Received: 5 December 2018 | Revised: 20 March 2019 | Accepted: 21 March 2019

DOI: 10.1002/yea.3391

RESEARCH ARTICLE

WILEY Yeast 2019

Evolutionary journey and characterisation of a novel pan-gene associated with beer strains of *Saccharomyces cerevisiae*

Nikolaos Vakirlis[1] | Chandre Monerawela[1] | Gavin McManus[2] | Orquidea Ribeiro[1] | Aoife McLysaght[1] | Tharappel James[1] | Ursula Bond[1]

[1] School of Genetics and Microbiology, Trinity College Dublin, College Green, Dublin 2, Ireland

[2] School of Biochemistry and Immunology, Trinity College Dublin, College Green, Dublin 2, Ireland

Correspondence
Ursula Bond, School of Genetics and Microbiology, Trinity College Dublin, College Green, Dublin 2, Ireland.
Email: ubond@tcd.ie

Funding information
European Commission, Marie Skłodowska-Curie Innovative Training Network award, Grant/Award Number: 764364; Trinity College Dublin, Grant/Award Number: 1592 award

Abstract

The sequencing of over a thousand *Saccharomyces cerevisiae* genomes revealed a complex pangenome. Over one third of the discovered genes are not present in the *S. cerevisiae* core genome but instead are often restricted to a subset of yeast isolates and thus may be important for adaptation to specific environmental niches. We refer to these genes as "pan-genes," being part of the pangenome but not the core genome.

Here, we describe the evolutionary journey and characterisation of a novel pan-gene, originally named hypothetical (*HYPO*) open-reading frame. Phylogenetic analysis reveals that *HYPO* has been predominantly retained in *S. cerevisiae* strains associated with brewing but has been repeatedly lost in most other fungal species during evolution. There is also evidence that *HYPO* was horizontally transferred at least once, from *S. cerevisiae* to *Saccharomyces paradoxus*. The phylogenetic analysis of *HYPO* exemplifies the complexity and intricacy of evolutionary trajectories of genes within the *S. cerevisiae* pangenome.

To examine possible functions for Hypo, we overexpressed a *HYPO-GFP* fusion protein in both *S. cerevisiae* and *Saccharomyces pastorianus*. The protein localised to the plasma membrane where it accumulated initially in distinct foci. Time-lapse fluorescent imaging revealed that when cells are grown in wort, Hypo-gfp fluorescence spreads throughout the membrane during cell growth. The overexpression of Hypo-gfp in *S. cerevisiae* or *S. pastorianus* strains did not significantly alter cell growth in medium-containing glucose, maltose, maltotriose, or wort at different concentrations.

KEYWORDS

beer yeasts, evolution, gene loss, horizontal gene transfer, pan-gene, taxonomically-restricted gene

1 | INTRODUCTION

Saccharomyces cerevisiae is the workhorse of many economically important industries, being used, for example, in the production of wine, beer, and bread (Pretorius & Boeke, 2018). The first genome sequence for *S. cerevisiae* was obtained in 1996 (Goffeau et al., 1996), and subsequent analysis predicted the presence of 6,604 potential open-reading frames (ORFs) of which 79% have now been verified, 11% remain uncharacterised, and 10% are considered dubious (https://www.yeastgenome.org). Despite the immense achievements in gene annotation over the past two decades, more than 1,800 ORFs have yet to be assigned to a biological process, and furthermore, the molecular function of more than 2,600 ORFs remains unknown (https://www.yeastgenome.org). *S. cerevisiae* isolates have

been found in different geographical, industrial, and clinical environments, and the genomes of over 1,000 *S. cerevisiae* strains have been sequenced with the aim of more fully understanding the relationships between the many strains and how this species evolved to adapt to different environmental niches (Bergström et al., 2014; Borneman et al., 2011; Gallone et al., 2016; Gonçalves, Rodrigues de Sousa, & Spencer-Martins, 2000; Peter et al., 2018; Strope et al., 2015; Zhu, Sherlock, & Petrov, 2016).

A common problem encountered in comparative genome analysis is the identification of genes within the newly sequenced genomes that do not map to the core genome, exemplified by the reference *S. cerevisiae* genome of strain *S288c* (https://www.yeastgenome.org). Such genes represent additional genetic material that forms part of the pangenome of the species and are here referred to as pan-genes. Pan-genes may arise from (a) introgressions, (b) horizontal gene transfer (HGT) from closely related species, (c) polymorphisms resulting in loss of start codons, gain of stop codons, or frameshifts within the ORF, or (d) de novo gene evolution, in one or the other strains being compared (Borneman, Forgan, Kolouchova, Fraser, & Schmidt, 2016; Marsit et al., 2017). Additionally, because pan-genes are generally located at the ends of chromosomes (Peter et al., 2018), the absence of such genes in a subset of strains may arise due to difficulty in assembling subtelomeric regions due to the presence of repetitive and non-unique sequences or due to loss of genetic material (Borneman et al., 2011). A recent comparative genome analysis of over a thousand *S. cerevisiae* strains indicates that the pangenome consists of 7,796 ORFs of which 4,940 are core ORFs, found in all strains, with the remaining 2,856 ORFs (pan-genes) variable between strains (Peter et al., 2018). Almost 50% of the variable pan-genes were accounted for by introgressions from *Saccharomyces paradoxus* or by HGT from *Torulaspora* and *Zygosaccharomyces* species. A significant proportion of the pangenome however is of unknown origin. De novo emergence of ORFs from previously nongenic sequences is a plausible evolutionary explanation for the origin of such ORFs (McLysaght & Guerzoni, 2015; Schlotterer, 2015). The prevalence of de novo gene emergence is well documented at the interspecific level in yeasts (Vakirlis et al., 2018), although relatively little is known about the intraspecific dynamics. Other evolutionary forces could have also played a role in shaping the *S. cerevisiae* pangenome; for example, completely novel ORFs can arise as a result of fast sequence divergence coupled with recurrent losses in multiple lineages (Tautz & Domazet-Loso, 2011).

Up to 40% of the variable pan-genes encode proteins with enrichments in gene ontology terms associated with biological processes such as cell–cell interactions, secondary metabolism, and stress responses, leaving a large portion of the variable pan-genes as uncharacterised (Peter et al., 2018). The analysis of the uncharacterised portion of the pangenome will enhance our understanding of the evolutionary relationship between strains of *S. cerevisiae* and may uncover novel gene functions related to specific environmental niches.

Here, we set out to examine the origins and functions of one such pan-gene, currently dubbed *HYPO*, as it is a hypothetical ORF with the potential to encode for a protein of 41 kDa, to which no biological or molecular function as been assigned. The ORF was originally identified in the genome of *Saccharomyces pastorianus*, a hybrid of *S. cerevisiae* and *Saccharomyces eubayanus*, which is used in the production of lager-type beers (Monerawela & Bond, 2018; Monerawela, James, Wolfe, & Bond, 2015; Nakao et al., 2009). An initial phylogenetic analysis revealed that the gene was also present in several *S. cerevisiae* strains used in ale and stout production and in a few environmental isolates but was absent from most laboratory and wine strains (Monerawela & Bond, 2017a; Monerawela et al., 2015). Furthermore, *HYPO*-like sequences were identified in only a few other *Saccharomyces stricto sensu* species; however, homologues were not found in any other prokaryotic or eukaryotic genomes available at that time. Thus, *HYPO* is an example of a pan-gene that appears to be restricted to a very narrow range of *Saccharomyces* species and which may confer specific phenotypes to yeasts growing in restricted niches.

With the expanding availability of genomes for *Saccharomyces* sp. and specifically of *S. cerevisiae*, we conducted a wider phylogenetic analysis to trace the evolutionary journey of *HYPO*. Additionally, we analysed potential functions for the encoded protein. We show that the gene is preferentially retained in yeast strains associated with brewing but has been extensively independently lost in multiple other lineages during evolution. The protein encoded by *HYPO* locates to the plasma membrane, where it accumulates in distinct foci. Time-lapse fluorescent cell imaging showed that Hypo fluorescence spreads across the membrane when cells are grown in wort.

2 | MATERIALS AND METHODS

2.1 | Yeast strains and culturing

S. cerevisiae strain S-150B (*Mat a*, *ura* 3-52, *leu* 2-112, *his* 3-52, and *trp* 1-1) was originally obtained from Dr. Jean Beggs, Edinburgh University). The *S. pastorianus* strain CMBS-33 is an allotetraploid lager yeast (Group II) and was obtained from the Belgian Culture Collection (James, Usher, Campbell, & Bond, 2008). Yeast strains were cultivated in YEP (10 g/L yeast extract and 20 g/L bacto peptone) or YNB medium (1.7 g/L yeast nitrogen base, 5.0 g/L ammonium sulfate, and 0.59 g/L synthetic drop-out medium without leucine and uracil). The YEP and YNB media were supplemented with glucose, maltose, maltotriose, or wort at different concentrations. Spray-dried wort (Browland, Belgium) containing a mixture of maltose, glucose, maltotriose, sucrose, and fructose in an approximate ratio of 6:2:1:0.2:0.2, respectively, was prepared as a 40% solution. The specific gravity of the wort was measured using a refractometer. A final concentration of 12°Brix corresponds to a specific gravity of 1.049. For growth of *S. cerevisiae* strains, containing the pGREG600 vectors in YNB medium, 0.01 g/L leucine was also added, whereas for strains containing the pRSH vector, both uracil and leucine (0.01 g/L) and 200 μg/mL hygromycin were added. The cultures were incubated at 30°C, unless otherwise specified. Growth curves were carried out in

a volume of 200 µl in 96-well plates, and the optical density was measured using a Thermo Fisher Multiskan Ascent spectrophotomoter.

3 | PHYLOGENETIC ANALYSIS

3.1 | Data collection

S. cerevisiae genomes (485 as of October 2017) and all available bacteria and fungi genomes were downloaded from GenBank (https://www.ncbi.nlm.nih.gov/genbank/). Genomes assembled from PacBio reads from Yue et al. (2017) were downloaded from https://yjx1217.github.io/Yeast_PacBio_2016/data/. The 1,011 genome data (Peter et al., 2018) were downloaded from http://1002genomes.u-strasbg.fr/files/. Information about the niche and isolate source of each *S. cerevisiae* strain was compiled from the supplementary tables of sequencing project studies (Almeida et al., 2015; Gallone et al., 2016; Strope et al., 2015; Wang, Liu, Liti, Wang, & Bai, 2012) and from the "isolate source" field in National Center for Biotechnology Information (NCBI) species information. Niche designations were simplified into nine groups for Figure 2.

3.2 | Similarity searches

Similarity searches were performed by BLAST (Altschul et al., 1997). TBLASTN with default settings and an Evalue threshold of 0.001 was used to look for homologues of *HYPO* and *YPL257W* in *S. cerevisiae*, fungi, and bacterial genomes. BLASTP with the same Evalue threshold was used to look for similar proteins in NCBI's NR database UniProt (https://www.uniprot.org/) and MycoCosm fungal database (https://genome.jgi.doe.gov/programs/fungi/index.jsf). Similar protein sequences of *HYPO* were also searched using phmmer of HMMER online (HMMER.org; Potter et al., 2018) against the reference proteomes.

3.2.1 | *HYPO* strains tree

A multiple alignment of 259 *HYPO* nucleotide sequences from strains where the coding sequence (CDS) is intact or disrupted and the *HYPO* homologues from *S. paradoxus* (strains YPS138 and UFRJ50816) and *Saccharomyces mikatae* (strain IFO 1815) was generated with the multiple alignment software MAFFT (https://mafft.cbrc.jp/). Positions with more than 5% gaps were removed. A phylogenetic tree was reconstructed using PhyML (Guindon, Delsuc, Dufayard, & Gascuel, 2009) with the following parameters: "-d nt -m HKY85 -v e -o tlr -c 4 -a e -b 100 -f e."

3.2.2 | *HYPO* outgroup tree

A multiple alignment consisting of *HYPO* nucleotide sequences from six representative *S. cerevisiae* strains and 22 outgroup species was generated with MAFFT (https://mafft.cbrc.jp/). A phylogenetic tree was reconstructed with PhyML with the following parameters: "-d nt -m HKY85 -v e -o tlr -c 4 -a e -b 100 -f e."

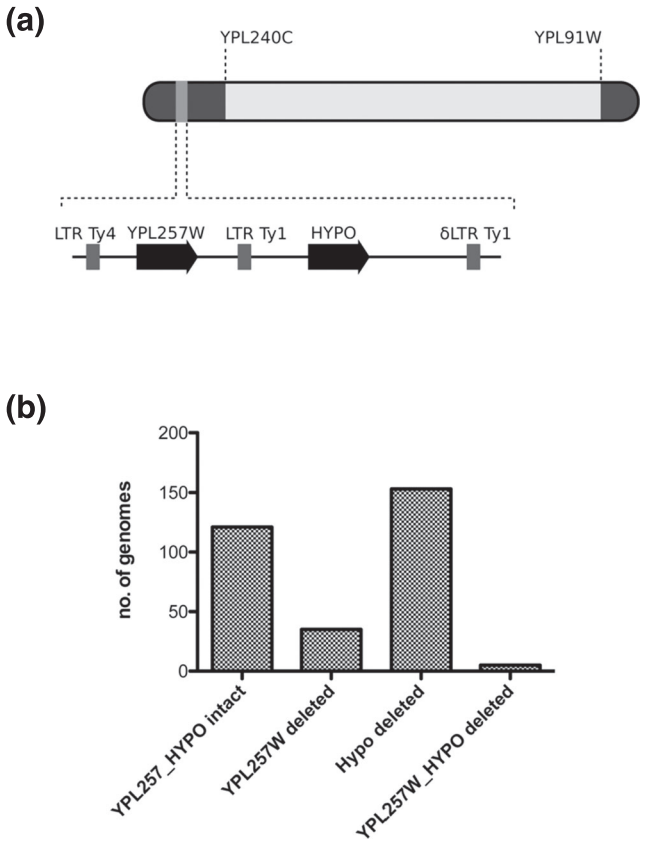

FIGURE 1 Location of the *HYPO* open-reading frame on *Saccharomyces pastorianus* chromosome XVI. (a) *HYPO* is located at the left subtelomeric region of the hybrid chromosome XVI in *S. pastorianus*, composed of *Saccharomyces cerevisiae* DNA (black) and *Saccharomyces eubayanus* DNA (grey). The hypothetical open-reading frame is located next to YPL257W and flanked by long terminal repeats of Ty1 elements. (b) Frequency of loss of *HYPO* and *YPL257W* in *S. cerevisiae* strains. The number of genomes in which *HYPO* and *YPL257W* are retained intact, lost individually or in combination

3.3 | *HYPO* ORF characterisation

The G + C percentage content of the *HYPO* ORF was calculated with a custom Python script. Transmembrane domains (TMDs) were predicted with Phobius (http://phobius.sbc.su.se/) and TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) using normal prediction parameters for Phobius and the extensive format for TMHMM. The membrane topology for *HYPO* was examined in TOPCONS (www.topcons.cbr.su.se/), which generates a consensus topology by comparing topologies generated independently by different methods (Bernsel, Viklund, Hennerdal, & Elofsson, 2009) and TMpred (www.embnet.vital-it.ch).

HMMER online (www.HMMER.org) was also used to search for protein domain matches to the protein family Pfam database using

hmmscan. ScanProsite was used to scan the hypo protein sequence for motifs using default parameters. Alignments were performed with MAFFT (Katoh & Standley, 2013). Codon-aware alignments for determining the ratio of non-synonymous and synonymous substitutions (d_N/d_S) were performed with MACSE software (Ranwez, Harispe, Delsuc, & Douzery, 2011) with default settings. d_N/d_S ratios were calculated using *yn00* from PAML (Yang, 2007). Nucleotide diversity and Tajima's *D* were calculated with MEGA7 (Kumar, Stecher, & Tamura, 2016). Gap removal and sequence identity calculation from pairwise nucleotide sequence alignments were performed using custom Python scripts.

3.4 | *HYPO* cloning

The *HYPO* gene sequences were obtained from the *S. pastorianus* Group II Weihenstephan WS34/70 genome (accession number ABPO01000069.1, coordinates 26652–25534; Nakao et al., 2009), and the DNA sequence was amplified from *S. pastorianus* strain CMBS-33 genomic DNA as a single fragment using primers Hypo1_F and Hypo1_R (Table S1). The amplified fragment was cloned into the vector pGREG600 (*URA3*, *pGAL1*, *HIS3*, *GFP*, *tCYC1*, *pAgTEF-kanMX-tAgTEF*, *CEN6*, *ARSH4*, *AmpR*, *ori*) upstream of the pre-existing *GFP* ORF (Jansen, Wu, Schade, Thomas, & Whiteway, 2005). Briefly, the vector was digested with SalI, which removes the *HIS3* gene. The *HYPO* gene fragment was inserted into the SalI-digested pGREG600 vector via homologous recombination in vivo in *S. cerevisiae* S-150B as previously described (Kricka, James, Fitzpatrick, & Bond, 2015). The amplified DNA contained a 5′ end extension homologous to a site 35-bp downstream from the *GAL* promoter whereas the 3′ end of the amplified *HYPO* fragment contained a 3′ end extension (35 bp) homologous with the start of the *GFP* ORF (Table S1), thus generating a *HYPO-GFP* fusion under the control of a GAL promoter.

A gene cassette consisting of a *PGK* promoter, a *HYPO-GFP* fusion gene, and a *CYC1* terminator was amplified in three overlapping DNA fragments. Fragment 1, consisting of the *PGK* promoter, was amplified from pGREG586-bgl1 (Fitzpatrick, Kricka, James, & Bond, 2014) using the primers PGK1_F and PGK1_R (Table S1). Fragment 2, consisting of the *HYPO* ORF minus the stop codon, was amplified from Group II lager yeast CMBS-33 genomic DNA using the primers Hypo2_F and Hypo2_R (Table S1). Lastly, Fragment 3, consisting of the *GFP* ORF and the *CYC1* terminator, was amplified from pGREG600 using primers GFPCYC1_F and GFPCYC1_R (Table S1). The three PCR products contained 35-bp overlapping homologies to allow assembly into a contiguous DNA fragment, and Fragments 1 and 3 contained 35-bp extensions, at the 5′ and 3′ ends, respectively, homologous to the vector, allowing for insertion into the high copy number vector pRSH42H at the Sal1 site via homologous recombination in vivo in *S. cerevisiae* as previously described (Kricka et al., 2015). The pRS42H vector (Taxis & Knop, 2006) contains a hygromycin resistance gene cassette (hphNT1) for selection. The *PGK* promoter was replaced by the native *HYPO* promoter DNA sequences (967 bp), which were amplified from *S. pastorianus* CMBS-33 genomic DNA using the primers HYPOprom_F and HYPOprom_R (Table S1) by homologous recombination in vivo in *S. cerevisiae* as described above.

A *GFP* gene, expressed from the *PGK* promoter, was also cloned into the pRS42H vector as two overlapping PCR products. Briefly, Fragment 1, consisting of DNA sequences for the *PGK* promoter, was amplified using the primers PGK1_F and PGK_GFP_R (Table S1). Fragment 2, consisting of the *GFP* ORF and the *CYC1* terminator, was amplified using the primers GFP2_F and GFPCYC1_R (Table S1). The 5′ and 3′ ends of Fragment 1 contained homology to the pRS42H vector and Fragment 2, respectively, and the 5′ and 3′ ends of Fragment 2 were homologous to Fragment 1 and the pRS42H vector, respectively. The *PGK-GFP-CYC* gene cassette was cloned into SalI-digested pRS42H via homologous recombination in vivo in the *S. cerevisiae* strain S-150B.

The pRS42H plasmids containing the *PGK-HYPO-GFP* and *PGK-GFP* gene cassettes were subsequently introduced into the Group II *S. pastorianus* strain CMBS-33 as previously described (Kricka et al., 2015).

3.5 | Fluorescence imaging

Yeast strains expressing the *HYPO-GFP* fusions gene or *GFP* alone were cultured in YEPD (20 g/L) at 30°C or wort (12°Brix) at 13°C for 16 hr. A 1 ml of the overnight cultures was washed twice with dH_2O, then resuspended in 100 µl of 4% formaldehyde to fix the cells. Cells were vortexed and left at room temperature for 15 min. The cells were centrifuged, and the supernatant was removed. Cells were washed in 1 ml of KPO_4/sorbitol (0.1M KPO_4 and 1.2M sorbitol) and resuspended in 500 µl of Mowiol (Sigma-Aldrich.com) and thoroughly mixed. The resuspended cells were placed under a coverslip on a glass slide and stored at 4°C overnight. The samples were viewed using a Nikon eclipse E400 microscope and Nikon digital camera DXM1200 using light microscopy and a fluorescein isothiocyanate (FITC) filter (excitation of 490 nm and emission of 525 nm). Images were taken using a 100 x objective magnification.

Confocal microscopy imaging was performed at the Microscopy and Imaging Facility within the School of Biochemistry and Immunology at Trinity College Dublin. *S. cerevisiae* cells grown overnight in wort (12°Brix) at 13°C were fixed with formaldehyde as described above and were imaged using a Leica SP8-gated STED confocal microscope.

3.5.1 | Time-lapse imaging

S. cerevisiae cells containing *HYPO-GFP* were grown overnight in YEPD, washed, and resuspended in 1 ml of dH_2O. Cells (15 µl) were placed in a chamber slide and overlaid with a preformed slab of wort agar (12°Brix). The agar was overlaid with mineral oil to create a microaerophilic environment, and images were taken every 20 min for 8 hr with an Olympus FV1000 point-scanning confocal microscope.

4 | RESULTS

4.1 | The structure and chromosome location of *HYPO* in *S. pastorianus*

The *HYPO* gene was originally identified in the *S. pastorianus* strain Weihenstephan 34/70 (WS34/70), a hybrid yeast arising from multiple hybridisation events between *S. cerevisiae* isolates and *S. eubayanus* (Monerawela & Bond, 2017b; Nakao et al., 2009). The gene is located at the left subtelomeric region of a hybrid version of chromosome XVI that contains regions of *S. cerevisiae* and *S. eubayanus* and which arose as a result of recombination between the homologous chromosomes of the parental species (Figure 1a). *HYPO* maps to the *S. cerevisiae* portion of chromosome XVI and is located adjacent to *YPL257W*, an uncharacterized ORF that is also present in the *S. cerevisiae* reference genome (Figure 1a). The *HYPO* ORF is 1,119 nt in length and has a GC content of 44.1% compared with the *S. cerevisiae* genome-wide average for CDSs of 39.9%. The Codon Adaptation Index for *HYPO* is 0.115, compared with the *S. cerevisiae* genome-wide average for ORFs of 0.18. The region contains several LTRs, which flank both *HYPO* and the adjacent gene *YPL257W*. Interestingly, the gene *YPL257W* in *S. pastorianus* shares only 84.97% sequence identity with the *S. cerevisiae* reference genome *S288C* whereas the *HYPO* gene is not present in the reference genome.

4.2 | *HYPO* is predominantly retained in beer strains of *S. cerevisiae*

To determine the distribution of *HYPO* within *S. cerevisiae* strains, we searched for *HYPO* within the *S. cerevisiae* genomes available in NCBI's GenBank, using TBLASTN, and found a match in 260 out of a total of 485 genomes. To confirm that this proportion is realistic, the search was repeated in the most recent genome resource of more than a thousand *S. cerevisiae* strains (Peter et al., 2018), which returned a match for *HYPO* in 500/1,011 genomes. Thus, *HYPO*-like sequences are found in approximately 50% of the currently sequenced *S. cerevisiae* strains. As in *S. pastorianus*, *HYPO* lies adjacent to *YPL257W* in *S. cerevisiae* strains. In strains lacking *HYPO*, the length of the deleted genomic region appears to be constant with a few recurrent variations, mostly coinciding with the positions of the

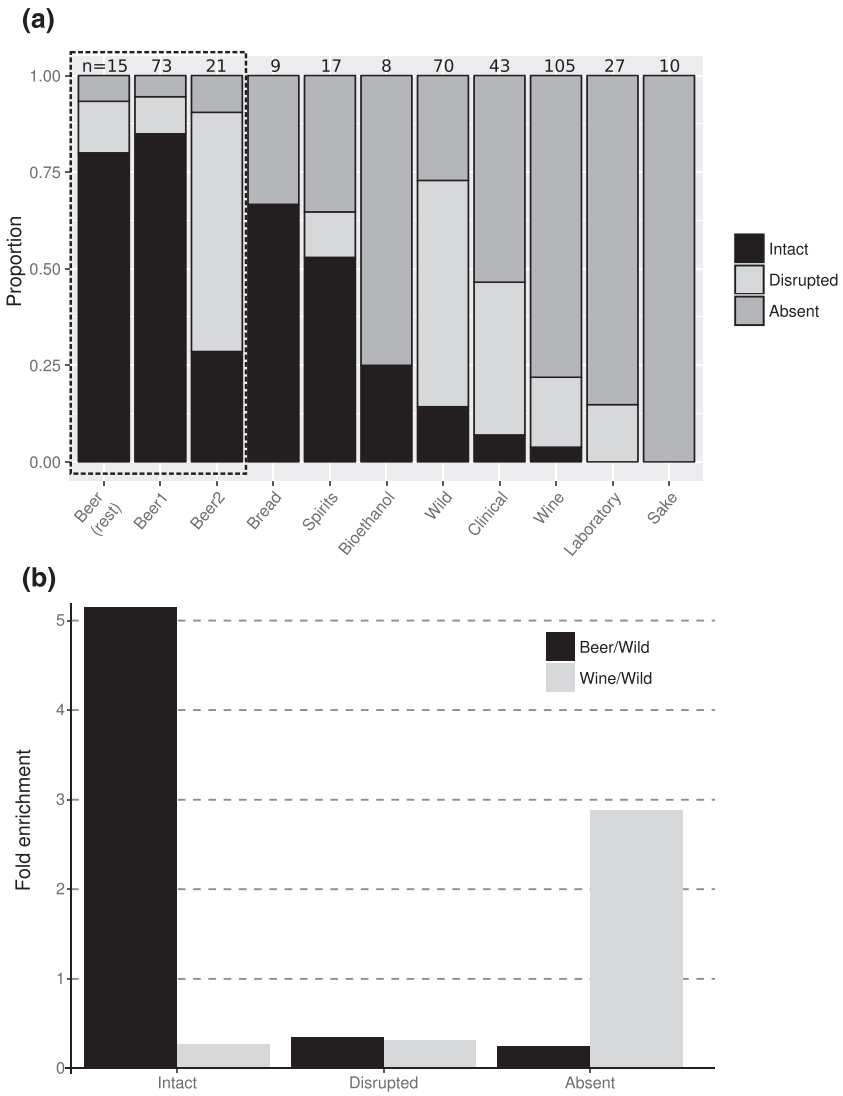

FIGURE 2 *HYPO* is specifically retained in beer *Saccharomyces cerevisiae* strains. (a) The relative proportion of the presence of *HYPO* in different classifications of *S. cerevisiae* strains is indicated. Intact open-reading frame (ORF): black, disrupted ORF; light grey, absent ORF; dark grey. The number of isolates for each *S. cerevisiae* class is shown above the columns. (b) Fold enrichment of *HYPO* in beer and wine strains. The proportion of beer and wine strains containing an intact or disrupted *HYPO ORF* or lacking *HYPO* relative to wild yeast isolates is shown

various LTRs that flank *HYPO* and its neighboring gene *YPL257W* (Figure 1b). This confirms that the region around *HYPO* is extremely dynamic, as has often been observed for near-subtelomeric regions.

In the 260 *S. cerevisiae* genomes where *HYPO* was examined in detail, the ORF is (a) intact in 132, (b) disrupted by a premature stop codon in 100, (c) frameshifted in 26, and (d) truncated in two (Figure 2a). In the latter category, the TBLASTN search returned a partial match to *HYPO* in which the C-terminal end of the protein was missing. Thus, only 27% of the 485 *S. cerevisiae* strains examined have the potential to encode for a functional hypo protein. There is an enrichment for the presence of intact *Hypo* among beer strains, compared with presence-absence in wild strains (chi-squared test, *p* value = 3.9e-14; Figure 2b). Among the beer strains, *HYPO* is predominantly found in strains from the Beer1 group, where the ORF is intact in 62/73 strains, compared with Beer2 group where only six of 21 strains contain an intact ORF (Figure 2a; Gallone et al., 2016). The presence of an intact *HYPO* ORF in strains also clusters geographically with almost all such strains originating from Britain, Belgium/Germany, and the United States (69/72) and not with beer strains from unknown locations (11/37). Furthermore, 34/35 *S. cerevisiae* strains categorised as containing mixed or mosaic genomes contain intact *HYPO* (Table S2). Intact *HYPO* is rarely encountered in wine strains (Figure 2a) where the vast majority of strains either lack *HYPO* or contain a disrupted ORF (chi-squared test, *p* value = 6.7e-11 for absence of *HYPO* in wine yeasts relative to wild yeasts; Figure 2b). Mapping the state of the *HYPO* ORF onto the *S. cerevisiae* strain phylogeny tree (Gallone et al., 2016) reveals that *HYPO* has been lost (by either gene deletion or inactivation) in multiple lineages of *S. cerevisiae* suggestive of multiple independent events (Figure S1).

HYPO sequences display low nucleotide diversity (average π = 0.001), which is within the normal range of diversity between *S. cerevisiae* strains. Analysis of the nucleotide diversity of *HYPO* reveals that the ORF appears to be evolving nonneutrally (Tajima's *D* score = −2 including both intact and disrupted *HYPO* sequences in the alignment and *D* = −1.57 for intact *HYPO* sequences alone). Furthermore, the average pairwise d_N/d_S ratios of 0.526 (intact + disrupted), 0.44 (intact), and 0.54 (disrupted) are indicative of purifying selection. It is therefore very likely that *HYPO* has been under negative selection and thus functional until very recently in the lineages where it has been lost.

4.3 | Origin and evolution of *HYPO*

We previously identified a few *HYPO*-like sequences in several *S. stricto sensu* yeast species but none outside of this classification (Monerawela et al., 2015). We reevaluated this finding in light of the expanding availability of yeast genome sequences. An exhaustive similarity search (see methods) identified sequences similar to *HYPO* in only 25 fungal species outside of the *Saccharomyces* genus, all belonging to the yeast genera *Pezizomycotina* and *Saccharomycotina* (Figure 3), the two major divisions of the fungal group *Ascomycota*. There was also a 100% identity match to less than half of the query sequence to a sequence in the *Entomophthora muscae* genome, a pathogenic fungus that infects many Diptera species. This is likely due to contamination rather than true homology.

Codon-aware alignment and d_N/d_S calculation suggest that the sequences have been evolving under purifying selection (mean pairwise d_N/d_S = 0.24, see methods). However, because the sequences have accumulated many mutations (the d_S is >3 in the majority of pairwise comparisons), this result should be taken with caution. The extremely sparse yet polyphyletic presence of the gene in yeasts raises the question of its evolution, and so, we looked for evidence of HGT.

We reconstructed a phylogenetic tree based on the multiple alignment of the *HYPO* sequence from the 25 fungal species containing *HYPO* sequences and some representative sequences of the *Saccharomyces* genus (Figure 3a). The tree topology does not suggest HGT between genera, and the two main phylogenetic groups are clearly separated, with the exception of *Kluyveromyces wickerhamii* and *Pichia membranifaciens* (Figure 3a). These two species however represent very deep branches within the *Saccharomycotina*, and thus, their phylogenetic placement can be problematic. As differences in G + C content can reflect horizontal acquisition, we looked at the G + C content of the *HYPO* homologues in the different species, compared with the rest of the CDS in the genomes (Figure 3b). We observed that the *HYPO* gene sequence is not significantly different from the genome average (paired *t* test, *p* value = 0.1). We therefore suggest that the peculiar phyletic pattern of presence-absence of *HYPO* is not due to horizontal "hopping" across species but rather to extensive loss events in multiple lineages during evolution.

There is nevertheless evidence of *HYPO* being horizontally transmitted at least once, within the *Saccharomyces*. It was recently suggested that *HYPO* is among the many genes that were introgressed into the *S. cerevisiae* pangenome from its wild relative, *S. paradoxus* (Peter et al., 2018). Our data, however, indicate that the transfer occurred in the opposite direction. DNA sequences, with >98% identity to *S. cerevisiae HYPO*, were identified in only two strains of *S. paradoxus* (YPS138 and UFRJ50816), although the ORF is disrupted by a premature stop codon in both. The presence of *HYPO* in only these two strains was also confirmed using data from Yue et al. (2017). The two *S. paradoxus* matches form a small monophyletic group with a wild *S. cerevisiae* isolate from cherry tree gum (see Figures 3a and S2). They are thus not an out-group to the entire *S. cerevisiae* population, unlike, for example, the match in *S. mikatae* (Figure 3a). Alignment of the region of the match ±8 kb from one of the *S. paradoxus* strains containing *HYPO* to the cherry tree gum *S. cerevisiae* isolate clearly identifies the chromosome segment that was transferred/introgressed (Figure 3c). The region is approximately the same as the one that has been repeatedly lost at the intraspecific level in *S. cerevisiae* (Figure 1a,b). Aligning this the same region to a more distantly related *S. cerevisiae* strain reveals a slight increase in sequence divergence within the putatively transferred region (Figure 3c dashed line), thus providing supporting evidence for the HGT of this region from the *S. cerevisiae* strain YJM1399 to *S. paradoxus* and not vice versa. Taking these together, these findings

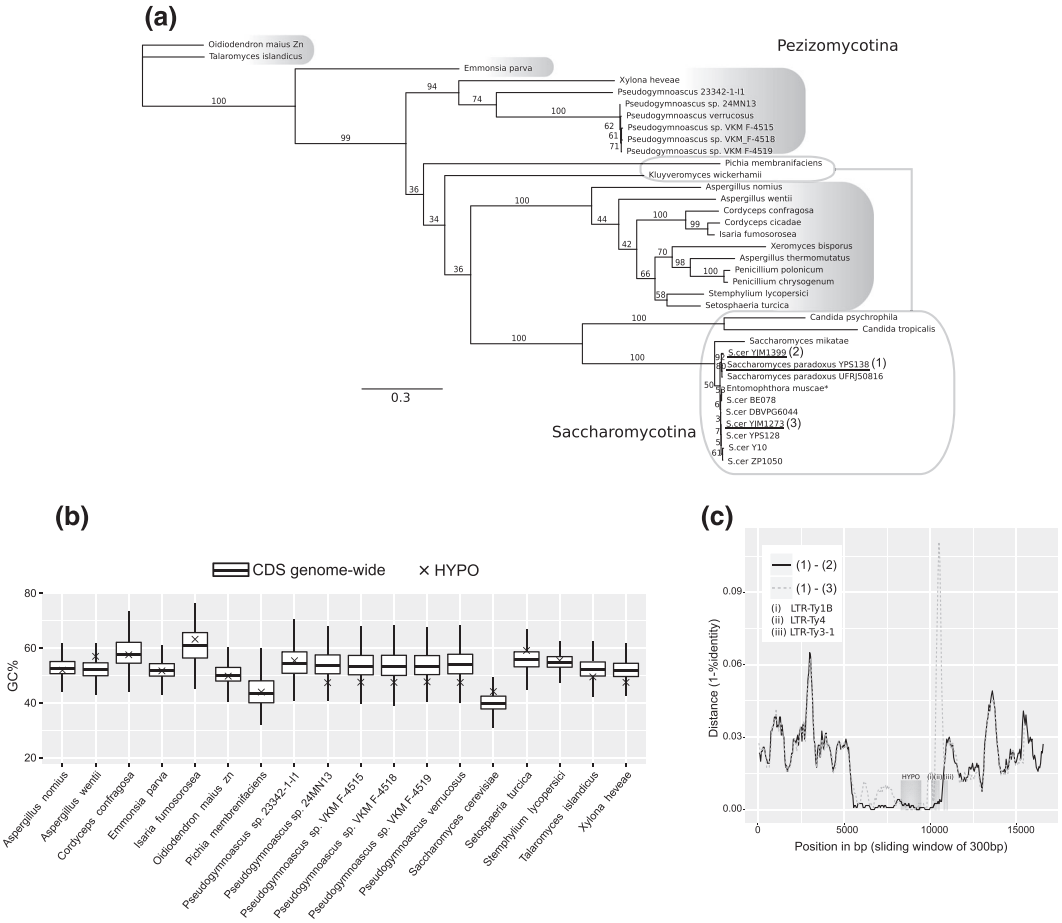

FIGURE 3 *HYPO* phylogeny suggests multiple losses and a recent horizontal gene transfer. (a) Phylogenetic tree of *HYPO* in seven *Saccharomyces cerevisiae* strains and all outgroup *Ascomycota* species containing the *HYPO* gene. Leaves are coloured according to their phylum. The *Entomophthora muscae* 100% identity match, likely due to contamination, is shown with an asterisk. Bootstrap values are shown on internal branches. (b) Box plots of GC% distributions of coding sequence in the outgroup species in which the *HYPO* open-reading frame is intact and GC% of *HYPO*. (c) Genetic distance between *Saccharomyces paradoxus* YPS138 (1) and two *S. cerevisiae* strains (YJM1399 [2] and Y10 [3]), in the ±8-kb region around *HYPO*. Identity was calculated over 300-bp windows with a step of 50 bp

show the complex evolutionary history of *HYPO*, which was initially lost in the ancestor of *S. paradoxus* strains and then regained horizontally from a wild *S. cerevisiae* strain.

4.4 | Functional analysis of *HYPO*: Cellular localisation

We had previously shown that the gene *HYPO* is expressed in *S. pastorianus* under fermentation conditions (Monerawela et al., 2015). To search for clues to possible functions for the encoded protein, we searched for functional protein domains or motifs within *HYPO*. The *HYPO* ORF potentially encodes for a protein of 41 kDa (373 aa) and is predicted to contain up to three TMDs, at amino acid positions 15–37, 94–111, and 132–152 (Figure 4a). Membrane topology prediction software (TOPCON) suggests a consensus topology in which only the first TMD spans the membrane, resulting in a topology whereby the N-terminus of the protein is cytoplasmic and the C-terminus is external to the cell (Figure 4a). Interestingly, in 17 of the 25 fungal species that possess a *HYPO* gene (Figure 3a), the ORF is intact and encodes a protein of similar length to *S. cerevisiae* (average of 384 aa). Furthermore, 12 of these ORFs encode two to three predicted TMDs in an arrangement similar to the *S. cerevisiae HYPO*, hinting towards a conserved protein structure (Figure 4b). A protein domain search identified matches to MASE1 (membrane-associated sensor) domains found in bacterial histidine kinases, diguanylate cyclases, and other bacterial signalling proteins. Although the matches were not deemed statistically significant, the fact that the matching regions overlap with the predicted TMDs raises the likelihood that the protein has a membrane-related function. Other than the identification of potential TMDs, the bioinformatics searches did not identify any protein homologues with characterised functions, and no further hints as to the function of the protein could be gleaned from this analysis.

We generated a *HYPO-GFP* fusion gene cassette using either the *HYPO* native promoter and 5′ UTR sequences from *S. pastorianus* or the constitutively expressed promoter from the phosphoglycerate kinase (*PGK*) gene and its 5′ UTR, to identify the cellular location

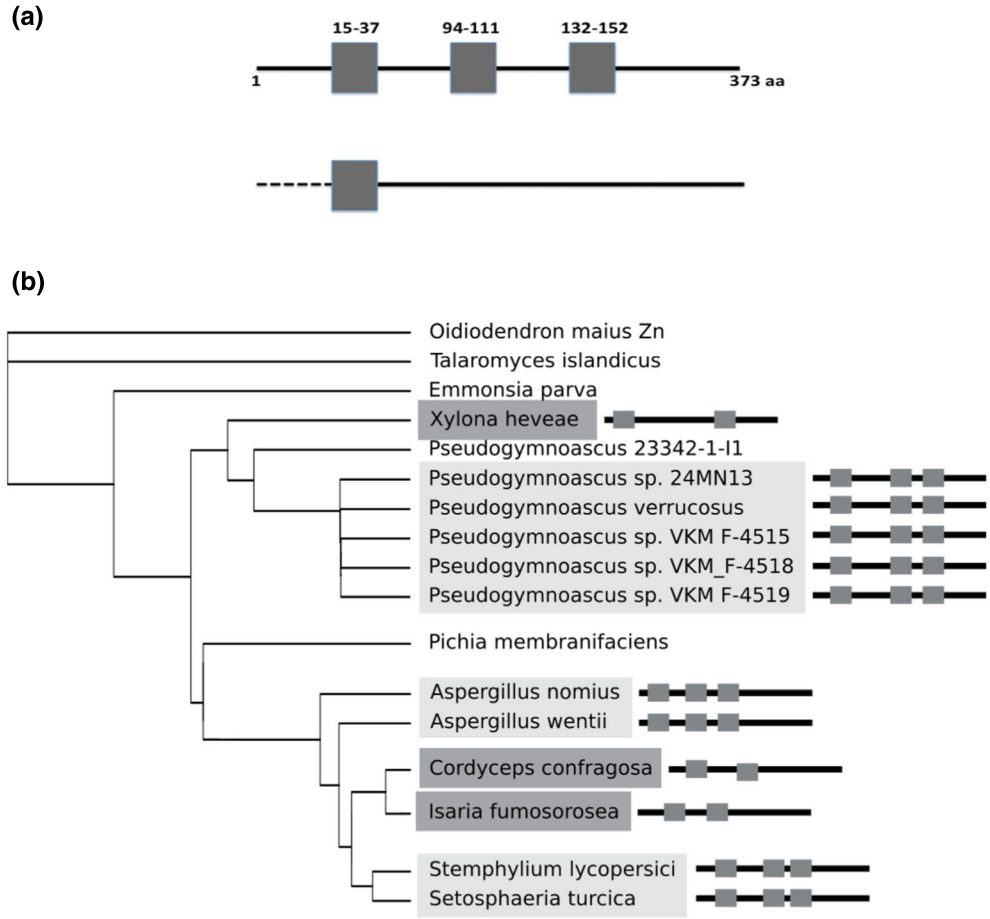

FIGURE 4 Predicted protein motifs in Hypo. (a) Protein motif and domain searches predict that Hypo contains three putative transmembrane domains (TMDs; grey boxes). The amino acid positions of the TMDs are shown. The topology of Hypo within the membrane was predicted using the programme TOPCONS. Dashed line, cytoplasmic; grey box, transmembrane domain; black line, extracellular. (b) The putative number and arrangement of the TMDs (grey) are shown to the right of fungal species that contain an intact *HYPO* open-reading frame

and possible functions for the encoded protein. The *HYPO-GFP* gene cassettes were introduced into the *S. pastorianus* strain CMBS-33, which already contained a native copy of *HYPO* under its own promoter and also into *S. cerevisiae* S-150B, which does not contain a native *HYPO* gene. When expressed from its native promoter on the introduced plasmid, very weak fluorescence was observed in both *S. pastorianus* and *S. cerevisiae* strains (data not shown), and a much higher level of fluorescence was detected when the gene was expressed from the stronger *PGK* promoter (Figure 5a); therefore, all subsequent experiments were conducted with the *PGK-HYPO-GFP* construct.

Fluorescence microscopy revealed that the fusion protein appeared as punctate patches at the cell periphery when *S. cerevisiae* cells were grown in YEPD (Figure 5a). When *S. cerevisiae* cells were grown in wort, the maltose-based medium used for beer fermentations, the intensity of the staining pattern for HYPO-GFP appeared to increase, and in addition to the punctate staining pattern, a diffuse staining was also apparent at the periphery of the cells (Figure 5b). Expression of GFP alone produced an even fluorescence distribution within the cell with clear nuclear exclusion (Figure 5c,d). The same fluorescence pattern for *HYPO-GFP* was also observed in the lager yeast strain CMBS-33 when grown in YEPD and wort (data not shown). Confocal fluorescence microscopy indicated that the HYPO-GFP appears to localise at the cell surface, most likely the cell membrane (Figure 5e). Time-lapse images of *S. cerevisiae* cells grown in wort reveal that Hypo-gfp first appears in the cell as a single punctate dot. As the cells grow, the dot grows in size and spreads out across the cell membrane to form the patchwork pattern observed by fluorescence microscopy (Supplemental material, time lapse.mov).

4.5 | Growth of *S. cerevisiae* and *S. pastorianus* strains overexpressing HYPO-GFP

To determine if the overexpression of Hypo-gfp affects the cell growth kinetics, *S. cerevisiae* strain S-150B and *S. pastorianus* strain

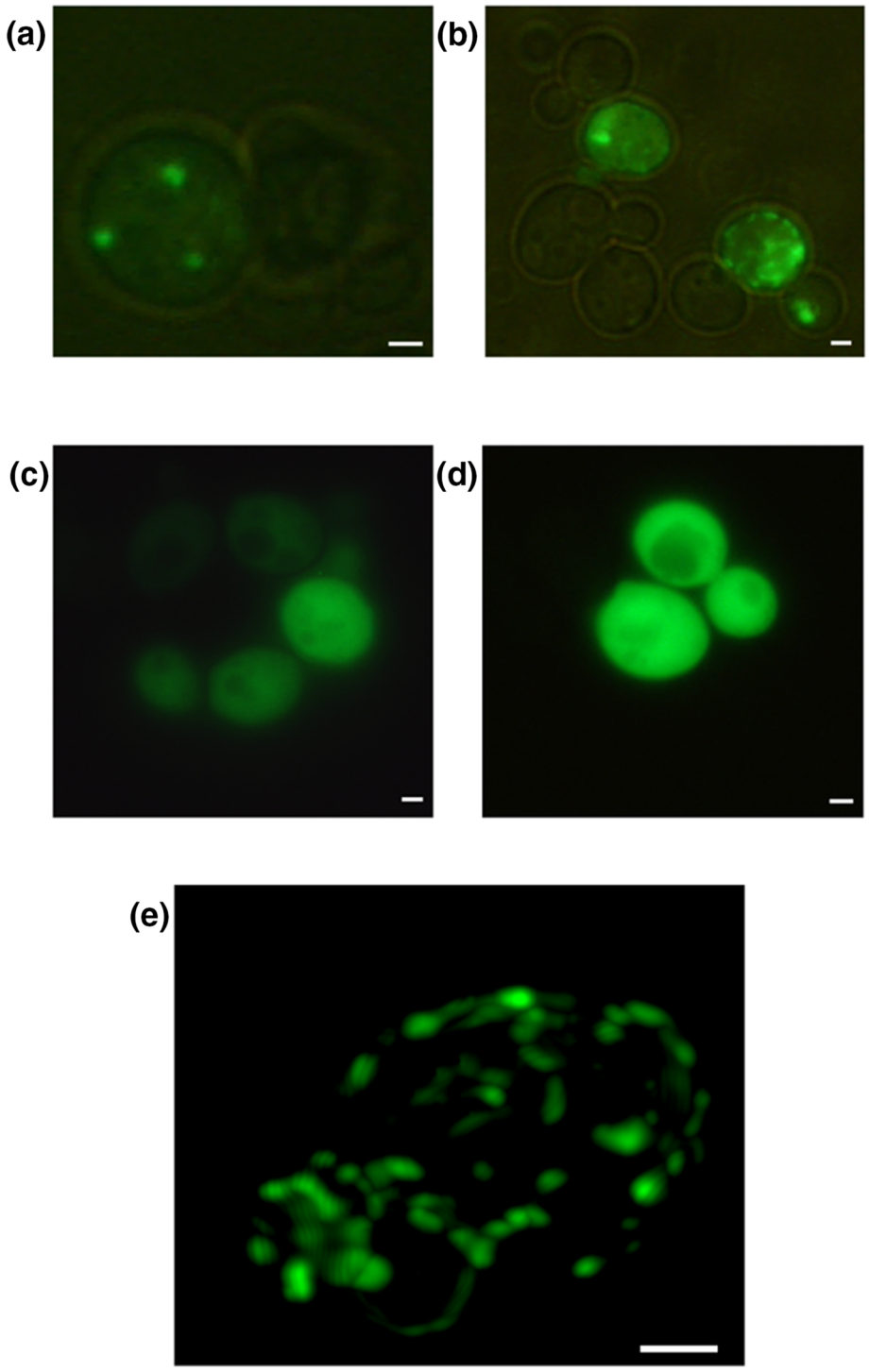

FIGURE 5 Cellular localisation of Hypo-gfp. *Saccharomyces cerevisiae* (S-150B) cells expressing the gene cassette (a,b) *PGK-HYPO-GFP* or (c,d) *PGK-GFP* were grown in (5a and 5c) YEPD at 30°C or (5b and 5d) 12°B wort at 13°C for 16 hr. Hypo-gfp was visualised using a Nikon eclipse E400 phase-contrast microscope using a fluorescein isothiocyanate (FITC) filter. (e) Confocal microscopic image of expression of Hypo-gfp. *S. cerevisiae* cells were grown in 12% wort, fixed with formaldehyde, and imaged using a Leica SP8-gated STED confocal microscope

CMBS-33 expressing *HYPO-GFP* from the *PGK* promoter were grown in YEP medium containing either glucose or maltose at low (2%) and high (10%) concentrations, and the growth of the strains was monitored over several days. As a control, growth of the strains expressing GFP alone was examined under the same growth conditions. The growth of the *S. cerevisiae* strain, which does not contain an endogenous copy of *HYPO*, was not significantly affected at either concentration of glucose (Figure 6a). The *S. cerevisiae* strain did not grow at either concentration of maltose, and the presence of *HYPO* did not significantly alter the growth pattern (Figure 6b).

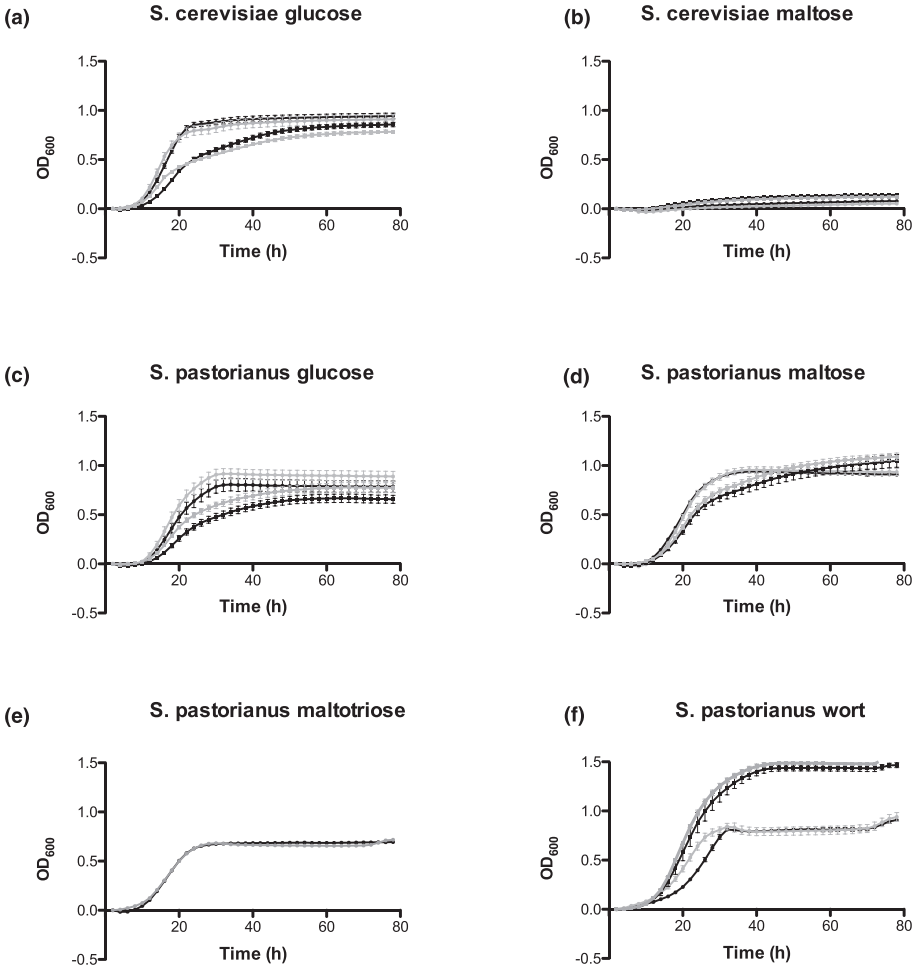

FIGURE 6 Growth of *Saccharomyces cerevisiae* and *Saccharomyces pastorianus* strains overexpressing HYPO-GFP. *S. cerevisiae*, S-150B (panels a, b), and *S. pastorianus*, CMBS-33 (panels c-f) expressing the *PGK-HYPO-GFP* or *PGK-GFP* gene cassettes were grown in YNB containing 2%; ● or 10%; ■ glucose (6a and 6c), 2%; ● or 10%; ■ maltose (6b and 6d), 1.3%; ● maltotriose (e) or in 2%; ● or 12%; ■ wort (f). *HYPO-GFP*; black lines, *GFP*; grey lines. The cultures were incubated at 30°C, and the optical density of the cultured measured at 1-hr intervals. Data represent the mean of triplicate biological samples, and the error bars show the standard error of the mean

The overexpression of *HYPO-GFP* in *S. pastorianus*, which does contain an endogenous copy of *HYPO*, did not significantly affect the growth of cells in glucose or maltose at either concentration; however, growth in 10M glucose and 10M maltose was slightly impeded (Figure 6c,d). This slight growth inhibition at higher concentrations of sugars was observed in several independent experiments (data not shown). Similarly, we also did not observe any significant growth differences in cells overexpressing Hypo-gfp in *S. pastorianus* grown in maltotriose (Figure 6e) in nor 2% or 12% wort (Figure 6f), although the slight impedance of growth observed in high glucose and maltose was also evident in 12% wort and at the early phase of growth in 2% wort.

5 | DISCUSSION

5.1 | Evolutionary journey of *HYPO*

The recent discovery of the expanded repertoire of ORFs within the pangenome of *S. cerevisiae* presents both challenges and opportunities to understanding the evolution and the unique physiological properties of yeast isolates. Phylogenetic and comparative genome analyses of over a thousand isolates of *S. cerevisiae* paint a picture of genetic acquisition from introgressions, HGT, and admixture leading to distinct lineages often associated with specific environmental and/or geographic location (Borneman et al., 2016; Gallone et al., 2016; Gonçalves et al., 2016; Peter et al., 2018). Thus, wine yeasts can clearly be distinguished from beer yeasts, with each lineage having acquired distinct gene alleles, allowing for adaptation to that specific environment. Our current usage of the *S. cerevisiae* reference genome, *S288C*, which lacks up to 36% of the total pangenome, places limits on our understanding of the biology of *S. cerevisiae*.

Here, we describe an approach to characterising the biological role of pan-genes for which few or no homologues can be identified in other species and for which protein domain or motif analysis reveals no clue as to the function of the gene product. Using both phylogenetic and functional analysis, we characterised the pan-gene, dubbed *HYPO*, which was originally identified as a lager-specific gene in

S. pastorianus (Nakao et al., 2009). The gene was subsequently identified in a subset of *S. cerevisiae* strains (Monerawela et al., 2015). With no similarity at the DNA or protein level to proteins with a known biological or molecular function, characterisation of *HYPO* poses many challenges.

Comparative genome analysis of ~500 *S. cerevisiae* isolates revealed a pattern of retention of *HYPO* in beer strains of yeasts and a specific absence in wine yeasts, suggestive of a possible role for the protein in the beer fermentation environment. A wider phylogenetic analysis revealed that *HYPO* is restricted to the fungal kingdom and even among the fungi, *HYPO* homologues are only present in a small number of species, mainly restricted to *Pezizomycotina* and *Saccharomycotina*. Furthermore, within these subphyla, only 17 species contain an intact ORF of *HYPO*. On the basis of the distribution and range of *HYPO*-similar sequences, it appears that the gene is specific to the *Ascomycota*. However, because the *Ascomycota* is a deep-branching group within fungi, it is possible that homologues of *HYPO*, which have diverged beyond any recognizable sequence similarity, could exist outside of this phylum. Alternatively, *HYPO* could have emerged entirely de novo, from a previously nongenic sequence, at the origin of the phylum. Whichever of the two scenarios that have occurred, the outcome was the emergence of a novel taxonomically restricted gene.

The fact that *HYPO* is found in a limited number of genomes within the *Ascomycota* raises further interesting questions regarding its evolution. Even assuming the most parsimonious scenario of origination at the base of the *Ascomycota*, there are at least a thousand species (represented by at least one genome in GenBank) where the gene was not found. This number may be even larger if we take into account the additional species with proteomes in the UniProt database, NCBI's NR database, and the MycoCosm fungal database, which we also searched for homologues. Thus, *HYPO* has been retained in the genomes of fewer than 2% of the *Ascomycota*. The pattern of *HYPO* conservation becomes even more peculiar when we consider that the gene is not found even in well-sampled yeast genera closely related to the *Saccharomyces*, such as the *Lachancea* (Vakirlis et al., 2016).

Two processes have likely contributed to this observed distribution of *HYPO*, namely, fast divergence along some lineages and multiple independent losses. The presence of *Ty* element remnants flanking *HYPO* and its adjacent gene *YPL257W* may point to a mechanism of acquisition and loss of these genes. The pattern of loss or retention of *HYPO* and *YPL257W* reveals that the two ORFs are most often lost independently and are less likely to be lost together. The genome region acquired by *S. paradoxus* from the cherry tree gum *S. cerevisiae* isolate is demarcated by the δLTR *Ty1* site that lies to right of the *HYPO* gene on chromosome XVI. Thus, it is possible that the transposition events may account for the loss of *HYPO* from so many *S. cerevisiae* strains and for the HGT into *S. paradoxus*.

Our analysis of the distribution of *HYPO* within the *Ascomycota* shows that a combination of evolutionary events can produce patterns of sparse presence that could, in extreme cases, potentially result in erroneous estimates of gene age and provides food for thought in the ongoing debate on gene age estimation (Domazet-Loso et al., 2017; Liebeskind, McWhite, & Marcotte, 2016; Moyers & Zhang, 2015). Furthermore, our finding of the recent HGT from an *S. cerevisiae* strain to two *S. paradoxus* strains raises the possibility on ongoing interchanging of genetic material between these two species.

5.2 | Characterisation of *HYPO*

The finding that the gene *HYPO* is predominantly retained in beer strains of *S. cerevisiae* as well as in strains of *S. pastorianus* used in lager production suggested that the presence of *HYPO* may confer some advantage to yeasts in the environmental conditions experienced in beer fermentations. To assess possible biological functions of *HYPO*, we generated a GFP fusion protein and examined the cellular localization pattern of the Hypo-gfp protein. The fusion protein produced a punctate staining pattern, which based on confocal imaging appears to localise to the cell membrane. Time-lapse fluorescence microscopy revealed that Hypo-gfp expression begins as a single punctate spot before spreading into multiple spots as the cells grow and divide. The punctate pattern was attributed to the Hypo protein as Gfp alone produces a diffuse-staining pattern. Additionally, this punctate pattern could not be a consequence of the overexpression of *HYPO* from a *PGK* promoter as the same pattern was observed when *HYPO-GFP* was expressed from its native promoter, which produces much lower expression levels. Furthermore, *GFP* fusions to other genes did not produce the same punctate patterns (W. Kricka and J. Fitzpatrick, personal communications). Interestingly, similar punctate staining patterns have been observed for other membrane-associated proteins such as the methionine permease, Mup1p, and the arginine permease, Can1p (Busto et al., 2018). Several models such as the picket fence, the lipid raft, and hydrophobic mismatch have been proposed to explain the segregation of membrane-associated proteins and lipids into such plasma membrane microdomains (Jensen & Mouritsen, 2004; Kusumi et al., 2012; Simons & Sampaio, 2011).

Analysis of the growth of yeast strains overexpressing *HYPO-GFP* in media containing different sugars as a sole carbohydrate source did not indicate any growth advantage. We did observed a slight but not significant growth inhibition when *S. pastorianus* strains, overexpressing *HYPO-GFP*, were grown in high sugar concentrations, be it, glucose, maltose, or wort, similar to concentrations used in industrial fermentations. Given the complex composition of wort and the many hexose and maltose transporters that have been uncovered in *S. cerevisiae* and *S. pastorianus* (Brown, Murray, & Verstrepen, 2010; Dietvorst, Walsh, van Heusden, & Steensma, 2010; Henderson & Baldwin, 2012; Horak, 2013; Vidgren & Londesborough, 2018; Vidgren, Ruohonen, & Londesborough, 2005), further research will be required to determine if *HYPO* plays a role in sugar uptake in yeast cells and to understand the biological advantage to retaining the gene in beer and lager yeasts.

ACKNOWLEDGEMENTS

Support for this research was obtained by way of a Ph.D. 1592 scholarship to C. M. supervised by U. B. and by a Marie Skłodowska-Curie Innovative Training Network award (Project Number 764364) to U. B. We thank Brigida Gallone and Kevin Verstrepen for access to the genome data from the paper (Gallone et al., 2016).

CONFLICT OF INTERESTS

There are no conflict of interests associated with the research reported in this manuscript.

ORCID

Nikolaos Vakirlis https://orcid.org/0000-0001-7606-6987
Aoife McLysaght https://orcid.org/0000-0003-2552-6220
Tharappel James https://orcid.org/0000-0002-0062-4468
Ursula Bond https://orcid.org/0000-0003-2877-6460

REFERENCES

Almeida, P., Barbosa, R., Zalar, P., Imanishi, Y., Shimizu, K., Turchetti, B., … Sampaio, J. P. (2015). A population genomics insight into the Mediterranean origins of wine yeast domestication. *Molecular Ecology*, *24*, 5412–5427. https://doi.org/10.1111/mec.13341

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, *25*, 3389–3402. https://doi.org/10.1093/nar/25.17.3389

Bergström, A., Simpson, J. T., Salinas, F., Barré, B., Parts, L., Zia, A., … Liti, G. (2014). A high-definition view of functional genetic variation from natural yeast genomes. *Molecular Biology and Evolution*, *31*, 872–888. https://doi.org/10.1093/molbev/msu037

Bernsel, A., Viklund, H., Hennerdal, A., & Elofsson, A. (2009). TOPCONS: Consensus prediction of membrane protein topology. *Nucleic Acids Research*, *37*, W465–W468. https://doi.org/10.1093/nar/gkp363

Borneman, A., Desany, B., Riches, D., Affourtit, J., Forgan, A., Pretorius, I., … Chambers, P. (2011). Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. *PLoS Genetics*, *7*, e1001287. https://doi.org/10.1371/journal.pgen.1001287

Borneman, A. R., Forgan, A. H., Kolouchova, R., Fraser, J. A., & Schmidt, S. A. (2016). Whole genome comparison reveals high levels of inbreeding and strain redundancy across the spectrum of commercial wine strains of *Saccharomyces cerevisiae*. *G3-Genes Genomes Genetics*, *6*, 957–971.

Brown, C. A., Murray, A. W., & Verstrepen, K. J. (2010). Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Current Biology*, *20*, 895–903. https://doi.org/10.1016/j.cub.2010.04.027

Busto, J. V., Elting, A., Haase, D., Spira, F., Kuhlman, J., Schafer-Herte, M., & Wedlich-Soldner, R. (2018). Lateral plasma membrane compartmentalization links protein function and turnover. *The EMBO Journal*, *37*, e99473. https://doi.org/10.15252/embj.201899473

Dietvorst, J., Walsh, M. C., van Heusden, G. P., & Steensma, H. Y. (2010). Comparison of the MTT1- and MAL31-like maltose transporter genes in lager yeast strains. *FEMS Microbiology Letters*, *310*, 152–157. https://doi.org/10.1111/j.1574-6968.2010.02056.x

Domazet-Loso, T., Carvunis, A. R., Alba, M. M., Sestak, M. S., Bakaric, R., Neme, R., & Tautz, D. (2017). No evidence for phylostratigraphic bias impacting inferences on patterns of gene emergence and evolution. *Molecular Biology and Evolution*, *34*, 843–856. https://doi.org/10.1093/molbev/msw284

Fitzpatrick, J., Kricka, W., James, T. C., & Bond, U. (2014). Expression of three *Trichoderma reesei* cellulase genes in *Saccharomyces pastorianus* for the development of a two-step process of hydrolysis and fermentation of cellulose. *Journal of Applied Microbiology*, *117*, 96–108. https://doi.org/10.1111/jam.12494

Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., … Verstrepen, K. J. (2016). Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell*, *166*, 1397–1410. https://doi.org/10.1016/j.cell.2016.08.020

Goffeau, A., Barrell, B., Bussey, H., Davis, R., Dujon, B., Feldmann, H., … Johnston, M. (1996). Life with 6000 genes. *Science*, *274*, 546–567. https://doi.org/10.1126/science.274.5287.546

Gonçalves, M., Pontes, A., Almeida, P., Barbosa, R., Serra, M., Libkind, D., … Sampaio, J. (2016). Distinct domestication trajectories in top-fermenting beer yeasts and wine yeasts. *Current Biology*, *26*, 1–12.

Gonçalves, P., Rodrigues de Sousa, H., & Spencer-Martins, I. (2000). FSY1, a novel gene encoding a specific fructose/H(+) symporter in the type strain of *Saccharomyces carlsbergensis*. *Journal of Bacteriology*, *182*, 5628–5630. https://doi.org/10.1128/JB.182.19.5628-5630.2000

Guindon, S., Delsuc, F., Dufayard, J. F., & Gascuel, O. (2009). Estimating maximum likelihood phylogenies with PhyML. *Methods in Molecular Biology*, *537*, 113–137. https://doi.org/10.1007/978-1-59745-251-9_6

Henderson, P. J., & Baldwin, S. A. (2012). Structural biology: Bundles of insights into sugar transporters. *Nature*, *490*, 348–350. https://doi.org/10.1038/490348a

Horak, J. (2013). Regulations of sugar transporters: Insights from yeast. *Current Genetics*, *59*, 1–31. https://doi.org/10.1007/s00294-013-0388-8

James, T., Usher, J., Campbell, S., & Bond, U. (2008). Lager yeasts possess dynamic genomes that undergo rearrangements and gene amplification in response to stress. *Current Genetics*, *53*, 139–152. https://doi.org/10.1007/s00294-007-0172-8

Jansen, G., Wu, C., Schade, B., Thomas, D. Y., & Whiteway, M. (2005). Drag&Drop cloning in yeast. *Gene*, *344*, 43–51. https://doi.org/10.1016/j.gene.2004.10.016

Jensen, M. O., & Mouritsen, O. G. (2004). Lipids do influence protein function—The hydrophobic matching hypothesis revisited. *Biochimica et Biophysica Acta*, *1666*, 205–226. https://doi.org/10.1016/j.bbamem.2004.06.009

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*, 772–780. https://doi.org/10.1093/molbev/mst010

Kricka, W., James, T. C., Fitzpatrick, J., & Bond, U. (2015). Engineering *Saccharomyces pastorianus* for the co-utilisation of xylose and cellulose from biomass. *Microbial Cell Factories*, *14*, 61. https://doi.org/10.1186/s12934-015-0242-4

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*, 1870–1874. https://doi.org/10.1093/molbev/msw054

Kusumi, A., Fujiwara, T. K., Chadda, R., Xie, M., Tsunoyama, T. A., Kalay, Z., … Suzuki, K. G. (2012). Dynamic organizing principles of the plasma membrane that regulate signal transduction: Commemorating the fortieth anniversary of Singer and Nicolson's fluid-mosaic model. *Annual Review of Cell and Developmental Biology*, *28*, 215–250. https://doi.org/10.1146/annurev-cellbio-100809-151736

Liebeskind, B. J., McWhite, C. D., & Marcotte, E. M. (2016). Towards consensus gene ages. *Genome Biology and Evolution*, *8*, 1812–1823. https://doi.org/10.1093/gbe/evw113

Marsit, S., Leducq, J. B., Durand, E., Marchant, A., Filteau, M., & Landry, C. R. (2017). Evolutionary biology through the lens of budding yeast comparative genomics. *Nature Reviews Genetics*, *18*, 581–598. https://doi.org/10.1038/nrg.2017.49

McLysaght, A., & Guerzoni, D. (2015). New genes from non-coding sequence: The role of de novo protein-coding genes in eukaryotic evolutionary innovation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*, 20140332. https://doi.org/10.1098/rstb.2014.0332

Monerawela, C., & Bond, U. (2017a). Brewing up a storm: The genomes of lager yeasts and how they evolved. *Biotechnology Advances*, *35*, 512–519. https://doi.org/10.1016/j.biotechadv.2017.03.003

Monerawela, C., & Bond, U. (2017b). Recombination sites on hybrid chromosomes in *Saccharomyces pastorianus* share common sequence motifs and define a complex evolutionary relationship between group I and II lager yeasts. *FEMS Yeast Research*, *17*, fox047.

Monerawela, C., & Bond, U. (2018). The hybrid genomes of *Saccharomyces pastorianus*: A current perspective. *Yeast*, *35*, 39–50. https://doi.org/10.1002/yea.3250

Monerawela, C., James, T. C., Wolfe, K. H., & Bond, U. (2015). Loss of lager specific genes and subtelomeric regions define two different *Saccharomyces cerevisiae* lineages for *Saccharomyces pastorianus* Group I and II strains. *FEMS Yeast Research*, *15*, fou008.

Moyers, B. A., & Zhang, J. (2015). Phylostratigraphic bias creates spurious patterns of genome evolution. *Molecular Biology and Evolution*, *32*, 258–267. https://doi.org/10.1093/molbev/msu286

Nakao, Y., Kanamori, T., Itoh, T., Kodama, Y., Rainieri, S., Nakamura, N., … Ashikari, T. (2009). Genome sequence of the lager brewing yeast, an interspecies hybrid. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, *16*, 115–129.

Peter, J., De Chiara, M., Friedrich, A., Yue, J. X., Pflieger, D., Bergstrom, A., … Schacherer, J. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*, *556*, 339–344. https://doi.org/10.1038/s41586-018-0030-5

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., & Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Research*, *46*, W200–W204. https://doi.org/10.1093/nar/gky448

Pretorius, I. S., & Boeke, J. D. (2018). Yeast 2.0—Connecting the dots in the construction of the world's first functional synthetic eukaryotic genome. *FEMS Yeast Research*, *18*, pii: 4939478.

Ranwez, V., Harispe, S., Delsuc, F., & Douzery, E. J. (2011). MACSE: Multiple alignment of coding sequences accounting for frameshifts and stop codons. *PLoS ONE*, *6*, e22594. https://doi.org/10.1371/journal.pone.0022594

Schlotterer, C. (2015). Genes from scratch—The evolutionary fate of de novo genes. *Trends in Genetics*, *31*, 215–219. https://doi.org/10.1016/j.tig.2015.02.007

Simons, K., & Sampaio, J. L. (2011). Membrane organization and lipid rafts. *Cold Spring Harbor Perspectives in Biology*, *3*, a004697.

Strope, P. K., Skelly, D. A., Kozmin, S. G., Mahadevan, G., Stone, E. A., Magwene, P. M., … McCusker, J. H. (2015). The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Research*, *25*, 762–774. https://doi.org/10.1101/gr.185538.114

Tautz, D., & Domazet-Loso, T. (2011). The evolutionary origin of orphan genes. *Nature Reviews Genetics*, *12*, 692–702. https://doi.org/10.1038/nrg3053

Taxis, C., & Knop, M. (2006). System of centromeric, episomal, and integrative vectors based on drug resistance markers for *Saccharomyces cerevisiae*. *BioTechniques*, *40*, 73–78. https://doi.org/10.2144/000112040

Vakirlis, N., Hebert, A. S., Opulente, D. A., Achaz, G., Hittinger, C. T., Fischer, G., … Lafontaine, I. (2018). A molecular portrait of de novo genes in yeasts. *Molecular Biology and Evolution*, *35*, 631–645. https://doi.org/10.1093/molbev/msx315

Vakirlis, N., Sarilar, V., Drillon, G., Fleiss, A., Agier, N., Meyniel, J. P., … Fischer, G. (2016). Reconstruction of ancestral chromosome architecture and gene repertoire reveals principles of genome evolution in a model yeast genus. *Genome Research*, *26*, 918–932. https://doi.org/10.1101/gr.204420.116

Vidgren, V., & Londesborough, J. (2018). Overexpressed maltose transporters in laboratory and lager yeasts: Localization and competition with endogenous transporters. *Yeast*, *35*, 567–576. https://doi.org/10.1002/yea.3322

Vidgren, V., Ruohonen, L., & Londesborough, J. (2005). Characterization and functional analysis of the MAL and MPH loci for maltose utilization in some ale and lager yeast strains. *Applied and Environmental Microbiology*, *71*, 7846–7857. https://doi.org/10.1128/AEM.71.12.7846-7857.2005

Wang, Q. M., Liu, W. Q., Liti, G., Wang, S. A., & Bai, F. Y. (2012). Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Molecular Ecology*, *21*, 5404–5417. https://doi.org/10.1111/j.1365-294X.2012.05732.x

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088

Yue, J. X., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., … Liti, G. (2017). Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nature Genetics*, *49*, 913–924. https://doi.org/10.1038/ng.3847

Zhu, Y. O., Sherlock, G., & Petrov, D. A. (2016). Whole genome analysis of 132 clinical *Saccharomyces cerevisiae* strains reveals extensive ploidy variation. *G3-Genes Genomes Genetics*, *6*, 2421–2434. https://doi.org/10.1534/g3.116.029397

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Vakirlis N, Monerawela C, McManus G, et al. Evolutionary journey and characterisation of a novel pan-gene associated with beer strains of *Saccharomyces cerevisiae*. *Yeast*. 2019;36:425–437. https://doi.org/10.1002/yea.3391