

# Genetic basis for probiotic yeast phenotypes revealed by nanopore sequencing

Joseph H. Collins,<sup>1,‡</sup> Lohith Kunyeyit,<sup>2,3,‡</sup> Sarah Weintraub,<sup>4</sup> Nilesh Sharma,<sup>1</sup> Charlotte White,<sup>1</sup> Nabeeha Haq,<sup>5</sup> K.A. Anu-Appaiah,<sup>3</sup> Reeta P. Rao,<sup>2,\*</sup> Eric M. Young<sup>1,\*</sup>

<sup>1</sup>Department of Chemical Engineering, Worcester Polytechnic Institute, Worcester, MA 01609, USA

<sup>2</sup>Department of Biology and Biotechnology, Worcester Polytechnic Institute, Worcester, MA 01609, USA

<sup>3</sup>Department of Microbiology and Fermentation Technology, CSIR—Central Food Technological Research Institute (CFTRI), Mysore, Karnataka 570020, India

<sup>4</sup>Bioinformatics and Computational Biology, Worcester Polytechnic Institute, Worcester, MA 01609, USA

<sup>5</sup>Department of Biology, Brandeis University, Waltham, MA 02453, USA

\*Corresponding author: GP 4003, WPI-Life Sciences & Bioengineering Center, Gateway Park, 60 Prescott St, Worcester, MA 01605, USA. Email: emyoung@wpi.edu;

\*Corresponding author: GP 4014, WPI-Life Sciences & Bioengineering Center, Gateway Park, 60 Prescott St, Worcester, MA 01605, USA. Email: rpr@wpi.edu

‡These authors contributed equally to this work.

## Abstract

Probiotic yeasts are emerging as preventative and therapeutic solutions for disease. Often ingested via cultured foods and beverages, they can survive the harsh conditions of the gastrointestinal tract and adhere to it, where they provide nutrients and inhibit pathogens like *Candida albicans*. Yet, little is known of the genomic determinants of these beneficial traits. To this end, we have sequenced 2 food-derived probiotic yeast isolates that mitigate fungal infections. We find that the first strain, KTP, is a strain of *Saccharomyces cerevisiae* within a small clade that lacks any apparent ancestry from common European/wine *S. cerevisiae* strains. Significantly, we show that *S. cerevisiae* KTP genes involved in general stress, pH tolerance, and adherence are markedly different from *S. cerevisiae* S288C but are similar to the commercial probiotic yeast species *S. boulardii*. This suggests that even though *S. cerevisiae* KTP and *S. boulardii* are from different clades, they may achieve probiotic effect through similar genetic mechanisms. We find that the second strain, ApC, is a strain of *Issatchenkia occidentalis*, one of the few of this family of yeasts to be sequenced. Because of the dissimilarity of its genome structure and gene organization, we infer that *I. occidentalis* ApC likely achieves a probiotic effect through a different mechanism than the *Saccharomyces* strains. Therefore, this work establishes a strong genetic link among probiotic *Saccharomycetes*, advances the genomics of *Issatchenkia* yeasts, and indicates that probiotic activity is not monophyletic and complimentary mixtures of probiotics could enhance health benefits beyond a single species.

**Keywords:** *Saccharomyces cerevisiae* strain KTP, *Issatchenkia occidentalis* strain ApC, probiotic traits associated genes, *FLO* genes

## Introduction

Probiotic microbes are present in many fermented and cultured products across diverse cultures. Several have been isolated and marketed as probiotic supplements. To be classified as a probiotic, a microbe must exhibit beneficial effects and be properly identified by phenotypic and genomic methods (Reuter et al. 2002). According to Qualified Presumption of Safety (QPS) developed by European Food Safety Authority (EFSA), definition of the taxonomy of a microorganism in feed and food application is a major safety parameter in the selection process (Sanders et al. 2010). In other words, both the phenotype and genotype of a microbe must be defined before it can be called a probiotic. We have recently established a whole genome sequencing (WGS) pipeline called Prymetime (Collins et al. 2021). This tool can achieve higher genome contiguity and accuracy than previous approaches. Therefore, applying Prymetime to probiotic yeast, WGS could improve taxonomic classification and provide insight into the genomic underpinnings of probiotic microbes. This could then be leveraged to genetically engineer targeted probiotic solutions for

human health. Here, we apply Prymetime to 2 recently isolated yeast strains with beneficial properties.

Probiotic yeasts uniquely combine human health benefits and tolerance phenotypes that enable survival in the gastrointestinal tract (Kuneyit et al. 2023). They can produce beneficial metabolites and inhibit bacterial and fungal pathogens (Foligne et al. 2010; Kuneyit et al. 2021). They also can survive at human body temperature, withstand acidic and alkaline pHs similar to the digestive tract, and tolerate constituents of the digestive system like bile, gastric enzymes, and pancreatic enzymes (Tiago et al. 2009; Chen et al. 2010; Somplang and Piyadeatsoontorn 2016). They also can adhere to gut epithelial cells (O'Mahony et al. 2005; Lohith and Anu-Appaiah 2018). Perhaps unsurprisingly, probiotic yeasts are frequently found in fermented foods and beverages. Although several yeasts are known to have probiotic properties, only *Saccharomyces boulardii* has been commercialized and is prescribed to control and prevent gastrointestinal complications (McFarland 2010; Offei et al. 2019).

WGS is the key to enabling yeast classification and domestication. For example, WGS of *S. boulardii* revealed that it was in fact a

Received: January 31, 2023. Accepted: April 14, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of The Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

strain of *S. cerevisiae* with a few notable differences in galactose metabolism and flocculation genes (Khatrri et al. 2013, 2017). This led to the expansion of genetic tools for *S. boulardii*, opening exciting possibilities for interrogating genotype-phenotype connections, and creating designer probiotics by CRISPR-mediated genome editing (Ansari et al. 2019; Durmusoglu et al. 2021).

Recently, 2 yeasts isolated from fermented beverages were shown to have probiotic effects against *Candida albicans* and non-*albicans Candida* strains (Lohith and Anu-Appaiah 2018; Kunyeit et al. 2019; Kunyeit et al. 2021). The first strain, KTP, was isolated from a fermented beverage made from coconut sap (coconut toddy). The second strain, ApC, was isolated from fermented apple juice. Both were shown to control *C. albicans* filamentation and adhesion properties (Lohith and Anu-Appaiah 2018; Kunyeit et al. 2021). These strains were also shown to limit adhesion, filamentation, and biofilm formation of several non-*albicans Candida* species, including *C. tropicalis*, *C. krusei*, *C. glabrata*, *C. parapsilosis*, and *C. auris* (Kunyeit et al. 2019). Yeasts isolated from fermented foods have been shown to be resistant to several stressors such as temperature, osmotic stress, oxidative stress, pHs and antimicrobials making them suitable for probiotic applications (Kunyeit et al. 2023). As evidence, 2 yeasts reported here also survived exposure to the harsh conditions of the gastrointestinal tract tested in simulated gastric and bile juices as well as attachment to Caco-2 epithelial cells, in *ex vivo* conditions (Lohith and Anu-Appaiah 2018). This current report explores the genomic underpinning of the beneficial traits for these two strains. Together these findings support the use of these yeasts as probiotics.

The genomic sequence of ApC reveals that it is *Issatchenkia occidentalis*, a non-*Saccharomyces* yeast with very few genomes reported. *I. occidentalis* strain ApC belongs to the *Pichiaceae* family and it exhibits phenotypic traits such as the ability to attach to the intestinal epithelia, tolerance to digestive juices, and production of extracellular enzymes. These traits make it a novel and attractive probiotic strain. A clear connection between the probiotic phenotypes of *I. occidentalis* ApC and its genotype is limited until more there is a more robust collection of non-*Saccharomyces* genomes. The novel genome presented here expands that collection and provides a benchmark dataset for future studies. In contrast, the genome sequence of KTP reveals that it is a strain of *S. cerevisiae* with a robust collection of reference genomes allowing whole genome analysis as well as investigation of individual genes previously implicated in probiotic phenotypes. Thus, this study expands the genomic information available for probiotic yeasts and sets a strong foundation for future comparative genomic and genetic studies enabled by WGS.

## Materials and methods

### Strain isolation and growth

The *S. cerevisiae* strain KTP and *I. occidentalis*, ApC, were originally isolated from coconut toddy and fermented apple juice, respectively. Both strains were grown in yeast extract peptone dextrose (YPD) media at 30°C overnight.

### Scanning electron microscopy

Twenty-four-hour-old yeast cells were harvested by centrifugation and washed 3 times with phosphate buffer saline (pH 7.4). Washed cells were then fixed in 2.5% glutaraldehyde and set aside overnight at 4°C. The fixed cells were washed with phosphate buffer saline and dehydrated by sequential exposure to 10, 30, 50, 80, 90, and 100% ethanol. The samples were coated with conductive

layer of sputtered gold and examined under scanning electron microscope (LEO 435 VP LEO Electron microscopy, Cambridge, UK), and images were captured (Ruthu et al. 2014).

### Quantification of alcohol content

Strains KTP and ApC were inoculated in YPD media containing 10 and 20% glucose and incubated for 30 h at 28°C. Two milliliters supernatant was diluted with milliQ water (q.s. 50 mL) and distilled at 55°C. Fifteen milliliters of the distillate was collected in a graduated tube, and 2.5 mL of distillate was used to estimate alcohol content. Absorbance was read at 600 nm. For standard graph, range of 0–15% of absolute ethanol was used (Caputi et al. 1968; Vijayalaxmi et al. 2013).

### Genomic DNA isolation and sequencing

High-molecular weight genomic DNA was isolated based on a modified version of Promega's Genomic DNA Isolation Kit (Promega, A1120) 21. Nanopore reads were prepared for sequencing using the Rapid Barcoding Kit (ONT, SQK-RBK004). Illumina reads were prepared for sequencing using the Nextera DNA Flex Library Prep Kit (Illumina, 20018704) along with the Nextera DNA CD Indexes (Illumina, 20018707). Nanopore sequencing was started using the MinKNOW software from ONT with the default settings. The resulting fastq files were demultiplexed using EPI2ME (Metrichor, Oxford, UK). Illumina sequencing was started by using the native Local Run Manager on the iSeq 100 machine. A GENERATEFASTQ run was initiated and run with the parameters Read Type: Paired End, Read Lengths: 151, and Index Reads: 2. Reads were demultiplexed using the native software on the iSeq machine.

### De novo genome assembly and annotation

The *S. cerevisiae*, KTP, and *I. occidentalis*, ApC, genomes were assembled using the Prymetime (v0.2) pipeline (Collins et al. 2021), which uses both Nanopore and Illumina reads. The resulting assemblies were annotated using Augustus v3.2.3 (Stanke and Morgenstern 2005). Augustus requires a probable organism to be designated when processing a genome. Therefore, the probable organism for *S. cerevisiae* KTP was *S. cerevisiae*, while the probable organism for *I. occidentalis* ApC was *Pichia stipitis*.

### Genome assessment

The assembly statistics number of contigs, N50, assembly length, and GC percentage were generated using QUAST v5.0.0 (Gurevich et al. 2013). Completeness of the genome assemblies was calculated using BUSCO v4.0.6 with the *Saccharomycetes* database (Simao et al. 2015).

### Phylogenomics

The *S. cerevisiae* KTP strain was placed in the global *S. cerevisiae* phylogenetic tree using data from the *S. cerevisiae* 100 genomes project (Strope et al. 2015). The project used a set of 16 conserved regions to make a phylogenetic tree. KTP's corresponding 16 conserved sequences were extracted from the assembly using BLASTN (Johnson et al. 2008) and then concatenated into one fasta file. MAFFT v7.464 was used to create a multiple sequence alignment, with the phylogenetic tree constructed using the UPGMA setting (Kato et al. 2019). The ApC strain was placed in the *Pichiaceae* family phylogenetic tree using BUSCO and the BUSCO-phylogenomics utility script ([https://github.com/jamiecmg/BUSCO\\_phylogenomics](https://github.com/jamiecmg/BUSCO_phylogenomics)). BUSCO v4.0.6 with the *Saccharomycetes* database was run on 53 publicly available *Pichiaceae* family genome assemblies, the ApC assembly, and

the *S. cerevisiae* S288C assembly (outgroup). The BUSCO-phylogenomics script was used to construct a supermatrix alignment of highly conserved BUSCO families, followed by a Maximum Likelihood and Bayesian phylogenetic reconstruction on the supermatrix (McGowan et al. 2020). FigTree v1.4.2 was used to visualize the phylogenetic trees (<http://tree.bio.ed.ac.uk/software/figtree/>).

## Comparative genomics

The KTP assembly was compared to two other assemblies: *S. cerevisiae* S288C (Goffeau et al. 1996) and *S. boulardii* unique28 (Khatri et al. 2017). A list of probiotic-related proteins from *S. cerevisiae* S288C was obtained from the Yeast Genome Database (Cherry et al. 2012). BLASTP v2.6.0 was used to search for hits to these proteins in predicted proteomes of *S. cerevisiae* KTP and *S. boulardii* unique28. BLASTP was run with the 2/9 parameters “-task blastp -outfmt ‘6 qseqid sseqid pident qcovs’ -evalue 10e-5”. Protein alignments were run with Clustal Omega v1.2.4 with the default parameters (Sievers et al. 2011). ProgressiveMauve was used to align the S288C, KTP, and unique28 assemblies with the default parameters (Darling et al. 2010). Shared and unique orthologous groups among the S288C, KTP, and unique28 proteomes were identified using OrthoVenn2 (Xu et al. 2019). Proteins from the unique KTP orthologous group were first extracted and then BLASTP was used to identify the closest strain hit.

## Results and discussion

### Morphology

Before sequencing, we characterized colony and cell morphology of the 2 strains. Colonies of KTP are white and smooth while ApC forms rough colonies (Figs. 1a and 2a, respectively). Scanning electron microscopy shows that KTP cells are ovoid with evident buds, very similar to *S. cerevisiae* (Fig. 1b). ApC cells are more rod-like with buds occurring primarily at the end of the rods (Fig. 2b). These results corroborated the initial classification of KTP in *Saccharomyces* and ApC in *Issatchenkia*.

### Sequencing and genotyping with ITS sequences

Genomic DNA from KTP and ApC was isolated and sequenced using both short read (Illumina iSeq) and long read (Oxford Nanopore MinION) sequencing. The reads were assembled using Pymtime v0.2 (Collins et al. 2021) and annotated with Augustus v3.2.3 (Stanke and Morgenstern 2005). Genome assembly and annotation statistics are reported in Supplementary Table 1, and the raw data, assemblies, and annotation are published under BioProject PRJNA869102 for KTP PRJNA869107 for ApC (see Data Availability).

To initially classify the strains, we genotyped the strains using internal transcribed spacer (ITS) sequences. The method described by Kurtzman uses PCR of the ITS region followed by Sanger sequencing and BLAST to find related organisms. Here, we extracted the ITS regions of KTP and ApC from each genome assembly and input them into BLAST. The results are reported in Supplementary Table 1. The strain KTP is convincingly *S. cerevisiae*—the top 10 BLAST hits were all *S. cerevisiae* strains with percent identity above 99.5%. The species of the ApC strain was less clear—the top 10 BLAST hits were from several *Pichia* species. ApC was previously identified as *P. occidentalis* (now *I. occidentalis*) (GeneBank no.: KF551991.1) (Lohith and Anu-Appaiah 2018), but this initial genotyping observed that other *Pichia* species returned higher BLAST scores than *P. occidentalis*. These results provided

enough information to further refine the phylogenetic position of each strain using the whole genome.

### Genotyping KTP and ApC using the whole genome

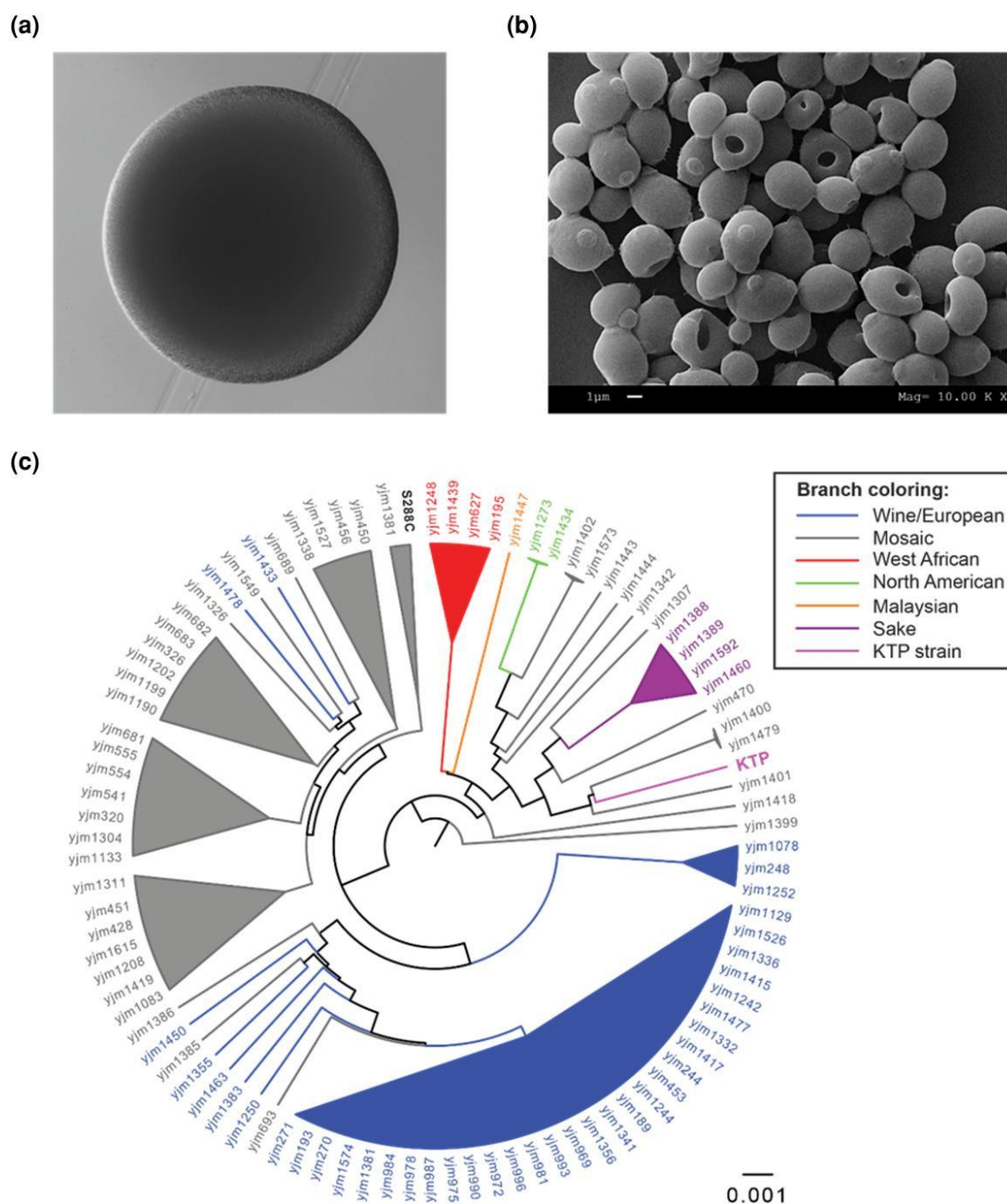
Since KTP was a candidate *Saccharomyces* strain, its phylogenetic position could be established using the *S. cerevisiae* 100 genomes project data as a reference (Strope et al. 2015). The *S. cerevisiae* 100 genomes project used a set of 16 conserved regions—one from each chromosome—to construct a global phylogenetic tree. To position KTP, these conserved regions were extracted from the KTP genome assembly and added to the 100 genomes project data (Fig. 1c). KTP is related to the Mosaic group, which has ancestry from two or more populations. KTP was closest to the yjm1400, yjm1479, and yjm1401 strains, which had significant ancestry from Sake, North American, and Malaysian strains. Unlike many of the strains tested in the 100 yeast genomes project, these three strains do not have any Wine/European ancestry.

ITS genotyping at least showed that ApC belonged in the *Pichiaceae* family. The *Pichiaceae* family yeasts are commonly found in spontaneous fermentation, as a result it commonly exists in several traditional fermented foods (Kunyeit et al. 2023). *Pichia* strains are used extensively in biotechnological applications, yet only a few such as *P. pastoris* and *P. kudriavzevii* are studied at the genomic level (Kuberl et al. 2011; Douglass et al. 2018). Furthermore, some *Pichiaceae* family shows a significant genome similarity with other yeast species. For example, a comparative genome analysis of *P. kudriavzevii* revealed 99.6% genome identity with the pathogenic yeast, *C. krusei* (Douglass et al. 2018). These initial findings underscore the need for a larger pool of genomic information to better classify the *Pichiaceae* family. Yet, with the current available data, we were able to conduct a phylogenetic analysis with 53 publicly available *Pichiaceae* family genome assemblies (Piskur et al. 2012; Ravin et al. 2013; Strope et al. 2015; Riley et al. 2016; Roach and Borneman 2020). Without predefined regins for classification, the analysis on all strains was carried out with essential genes determined by BUSCO (Simao et al. 2015). To do this, a utility script was used to create a supermatrix alignment of highly conserved BUSCO families and produce a phylogenetic reconstruction ([https://github.com/jamiemcg/BUSCO\\_phylogenomics](https://github.com/jamiemcg/BUSCO_phylogenomics)). The reconstruction and visualization of the phylogenetic tree (Fig. 2c) indicates that ApC is in fact closest to *I. occidentalis*. This highlights the usefulness of accurate tools like Pymtime for analysis and annotation of whole genome sequences to corroborate and resolve the taxonomic classification of yeasts. While probiotic attributes of non-*Saccharomyces* yeasts such as *I. occidentalis* have been studied extensively using in-vitro, ex-vivo, and preclinical models (Kunyeit et al. 2023), there is not a suitable non-*Saccharomyces* reference strain to aid in the analysis of the ApC genome. However, we were able to further analyze *S. cerevisiae* KTP in comparison to the well-known commercialized *Saccharomyces* probiotic yeast, *S. boulardii*, as several genes and pathways have been implicated in its probiotic traits.

### Comparative genome analysis reveals conserved and unique KTP genome features

After classification, we compared the *S. cerevisiae* KTP genome to the probiotic strain *S. boulardii* unique28 and the laboratory strain *S. cerevisiae* S288c. Each genome used in the comparison is nearly complete as estimated by BUSCO. The *S. cerevisiae* KTP assembly has 99.5% of single-copy BUSCOs, while the *S. cerevisiae* S288c assembly has 99.4%, and the *S. boulardii* unique28 assembly has 99.2% (Fig. 3a). We then analyzed orthologs using the protein clustering algorithm OrthoVenn2. We found 5081 orthologs shared



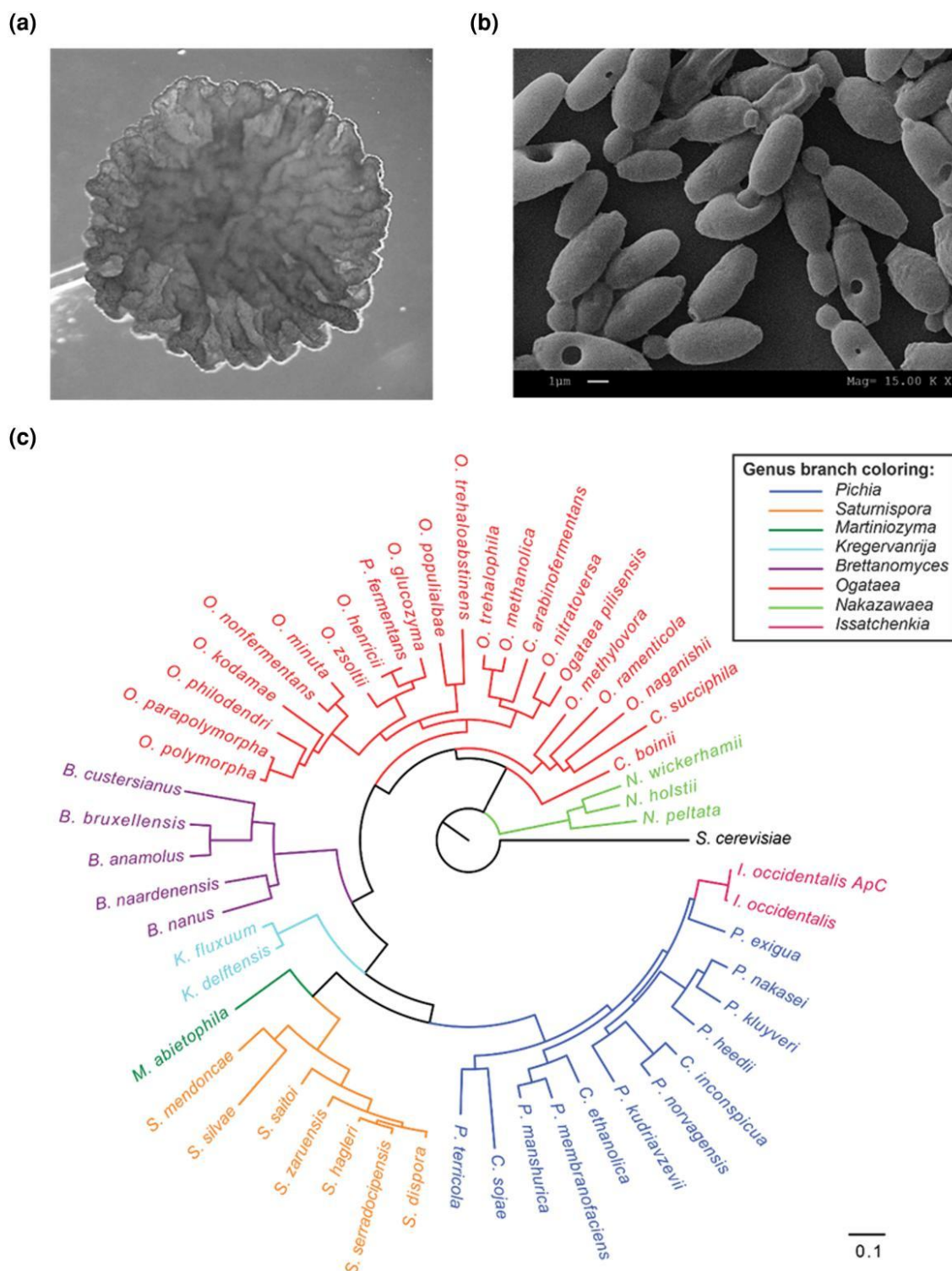


**Fig. 1.** Analysis and placement of *S. cerevisiae* KTP. a) A colony of *S. cerevisiae* KTP imaged under a microscope after 24 h incubation on a YPD agar plate. b) Scanning electron micrograph of *S. cerevisiae* KTP. c) Global phylogenetic position of the KTP strain in comparison to other *S. cerevisiae* strains.

among the three yeast strains and KTP had 43 unique sequences, S288c had 42 unique sequences, and unique28 had 8 unique sequences (Fig. 3b). We then compared the genome structures of the three strains using progressive Mauve 2.3.1. These analyses show that the nucleotide alignment of these three strains is very similar (represented as colored blocks in Fig. 3c) and support the argument that *S. boulardii* is a strain of *S. cerevisiae*. Furthermore, the alignment also shows that the genome configuration of *S. cerevisiae* KTP is more like S288c than *S. boulardii* unique28—particularly the pink block beginning at 6 MB is truncated only in *S. b. unique28* (Fig. 3c).

Unique orthologs are of particular interest because they could be the reason *S. cerevisiae*, KTP exhibits probiotic effects not seen

in S288c. Therefore, to classify these orthologs we performed BLASTp of the unique proteins in KTP. This analysis produced many hits across several species of yeasts. The top five gene description hits were heat shock protein Ssb1, elongation factor 1- $\alpha$ , amidase, Ato3p, and glyceraldehyde 3-phosphate dehydrogenase (Supplementary Table 2). These were compiled into a word cloud to show the number of hits for each gene (Fig. 3d). For Ssb1, which had five hits, we analyzed whether this was an assembly artifact, but each gene had read coverage consistent with the rest of the genome, ruling out assembly artifacts (Supplementary Fig. 1). Ssb1 is member of HSP70 family, which are highly abundant heat shock proteins unique to fungi (Dombek et al. 2004). The other genes are associated in translation and



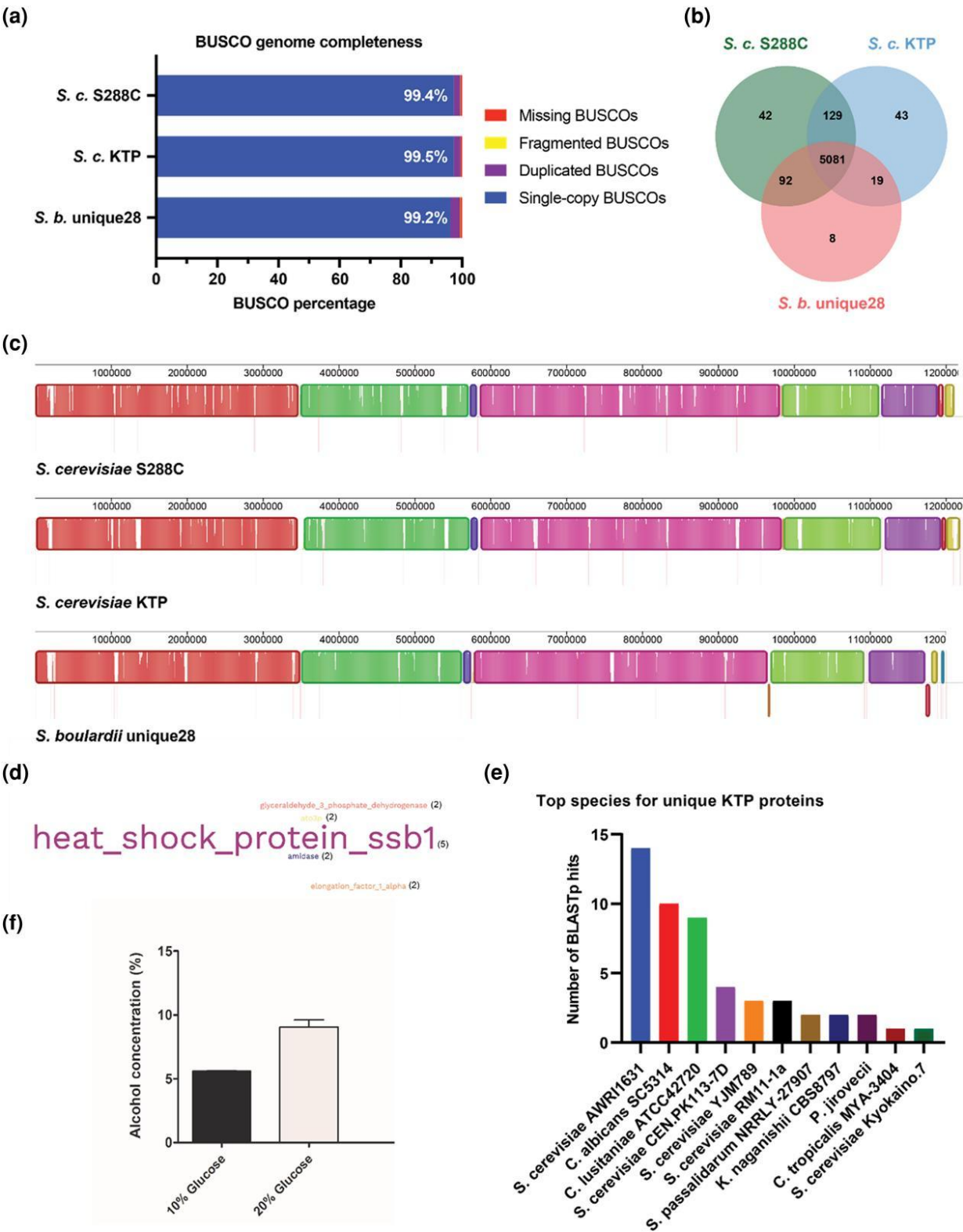
**Fig. 2.** Analysis and placement of *I. occidentalis* ApC. a) A colony of *I. occidentalis* ApC imaged under a microscope after 24 h incubation on a YPD agar plate. b) Scanning electron micrograph of *I. occidentalis* ApC. c) Phylogenetic position of *I. occidentalis* ApC in the Pichiaceae family.

central carbon metabolism. This indicates that while its genome structure may be similar to S288c, KTP could have key differences in tolerance, gene expression, and metabolism that contribute to the probiotic effect.

We further sorted the BLASTp hits of the unique genes by species, showing that many of the unique proteins are typically

shared among wine yeasts such as *S. cerevisiae* AWRI1631, followed by two *Candida* species, and also *S. cerevisiae*, strains CEN.PK113-7D, YJM989 and RM11-1a (Fig. 3e).

An ethanol concentration greater than eight percent (>8%) is a key parameter for a yeast strain to be used in winemaking (Fleet 2008) and *S. cerevisiae* AWRI1631 produces 12% ethanol when

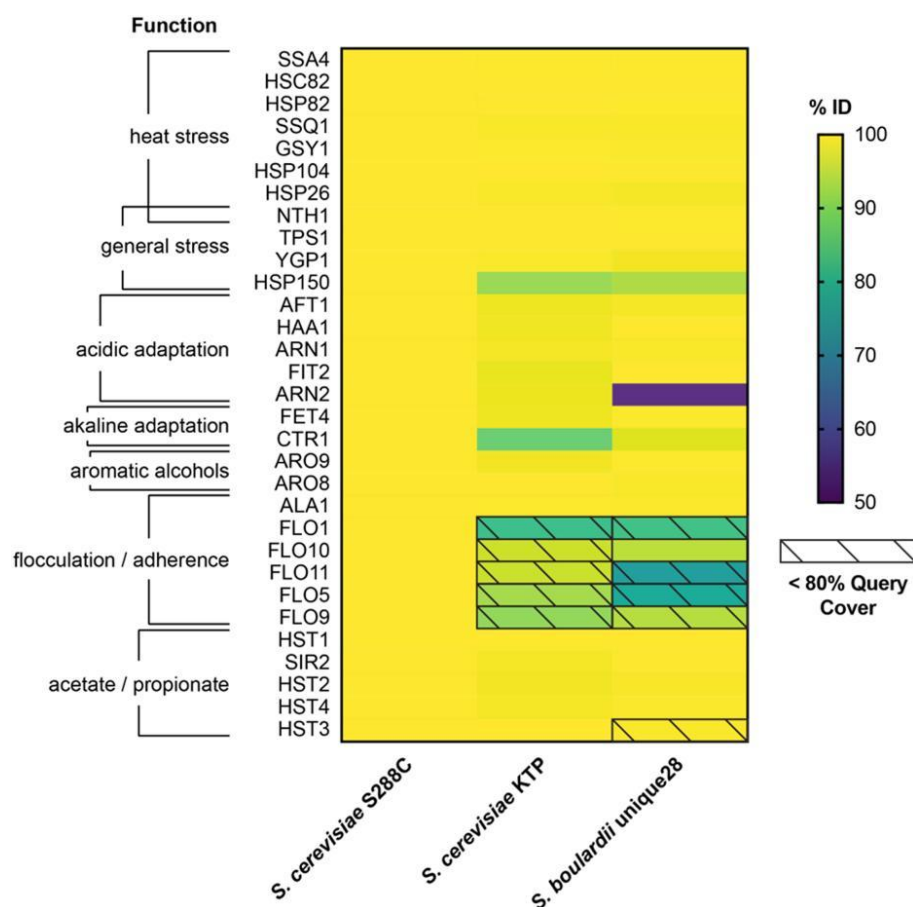


**Fig. 3.** Comparing the genomes of *S. cerevisiae* S288c, *S. boulardii* unique28, and *S. cerevisiae* KTP. a) BUSCO genome completeness assessment of the 3 strains. The white text shows the completeness score for each assembly. b) Venn diagram of orthologous proteins shared or unique to the 3 strains. c) Whole genome alignment with ProgressiveMauve. Colored blocks indicate regions of high similarity. Red vertical lines designate a new contig in the genome assembly. Blocks below the center line are aligned sequences in the reverse direction. White vertical lines within blocks represent the localized areas that have not aligned. d) Word cloud of the top BLASTp descriptions for the unique KTP proteins. e) Number of BLASTp hits for the unique KTP proteins according to the top strain found. f) Ethanol production of *S. cerevisiae* KTP after 30 h of fermentation in YP with either 10% or 20% glucose.

grown in medium containing 20% glucose (Contreras et al. 2014). Therefore, we measured the ethanol content of *S. cerevisiae* KTP. We found that it produces  $5.58 \pm 0.09\%$  and  $9.07 \pm 0.92\%$  ethanol

in medium containing 10 and 20% glucose, respectively (Fig. 3f). This result suggests that *S. cerevisiae* KTP is suitable for alcohol fermentation though that is not the focus of our study.





**Fig. 4.** Percent identity to proteins implicated in probiotic phenotypes. BLASTp hits from the *S. cerevisiae* KTP and *S. boulardii* unique28 strains with less than 80% query coverage are shown with black diagonal lines.

### *S. cerevisiae* KTP and *S. boulardii* share genes implicated in probiotic phenotypes

Several probiotic phenotypes have known genotypes. These include tolerance, adherence, and metabolite biosynthesis. The conditions of the gastrointestinal tract require tolerance to pH, both acidic and alkaline, as well as tolerance to a temperature of 37°C. Adherence to intestinal epithelial cells is necessary because the natural flow of material through the gastrointestinal tract can quickly clear non-adherent microbes. Biosynthesis of certain metabolites like aromatic alcohols and short chain fatty acids has been associated with several functional attributes in probiotic microbes. Particularly, acetate and propionate can decrease inflammation of the colon, control the secretory activity of gut by modulating enteric nervous system, and improve gut immunity (Koh et al. 2016). Further, the aromatic alcohols tryptophol and phenylethanol were shown to inhibit filamentation of *C. albicans* (Kunyeit et al. 2021). It has already been shown that *S. cerevisiae* KTP exhibits pH and heat tolerance as well as adherence to intestinal cells (Lohith and Anu-Appaiah 2018). Therefore, we compiled a list of 31 genes that have been implicated in stress tolerance, adherence, and biosynthesis and compared these genes across *S. cerevisiae* KTP, *S. cerevisiae* S288c, and *S. boulardii* unique28 (Fig. 4).

The microenvironment of the human gastrointestinal presents conditions that inhibit growth of microorganisms. However, several microorganisms, including *S. cerevisiae*, can survive adverse conditions of the gut by regulating genes that modulate their response to stresses such as extreme temperature, osmolarity,

and availability of key micronutrients such as essential trace elements. We identified 18 genes implicated in heat and pH tolerance. The heat response genes included HSP26, SSA4, HSP82, HSC82, HSP104, and GSY1, the general stress regulators TPS1 and NTH1, and the long-term heat stress response gene SSQ1 (Auesukaree et al. 2012; Chamnipa et al. 2018). The acidic pH tolerance genes included were stress response genes YGP1 and HSP150, the metal metabolism genes FIT2, ARN1, and ARN2 (Kawahata et al. 2006), and AFT1 and HAA1 (Haitani et al. 2012). The alkaline pH tolerance genes included were FET4 and CTR1 (Serrano et al. 2004) (Fig. 4). Large scale studies have identified these genes to be involved in the regulation of general stress responses including metal homeostasis (Yun et al. 2000a, 2000b). We believe that these genes likely play a role in modulating the adaptation of food-derived yeasts in the human gastrointestinal tract.

We identified 6 genes involved in adherence. These primarily consisted of the FLO family—FLO1, FLO5, FLO9, FLO10, and FLO11—that encode for cell-wall associated surface proteins and regulate the cell surface properties of *S. cerevisiae* in genetic and epigenetic level (Halme et al. 2004; Hope and Dunham 2014). ALA1 was also included because it has been implicated in the ability to bind extracellular matrices (Gaur and Klotz 1997). The FLO family is of particular interest because these genes have been shown to have a large impact on probiotic characteristics. FLO1, FLO5, FLO9, and FLO10 are responsible for cell-to-cell adhesion as well as adhesion to abiotic surfaces in *S. cerevisiae* (Teunissen and Steensma 1995). FLO1, FLO5 and FLO9 are important for yeast bio-film formation (Yang et al. 2018). FLO11 is involved in adhesion to

agar and abiotic surfaces, sliding motility, filament formation, invasive growth, and substrate adhesion (Guo *et al.* 2000; Halme *et al.* 2004; Bayly *et al.* 2005). Higher expression of *FLO1*, *FLO5* and *FLO11* in *S. cerevisiae* has also been shown to enhance thermotolerance and viability (Vergara-Alvarez *et al.* 2019) (Fig. 4).

We identified 7 genes involved in metabolic pathways relevant to probiotic yeasts. The genes included from acetate and propionate metabolism were *SIR2*, *HST1*, *HST2*, *HST3*, and *HST4* (Starai *et al.* 2003). The genes *ARO8* and *ARO9* were included because they participated in aromatic alcohol biosynthesis (tryptophol and phenylethanol) (Chen and Fink 2006) (Fig. 4).

Each of the 32 genes were then found in *S. cerevisiae* KTP, *S. cerevisiae* S288c, and *S. boulardii* unique 28 using BLASTp. The percent identity of the top protein hit to the query is shown in Fig. 4. Interestingly, both KTP and *S. boulardii* unique28 have less than 80% query coverage for *FLO1*, *FLO10*, *FLO5*, *FLO9*, and *FLO11*. The observed differences are not due to assembly, as read coverage is consistent across these five genes (Supplementary Fig. 2). Protein sequence alignment of the KTP and S288c sequences for each are shown in Supplementary Figs. 3, 4, 5, 6 and 7. In addition to the *FLO* family, the genes encoding the heat shock protein *HSP150* and the alkaline adaptation gene *CTR1* had a lower query coverage; however, these may be due to assembly artifacts (Supplementary Fig. 2).

## Conclusion

These results show that WGS is vital for accurate taxonomic classification and genome analysis of probiotic yeasts. Using an accurate genome assembly, we were able to determine that *S. cerevisiae* KTP is from a different *Saccharomyces* lineage than *S. boulardii*, yet they both share mutations in flocculation genes, suggesting at a convergent evolutionary strategy for probiotic mechanisms. We also showed that *S. cerevisiae* KTP shares several gene classes with wine yeasts and produces ~8% ethanol. We also showed that *I. occidentalis* ApC is from a branch of nonconventional yeasts with few probiotic strains recognized. Its genome is significantly different from the other probiotic strains, suggesting it may have a different probiotic strategy. This is an area for future probiotic research as more genomes of non-*Saccharomyces* probiotic yeasts become available. Overall, this study provides a blueprint for WGS of promising yeast strains that could have medicinal or industrial benefit and demonstrates how analysis of accurate genomic information can yield insight into phenotype-genotype relationships.

## Data availability

*S. cerevisiae* KTP has been assigned the NCBI BioSample accession SAMN30285631, and all data are available under BioProject PRJNA869102. The whole genome assembly can be accessed with accession JANQBH000000000. The raw reads are available at the NCBI Sequence Read Archive (SRA)—nanopore reads are available under accession SRR21031641 and illumina reads are available under accession SRR21031642. *I. occidentalis* has been assigned the NCBI BioSample accession SAMN30286403, and all data are available under BioProject PRJNA869107. The whole genome assembly can be accessed with accession JANQBI000000000. The raw reads are available at the NCBI SRA—nanopore reads are available under accession SRR21034916 and illumina reads are available under accession SRR21034917.

Supplemental material available at G3 online.

## Acknowledgements

We thank the Director, CSIR-Central Food Technological Research Institute (CFTRI) for encouragement and research support. L.K. is grateful to INSPIRE program, Department of Science and Technology, Government of India, and Fulbright-Nehru doctoral fellowship, United States-India Education Foundation (USIEF), India for the financial support for his doctoral research.

## Funding

Department of Science and Technology, Ministry of Science and Technology, India, award number DST/INSPIRE Fellowship/2013/553 and United States-India Education Foundation, award number 2310/DR/2018-2019 to L.K. This work is partially supported by National Center for Complementary and Integrative Health, grant, 1R15AT009926-01 to R.P.R.

## Conflicts of interest

The authors declare no conflict of interest.

## Literature cited

- Ansari JM, Colasacco C, Emmanouil E, Kohlhepp S, Harriott O. Strain-level diversity of commercial probiotic isolates of *Bacillus*, *Lactobacillus*, and *Saccharomyces* species illustrated by molecular identification and phenotypic profiling. *PLoS One*. 2019;14(3): e0213841. doi:10.1371/journal.pone.0213841.
- Auesukaree C, Koedrih P, Saenpayavai P, Asvarak T, Benjaphokee S, Sugiyama M, Kaneko Y, Harashima S, Boonchird C. Characterization and gene expression profiles of thermotolerant *Saccharomyces cerevisiae* isolates from Thai fruits. *J Biosci Bioeng*. 2012;114(2):144–149. doi:10.1016/j.jbiosc.2012.03.012.
- Bayly JC, Douglas LM, Pretorius IS, Bauer FF, Dranginis AM. Characteristics of flo11-dependent flocculation in *Saccharomyces cerevisiae*. *FEMS Yeast Res*. 2005;5(12):1151–1156. doi:10.1016/j.femsyr.2005.05.004.
- Caputi A, Ueda M, Brown T. Spectrophotometric determination of ethanol in wine. *Am J Enol Vitic*. 1968;19(3):160–165. doi:10.5344/ajev.1968.19.3.160.
- Chamnira N, Thanonkeo S, Klanrit P, Thanonkeo P. The potential of the newly isolated thermotolerant yeast *Pichia kudriavzevii* rz8-1 for high-temperature ethanol production. *Braz J Microbiol*. 2018;49(2):378–391. doi:10.1016/j.bjm.2017.09.002.
- Chen H, Fink GR. Feedback control of morphogenesis in fungi by aromatic alcohols. *Genes Dev*. 2006;20(9):1150–1161. doi:10.1101/gad.1411806.
- Chen LS, Ma Y, Maubois JL, He SH, Chen LJ, Li HM. Screening for the potential probiotic yeast strains from raw milk to assimilate cholesterol. *Dairy Sci Technol*. 2010;90(5):537–548. doi:10.1051/dst/2010001.
- Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, et al. *Saccharomyces* genome database: the genomics resource of budding yeast. *Nucleic Acids Res*. 2012;40(D1):D700–D705. doi:10.1093/nar/gkr1029.
- Collins JH, Keating KW, Jones TR, Balaji S, Marsan CB, Como M, Newlon ZJ, Mitchell T, Bartley B, Adler A, et al. Engineered yeast genomes accurately assembled from pure and mixed samples. *Nat Commun*. 2021;12(1):1485. doi:10.1038/s41467-021-21656-9.
- Contreras A, Hidalgo C, Henschke PA, Chambers PJ, Curtin C, Varela C. Evaluation of non-*Saccharomyces* yeasts for the reduction of



- alcohol content in wine. *Appl Environ Microbiol.* 2014;80(5):1670–1678. doi:10.1128/AEM.03780-13.
- Darling AE, Mau B, Perna NT. Progressivemaue: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One.* 2010;5(6):e11147. doi:10.1371/journal.pone.0011147.
- Dombek KM, Kacherovsky N, Young ET. The reg1-interacting proteins, bmh1, bmh2, ssb1, and ssb2, have roles in maintaining glucose repression in *Saccharomyces cerevisiae*. *J Biol Chem.* 2004;279(37):39165–39174. doi:10.1074/jbc.M400433200.
- Douglass AP, Offei B, Braun-Galleani S, Coughlan AY, Martos AA, Ortiz-Merino RA, Byrne KP, Wolfe KH. Population genomics shows no distinction between pathogenic *Candida krusei* and environmental *Pichia kudriavzevii*: one species, four names. *PLoS Pathog.* 2018;14(7):e1007138. doi:10.1371/journal.ppat.1007138.
- Durmusoglu D, Al'Abri IS, Collins SP, Cheng J, Eroglu A, Beisel CL, Crook N. In situ biomanufacturing of small molecules in the mammalian gut by probiotic *Saccharomyces boulardii*. *ACS Synth Biol.* 2021;10(5):1039–1052. doi:10.1021/acssynbio.0c00562.
- Fleet GH. Wine yeasts for the future. *FEMS Yeast Res.* 2008;8(7):979–995. doi:10.1111/j.1567-1364.2008.00427.x.
- Foligne B, Dewulf J, Vandekerckove P, Pignede G, Pot B. Probiotic yeasts: anti-inflammatory potential of various non-pathogenic strains in experimental colitis in mice. *World J Gastroenterol.* 2010;16(17):2134–2145. doi:10.3748/wjg.v16.i17.2134.
- Gaur NK, Klotz SA. Expression, cloning, and characterization of a *Candida albicans* gene, ala1, that confers adherence properties upon *Saccharomyces cerevisiae* for extracellular matrix proteins. *Infect Immun.* 1997;65(12):5289–5294. doi:10.1128/iai.65.12.5289-5294.1997.
- Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, et al. Life with 6000 genes. *Science.* 1996;274(5287):546, 563–567. doi:10.1126/science.274.5287.546.
- Guo B, Styles CA, Feng Q, Fink GR. A *Saccharomyces* gene family involved in invasive growth, cell-cell adhesion, and mating. *Proc Natl Acad Sci U S A.* 2000;97(22):12158–12163. doi:10.1073/pnas.220420397.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. Quast: quality assessment tool for genome assemblies. *Bioinformatics.* 2013;29(8):1072–1075. doi:10.1093/bioinformatics/btt086.
- Haitani Y, Tanaka K, Yamamoto M, Nakamura T, Ando A, Ogawa J, Shima J. Identification of an acetate-tolerant strain of *Saccharomyces cerevisiae* and characterization by gene expression analysis. *J Biosci Bioeng.* 2012;114(6):648–651. doi:10.1016/j.jbiosc.2012.07.002.
- Halme A, Bumgarner S, Styles C, Fink GR. Genetic and epigenetic regulation of the flo gene family generates cell-surface variation in yeast. *Cell.* 2004;116(3):405–415. doi:10.1016/S0092-8674(04)00118-7.
- Hope EA, Dunham MJ. Ploidy-regulated variation in biofilm-related phenotypes in natural isolates of *Saccharomyces cerevisiae*. *G3 (Bethesda).* 2014;4(9):1773–1786. doi:10.1534/g3.114.013250.
- Johnson M, Zaretskaya I, Raytselis Y, Merezhuik Y, McGinnis S, Madden TL. Ncbi blast: a better web interface. *Nucleic Acids Res.* 2008;36(Web Server):W5–W9. doi:10.1093/nar/gkn201.
- Katoh K, Rozewicki J, Yamada KD. Mafft online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinformatics.* 2019;20(4):1160–1166. doi:10.1093/bib/bbx108.
- Kawahata M, Masaki K, Fujii T, Iefuji H. Yeast genes involved in response to lactic acid and acetic acid: acidic conditions caused by the organic acids in *Saccharomyces cerevisiae* cultures induce expression of intracellular metal metabolism genes regulated by aft1p. *FEMS Yeast Res.* 2006;6(6):924–936. doi:10.1111/j.1567-1364.2006.00089.x.
- Khatri I, Akhtar A, Kaur K, Tomar R, Prasad GS, Ramya TNC, Subramanian S. Gleaning evolutionary insights from the genome sequence of a probiotic yeast *Saccharomyces boulardii*. *Gut Pathog.* 2013;5(1):30. doi:10.1186/1757-4749-5-30.
- Khatri I, Tomar R, Ganesan K, Prasad GS, Subramanian S. Complete genome sequence and comparative genomics of the probiotic yeast *Saccharomyces boulardii*. *Sci Rep.* 2017;7(1):371. doi:10.1038/s41598-017-00414-2.
- Koh A, De Vadder F, Kovatcheva-Datchary P, Backhed F. From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. *Cell.* 2016;165(6):1332–1345. doi:10.1016/j.cell.2016.05.041.
- Kubler A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, Jaenicke S, Brinkrolf K, Goesmann A, Szczepanowski R, et al. High-quality genome sequence of *Pichia pastoris* CBS7435. *J Biotechnol.* 2011;154(4):312–320. doi:10.1016/j.jbiotec.2011.04.014.
- Kunyeit L, Kurrey NK, Anu-Appaiah KA, Rao RP. Probiotic yeasts inhibit virulence of non-albicans *Candida* species. *mBio.* 2019;10(5):e02307–19. doi:10.1128/mBio.02307-19.
- Kunyeit L, Kurrey NK, Anu-Appaiah KA, Rao RP. Secondary metabolites from food-derived yeasts inhibit virulence of *Candida albicans*. *mBio.* 2021;12(4):e0189121. doi:10.1128/mBio.01891-21.
- Kunyeit L, Rao RP, Anu-Appaiah KA. Yeasts originating from fermented foods, their potential as probiotics and therapeutic implication for human health and disease. *Crit Rev Food Sci Nutr.* 2023. doi.org/10.1080/10408398.2023.2172546.
- Lohith K, Anu-Appaiah KA. Antagonistic effect of *Saccharomyces cerevisiae* K1P and *Issatchenkia occidentalis* ApC on hyphal development and adhesion of *Candida albicans*. *Med Mycol.* 2018;56(8):1023–1032. doi:10.1093/mmy/myx156.
- McFarland LV. Systematic review and meta-analysis of *Saccharomyces boulardii* in adult patients. *World J Gastroenterol.* 2010;16(18):2202–2222. doi:10.3748/wjg.v16.i18.2202.
- McGowan J, O'Hanlon R, Owens RA, Fitzpatrick DA. Comparative genomic and proteomic analyses of three widespread *Phytophthora* species: *Phytophthora chlamydospora*, *Phytophthora gonapodyides* and *Phytophthora pseudosyringae*. *Microorganisms.* 2020;8(5):653. doi:10.3390/microorganisms8050653.
- O'Mahony L, McCarthy J, Kelly P, Hurley G, Luo F, Chen K, O'Sullivan GC, Kiely B, Collins JK, Shanahan F, et al. Lactobacillus and bifidobacterium in irritable bowel syndrome: symptom responses and relationship to cytokine profiles. *Gastroenterology.* 2005;128(3):541–551. doi:10.1053/j.gastro.2004.11.050.
- Offei B, Vandecruys P, De Graeve S, Foulquie-Moreno MR, Thevelein JM. Unique genetic basis of the distinct antibiotic potency of high acetic acid production in the probiotic yeast *Saccharomyces cerevisiae* var. *boulardii*. *Genome Res.* 2019;29(9):1478–1494. doi:10.1101/gr.243147.118.
- Piskur J, Ling Z, Marcet-Houben M, Ishchuk OP, Aerts A, LaButti K, Copeland A, Lindquist E, Barry K, Compagno C, et al. The genome of wine yeast *Dekkera bruxellensis* provides a tool to explore its food-related properties. *Int J Food Microbiol.* 2012;157(2):202–209. doi:10.1016/j.ijfoodmicro.2012.05.008.
- Ravin NV, Eldarov MA, Kadnikov VV, Beletsky AV, Schneider J, Mardanova ES, Smekalova EM, Zvereva MI, Dontsova OA, Mardanov AV, et al. Genome sequence and analysis of methylotrophic yeast *Hansenula polymorpha* dl1. *BMC Genomics.* 2013;14(1):837. doi:10.1186/1471-2164-14-837.
- Reuter G, Klein G, Goldberg M. Identification of probiotic cultures in food samples. *Food Res Int.* 2002;35(2–3):117–124. doi:10.1016/S0963-9969(01)00172-7.

- Riley R, Haridas S, Wolfe KH, Lopes MR, Hittinger CT, Goker M, Salamov AA, Wisecaver JH, Long TM, Calvey CH, et al. Comparative genomics of biotechnologically important yeasts. *Proc Natl Acad Sci U S A*. 2016;113(35):9882–9887. doi:10.1073/pnas.1603941113.
- Roach MJ, Borneman AR. New genome assemblies reveal patterns of domestication and adaptation across *Brettanomyces* (dekkera) species. *BMC Genomics*. 2020;21(1):194. doi:10.1186/s12864-020-6595-z.
- Ruthu, Murthy PS, Rai AK, Bhaskar N. Fermentative recovery of lipids and proteins from freshwater fish head waste with reference to antimicrobial and antioxidant properties of protein hydrolysate. *J Food Sci Technol*. 2014;51(9):1884–1892. doi:10.1007/s13197-012-0730-z.
- Sanders ME, Akkermans LM, Haller D, Hammerman C, Heimbach JT, Hörmannspurger G, and Huys G. Safety assessment of probiotics for human use. *GUT Microbes*. 2010;1(3):164–185. doi:10.4161/gmic.1.3.12127.
- Serrano R, Bernal D, Simon E, Arino J. Copper and iron are the limiting factors for growth of the yeast *Saccharomyces cerevisiae* in an alkaline environment. *J Biol Chem*. 2004;279(19):19698–19704. doi:10.1074/jbc.M313746200.
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol Syst Biol*. 2011;7(1):539. doi:10.1038/msb.2011.75.
- Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210–3212. doi:10.1093/bioinformatics/btv351.
- Sornplang P, Piyadeatsoontorn S. Probiotic isolates from unconventional sources: a review. *J Anim Sci Technol*. 2016;58(1):26. doi:10.1186/s40781-016-0108-2.
- Stanke M, Morgenstern B. Augustus: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res*. 2005;33(Web Server):W465–W467. doi:10.1093/nar/gki458.
- Starai VJ, Takahashi H, Boeke JD, Escalante-Semerena JC. Short-chain fatty acid activation by acyl-coenzyme A synthetases requires sir2 protein function in *Salmonella enterica* and *Saccharomyces cerevisiae*. *Genetics*. 2003;163(2):545–555. doi:10.1093/genetics/163.2.545.
- Strope PK, Skelly DA, Kozmin SG, Mahadevan G, Stone EA, Magwene PM, Dietrich FS, McCusker JH. The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res*. 2015;25(5):762–774. doi:10.1101/gr.185538.114.
- Teunissen AW, Steensma HY. Review: the dominant flocculation genes of *Saccharomyces cerevisiae* constitute a new subtelomeric gene family. *Yeast*. 1995;11(11):1001–1013. doi:10.1002/yea.320111102.
- Tiago FCP, Martins FS, Rosa CA, Nardi RMD, Cara DC, Nicoli JR. Physiological characterization of non-*Saccharomyces* yeasts from agro-industrial and environmental origins with possible probiotic function. *World J Microbiol Biotechnol*. 2009;25(4):657–666. doi:10.1007/s11274-008-9934-9.
- Vergara-Alvarez I, Quiroz-Figueroa F, Tamayo-Ordóñez MC, Oliva-Hernández AA, Larralde-Corona CP, Narvaez-Zapata JA. Flocculation and expression of flo genes of a *Saccharomyces cerevisiae* mezcal strain with high stress tolerance. *Food Technol Biotechnol*. 2019;57(4):544–553. doi:10.17113/ftb.57.04.19.6063.
- Vijayalaxmi S, Anu Appaiah KA, Jayalakshmi SK, Mulimani VH, Sreeramulu K. Production of bioethanol from fermented sugars of sugarcane bagasse produced by lignocellulolytic enzymes of *Exiguobacterium* sp. Vsg-1. *Appl Biochem Biotechnol*. 2013;171(1):246–260. doi:10.1007/s12010-013-0366-0.
- Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, Zhang G, Gu YQ, Coleman-Derr D, Xia Q, et al. Orthovenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res*. 2019;47(W1):W52–W58. doi:10.1093/nar/gkz333.
- Yang L, Zheng C, Chen Y, Ying H. Flo genes family and transcription factor mig1 regulate *Saccharomyces cerevisiae* biofilm formation during immobilized fermentation. *Front Microbiol*. 2018;9:1860. doi:10.3389/fmicb.2018.01860.
- Yun CW, Ferea T, Rashford J, Ardon O, Brown PO, Botstein D, Kaplan J, Philpott CC. Desferrioxamine-mediated iron uptake in *Saccharomyces cerevisiae*. Evidence for two pathways of iron uptake. *J Biol Chem*. 2000a;275(14):10709–10715. doi:10.1074/jbc.275.14.10709.
- Yun CW, Tiedeman JS, Moore RE, Philpott CC. Siderophore-iron uptake in *Saccharomyces cerevisiae*. Identification of ferrichrome and fusarinine transporters. *J Biol Chem*. 2000b;275(21):16354–16359. doi:10.1074/jbc.M001456200.

Editor: N. Whiteman