50% or shared synteny of 90% of its members with the *S. cerevisiae* homolog.

species (supplementary fig. S3, Supplementary Material online). These genes may have experienced concerted evolution, which homogenizes several copies of a multi-gene subfamily via recombination, and this process may be adaptive when gene dosage is critical to function (Hurst and Smith 1998). Indeed, the overlapping functions of *TIR1* and *TIR2* demonstrated in our genetic complementation assays and the redundancies in our combinatorial mutant analyses support this model of dosage-dependent homogenization. Several members of the *Torulaspora* and *Zygotorulaspora* genera have one or more members of the *TIR1* subfamily, and these originate at the base of the *TIR1* clade, suggesting that this subfamily was present in the ZT parent of the WGD hybridization (Fig. 4: yellow taxa at base of *TIR1* clade).

Genes related to *TIP1* formed a clade with 25% bootstrap support, and all post-WGD species, except *Kazachstania rosinii*, *Kazachstania africana*, *Tetrapisispora* spp., and *Y. sinensis*, contain at least one homolog. This subfamily may also include the anCWMP genes from *Eremothecium* spp., which branched at the base of the *TIP1* clade, were found in the ancestral locus of all anCWMP genes, and had a bootstrap support of 14%. Similar to the logic applied for *TIR3* and *TIR1*, this result is consistent with the presence of *TIP1* in the KLE parent of the WGD event, which was then translocated after the WGD. We infer that *TIP1* was subsequently lost in the *Tetrapisispora*/*Yueomyces* clade.

Genes related to *DAN1* and *DAN4* formed a clade with 72% bootstrap support. Their long branch lengths suggest that these genes rapidly diverged from the other anCWMP genes, and most post-WGD species contain at least one homolog, except some *Nakaseomyces* and *Kazachstania* species. Genes of this subfamily are generally located in the subtelomeres of chromosomes or near the ends of small contigs. *Tetrapisispora* spp. and *Yueomyces sinensis* *DAN1* genes are informative exceptions because they are located near the putative ancestral locus of all anCWMP genes (blue-colored clades at base of *DAN1* subfamily in Fig. 4). No non-WGD species contain clear *DAN1* homologs. These data imply that the *DAN1* gene subfamily emerged after the WGD event and that subsequent translocation to subtelomeric regions of chromosomes did not occur in the *Tetrapisispora*/*Yueomyces* lineage. The *PAU* genes are also embedded within this clade. These serine/threonine-poor genes are only found in *Saccharomyces* spp., and they number between 9 and 24, depending upon the species. Like the closely related *DAN* genes, the *PAU* genes are also mostly found in the subtelomeric regions of chromosomes.

These classifications exclude a small set of well-supported clades that lacked clear relationships with *S. cerevisiae* subfamilies. These divergent homologs may represent additional subfamilies and include several clades of genes found in *Naumovozyma* spp., *Kazachstania* spp., and *Tetrapisispora*/*Yueomyces* spp. Due to the poor bootstrap values for these clades and their phylogenetic placement basal to well-supported subfamilies, it is unclear whether these are divergent homologs of recognized subfamilies or additional novel subfamilies.

Origins of the anCWMP Genes, Their Divergence, and Their Role in Anaerobic Growth

To better understand the evolution of the anCWMP gene family and their function, we further investigated this gene

family in the context of all budding yeast species and its relationship to the homolog *AFB1*. We found several homologs of the *AFB1* gene as significant hits while searching for homologs of the anCWMP gene family in budding yeasts. We did not find homologs of either *AFB1* or the anCWMP genes in species outside the budding yeast subphylum, nor in any of the more basal budding yeast major clades, such as Lipomycetaceae, Trigonopsidaceae, or Dipodascaceae/Trichomonascaceae. *AFB1* encodes an a-factor barrier protein that is believed to bind to a-factor that is secreted from *MAT***a** cells, but it is not required for mating (Huberman and Murray 2013). The anCWMP genes and *AFB1* share sequence similarity in their N-terminal structured regions, as well as a low-complexity serine/threonine-rich region and a putative GPI-anchoring C-terminal region. The precise role of the anCWMP genes during anaerobic growth may be related to the putative function of Afb1, which binds in the cell wall to isoprenoid-related molecules that are structurally similar sterols.

While *AFB1* is predominantly a single-copy gene when present in budding yeasts, the anCWMP genes are often multi-copy, with 46% of species that contain anCWMP genes possessing two or more. This multi-copy nature is likely indicative of lineage-specific amplifications: copy number varies widely between yeast clades, while anCWMP genes within yeast clades tend to cluster together, rather than with homologs from other yeast clades. Although six species within the Phaffomycetaceae, *Babjeviella inositovorans*, and the *Brettanomyces* spp. all contain four or more anCWMP genes, here we focused on the most striking gene family amplification that occurred within the post-WGD lineage of Saccharomycetaceae. Functional characterization of the anCWMP genes outside of *S. cerevisiae* will better illuminate the roles of these genes in budding yeasts. The amplifications within *Saccharomycetaceae* and *Brettanomyces* are particularly interesting because they coincide with the independent evolution of these clades' abilities to grow anaerobically (Visser et al. 1990).

Our genetic experiments in *S. cerevisiae* show that the *TIR1* and *TIR3* genes are the major cell wall mannoprotein contributors to anaerobic growth. Establishing the precise timing of divergence between these two genes is challenging due to the general difficulty in determining the relationships between subfamilies using the few alignable sites. Bootstrap supports were low on branches connecting subfamilies (Fig. 4A), and phylogenetic topology tests failed to reject any tree topologies (supplementary table S1, Supplementary Material online). We attempted to root the Saccharomycetaceae subtree using anCWMP genes from Phaffomycetaceae or *AFB1* genes. Much like our attempts to determine relationships between subfamilies, the two trees were inconsistent in root placement and tree topology, demonstrating the difficulty in reliably assessing the evolutionary relationships among the anCWMP subfamilies (supplementary fig. S4, Supplementary Material online). Nonetheless, we can conclude that the anCWMP genes in the Saccharomycetaceae arose via several duplication and divergence events.

We can also infer that the *TIR1* and *TIR3* genes diverged prior to the WGD event because the non-WGD species *Torulaspora microellipsioides* and *Torulaspora maleeae* both contain putative members of both subfamilies. This result implies that the ZT parent of the WGD allopolyploidization event contained both genes. Given the presence of *TIR3* homologs in multiple *Lachancea* spp., the KLE parent likely contained at least a *TIR3* gene as well. Future work will be needed to determine whether the non-WGD homologs of these genes can functionally replace their post-WGD counterparts and to contribute to our understanding of the timing of the functional divergence between *TIR1* and *TIR3*. The timing of critical functional divergence is also complicated by the absence of clear *TIR1* homologs in the *Tetrapisispora*/*Yueomyces* lineage, a majority of whose species we found to grow under anaerobic conditions (supplementary table S2, Supplementary Material online). The inability of *S. cerevisiae TIR3* to complement a *C. glabrata tir3Δ* mutant further complicates our understanding of how individual anCWMP genes contribute to anaerobic growth in different species (Inukai et al. 2015). Future work is needed to address whether the anCWMP genes we have found to be critical for anaerobic growth in *S. cerevisiae* are the same as those critical in other species. While the relationships we inferred between anCWMP genes in this study were based on the conserved N-terminal and C-terminal portions of the proteins, the length and composition of the serine/threonine-rich regions may also contribute to functionality and functional divergence (Gemayel et al. 2010; Boisramé et al. 2011).

The anCWMP genes present an interesting system for studying the birth-and-death model of evolution, in which multi-gene families experience recurring duplications followed by functional divergence, dosage changes, or gene loss events. While we focused here on *S. cerevisiae* and its close relatives, future genetic experiments in *Brettanomyces* under anaerobic growth conditions, as well as experiments to identify functions of the expanded gene family members in other yeast lineages, would further contribute to understanding the birth-and-death process, as well as the cryptic functions of these genes. One prime target for future study is the genus *Kazachstania*, where every anCWMP gene subfamily has experienced either lineage-specific amplification events, gene loss events, or both. The *PAU* genes of the genus *Saccharomyces* are another example of genes that have experienced a lineage-specific amplification within the genus followed by many loss events, and functional characterization of these genes, which number as many as 24 in *S. cerevisiae*, may be on the horizon with genetic tools, such as CRISPR-Cas9. The *PAU* genes are closely related to *DAN1*, and their lack of a serine/threonine-rich region or a GPI-anchoring signal sequence is a strong contrast with the anCWMP genes that raises the possibility of their own distinct functional roles in the cell. Further work in diverse budding yeast

species will continue to shed light on how duplication, functional divergence, and gene loss in the CMWP gene family have occurred in various yeast lineages and what phenotypic effects these processes have had.

Here, we have identified a minimal set of two anCWMP genes, *TIR3* and *TIR1*, that are necessary and sufficient for anaerobic growth, likely by supporting sterol transport in *S. cerevisiae*. This finding may facilitate engineering sterol uptake into naïve yeast species or into *S. cerevisiae* under conditions when sterol uptake is normally repressed. This minimal set of genes provides a simplified system for studying the function of cell wall mannoproteins, as well as the functional divergence that underlies their mutual necessity for anaerobic growth. While we identified a critical anaerobic role for these genes, understanding their function in obligate aerobic species will yield insights into the evolutionary origins of anaerobic growth. The contemporaneity of anCWMP gene family expansions and origins of anaerobic growth that have independently occurred within the Saccharomycetaceae and the distantly related genus *Brettanomyces* suggest that gene amplification may be a critical step in the evolution of anaerobic growth. The phylogenetic relationships and functional differentiation among *S. cerevisiae* anCWMP genes observed here for the first time shed vital light on the origins of this ecologically and industrially important trait and set the stage for broader investigations.

Materials and Methods

Strains, Media, and Oligonucleotides

All genetic manipulations were performed in the prototrophic *S. cerevisiae* S288C *MAT***α** strain (*SUC2 gal2 mal2 mel flo1 flo8-1 hap1 ho bio1 bio6*). Gene replacement mutants using the *kanMX*, *hygMX*, *natMX*, and *zeoMX* cassettes were selected on 200 mg/L G418 (US Biological Life Sciences), 300 mg/L hygromycin B (US Biological Life Sciences), 100 mg/L ClonNAT (WERNERBioAgentsGmbH), and 100 mg/L Zeocin (Invitrogen), respectively. Markerless *ura3*-Δ mutants were also made. The *tip1*-Δ markerless deletion was obtained by first replacing the *TIP1* CDS with the *URA3* marker, then counterselecting against the *URA3* marker with a markerless repair template. All yeast strains used in this study can be found in supplementary table S3, Supplementary Material online Oligonucleotide sequences for construction and screening of the mutants can be found in supplementary table S4, Supplementary Material online.

Strains were generally grown in synthetic complete (SC) consisting of 5 g/L ammonium sulfate, 1.72 g/L yeast nitrogen base, 2 g/L synthetic dropout mix, and 20 g/L glucose (all reagents from US Biological Life Sciences). SC without uracil was used for strains carrying *URA3*-selection plasmid, and the synthetic dropout mix for this medium lacked uracil. For counterselection against *URA3*, SC plates containing 1 g/L 5′-FOA and 50 mg/L additional uracil were used (US Biological Life Sciences). YPD medium contained 10 g/L yeast extract, 20 g/L peptone, and 20 g/L glucose.

Backcrossing and Collecting Haploids

The *MAT***α** *dan1-dan4Δ tir2-tir4Δ tir3Δ tir1Δ tip1Δ ura3Δ* strain and a *MAT***a** *met6Δ* strain were spotted together and grown overnight on a minimal medium plate (lacking methionine and uracil). Colonies that grew were restreaked to a minimal medium plate and confirmed by PCR to be diploid at the *MAT* locus. The diploid was then streaked to a GNA presporulation plate (50 g/L glucose, 30 g/L Difco nutrient broth, 10 g/L yeast extract, 20 g/L agar) overnight, followed by inoculation of cells to sporulation medium for 72 hours at room temperature. The mix of spores and unsporulated diploids was then centrifuged, and the pellet was incubated with yeast protein extraction reagent (Thermo Scientific) and vortexed, followed by several washes with sterile dH_2O. The cells were then plated to YPD, and colonies were recovered and screened by PCR at the *MAT* locus and all anCWMP loci to confirm ploidy and determine genotypes.

Anaerobic Growth Experiments

A portion of a yeast colony was picked from a YPD plate into YPD medium for overnight growth at 30°C. The next morning, the saturated culture was diluted 20-fold into SC medium and grown for four hours at 30°C. These actively growing cultures were then introduced into a Coy anaerobic chamber and diluted 50-fold into an anaerobic 96-well plate containing 200µl of anaerobic medium and placed on a Tecan Spark-Stacker. The plate reader read absorbance or optical density (OD) at 600 nm once every hour after 5 s of shaking. After 24 hours of anaerobic culturing, the cultures were diluted 50-fold for a second round of growth, and these data form the results of the anaerobic experiments. For strains carrying plasmids with *URA3* selection, all YPD and SC media were replaced with SC medium lacking uracil. All anaerobic media contained anaerobic supplements of 20µg/mL ergosterol and Tween80.

Generating the anCWMP Gene Sequences Within Saccharomycotina

The amino acid sequences of all seven anCMWP genes from *S. cerevisiae* were used to search the publicly available genome sequences of 332 yeast species via http://y1000plus.org/blast using an e-value cutoff of 0.01 (Shen et al. 2016; Priyam et al. 2019). These sequences were also used as queries against the NCBI non-redundant protein sequence (nr) database, but all significant BLAST hits outside of budding yeasts were determined to most likely be contamination of sequencing projects with yeast DNA, due to high identity to *PAU* genes from *S. cerevisiae* (Supplementary File 6). These sequences were also used as queries against 1,011 published *S. cerevisiae* genomes (Peter et al. 2018), and these results can be found in Supplementary File 7. Full gene sequences were manually extracted from the genome sequence files using coordinates from the blast outputs. Genes containing 'N' nucleotide calls due to scaffolding or sequencing errors and

multi-domain anCWMP genes were excluded from the gene list. The full-length sequences used in this study can be found in Supplementary File 1. Both directly neighboring genes in *S. cerevisiae* were used as queries to identify homologs in the target genomes, and a anCWMP gene was considered syntenic with the *S. cerevisiae* homolog if a shared neighboring gene was found within 10Kb of the anCWMP homolog (supplementary table S5, Supplementary Material online). Because of alignment difficulties presented by the serine/threonine-rich region, this region was removed by concatenating the N-terminal portion to the last thirty codons encoding the C-terminal portion (see Fig. 3*B*, sequences in Supplementary File 2, and supplementary fig. S5, Supplementary Material online). Amino acid sequences were aligned using MAFFT (Katoh and Standley 2013), and alignment positions with greater than 25% gaps were removed using Trimal version 3 (Capella-Gutiérrez et al. 2009). Maximum-likelihood phylogenies were constructed using RAxML v.8.2.11 with the PROTGAMMAAUTO parameter and 100 rapid bootstrap calculations (Stamatakis 2014). The resulting tree files in Newick format can be found in Supplementary File 3 and Supplementary File 4, with a name decoding key in Supplementary File 5. AU tests were performed using IQTREE v.1.6.8 (Nguyen et al. 2015).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Robert A. Sclafani for S288C and Trey K. Sato, Kaitlin J. Fisher, and lab members for helpful suggestions. This material is based upon work supported in part by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409; the National Science Foundation under Grant Nos. DEB-1442148 and DEB-2110403; and the USDA National Institute of Food and Agriculture (Hatch Project 1020204). C.T.H. is a H.I. Romnes Faculty Fellow, supported by the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation.

Data Availability

The data underlying this article are available in the article and in its online supplementary material.

References

Abramova N, Sertil O, Mehta S, Lowry CV. 2001. Reciprocal regulation of anaerobic and aerobic cell wall mannoprotein gene expression in *Saccharomyces cerevisiae*. *J Bacteriol*. **183**:2881–2887.

Alimardani P, Régnacq M, Moreau-Vauzelle C, Ferreira T, Rossignol T, Blondin B, Bergès T. 2004. *SUT1*-promoted sterol uptake involves the ABC transporter Aus1 and the mannoprotein Dan1 whose synergistic action is sufficient for this process. *Biochem J*. **381**: 195–202.

Boisramé A, Cornu A, Da Costa G, Richard ML. 2011. Unexpected role for a serine/threonine-rich domain in the *Candida albicans* Iff protein family. *Eukaryot Cell*. **10**:1317–1330.

Brown CA, Murray AW, Verstrepen KJ. 2010. Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Curr Biol*. **20**:895–903.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. **25**:1972–1973.

Cliften PF, Fulton RS, Wilson RK, Johnston M. 2006. After the duplication: gene loss and adaptation in *Saccharomyces* genomes. *Genetics*. **172**:863–872.

Dean EJ, Davis JC, Davis RW, Petrov DA. 2008. Pervasive and persistent redundancy among duplicated genes in yeast. *PLoS Genet*. **4**: e1000113.

Delneri D, Gardner DC, Oliver SG. 1999. Analysis of the seven-member *AAD* gene set demonstrates that genetic redundancy in yeast may be more apparent than real. *Genetics*. **153**:1591–1600.

Donzeau M, Bourdineaud J-P, Lauquin GJ-M. 1996. Regulation by low temperatures and anaerobiosis of a yeast gene specifying a putative GPI-anchored plasma membrane. *Mol Microbiol*. **20**: 449–459.

Galardini M, Busby BP, Vieitez C, Dunham AS, Typas A, Beltrao P. 2019. The impact of the genetic background on gene deletion phenotypes in *Saccharomyces cerevisiae*. *Mol Syst Biol*. **15**:e8831.

Gemayel R, Vinces MD, Legendre M, Verstrepen KJ. 2010. Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu Rev Genet*. **44**:445–477.

Gordon JL, Byrne KP, Wolfe KH. 2009. Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. *PLoS Genet*. **5**:e1000485.

Hagman A, Säll T, Compagno C, Piskur J. 2013. Yeast "make-accumulate-consume" life strategy evolved as a multistep process that predates the whole genome duplication. *PLoS One*. **8**:e68734.

Hickman MA, Rusche LN. 2007. Substitution as a mechanism for genetic robustness: the duplicated deacetylases Hst1p and Sir2p in *Saccharomyces cerevisiae*. *PLoS Genet*. **3**:e126.

Hittinger CT, Carroll SB. 2007. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature*. **449**:nature06151.

Huberman LB, Murray AW. 2013. Genetically engineered transvestites reveal novel mating genes in budding yeast. *Genetics*. **195**: 1277–1290.

Hughes GM, Boston ESM, Finarelli JA, Murphy WJ, Higgins DG, Teeling EC. 2018. The birth and death of olfactory receptor gene families in mammalian niche adaptation. *Mol Biol Evol*. **35**:1390–1406.

Hurst LD, Smith NGC. 1998. The evolution of concerted evolution. *Proc Biol Sci*. **265**:121–127.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet*. **11**:97–108.

Inukai T, Nagi M, Morita A, Tanabe K, Aoyama T, Miyazaki Y, Bard M, Nakayama H. 2015. The mannoprotein *TIR3* (CAGL0C03872g) is required for sterol uptake in *Candida glabrata*. *Biochim Biophys Acta*. **1851**:141–151.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al*. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**:583–589.

Katoh K, Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. **30**:772–780.

Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. 2002. Selection in the evolution of gene duplications. *Genome Biol*. **3**:research0008.
Kuang MC, Hutchins PD, Russell JD, Coon JJ, Hittinger CT. 2016. Ongoing resolution of duplicate gene functions shapes the diversification of a metabolic network. *ELife*. **5**:e19027.
Lipke PN, Ovalle R. 1998. Cell wall architecture in yeast: new structure and new challenges. *J Bacteriol*. **180**:3735–3740.
Luo Z, van Vuuren HJJ. 2009. Functional analyses of *PAU* genes in *Saccharomyces cerevisiae*. *Microbiology (Reading)*. **155**: 4036–4049.
Marcet-Houben M, Gabaldón T. 2015. Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLoS Biol*. **13**:e1002220.
Myers KS, Riley NM, MacGilvray ME, Sato TK, McGee M, Heilberger J, Coon JJ, Gasch AP. 2019. Rewired cellular signaling coordinates sugar and hypoxic responses for anaerobic xylose fermentation in yeast. *PLoS Genet*. **15**:e1008037.
Nei M, Gu X, Sitnikova T. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci*. **94**:7799–7806.
Nei M, Rooney AP. 2005. Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet*. **39**:121–152.
Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating Maximum-likelihood phylogenies. *Mol Biol Evol*. **32**:268–274.
Niimura Y, Nei M. 2003. Evolution of olfactory receptor genes in the human genome. *PNAS*. **100**:12235–12240.
Ohno S. 1970. *Evolution by gene duplication*. London, New York: Springer-Verlag.
Papay M, Klein C, Hapala I, Petriskova L, Kuchler K, Valachovic M. 2020. Mutations in the nucleotide-binding domain of putative sterol importers Aus1 and Pdr11 selectively affect utilization of exogenous sterol species in yeast. *Yeast*. **37**:5–14.
Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, Sigwalt A, Barre B, Freel K, Llored A, *et al*. 2018. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*. **556**:339–344.
Priyam A, Woodcroft BJ, Rai V, Moghul I, Munagala A, Ter F, Chowdhary H, Pieniak I, Maynard LJ, Gibbins MA, *et al*. 2019. Sequenceserver: a modern graphical user interface for custom BLAST databases. *Mol Biol Evol*. **36**:2922–2924.
Reiner S, Micolod D, Zellnig G, Schneiter R. 2006. A genomewide screen reveals a role of mitochondria in anaerobic uptake of sterols in yeast. *Mol Biol Cell*. **17**:90–103.
Shen X-X, Opulente DA, Kominek J, Zhou X, Steenwyk JL, Buh KV, Haase MAB, Wisecaver JH, Wang M, Doering DT, *et al*. 2018. Tempo and mode of genome evolution in the budding yeast subphylum. *Cell*. **175**:1533–1545.e20.
Shen X-X, Zhou X, Kominek J, Kurtzman CP, Hittinger CT, Rokas A. 2016. Reconstructing the backbone of the Saccharomycotina yeast phylogeny using genome-scale data. *G3 (Bethesda)*. **6**: 3927–3939.
Snoek ISI, Steensma HY. 2006. Why does *Kluyveromyces lactis* not grow under anaerobic conditions? Comparison of essential anaerobic genes of *Saccharomyces cerevisiae* with the *Kluyveromyces lactis* genome. *FEMS Yeast Res*. **6**:393–403.
Snoek ISI, Steensma HY. 2007. Factors involved in anaerobic growth of *Saccharomyces cerevisiae*. *Yeast*. **24**:1–10.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. **30**: 1312–1313.
Thompson DA, Roy S, Chan M, Styczynsky MP, Pfiffner J, French C, Socha A, Thielke A, Napolitano S, Muller P, *et al*. 2013. Evolutionary principles of modular gene regulation in yeasts. *Elife*. **2**:e00603.
Trimmer C, Keller A, Murphy NR, Snyder LL, Willer JR, Nagai MH, Katsanis N, Vosshall LB, Matsunami H, Mainland JD. 2019. Genetic variation across the human olfactory receptor repertoire alters odor perception. *PNAS*. **116**:9475–9480.
Van Mulders SE, Christianen E, Saerens SMG, Daenen L, Verbelen PJ, Willaert R, Verstrepen KJ, Delvaux FR. 2009. Phenotypic diversity of Flo protein family-mediated adhesion in *Saccharomyces cerevisiae*. *FEMS Yeast Res*. **9**:178–190.
Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, *et al*. 2022. Alphafold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res*. **50**:D439–D444.
Visser W, Scheffers WA, Batenburg-van der Vegte WH, van Dijken JP. 1990. Oxygen requirements of yeasts. *Appl Environ Microbiol*. **56**: 3785–3792.
Wieczorke R, Krampe S, Weierstall T, Freidel K, Hollenberg CP, Boles E. 1999. Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. *FEBS Lett*. **464**:123–128.
Wilcox LJ, Balderes DA, Wharton B, Tinkelenberg AH, Rao G, Sturley SL. 2002. Transcriptional profiling identifies two members of the ATP-binding cassette transporter superfamily required for sterol uptake in yeast. *J Biol Chem*. **277**:32466–32472.
Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*. **387**:708–713.