Applied Microbiology and Biotechnology (2024) 108:539
https://doi.org/10.1007/s00253-024-13382-1

APPLIED GENETICS AND MOLECULAR BIOTECHNOLOGY

Genomic characteristics and genetic manipulation of the marine yeast *Scheffersomyces spartinae*

Awkash Sharma[1] · Xing Liu[1] · Jun Yin[1] · Pei-Jing Yu[1] · Lei Qi[1] · Min He[1] · Ke-Jing Li[1] · Dao-Qiong Zheng[1]

Received: 19 September 2024 / Revised: 6 December 2024 / Accepted: 9 December 2024 / Published online: 19 December 2024

Abstract

The halotolerant yeast *Scheffersomyces spartinae*, commonly found in marine environments, holds significant potential for various industrial applications. Despite this, its genetic characteristics have been relatively underexplored. In this study, we isolated a strain of *S. spartinae* named YMxiao from seawater in Zhoushan City, China. Through scanning electron microscopy and flow cytometry, we characterized *S. spartinae* YMxiao cells as urn-shaped, demonstrating asymmetric division via budding, and possessing a diploid genome. Compared to the model yeast *Saccharomyces cerevisiae*, *S. spartinae* YMxiao exhibited greater tolerance to various stressful conditions. Furthermore, *S. spartinae* YMxiao was capable of utilizing xylose, mannitol, sorbitol, and arabinose as sole carbon sources for growth. We conducted whole-genome sequencing of *S. spartinae* YMxiao using a combination of Nanopore and Illumina technologies, resulting in a telomere-to-telomere complete genome assembly of 12 Mb. Genome annotation identified 5311 protein-coding genes, 214 tRNA genes, and 236 transposable elements distributed across 8 chromosomes. Comparative genomics between *S. spartinae* strains YMxiao and ARV011 revealed genomic variations and evolutionary patterns within this species. Notably, certain genes in *S. spartinae* strains were found to be under strong positive selection. Additionally, we developed a genetic manipulation protocol that successfully enabled gene knockouts in *S. spartinae*. Our findings not only enhance our understanding of the *S. spartinae* genome but also provide a foundation for future research into its potential biotechnological applications.

Key points

- *The unique phenotypes and genetic characteristics of S. spartinae were disclosed.*
- *Comparative genomics showed vast genetic variations between S. spartinae strains.*
- *Genetic manipulation protocol was established for S. spartinae strain.*

Keywords *S. spartinae* · Marine yeast · Comparative genomics · Genetic manipulation · Selection marker

Introduction

The single-cell organism yeast plays a pivotal role in the biotechnology field, particularly in the conversion of biomass into biofuels and other valuable chemicals. Among yeasts, *Saccharomyces cerevisiae* and its close relatives are renowned for their efficient fermentation of glucose and sucrose into ethanol and CO_2. However, *S. cerevisiae* lacks the ability to ferment various sugars, such as arabinose, mannitol, and xylose, which are present in lignocellulose (Papini et al. 2012; Kwak and Jin 2017). Non-*Saccharomyces* strains, such as *Scheffersomyces stipitis*, provide alternative metabolic pathways for the utilization of non-glucose sugars and the production of various fermentation products (Trichez et al. 2023; Barros et al. 2024). The diverse metabolic capabilities of non-conventional yeasts make them valuable for industrial applications, where they can be employed as pure or mixed cultures to achieve specific fermentation outcomes.

Scheffersomyces spartinae, an ascomycetous yeast species, belongs to the *Debaryomycetaceae* family within the *Ascomycota* phylum (Kurtzman and Suzuki 2010). Formerly classified under the *Candida* and *Pichia* genera, *S. spartinae*

Awkash Sharma and Xing Liu contributed equally to this work.

✉ Ke-Jing Li
22034192@zju.edu.cn

✉ Dao-Qiong Zheng
zhengdaoqiong@zju.edu.cn

[1] National Key Laboratory of Biobased Transportation Fuel Technology, Ocean College, Zhejiang University, Hangzhou 310027, China

was later reassigned to the *Scheffersomyces* genus, being considered an ancestral species of the *Scheffersomyces* clade (Kurtzman and Suzuki 2010). *S. spartinae* was detected in the biofilm of a water treatment plant near Botany Bay, New South Wales, Australia, where it contributed to blockages and its abundance was found to correlate positively with salinity levels, underscoring its adaptation to high salt concentrations (Gillings et al. 2006). *S. spartinae* had been also isolated from mangrove sediments (Li et al. 2019). This strain is capable of producing coenzyme Q9, a crucial component involved in various biological functions (Kurtzman and Suzuki 2010). Certain *S. spartinae* strain was found can perform aerobic decolorization, degradation, and detoxification of azo dyes (Tan et al. 2016). In addition, *S. spartinae* strains show promising potential as bio-control agents of gray mold on strawberries (Zou et al. 2021).

Recent developments in high-throughput sequencing technology and genomic analysis have facilitated deeper investigation into the genetic diversity, evolution, and metabolism of yeasts. Despite the potential of *S. spartinae* for biotechnological applications, genomic studies on this yeast species remain limited. So far, two *S. spartinae* strains ARV011 and NRRL Y-7322 have been sequenced by Illumina technology and their draft genome sequences have been published (Villarreal et al. 2021). The genome of *S. spartinae* was ~ 12 Mb, encoding more than 5000 genes. However, the functions of those genes were not clarified, mainly due to the lacking of genetic manipulation tools for this species. In this study, we studied the phenotypic traits of *S. spartinae* YMxiao by comparing the growth capability of this strain and the model organism *S. cerevsiae*. We also obtained a complete genome of *S. spartinae* YMxiao using a combination of Illumina and Nanopore sequencing technologies. Comparative genomic analyses of different *S. spartinae* strains elucidated the patterns of genome evolution of this species. Gene prediction and annotation facilitated the identification of genes responsible for the distinct phenotypes, such as carbon source utilization, observed between *S. spartinae* and *S. cerevisiae*. Additionally, this study established a gene deletion protocol for *S. spartinae*, providing a valuable reference for developing genetic manipulation techniques for non-model yeast strains.

Materials and method

Yeast strains and media

Two yeast strains, *S. cerevisiae* YJS329 (Zheng et al. 2012) and *S. spartinae* YMxiao, were used in this study. The strain YMxiao was isolated from the sea water of Dongji Island, Zhoushan city, China, in November, 2018. This strain has been stored in China Center for Type Culture Collection with an accession number of CCTCC AY 2022004. The 18S rDNA sequence of YMxiao was amplified using the primers 5′-GTAGTCATATGCTTGTCTC-3′ and 5′-TCCGCAGGTTCACCTACGGA-3′ and then sequenced. YPD medium was prepared with 20 g/L glucose, 20 g/L peptone, and 10 g/L yeast extract, adjusted to pH 7. Variants of the YPD medium included YPM (containing 20 g/L mannitol), YPA (containing 20 g/L arabinose), YPS (containing 20 g/L sorbitol), and YPX (containing 20 g/L xylose) as the respective carbon sources. For stress assays, YPD + H_2O_2 and YPD + NaCl media were prepared by adding H_2O_2 and NaCl to liquid YPD to final concentrations of 5 mM and 0.7 M, respectively.

Scanning electron microscope (SEM) analysis

A modified protocol based on Bang and Pazirandeh (1999) was employed to examine the morphological characteristics of *S. spartinae* YMxiao. The yeast cells were initially fixed with 2.5% glutaraldehyde in 0.1 M cacodylate buffer at pH 7.4 for 1 h at 4 °C. Subsequently, the cells were washed three times with 0.1 M cacodylate buffer and subjected to a series of ethanol washes (50%, 70%, 90%, and 100%) for 10 min each to achieve dehydration. The dehydrated cells were critically dried using CO_2 in a Bal-Tec CPD 030 critical point dryer (BAL-TEC Ltd., Balzers, Switzerland). These dried cells were then mounted onto SEM stubs using double-sided carbon tape and coated with gold–palladium using a sputter coater. Finally, the coated samples were visualized under high vacuum using a JEOL JSM-6010LA scanning electron microscope (JEOL Ltd., Tokyo, Japan) at an accelerating voltage of 15 kV.

Flow cytometry

The cells of *S. spartinae* YMxiao and *S. cerevisiae* strains YJSH1 (Zheng et al. 2012) and YJS329 were incubated in 5 mL YPD for 40 h. Approximately 5×10^6 cells were collected and fixed by 70% ethanol for 0.5 h. The cells were then treated with RNase A (0.5 mg/mL) for 4 h and proteinase K (2 mg/mL) for 1 h. Following treatment, the cells were stained with 10 µg/mL propidium iodide at 4 °C overnight. The ploidy of these yeast strains was monitored by the flow cytometry machine (BD Biosciences, San Jose, USA) as described in Zhu et al. (2024a, b).

Growth measurement

The *S. spartinae* and *S. cerevisiae* cells were incubated in 30 mL YPD, YPM, YPX, YPD + H_2O_2, YPD + NaCl, YPDpH3.5, and YPDpH8.5 with an initial OD_{600} of 0.05 for 48 h. The growth (OD_{600}) of yeast was determined using a Bio-Tek PowerWaveTM XS/XS2 plate reader (Bio-Tek

Instruments, Winooski, VT, USA). For each measurement, 200 µL cells were diluted and transferred to a 96-well plate reader. The medium without cell incubation was used as a control to establish a baseline.

Nanopore sequencing

Yeast cells were grown in 20 mL of YPD medium for 20 h. Following growth, cells were collected, and genomic DNA was extracted using a modified CTAB (cetyltrimethylammonium bromide) method (Qi et al. 2023). Nanopore sequencing (Oxford Nanopore Technologies, Oxford, UK) was performed as described in our previous study (Qi et al. 2023). Briefly, the genomic DNA was treated with NEBNext FFPE RepairMix M6630 (New England Biolabs, Ipswich, MA, USA) to repair single-strand nicks. The DNA ends were then repaired to form blunt ends using the NEBNext End Repair Module E6050 (New England Biolabs, Ipswich, MA, USA). DNA samples were purified with AMPure XP beads (Beckman Coulter, Brea, CA, USA). Each sample was barcoded using the Oxford Nanopore Native Barcoding Genomic DNA Kit EXP-NBD104 (Oxford Nanopore Technologies, Oxford, UK), followed by adapter ligation with the ligation-based library kit SQK-LSK109 (Oxford Nanopore Technologies, Oxford, UK). After final purification, multiple barcoded libraries were loaded onto MinION flow cells FLO-MIN106D R9.4.1 (Oxford Nanopore Technologies, Oxford, UK) and sequenced on a MinION sequencer MIN-101B (Oxford Nanopore Technologies, Oxford, UK).

Illumina sequencing

Cells of *S. spartinae* YMxiao were grown in 20 mL YPD medium for 20 h, after which they were collected for DNA extraction using the EZNA Yeast DNA Kit (Omega, Norcross, GA, USA). The genomic DNA was then subjected to Illumina sequencing on the NextSeq 500 sequencer (Illumina, San Diego, CA, USA) using a 2×150 bp paired-end indexing protocol. DNA libraries were constructed with the NEBNext® Ultra™ DNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA), following the manufacturer's instructions. Sequencing reads were processed using Cutadapt to remove adapters and low-quality reads (Martin 2011).

Genome assembly of *S. spartinae* YMxiao

The raw Nanopore reads were subjected to quality control using FastQC v0.11.8 (Andrews 2010) and adapter trimming and filtering of nanopore sequencing reads were performed by Porechop (Wick et al. 2018). The resulting high-quality reads were assembled with nanopore whole-genome sequencing data using NECAT (Chen et al. 2020). To further refine the assembly and correct errors, we employed the Pilon software tool, which utilizes a combination of Illumina reads and an existing assembly to perform error correction (Hu et al. 2020).

Gene prediction and annotation of assembled genome

The genome annotation of *S. spartinae* YMxiao was performed using the Augustus gene predictor which was trained using a set of curated proteins from related species *S. cerevisiae* (Stanke et al. 2006). tRNA-coding regions were predicted by the tRNAscan-SE program (Lowe and Chan 2016). The mitochondrial genome annotation was performed using MFannot (https://megasun.bch.umontreal.ca/apps/mfannot/) with the correction of open reading frame (ORF) structures. In order to identify transposable elements, the transposon annotation resonaTE tool was employed (Riehl et al. 2022). Transposon annotation resonaTE tool employs RepeatMasker to identify repeats and RepeatModeler to build repeat libraries from genome sequences, which are then used to classify transposable elements. Functional annotation of the predicted genes was performed by similarity using BLAST against public databases, including the NCBI's non-redundant protein (Nr) database, and Kyoto Encyclopedia of Genes and Genomes (KEGG). A BLAST search was conducted against the aforementioned databases with an *E*-value less than 1e-5 and a minimum alignment length percentage greater than 40% (identity $\geq 40\%$, coverage $\geq 40\%$) of the entire genome. The Gene Ontology (GO) annotations were performed with the default parameters of the program Blast2GO (Conesa et al. 2005), and GO-term classification was performed based on the Nr annotations.

Whole-genome alignment and comparison

To perform a comprehensive whole-genome alignment and comparison, we used the MUMmer4 (Marçais et al. 2018), specifically leveraging the NUCmer program, to align the genomes of *S. spartinae* YMxiao against *S. spartinae* ARV011. Using default parameters ensured optimal alignment accuracy. Post-alignment, structural variations, encompassing significant deletions and insertions, were identified with the show-diff utility within MUMmer. To detect the single-nucleotide variations (SNVs) and small insertions and deletions (InDels) between the two homologs of *S. spartinae* YMxiao, the fastq file of *S. spartinae* YMxiao was mapped to the assembled genome of *S. spartinae* YMxiao using BWA-MEM v0.7.17 (Li 2013). Subsequently, the bam files generated from the alignment were sorted and indexed using SAMtools v1.9 (Li et al. 2009). VarScan v2.4.4 software was used to call the SNVs and InDels (Koboldt et al. 2012). Finally, SnpEff was utilized for annotation of the filtered

variants using *S. spartinae* YMxiao as the reference genome (Cingolani et al. 2012).

Ka (non-synonymous substitutions) and Ks (synonymous substitutions) ratio calculation

The coding sequence (CDS) of *S. spartinae* ARV011 were obtained from the NCBI database (https://www.ncbi.nlm.nih.gov/bioproject/756327). The Ka/Ks ratio was subsequently calculated for each gene pair using the TBtools v1.108, which employs the KaKs Calculator v2.0 tool (http://evolution.genomics.org.cn/software.htm) with Model Selection set to "ModelAverage."

Genetic manipulation of *S. spartinae* YMxiao

The nourseothricin-resistant gene *NAT*, including the *TEF1* promoter and terminator, was amplified from the plasmid pAG25 (Goldstein and McCusker 1999) using primers dnatMX6-pUG72-s and dnatMX6-pUG72-a (Table 1). The *NAT* fragment was then cloned into the *Xba*I and *Sac*I sites of the plasmid pUG72 (Gueldener et al. 2002), creating a plasmid named pUG72-NATsc. To ensure the functionality of *NAT* gene in *S. spartinae*, the coding region was codon-optimized and synthesized de novo to replace the original *NAT* gene, using the restriction endonucleases *Nco*I and *Sph*I. The resulting plasmid was named pUG72-NATss.

To construct a gene deletion cassette, we first amplified three separate fragments: the upstream homologous arm, the selection marker gene, and the downstream homologous arm. These fragments were subsequently ligated together by overlap PCR using the primers shown in Table 1. Transformation of *S. spartinae* YMxiao with the resulting deletion cassettes was performed using the PEG/LiAc/ssDNA method (Gietz and Schiestl 2007). Transformants were selected on media containing 75 mg/mL nourseothricin and verified by PCR with the primers listed in Table 1.

Data availability

The 18S rDNA sequence of *S. spartinae* YMxiao and the sequence of plasmid pUG72-NATss has been deposited in the NCBI nucleotide database with the accession numbers PQ279892 and PQ336169, respectively. The Whole Genome Shotgun project of *S. spartinae* YMxiao has been deposited at DDBJ/ENA/GenBank under the accession JBJABQ000000000. The version described in this paper is version JBJABQ010000000.The raw data of next-generation sequencing were available at SRA database with the accession number of PRJNA1152126.

Results

Cell morphology, phylogenic analysis, and ploidy of *S. spartinae* YMxiao

To determine the evolutionary relationship between YMxiao and its related yeast strains, we constructed a phylogenetic tree based on 18S rDNA (Fig. 1A). The analysis revealed that YMxiao was grouped within the cluster of *S. spartinae* strains (Fig. 1A), supporting the taxonomic classification of strain YMxiao as a member of the *S. spartinae* species. The SEM result showed that the cells of *S. spartinae* YMxiao are urn-shaped, ranging from 20 to 40 μm in length, and exhibit slight tapering at one end, while maintaining a uniform size (Fig. 1B). The surface of the cells displayed a rough and uneven topography, which is attributed to the presence of various appendages, including bud scars, that give rise to a bumpy appearance (Fig. 1B).

To determine the ploidy of YMxiao, we cultured cells from this strain alongside diploid *S. cerevisiae* YJS329 in YPD medium for 40 h until they reached the stationary phase. At this stage, nearly all cells were arrested in the G1 phase, as indicated by the absence of budding (Fig. 1C). Using flow cytometry, we found that the DNA content of *S. spartinae* YMxiao was higher than that of the haploid *S. cerevisiae* YJSH1, but closely resembles the DNA content of *S. cerevisiae* YJS329 (Fig. 1D). Given that the genome size of *S. spartinae* is approximately 12 Mb, which is similar to that of *S. cerevisiae*, we infer that YMxiao is a diploid yeast strain.

Phenotypic comparison of *S. spartinae* YMxiao and *S. cerevisiae* YJS329

To determine the growth capability of *S. spartinae* under different conditions, the optical density (OD_{600}) of YMxiao cells were measured in YPD and YPD with certain stressors. In this experiment, an industrial *S. cerevesiae* strain, YJS329, was used as a control strain. In YPD at 30 °C, the max growth rates of YMxiao and YJS329 were 0.44 ± 0.04 and 0.35 ± 0.01 (Fig. 2A). The presence of 0.7 M NaCl decreased the max growth rate by 57% and 71% for YMxiao and YJS329, respectively (Fig. 2B). Although both strains were inhibited by NaCl, the biomass of YMxiao was 5.64-fold of that YJS329 at 24 h (Fig. 2B). Elevated temperature (37 °C) (Fig. 2C), high pH (Fig. 2D), and low pH (Fig. 2E) also led to decreased growth rates of both YMxiao and YJS329. The relative biomass (OD_{600} value of a stressful condition divided by that in YPD condition at 24 h) of YMxiao was 0.71-, 1.83-, and 0.92-fold of that of YJS329 under 37 °C, pH 8.5, and

Table 1 Primers used in this study

Primers	Sequences (5′−3′)	Purpose
dnatMX6-pUG72-s	GCTATACGAAGTTATTAGGTCTAGAAGCGA CATGGAGGCCCAGAATA	Amplification of *NAT* gene from pAG25
dnatMX6-pUG72-a	TTATATTAAGGGTTCTCGAGAGCTCACACTGGA TGGCGGCGTTAGTA	
vnatMX6-pUG72-s	GCTTCGTGGTCGTCTCGTA	Verification of the insertion of *NAT* gene into pUG72
vnatMX6-pUG72a	CGAGTCAGTGAGCGAGGAA	
dURA3up-s	ATGTTTCAATCCATGTCGTGAGATCCGTGCATT TGCTCTTTCGGTGTTAC	Amplification of *URA3* upstream homologous arm
dURA3up-a	ACAGCTTCTGCTATTGGAACTG	
dnatMX6-URA3-s	CTCAGAACGAGGAGAAATCAGTTCCAATAG CAGAAGCTGTGCGGCATCAGAGCAGATTG	Amplification of *NAT* gene from pUG72-NATss
dnatMX6-URA3-a	ATAATGCTTCGAGAAATTATCAGCGTCGTA AACCTTTCATCGAGTCAGTGAGCGAGGAA	
dURA3down-s	ATGAAAGGTTTACGACGCTGAT	Amplification of *URA3* downstream homologous arm
dURA3down-a	TGGAGCAATCCAATCGTTTATAGCTGCACC AGTGGATGCTATATTCACCA	
Primer S	CGTGAACGCTCCTGCTATTC	Verification of *URA3* deletion
Primer A	CCTACCACCATTTATTGCCAAT	
dADE2up-s	GCGCGATGAGATGAGGTAGCAATAAGATAT ACACCCTTCTACTAACTACC	Amplification of *ADE2* upstream homologous arm
dADE2up-a	TGATAGTCCAGGCAAATAGT	
dnatMX6-ADE2-s	ATCATCATTATTATATTGTATACTATTTGCCTG GACTATCAGCTGCAGGTCGACAACCC	Amplification of *NAT* gene from pUG72-NATss
dnatMX6-ADE2-a	CTTCTTAAACTTTACTGCTCAACACCATCAAGT GGATCTGATATCACC	
dADE2down-s	TGATGGTGTTGAGCAGTAAAGT	Amplification of *ADE2* downstream homologous arm
dADE2down-a	GGGTACCTCAGAAAAACAATATGGGGTCTT CCTGGTTGGACAA	
vADE2(natMX6)-s	GATGGAACCGCTTCAGAT	Verification of *ADE2* deletion
vADE2(natMX6)-a	TGCTACCGAGATCAACAAT	
ade2_d*ura3*-s	GAACGAGGAGAAATCAGTTCCAATAGCAGA AGCTGTTTATGCAGCACGGTTCTTCG	Construction of *URA3* deletion cassette with *ADE2* as a selection marker
ade2_d*ura3*-a	ATGCTTCGAGAAATTATCAGCGTCGTAAAC CTTTCATAGAGATTCTCGGAGTCAACACT	
vura3_ade2-s	AGTGTTGACTCCGAGAATCTCT	Verification of the replacement of *URA3* with *ADE2* gene
vura3_ade2-a	TGCCAATTCGATCTCAGGTATG	
dnat-g4934.t1-s	GTTTCTGAGTTGGCTGCTGGAGATGGTCGT TAGAACGCGGCTACAA	Amplification of *NAT* for g4934.t1 deletion cassette
dnat-g4934.t1-a	GAGTGTGGAATACCTTGTACGCGGAATACT ATAGGGAGACCGGCAGAT	
g4934.t1_up-s	CAGAACGCACGGTTCCAAG	Amplification of g4934.t1 upstream homologous arm
g4934.t1_up-a	CCATCTCCAGCAGCCAACT	
g4934.t1_down-s	GCGTACAAGGTATTCCACACTC	Amplification of g4934.t1 downstream homologous arm
g4934.t1_down-a	AATGCTGCCAAGGCTGCT	
overlap_nat_g4934.t1_up-s	CAAGTGAGAGCCAAACAAGAGA	Amplification of the deletion cassette for g4934.t1
overlap_nat_g4934.t1_down-a	GGTAGACAAGGAACTCCAAAGG	
vnat_g4934.t1_s1	CTCAACGGTTCAGAGACATCAC	Verification of the deletion of g4934.t1
vnat_g4934.t1_a1	GTAGCCGCGTTCTAACGATTA	
g4934.t1_up-s-150 bp	GAAGAAGGTGCAGGAGAAGAAG	Amplification of 150 bp g4934.t1 upstream arm
g4934.t1_up-a-150 bp	CCATCTCCAGCAGCCAACT	
g4934.t1_up-s-250 bp	TCACCAACAGACGTGTCAATG	Amplification of 250 bp g4934.t1 upstream arm
g4934.t1_up-a-250 bp	CCATCTCCAGCAGCCAACT	

Table 1 (continued)

Primers	Sequences (5′−3′)	Purpose
g4934.t1_down-s-150 bp	GCGTACAAGGTATTCCACACTC	Amplification of 150 bp g4934.t1 downstream arm
g4934.t1_down-a-150 bp	AACAATTTCGGAGACCAGCATT	
g4934.t1_down-s-250 bp	CGTACAAGGTATTCCACACTCC	Amplification of 250 bp g4934.t1 downstream arm
g4934.t1_down-a-250 bp	TTTATGTCGCCTTCCACCATAT	

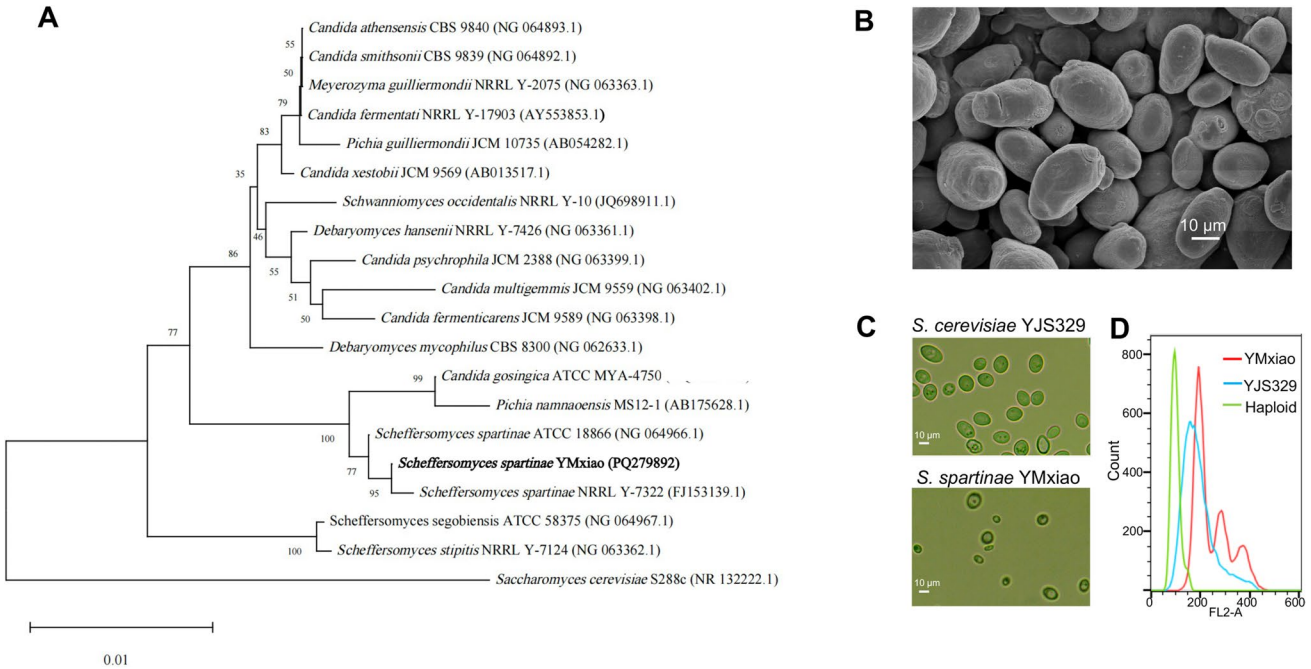

Fig. 1 Phylogenic analysis, cell morphology, and ploidy of *S. spartinae* YMxiao. **A** A phylogenetic tree showing the evolutionary relationship between YMxiao and its related yeast strains. **B** A scanning electron microscopy image of YMxiao yeast cells, providing a detailed view of their morphology with a 10-μm scale for reference. **C** Cell morphology of *S. cerevisiae* YJS329 and *S. spartinae* YMxiao under the microscope. **D** The relative DNA content of *S. spartinae* YMxiao and *S. cerevisiae* analyzed by flow cytometry. The haploid *S. cerevisiae* strain YJSH1 was used as a control

pH 3 conditions. The presence of 5 mM H_2O_2 resulted in a longer lag phase for both YMxiao and YJS329, but latter formed more biomass relative to the YPD condition (Fig. 2F). These results showed that *S. spartinae* YMxiao formed more biomass under multiple stressful conditions than *S. cerevesiae* strain YJS329. Additionally, we found that *S. spartinae* strain YMxiao has a superior capacity to utilize various less favorable carbon sources for growth compared to *S. cerevisiae* strain YJS329. Our results demonstrated that YMxiao produced significantly more biomass when grown on D-arabinose, D-xylose, sorbitol, and mannitol (Fig. 2G–J). Notably, YMxiao showed similar growth rates and biomass production in YPM (mannitol medium) as observed in YPD. Overall, these findings indicate that *S. spartinae* strain YMxiao exhibits enhanced growth and stress tolerance across a range of environmental conditions compared to *S. cerevisiae*.

Genome sequencing and annotation of *S. spartinae* YMxiao

Using a combination of Nanopore and Illumina sequencing technology, a telomere-to-telomere assembly of *S. spartinae* YMxiao genome was obtained. The genome was found to be distributed across 8 chromosomes with sizes ranging from 1.19 to 2.04 Mb, with a total size of 12.16 Mb. The GC content of the genome was found to be 38.7%, which is within the range observed in other yeast genomes such as *S. cerevisiae* (38.3%), *Kluyveromyces lactis* (38.7%), and *Debaryomyces hansenii* (36.3%) (Dujon et al. 2004). The annotation of the genome identified a total of 5305 protein coding ORFs (Table 2 and Supplemental Dataset S1), with an average length of 1577 bp. We found that 512 ORFs contain at least one intron sequence; and the average length of introns is 204 bp. The annotation also revealed the

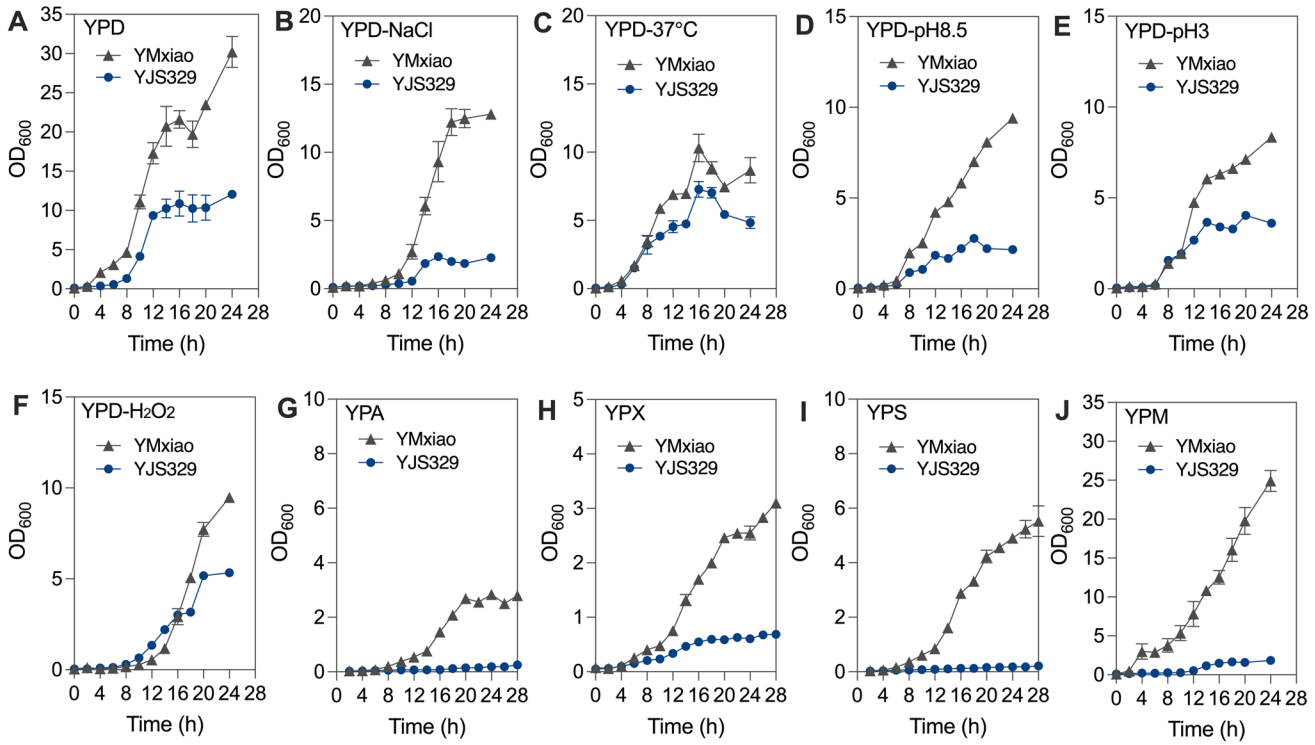

Fig. 2 The growth curves of yeast strains, YMxiao and YJS329, under different conditions. Their biomass was measured by the optical density at 600 nm (OD_{600}). In each figure, the growth of YMxiao is denoted by triangles, while YJS329 is represented by circles. The data points include error bars, indicating variability or standard deviation. **A** Normal YPD. **B** YPD with 0.7 M NaCl. **C** YPD incubated at 37 °C. **D** YPD was adjusted to pH 8.5. **E** YPD was adjusted to pH 3. **F** YPD with 5 mM H_2O_2. The dextrose in YPD was replaced by **G** arabinose, **H** xylose, **I** sorbitol, and **J** mannitol

Table 2 The profile of the *S. spartinae* YMxiao genome

Chromosome	Length (bp)	ORFs	DNA transposon	Retrotransposon	tRNA
chr1	2,037,313	889	29	8	26
chr2	1,821,069	831	27	6	31
chr3	1,681,693	723	20	6	15
chr4	1,482,941	648	22	7	33
chr5	1,383,034	619	18	1	29
chr6	1,274,463	555	21	9	42
chr7	1,192,923	508	13	3	21
chr8	1,290,903	536	25	7	17
Mitochondria	38,075	24	-	-	25
Total	12,164,339	5309	175	47	214

presence of 214 tRNA genes on 8 chromosomes (Table 2). In addition to protein-coding genes and tRNA genes, we identified a total of 222 transposable elements (TEs) on the 8 chromosomes, 175 were DNA transposons, and 46 were retrotransposons (Table 2). The most abundant type of DNA transposons is Zator (109) elements, while retrotransposons mainly consist of Copia (11) and Gypsy (35). The rDNA repeats were located at ~ 836 kb on chromosome 8 (2 copies of rDNA repeats are presented on the assembled genome). An analysis of the coverage of the rDNA region suggested the presence of approximately 60 copies of the rDNA repeat in the diploid YMxiao genome. The telomeres of YMxiao consist of ~ 300 bp of degenerate repeats with the consensus sequence TTTTAGGGTTTCGGTGTGCGGGGCTCT TGTG repeat units. This sequence is different from that (5′-[$(TG)_{0–6}$TGGGTGTG(G)]$_n$−3) observed in *S. cerevisiae* (Förstemann and Lingner 2001). Through mapping the reads of next-generation sequencing onto the assemble genome of YMxiao, we did not identify heterozygous regions across its 8 chromosomes. This result suggests that the mutations on *S. spartinae* YMxiao genome were homogenized through DNA homologous recombination during meiosis process.

The genes associated with carbon source utilization in *S. spartinae* YMxiao

In Fig. 2, we showed that *S. spartinae* YMxiao could utilize D-xylose, D-arabinose, sorbitol, and mannitol for growth. Below, we explored the genes contributing to its ability of carbon metabolism. D-xylose reductase (XR)-xylitol dehydrogenase (XDH) pathway is the most common

xylose utilization pathway in fungi (Fig. 3). XR is an enzymatic catalyst that mediates the conversion of D-xylose to xylitol using NADPH as a cofactor (Atzmüller et al. 2020). The annotation result shows the gene g3620.t1 in YMxiao encodes XR (Fig. 3). The transformation of xylitol to xylulose is mediated by XDH utilizing NAD^+ as a cofactor, encoded by the gene g5233.t1 in YMxiao (Fig. 3). The D-xylulose produced is then phosphorylated to D-xylulose-5-phosphate by xylulokinase that should be encoded by gene g2197.t1 (Fig. 3). After being transported into fungi cells, the D-arabinose can be converted to D-arabinitol by XR. Two metabolic enzymes have been identified for D-arabinitol: D-arabinitol 2-dehydrogenase (EC 1.1.1.250) and D-arabinitol 4-dehydrogenase (EC 1.1.1.11), which convert D-arabinitol to D-ribulose and D-xylulose, respectively. In YMxiao, g1157.t1 was identified encoding D-arabinitol 2-dehydrogenase, while none gene was predicted to encode D-arabinitol 4-dehydrogenase. Sorbitol dehydrogenase converts sorbitol to fructose that can be further phosphorylated and enter into glycolysis. In YMxiao, g3628.t1 and g5233.t1 were identified encoding sorbitol dehydrogenase (Supplemental Dataset S1 and Fig. 3). The first step in mannitol metabolism is its conversion of mannitol to mannose, a process catalyzed by mannitol dehydrogenase (Meena et al. 2015). This enzyme, which exhibits oxidoreductase activity, oxidizes mannitol using $NADP^+$ as a cofactor, producing mannose and NADPH. The genes g4934.t1 and g4935.t1 in *S. spartinae* YMxiao were predicted being encoding probable NADP-dependent mannitol dehydrogenase (Fig. 3). Mannose can be converted to mannose-6-phosphate by hexokinase putatively encoded by g3186.t1, g3545.t1, and g3603.t1 in YMxiao. The gene g2779.t1 encodes mannose phosphate isomerase that catalyze the conversion of mannose-6-phosphate to fructose 6-phosphate, which enters into central metabolic pathways (Fig. 3). Theses analysis confirmed the capability of multiple monosaccharide utilization in *S. spartinae* YMxiao at the genetic level.

Comparative genomics of different *S. spartinae* strains

S. spartinae ARV011 was isolated from marine water and sediment collected from the Great Sippewissett Marsh in Falmouth, Massachusetts (Villarreal et al. 2021). The genome of ARV011 was sequenced by an Illumina platform and assembled by SPAdes genome assembler, resulting in an uncompleted genome consists of 103 contigs (Villarreal et al. 2021). Using sequence alignment software Mummer, we identified 229,276 SNVs and 21,987 InDels between the chromosomes of YMxiao and ARV011 (Supplemental Dataset S2), but no large (> 50 kb) deletions, duplications, and inversions were detected. The

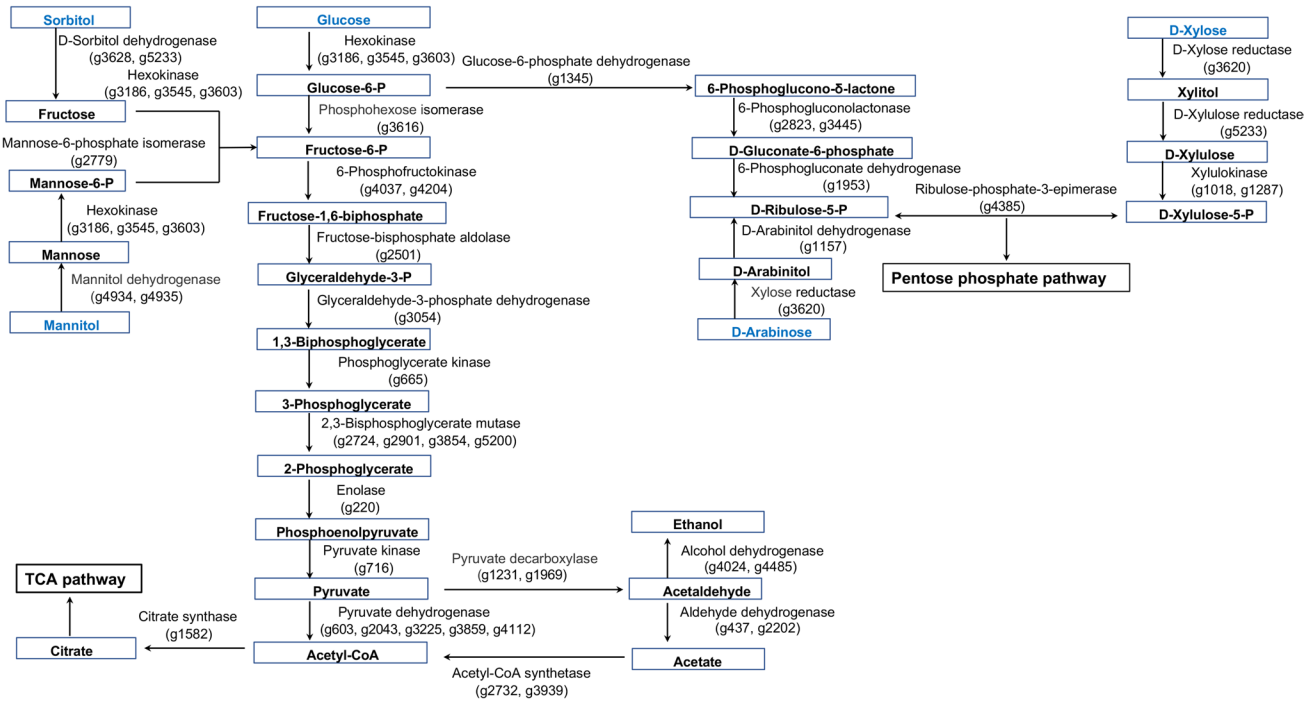

Fig. 3 Overview of genes involved in central carbon metabolic pathways. This schematic diagram represents the interconversion of various carbohydrates, including glucose, sorbitol, mannitol, D-xylose, and D-arabinose, into central metabolic intermediates, highlighting their integration into the pentose phosphate, glycolytic, and tricarboxylic acid (TCA) pathways, as well as the fermentation process leading to ethanol and acetate production. The genes encoding the enzymes in *S. spartinae* YMxiao are shown in the bracket

calculated frequencies of SNVs and InDels were 18.8 and 1.8 per kilobase of the *S. spartinae* genome, respectively. Nearly half (47.5%) of these point mutations were located at intragenic regions, which account up to 68.8% of the whole genome. This result suggested that the intergenic regions are more likely to accumulate mutations between *S. spartinae* strains. Analysis of the spectrum of SNVs indicated that C > T or G > A was the most frequent class of base substitution (36.6%), followed by A > G or T > C substitutions (27.3%). The transitions/transversions ratio was 1.77. Using the mutation annotation software snpEff, we found that more than half (67%) of the intragenic SNVs did not cause amino acids changes (synonymous variants) and are thus often considered neutral in terms of protein function. The 38,261 missense mutations are distributed across 5000 ORFs. Hundreds of those point mutations would also affect the start/stop codes gain or loss (Supplemental Dataset S2). Among them, 143 SNVs resulted in nonsense mutations which introduce premature stop codons and can potentially truncate the resulting protein. We also identified 1607 InDels that could cause frameshifts in ORFs. Such frameshifts are known to produce a completely different amino acid sequence downstream of the mutation site, often leading to a loss of normal protein function.

We analyzed 4956 gene pairs between *S. spartinae* strains YMxiao and ARV_011 and successfully calculated Ka/Ks ratios for 3078 of them. A significant 98.5% (3031 out of 3078) of these gene pairs exhibited Ka/Ks values less than 1, indicative of purifying selection. Conversely, we identified 44 gene pairs with Ka/Ks values exceeding 1, revealing positive selection (Supplemental Dataset S3). Annotation of the positively selected genes revealed that the gene g2438.t1 exhibited an exceptionally high Ka/Ks ratio of 8.388 (Supplemental Dataset S3). This gene encodes cysteine-rich, acidic integral membrane protein, but its biological function was not known. Furthermore, the gene g3519.t1, annotated as "bud emergence protein 1," displayed a Ka/Ks ratio of 5.818 (Supplemental Dataset S3). In *S. cerevisiae*, this protein was proven to be involved in establishing cell polarity and morphogenesis (Madden and Snyder 1998). Additional genes, such as g3875.t1 and g3735.t1, with Ka/Ks ratios of 3.255 and 2.185 respectively, were presumed to encode probable quinate permease and casein kinase 1 (Supplemental Dataset S3). These results suggest that genes involved in transmembrane transport and cellular metabolism have experienced stronger selection pressure compared to other genes. Although comparative genomics have revealed abundant genetic variations among *S. spartinae* strains, the ways in which these variations affect phenotypic and physiological differences still require further verification. Genetic manipulation in *S. spartinae* serves as a crucial foundation for such verification, and this was explored below.

Genetic manipulation of *S. spartinae* YMxiao

NAT gene could be used as a drug-resistant marker in *S. spartinae*

In yeast, a classic approach to knocking out a functional gene involves using a deletion cassette composed of a selection marker and flanking homologous arms. Geneticin (G418), nourseothricin, hygromycin, and Zeocin are commonly used selection antibiotics in yeast genetic manipulations (Zheng et al. 2011; Zhu et al. 2024a, b). To assess their applicability for *S. spartinae* transformation, we first evaluated the sensitivity of *S. spartinae* YMxiao to these antibiotics. Our results revealed that YMxiao was more resistant to 200 μg/mL geneticin, 150 μg/mL hygromycin, and 150 μg/mL Zeocin compared to *S. cerevisiae* (Fig. 4A). Conversely, YMxiao was sensitive to 75 μg/mL nourseothricin (Fig. 4A). This prompted us to investigate whether the *NAT* gene could effectively confer nourseothricin resistance in *S. spartinae*. Because the CUG codon is reassigned from Leu to Ser in the "CTG" yeast clade (including *Scheffersomyces* strains) (Papon et al. 2014), we de novo synthesized the *NAT* gene with the original CTG codon replaced by TTG, CTT, or CTA. *TEF1* promoter and terminator were used to regulate the expression of *NAT* gene (Fig. 4B) (Goldstein and McCusker 1999). We amplified the 500-bp upstream and 500-bp downstream sequences of the *URA3* gene (on chromosome 2) to construct a deletion cassette, incorporating the modified *NAT* gene through overlap PCR (Fig. 4B). This resulting cassette was transformed into *S. spartinae* YMxiao cells, and transformants were screened on plates containing 75 μg/mL nourseothricin. This transformation was repeated three times, yielding an average transformation rate of 0.9×10^{-6} transformants per cell. PCR experiments using the primers shown in Fig. 4B confirmed that 88 out of 90 transformants had integrated the deletion cassette at the correct sites. This demonstrates that the modified *NAT* gene effectively allows for the selection of genetically modified transformants in nourseothricin-containing media.

500-bp homologous arms have effectively facilitated homologous recombination in *S. spartinae*

To determine if other *S. spartinae* genes can be replaced by the *NAT* gene, we constructed a deletion cassette for the gene g4934.t1 by overlap PCR using the primers in Table 1. We evaluated the transformation efficiency with different lengths of homologous arms. Our results showed that using 1000-bp and 500-bp homologous arms achieved transformation rates of 3.4×10^{-6} and 1.2×10^{-6} transformants per cell, respectively (Fig. 4C). In contrast, 100-bp and 250-bp homologous arms did not yield any transformants on the selection plates (Fig. 4C). These findings indicate that 500-bp homologous

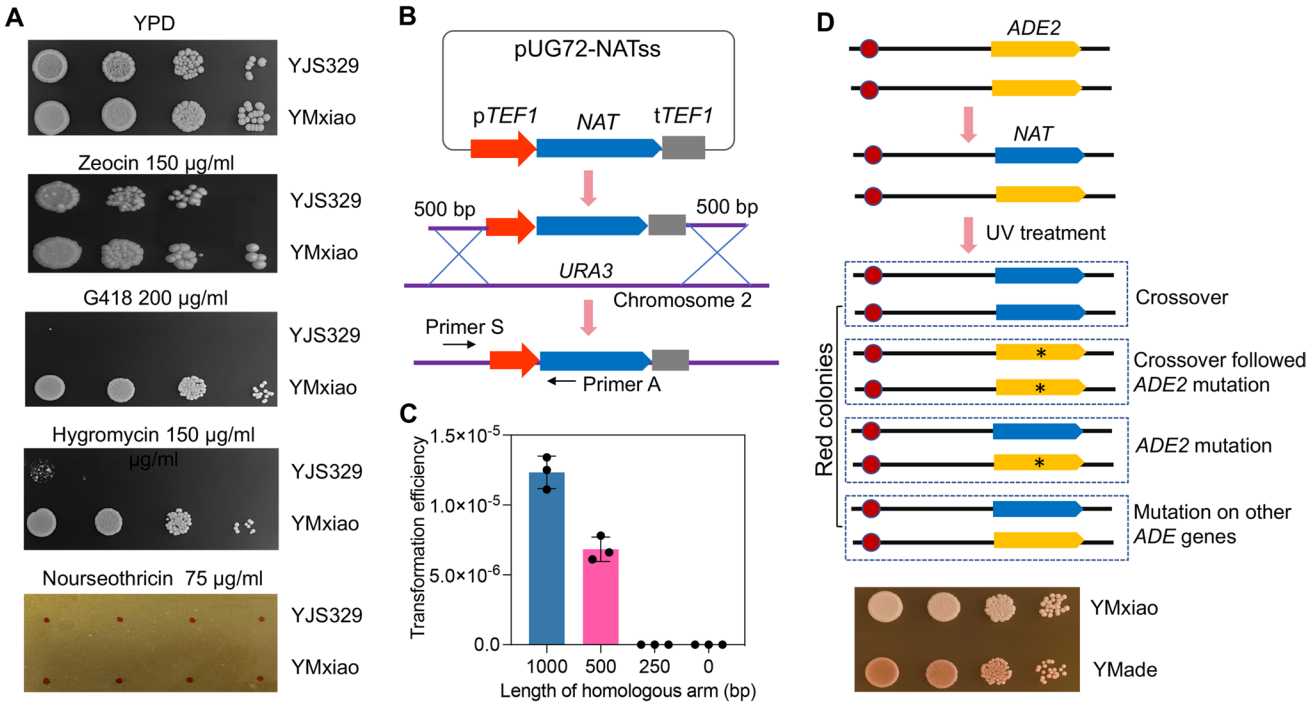

Fig. 4 Development of genetic manipulation methods in *S. spartinae*. **A** Comparison of drugs resistance between *S. spartinae* YMxiao and *S. cerevisiae* YJS329. **B** The plasmid pUG72-NATss was constructed to facilitate the gene knockout in *S. spartinae*. The upstream homologous arm of a target gene (*URA3*), *NAT* gene, and the downstream homologous arm of a target gene were ligated to a deletion cassette by overlap PCR. Primers S and A (Table 1) were used to verify the *URA3* deletion. **C** Different transformation efficiencies resulted from the various lengths of homologous arms of gene g4934.t1 knockout cassette. **D** Generation of a diploid *S. spartinae ade2/ade2* mutant. One copy of *ADE2* gene in YMxiao was deleted using *NAT* gene as the selection mark. The resulted mutant was treated with UV to stimulate chromosomal crossover and mutation. One red colony appeared on YPD plates after UV treatment, which was named as YMade, lost the *NAT* gene due to crossover followed the *ADE2* gene was mutated

arms can effectively facilitate gene knockout in *S. spartinae* via homologous recombination, with longer arms proving to be more effective.

Generation of an *ade2* mutant of *S. spartinae* strain

The *ADE2* gene encodes the enzyme phosphoribosylaminoimidazole carboxylase, which is involved in the adenine biosynthesis pathway (Gedvilaite and Sasnauskas 1994). A notable feature of the *ADE2* gene is its role as a reporter in yeast genetics (Ugolini and Bruschi 1996; Zheng et al. 2016). When *ADE2* is non-functional due to mutations or deletions, yeast colonies typically appear pink or red on medium containing limiting adenine (Zhang et al. 2022). To obtain an *ade* mutant of *S. spartinae*, we first constructed a *ADE2* deletion cassette (ADE2upstream arm-*NAT*-ADE-2downstream arm) to delete one copy of the *ADE2* gene on chromosome 6. This resulted in a transformant with one wild-type *ADE2* allele and one *NAT* allele on the opposite homolog (Fig. 4D). Following this, UV radiation was used to stimulate chromosome crossover between centromere and *ADE2* to eliminate the remaining copy of the *ADE2* gene in the daughter cells (Fig. 4D). After UV treatment, we selected for red colonies on YPD plates and observed a frequency of 4×10^{-5} red colonies per cell. PCR amplification and Sanger sequencing of the *ADE2* ORF in 8 red colonies revealed that 2 of them contained base substitutions that led to missense mutations in the *ADE2* gene. Among these 2 *ade* mutants, one had lost the *NAT* gene due to chromosomal crossover followed *ADE2* was mutated (the T at position 1042 bp was substituted to C), and was designated as YMade (Fig. 4D). One red mutant lost the entire *ADE2* allele due to a crossover event, resulting in the strain harboring two copies of the *NAT* gene (Fig. 4D). The other 4 red mutants likely have mutations in other genes involved in adenine synthesis. Whole-genome sequencing of 2 of these 4 mutants revealed nonsense mutations in either *ADE1* or *ADE8*.

To validate the function of *ADE2* in gene knockout in *S. spartinae* YMxiao, we constructed a *URA3* deletion cassette containing the *ADE2* gene, using the primers listed in Table 1. Although we successfully knocked out one copy of *URA3*, attempts to replace the second *URA3* copy with *NAT* did not produce the expected uracil synthesis-defective mutant. This failure suggests that *URA3* may be essential

for cell viability in this strain, as complete deletion could be lethal. In contrast, in *S. cerevisiae*, *URA3* is non-essential because the organism can import uracil via the membrane transport protein Fur4 (Fujimura 1998). We could not identify a homologous Fur4 protein encoding gene in the genome of *S. spartinae* YMxiao, which may explain why *URA3* is essential in this strain.

Overall, our results demonstrated that code-optimized *NAT* genes with > 500-bp homologous arms were effective for gene knockout in *S. spartinae*. Additionally, we successfully obtained a diploid *ade/ade* mutant of *S. spartinae* YMade that can utilize both *ADE2* and *NAT* as selection markers, which facilitates genetic manipulation in this yeast species.

Discussion

Previous studies have demonstrated several valuable traits in *S. spartinae*, including salt tolerance, coenzyme Q9 production, azo dye decolorization, and biocontrol capabilities against gray mold in strawberries (Kurtzman and Suzuki 2010; Tan et al. 2016; Zou et al. 2021; Breyer et al. 2023). Despite these promising characteristics, the genetic features of *S. spartinae* remain underexplored, highlighting a significant gap in our understanding of this yeast's full potential. This study investigates the phenotypic and genomic characteristics of the diploid yeast *S. spartinae* and delves into its genetic manipulation. Our main findings include the following: (1) the diploid *S. spartinae* exhibited greater resistance to certain stressful conditions compared to *S. cerevisiae* and was capable of growing using multiple monosaccharides as its sole carbon source. (2) We obtained a complete genome of *S. spartinae* YMxiao that encodes over 5300 protein-encoding genes. (3) The patterns of genomic variations among different *S. spartinae* strains were disclosed. (4) We successfully established a genetic manipulation system to knockout the genes in *S. spartinae* YMxiao. Below we will discuss the implications of our results.

S. spartinae strains, including YMxiao, were typically isolated from marine environments (Villarreal et al. 2021). The higher biomass formation capability of YMxiao compared to *S. cerevisiae* YJS329 under alkaline conditions (Fig. 2G) suggests an evolutionary adaptation to its natural habitat. Additionally, our results indicated that *S. spartinae* can metabolize a diverse range of carbon substrates, including arabinose, sorbitol, and mannitol, which are not utilized by *S. cerevisiae* (Fig. 2). The ability to utilize these non-optimal carbon sources has economic significance for the industrial production of valuable products. For instance, the marine oleaginous yeast *Rhodosporidiobolus fluvialis* strain Y2 can utilize mannitol, a major carbohydrate found in brown macroalgae, to produce lipids (Nakata et al. 2020). We also found that genes involved in coenzyme Q9 synthesis, particularly *COQ9*, are present in the YMxiao genome (Supplemental Dataset S1), suggesting that this strain may be capable of producing this coenzyme. In Supplemental Dataset S4, we demonstrated that *S. spartinae* YMxiao could effectively consume 90 g/L glucose and produce 41.1 g/L ethanol within 40 h, with a yield of 0.46 g ethanol/g glucose. This yield is very close to that observed in the bioethanol production microbe, *S. cerevisiae* (Zheng et al. 2012). Given the high NaCl tolerance of *S. spartinae*, this yeast may be a promising candidate for bioethanol production using seawater instead of freshwater.

In combination with next-generation sequencing and nanopore technology, we obtained a complete genome of *S. spartina*, including telomere-to-telomere assembled chromosomes and the mitochondrial sequence. We found the genome size, average ORF length, GC content, and the number of introns containing genes of *S. spartinae* are similar to that in the classical yeast *S. cerevisiae* (Goffeau et al. 1996). In addition, we revealed that the diploid genome of *S. spartinae* YMxiao is highly homozygous, indicating this strain underwent frequent meiosis in its natural habits. Genomic annotation allowed the identification of genes involved in monosaccharide utilization (Fig. 3). We found that certain enzymes in monosaccharide metabolism were encoded by more than one gene. For instance, genes g4934.t1 and g4935.t1 in *S. spartinae* YMxiao were predicted to code NADP-dependent mannitol dehydrogenase. The relative contributions of these two genes to mannitol metabolism and how they were regulated at the transcription level remains to be determined. Comparative genomics revealed abundant point mutations, including SNVs and InDels between *S. spartinae* strains (Supplemental Dataset S2). The frequency of these point mutations was higher than that among *S. cerevisiae* strains (Zhang et al. 2016; Lee et al. 2022 O'Donnell et al. 2023), suggesting a longer divergence time among *S. spartinae* strains. We identified numerous point mutations, including 143 premature stop codons resulting from SNVs and 1607 frameshift InDels, which likely contribute to the phenotypic diversity between *S. spartinae* strains. C > T or G > A substitutions were identified as the most frequent types of base substitutions, likely due to DNA methylation and spontaneous base deamination (Chatterjee and Walker 2017). Specifically, cytosine (C) can be methylated by DNA methyltransferase and *S*-adenosylmethionine to form 5-methylcytosine (5mC). Deamination of 5mC, which involves the loss of its exocyclic amine group, converts it to T. Additionally, cytosine can lose its exocyclic amine to become U, which pairs with adenine A, leading to C > T substitutions during subsequent rounds of DNA replication. Overall, the high-quality genome assemble of *S. spartinae* provides a foundation for further exploration and comparative genomics studies of yeasts.

One significant advantage of using *S. cerevisiae* as a model organism is its exceptional ease of molecular genetic manipulation (Ostergaard et al. 2000). Specifically, in *S. cerevisiae*, gene knockouts can be efficiently achieved using knockout cassettes with homologous arms as short as 40 bp (Gueldener et al. 2002). This efficiency is attributed to the species' robust homologous recombination capabilities. In contrast, other yeast species, such as the classical marine yeast *Yarrowia lipolytica*, which is employed in the production of high-value biochemicals, necessitate significantly longer homologous arms for effective gene knockout or DNA integration (Abdel-Mawgoud and Stephanopoulos 2020; Ji et al. 2020; Cui et al. 2021). Moreover, the pronounced non-homologous end-joining capability in *Y. lipolytica* facilitates the integration of knockout cassettes into non-target genomic locations, even when the homologous arms are longer than 1 kb (Abdel-Mawgoud and Stephanopoulos 2020; Ji et al. 2020; Cui et al. 2021). In our study, we found that homologous arms exceeding 500 bp were sufficient for targeted gene knockout in *S. spartinae* without the aid of other gene editing tools, such as the clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 system. This suggests that the homologous recombination efficiency of this strain is higher than that of *Y. lipolytica*.

A notable challenge in the development of genetic manipulation methods is the robustness of *S. spartinae*, which confers higher resistance to antibiotics such as geneticin, Zeocin, and hygromycin, commonly used in the screening of yeast transformants (Fig. 4A). Furthermore, the unique codon usage pattern of *S. spartinae* required codon optimization for the antibiotic-resistant reporter genes. Our results indicate that the codon-optimized *NAT* gene is an effective selectable marker for *S. spartinae*, establishing a crucial foundation for subsequent genetic studies of this species. Additionally, we successfully developed a diploid *ade2*-deficient *S. spartinae* strain by employing a combination of gene knockout techniques and UV-induced chromosomal crossover (Fig. 4D). The functional deficiency of Ade2 results in the inability of the strain to grow on media lacking adenine and induces red colony formation under adenine-limited conditions (Zhang et al. 2022). This phenotype facilitates visual selection and enhances the efficiency of genetic manipulation experiments. Our results also demonstrated that UV treatment, combined with screening based on colony color changes, is effective for isolating other *ade* mutants, such as *ade1* and *ade8*. The genes *ADE1* and *ADE8* may be more convenient to clone and integrate into the genome compared to *ADE2*, as their CDS length is significantly shorter than that of *ADE2*. The *URA3* gene, unlike in *S. cerevisiae* (Li et al. 2024) and *Yarrowia lipolytica* (Cui et al. 2021), is crucial for the survival of *S. spartinae*. One explanation for its essentiality is likely that *S. spartinae* lacks an uracil membrane transporter, which is encoded by the *FUR4* gene in *S. cerevisiae*. The expression of a heterozygous *FUR4* gene in *S. spartinae* may enable the deletion of both copies of *URA3* in this species, allowing *URA3* to be used as a reporter gene. Overall, the successful application of the codon-optimized *NAT* for gene deletion and the construction of *ade* mutants in *S. spartinae* provides valuable insights for developing genetic manipulation protocols in robust diploid wild-type yeasts.

In summary, this study explored the phenotypic and genomic characteristics of *S. spartinae* strains and established a genetic manipulation protocol for this species, which contributes to the genetic knowledge of yeasts and pave the way for future investigations on its biotechnological applications.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00253-024-13382-1.

Author contribution AS, KL, and DZ conceived and designed research. AS, XL, JY, PY, LQ, and MH conducted experiments and analyzed data. AS, KL, XL, and DZ wrote the manuscript. All authors read and approved the manuscript.

Funding This study was supported by the National Natural Science Foundation of Zhejiang Province (LDT23D06022D06), the Fundamental Research Funds for the Central Universities (226–2024-00019), and the Key Research and Development Program of Hainan Province (ZDYF2024SHFZ046).

Declarations

Ethics approval This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest The authors declare no competing interests.

References

Abdel-Mawgoud A, Stephanopoulos G (2020) Improving CRISPR/Cas9-mediated genome editing efficiency in *Yarrowia lipolytica* using direct tRNA-sgRNA fusions. Metab Eng 62:106–115

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge

Atzmüller D, Ullmann N, Zwirzitz A (2020) Identification of genes involved in xylose metabolism of *Meyerozyma guilliermondii* and their genetic engineering for increased xylitol production. AMB Exp 10:78

Bang S, Pazirandeh M (1999) Physical properties and heavy metal uptake of encapsulated *Escherichia coli* expressing a metal binding gene (NCP). J Microencapsul 16:489–499

Barros KO, Mader M, Krause DJ, Pangilinan J, Andreopoulos B, Lipzen A, Mondo SJ, Grigoriev IV, Rosa CA, Sato TK (2024) Oxygenation influences xylose fermentation and gene expression in the yeast genera *Spathaspora* and *Scheffersomyces*. Biotechnol Biof Biop 17:20

Breyer E, Espada-Hinojosa S, Reitbauer M, Karunarathna SC, Baltar F (2023) Physiological properties of three pelagic fungi isolated from the Atlantic Ocean. J Fungi 9:439

Chatterjee N, Walker GC (2017) Mechanisms of DNA damage, repair, and mutagenesis. Environ Mol Mutagen 58:235–263

Chen Y, Nie F, Xie S-Q, Zheng Y-F, Bray T, Dai Q, Wang Y-X, Xing J-f, Huang Z-J, Wang D-P (2020) Fast and accurate assembly of Nanopore reads via progressive error correction and adaptive read selection. BioRxiv: 2020.2002. 2001.930107

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly 6:80–92

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676

Cui Z, Zheng H, Zhang J, Jiang Z, Zhu Z, Liu X, Qi Q, Hou J (2021) A CRISPR/Cas9-mediated, homology-independent tool developed for targeted genome integration in *Yarrowia lipolytica*. Appl Environ Microbiol 87:e02666-e2620

Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, de Montigny J, Marck C, Neuvéglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich J-M, Beyne E, Bleykasten C, Boisramé A, Boyer J, Cattolico L, Confanioleri F, de Daruvar A, Despons L, Fabre E, Fairhead C, Ferry-Dumazet H, Groppi A, Hantraye F, Hennequin C, Jauniaux N, Joyet P, Kachouri R, Kerrest A, Koszul R, Lemaire M, Lesur I, Ma L, Muller H, Nicaud J-M, Nikolski M, Oztas S, Ozier-Kalogeropoulos O, Pellenz S, Potier S, Richard G-F, Straub M-L, Suleau A, Swennen D, Tekaia F, Wésolowski-Louvel M, Westhof E, Wirth B, Zeniou-Meyer M, Zivanovic I, Bolotin-Fukuhara M, Thierry A, Bouchier C, Caudron B, Scarpelli C, Gaillardin C, Weissenbach J, Wincker P, Souciet J-L (2004) Genome evolution in yeasts. Nature 430:35–44

Förstemann K, Lingner J (2001) Molecular basis for telomere repeat divergence in budding yeast. Mol Cell Biol 21:7277–7286

Fujimura H (1998) Growth inhibition of *Saccharomyces cerevisiae* by the immunosuppressant leflunomide is due to the inhibition of uracil uptake via Fur4p. Mol Gen Genet 260:102–107

Gedvilaite A, Sasnauskas K (1994) Control of the expression of the *ADE2* gene of the yeast *Saccharomyces cerevisiae*. Curr Genet 25:475–479

Gietz RD, Schiestl RH (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc 2:31–34

Gillings M, Holley M, Selleck M (2006) Molecular identification of species comprising an unusual biofilm from a groundwater treatment plant. Biofilms 3:19–24

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M (1996) Life with 6000 genes. Science 274:546–567

Goldstein AL, McCusker JH (1999) Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. Yeast 15:1541–1553

Gueldener U, Heinisch J, Koehler G, Voss D, Hegemann J (2002) A second set of *loxP* marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. Nucleic Acids Res 30:e23

Hu J, Fan J, Sun Z, Liu S (2020) NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics 36:2253–2255

Ji Q, Mai J, Ding Y, Wei Y, Ledesma-Amaro R, Ji X-J (2020) Improving the homologous recombination efficiency of *Yarrowia lipolytica* by grafting heterologous component from *Saccharomyces cerevisiae*. Meta Eng Commun 11:e00152

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22:568–576

Kurtzman CP, Suzuki M (2010) Phylogenetic analysis of ascomycete yeasts that form coenzyme Q-9 and the proposal of the new genera *Babjeviella*, *Meyerozyma*, *Millerozyma*, *Priceomyces*, and *Scheffersomyces*. Mycoscience 51:2–14

Kwak S, Jin Y-S (2017) Production of fuels and chemicals from xylose by engineered *Saccharomyces cerevisiae*: a review and perspective. Microb Cell Fact 16:82

Lee TJ, Liu Y-C, Liu W-A, Lin Y-F, Lee H-H, Ke H-M, Huang J-P, Lu M-YJ, Hsieh C-L, Chung K-F (2022) Extensive sampling of *Saccharomyces cerevisiae* in Taiwan reveals ecology and evolution of predomesticated lineages. Genome Res 32:864–877

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079

Li C, Shi K, Zhang Y, Wang G (2019) *Nocardioides silvaticus* sp. nov., isolated from forest soil. Int J Syst Evol Microbiol 69:68–73

Li K-J, Qi L, Zhu Y-X, He M, Xiang Q, Zheng D-Q (2024) Spontaneous and environment induced genomic alterations in yeast model. Cell Insight 4:100209

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997

Lowe TM, Chan PP (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44:W54–W57

Madden K, Snyder M (1998) Cell polarity and morphogenesis in budding yeast. Annu Rev Microbiol 52:687–744

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A (2018) MUMmer4: a fast and versatile genome alignment system. PLoS Comp Biol 14:e1005944

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:10–12

Meena M, Prasad V, Zehra A, Gupta VK, Upadhyay RS (2015) Mannitol metabolism during pathogenic fungal-host interactions under stressed conditions. Front Microbiol 6:1019

Nakata S, Hio M, Takase R, Kawai S, Watanabe D, Hashimoto W (2020) Polyunsaturated fatty acids-enriched lipid from reduced sugar alcohol mannitol by marine yeast *Rhodosporidiobolus fluvialis* Y2. Biochem Bioph Res Co 526:1138–1142

O'Donnell S, Yue J-X, Saada OA, Agier N, Caradec C, Cokelaer T, De Chiara M, Delmas S, Dutreux F, Fournier T (2023) Telomere-to-telomere assemblies of 142 strains characterize the genome structural landscape in *Saccharomyces cerevisiae*. Nat Genet 55:1390–1399

Ostergaard S, Olsson L, Nielsen J (2000) Metabolic engineering of *Saccharomyces cerevisiae*. Microbiol Mol Biol Rev 64:34–50

Papini M, Nookaew I, Uhlén M, Nielsen J (2012) *Scheffersomyces stipitis*: a comparative systems biology study with the Crabtree positive yeast *Saccharomyces cerevisiae*. Microbial Cell Fac 11:136

Papon N, Courdavault V, Clastre M (2014) Biotechnological potential of the fungal CTG clade species in the synthetic biology era. Trends Biotechnol 32:167–168

Qi L, Sui Y, Tang X-X, McGinty RJ, Liang X-Z, Dominska M, Zhang K, Mirkin SM, Zheng D-Q, Petes TD (2023) Shuffling the yeast genome using CRISPR/Cas9-generated DSBs that target the transposable Ty1 elements. PLoS Genet 19:e1010590

Riehl K, Riccio C, Miska EA, Hemberg M (2022) TransposonUltimate: software for transposon classification, annotation and detection. Nucleic Acids Res 50:e64–e64

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B (2006) AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res 34:W435–W439

Tan L, He M, Song L, Fu X, Shi S (2016) Aerobic decolorization, degradation and detoxification of azo dyes by a newly isolated salt-tolerant yeast *Scheffersomyces spartinae* TLHS-SF1. Bioresour Technol 203:287–294

Trichez D, Steindorff AS, de Morais Júnior WG, Vilela N, Bergmann JC, Formighieri EF, Gonçalves SB, de Almeida JRM (2023) Identification of traits to improve co-assimilation of glucose and xylose by adaptive evolution of *Spathaspora passalidarum* and *Scheffersomyces stipitis* yeasts. Appl Microbiol Biotechnol 107:1143–1157

Ugolini S, Bruschi CV (1996) The red/white colony color assay in the yeast *Saccharomyces cerevisiae*: epistatic growth advantage of white *ade8-18*, *ade2* cells over red *ade2* cells. Curr Genet 30:485–492

Villarreal AR, Campbell DE, Webster SS, Hunter RC (2021) Draft genome sequence of *Scheffersomyces spartinae* ARV011, a marine yeast isolate. Microbiol Resour Ann 10:e00652-e621

Wick RR, Judd LM, Holt KE (2018) Deepbinner: demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. PLoS Compu Biol 14:e1006583

Zhang K, Zhang L-J, Fang Y-H, Jin X-N, Qi L, Wu X-C, Zheng D-Q (2016) Genomic structural variation contributes to phenotypic change of industrial bioethanol yeast *Saccharomyces cerevisiae*. FEMS Yeast Res 16:fov118

Zhang K, Sui Y, Li W-L, Chen G, Wu X-C, Kokoska RJ, Petes TD, Zheng D-Q (2022) Global genomic instability caused by reduced expression of DNA polymerase ε in yeast. P Natl Acad Sci U S A 119:e2119588119

Zheng D-Q, Wu X-C, Wang P-M, Chi X-Q, Tao X-L, Li P, Jiang X-H, Zhao Y-H (2011) Drug resistance marker-aided genome shuffling to improve acetic acid tolerance in *Saccharomyces cerevisiae*. J Ind Microbiol Biotechnol 38:415–422

Zheng D-Q, Wang P-M, Chen J, Zhang K, Liu T-Z, Wu X-C, Li Y-D, Zhao Y-H (2012) Genome sequencing and genetic breeding of a bioethanol *Saccharomyces cerevisiae* strain YJS329. BMC Genomics 13:479

Zheng D-Q, Zhang K, Wu X-C, Mieczkowski PA, Petes TD (2016) Global analysis of genomic instability caused by DNA replication stress in *Saccharomyces cerevisiae*. P Natl Acad Sci U S A 113:E8114–E8121

Zhu Y, Liu J, Sun L, Liu M, Qi Q, Hou J (2024a) Development of genetic markers in *Yarrowia lipolytica*. Appl Microbiol Biotechnol 108:14

Zhu Y-X, He M, Li K-J, Wang Y-K, Qian N, Wang Z-F, Sheng H, Sui Y, Zhang D-D, Zhang K (2024b) Novel insights into the effects of 5-hydroxymethfurural on genomic instability and phenotypic evolution using a yeast model. Appl Environ Microbiol 90:e01649-e1623

Zou X, Wei Y, Dai K, Xu F, Wang H, Shao X (2021) Yeasts from intertidal zone marine sediment demonstrate antagonistic activities against *Botrytis cinerea* in vitro and in strawberry fruit. Biol Control 158:104612