Genes to Cells

WILEY

Genes to Cells

ORIGINAL ARTICLE OPEN ACCESS

Advantages of Mutant Generation by Genome Rearrangements of Non-Conventional Yeast via Direct Nuclease Transfection

Arisa H. Oda[1,2] | Taishi Yasukawa[3] | Miki Tamura[1] | Ayumu Sano[3] | Naohisa Masuo[3] | Kunihiro Ohta[1,2,4]

[1]Department of Life Sciences, Graduate School of Arts & Sciences, the University of Tokyo, Tokyo, Japan | [2]Collaborative Research Institute for Innovative Microbiology, Tokyo, Japan | [3]Mitsubishi Corporation Life Sciences Limited, Tokyo, Japan | [4]The Universal Biology Institute of the University of Tokyo, Tokyo, Japan

Correspondence: Arisa H. Oda (odar@bio.c.u-tokyo.ac.jp)

Received: 12 December 2024 | **Revised:** 19 February 2025 | **Accepted:** 23 February 2025

Transmitting Editor: Takehiko Kobayashi

Funding: This work was partially supported by Japan Society for the Promotion of Science grant 24H01393, 23K04984, 19K16070 to A.H.O., 24K02068 to K.O., AMED grant number JP20wm0325003 to K.O., and Japan Science and Technology Agency CREST JPMJCR18S3 to K.O., and a research grant by Mitsubishi Corporation Life Sciences Limited, Japan to K.O.

Keywords: genome rearrangement | low pH resistance | mutagenesis | nonconventional yeast | NTG treatment | TAQing2.0 | torula yeast

ABSTRACT

We previously developed a genome engineering method (TAQing2.0) based on the direct delivery of DNA endonucleases into living cells, which induces genome rearrangements even in non-sporulating nonconventional yeasts without introducing foreign DNA. Using TAQing2.0 and conventional mutagenesis (by nitrosoguanidine), we obtained mutant asexual *Candida utilis* strains capable of growing under highly acidic conditions (pH 1.8). Whole genome resequencing revealed that the genomic sequences of mutants generated by both methods contain a negligible small population of unmappable sequences, suggesting that both types of mutants can be regarded as equivalent to naturally occurring mutants. TAQing2.0 mutants exhibit multiple genome rearrangements with few point mutations, whereas conventional mutagenesis produces numerous point mutations. This feature enabled us to easily identify candidate genes (e.g., *LYP1* homolog) responsible for acid resistance. TAQing2.0 is a powerful and versatile tool for mutant production and gene hunting without invasion of foreign DNA.

1 | Introduction

Conventional mutagenesis, genome editing, and genome shuffling technologies have been powerful tools for producing mutants with useful phenotypes. Conventional mutagenesis with drugs or radiation induces numerous base substitutions, i.e., single nucleotide variations (SNVs) and insertions/deletions (InDels), leading to genetically diverse pools. It has been used for a long time for mutant production and hunting causative genes important for altered useful phenotypes. However, in many cases, the mutant lines produced by conventional mutagenesis contain numerous point mutations at sites other than the gene of interest. It is therefore necessary to repeat time-consuming and laborious backcrossing to remove the effects of the off-target mutations.

We have developed the genome engineering technology named TAQing system based on induced large-scale genome rearrangements. In this system, DNA double-strand breaks (DSBs) are conditionally induced at multiple random

Arisa H. Oda and Taishi Yasukawa contributed equally to this study.

chromosomal locations, which lead to large-scale genome rearrangements often accompanied by phenotypic alterations (Muramoto et al. 2018; Tanaka et al. 2020; Yasukawa et al. 2022; Yone et al. 2022). In the original TAQing technology, the restriction enzyme TaqI (from *Thermus thermophilus* HB8 and recognizes the 4-base sequence "TCGA") is expressed in living cells. The temperature-dependent TaqI endonuclease in the cell is inactive at 30°C but is partially activated by shifting the temperature around 37°C–42°C to induce DSBs at TCGA sites. The resulting DSBs are repaired, followed by the generation of various types of genomic rearrangements. The TAQing system, compared to conventional mutagenesis methods, exhibits a lower incidence of point mutations. Hence, phenotypic changes caused by TAQing often arise due to copy number variations (CNVs), loss of heterozygosity (LOH), translocations (TLs), deletions (Dels), and other mechanisms. These characteristics enable us to easily identify the causative genes associated with the altered phenotypes of the mutants, without the need for time-consuming backcrossing (Yone et al. 2022). In comparison, restriction enzyme-mediated integration, referred to as the REMI method (Kuspa and Loomis 1992), enables linear plasmid DNA integration with a high proportion at genomic restriction sites by restriction enzymes, and this recombination process is intrinsically similar to that of the TAQing system. Therefore, this report supports the practical achievements of mutation occurrences reported in the TAQing system.

In addition, we developed a minimally invasive and versatile system, termed TAQing2.0, utilizing protein-transfection of DNA endonucleases (TaqI) facilitated by cell-penetrating peptides (Yasukawa et al. 2022). The event-triggering mechanism of this technology is based on the original TAQing system, i.e., it temporally activates intracellular TaqI by shifting 38°C to induce DSBs in alive cells. This means that the inducible mutation patterns of TAQing2.0 are similar to those of the TAQing system. On the other hand, the REMI and TAQing2.0 methods are mechanistically different in that the former uses foreign DNA and marker integration, while the latter needs neither of them. The TAQing2.0 system is applicable across a broad range of species that lack established gene expression vectors. Employing TAQing2.0, we successfully generated mutants of nonconventional yeasts, including Torula yeast (*Candida utilis*), which are typically inefficient in sporulation and thereby present challenges in identifying causative genes for altered phenotypes. *C. utilis* has been used commercially as food-grade yeast since it contains rich protein, minerals, and vitamins (Bekatorou et al. 2006) and its genome sequence has been studied (Tomita et al. 2012). Our previous study (Yasukawa et al. 2022) also estimated that the genome structure of *C. utilis* is triploid with six chromosomes, in which one copy of a ~700 kb segment of the right arm of chromosome II exists. Within wild-type triploid chromosome sets, two major alleles tend to have similar single nucleotide variations (SNVs). Thus, a basic platform for breeding *C. utilis* without foreign DNA integration has been established. However, no studies have been conducted exploring the industrial practicality of TAQing2.0 technology so far. In addition, as TAQing2.0 does not necessitate the transfection of foreign DNA/RNA fragments from heterologous species, the resulting mutant strains are considered non-genetically modified. This characteristic is advantageous for gaining social acceptance when mutant strains are used for the production of natural food and cosmetics. However, it remains to be tested whether the mutant strains obtained by TAQing2.0 are essentially equivalent to those produced by conventional methods in terms of non-contaminating foreign genetic information.

In this study, to compare the efficiency of obtaining mutants with desired traits and the easiness of causative gene identification between TAQing2.0 and conventional mutagenesis, we generated *C. utilis* strains with acid resistance using TAQing2.0 and a conventional mutagenesis, N-methyl-N′-nitro-N-nitrosoguanidine (NTG) treatment (Adelberg et al. 1965). Here, we evaluated the efficiency of mutant production, the spectrum of genetic alterations, and the population of foreign and unmappable DNA sequences between these two methodologies. TAQing2.0-generated mutants had an equivalent or lower frequency of foreign and unmappable DNA sequences than conventional mutagenesis. In addition, the incidence of point mutations was significantly lower with TAQing2.0 than with conventional mutagenesis, suggesting that phenotypic changes are mainly due to rearrangements such as LOHs and Dels. This property also helped us to identify candidate gene regions for acid resistance. Altogether, TAQing2.0 serves as a powerful genome engineering technique that does not require the incorporation of foreign genetic information. Additionally, it is an effective gene identification tool, even for species lacking established gene expression systems or the ability to crossbreed.

2 | Results

2.1 | Production of Acid-Resistant Torula Yeast (*C. utilis*) Strains

Wild-type Torula yeast cells cannot grow in highly acidic environments, such as at pH 2.0. However, Torula yeast strains that are resistant to acidic conditions have a significant advantage in industrial fermentation, as they can avoid contamination from other microorganisms. We attempted to obtain *C. utilis* mutants capable of proliferating under highly acidic (pH 1.8) conditions by applying TAQing2.0 or conventional NTG mutagenesis. A mutant library was generated by treating *C. utilis* cells at the early logarithmic phase once with either TAQing2.0 or NTG treatment. For the primary screening, the entire library was inoculated onto minimal synthetic defined (SD) medium agar plates adjusted to pH 1.8 with sulfuric acid, and colonies larger in size than those of the wild-type strain grown under the same acidic conditions were isolated. From the TAQing2.0-treated group and the NTG-treated group, 9 and 63 colonies were selected, respectively. These cells were then inoculated onto SD agar plates (pH 1.8), followed by single colony isolation. By this, we obtained 3 and 63 candidate strains from the TAQing2.0-treated and the NTG-treated groups, respectively (Table 1).

To select stable acid-resistant *C. utilis* strains, the above-mentioned candidate strains from the first screening were subjected to five passages of serial culturing in an SD liquid medium without selection pressure. Among these strains (hereafter referred to as "5 passages"), those retaining the acid resistance

TABLE 1 | Efficiency of low-pH resistant screening by TAQing2.0 and NTG mutagenesis.

		TAQing2.0	NTG mutagenesis
1st screening (2× low pH agar plates)	Library size (cells)	1.3×10^{5}	9.4×10^{6}
	Selected mutant (cells)	3	63
2nd screening (stable phenotype after 5×passages)	Population size (cells)	3	29
	Selected mutant (cells)	3	5
	Gained mutants/library size (%)	2.3×10^{-3}	1.2×10^{-4}

of the original mutant strains (referred to as "ancestors") were subjected to the secondary screening.

Figure 1a shows the growth of the ancestors and the five passages in an SD liquid medium at pH 2.0. When the growth of the three candidate TAQing2.0-treated mutants (referred to as "TAQed m1, m2, and m3") was examined, we confirmed that all of the 5 passage descendants as well as their ancestors grew under acidic conditions (Table S1). Within the NTG-treated candidate group, randomly picked 29 mutants were examined in SD liquid medium at pH 2.0, and five strains #3, #9, #11, #17, and #24 (referred to as "NTGed m3, m9, m11, m17, and m24") were selected, whose 5 passage descendants showed the same growth rate as their ancestors (Figures 1a and S1, Table S1). We also conducted spot assays on SD agar plates adjusted to pH 1.8 with sulfuric acid or hydrochloric acid. All strains demonstrated acid resistance to hydrochloric acid (Figure 1b) as well as sulfuric acid, suggesting that these strains may have enhanced resistance to protons rather than to a specific anion. Interestingly, the growth rate of TAQed m1 was not much different from that of WT under the non-acidic condition (pH 4.6), while that of TAQed m2 and m3 was significantly slow. All of the NTG-treated mutants showed a slight but significant delay (Figure S2).

Table 1 indicates the success rate of obtaining acid-resistant *C. utilis* mutants from the TAQing2.0-treated and the NTG-treated groups. The frequency was 2.3×10^{-3}% for the TAQing2.0-treated group and 1.2×10^{-4}% for the NTG-treated group.

2.2 | Possibility of Foreign DNA Sequence Contamination

For all strains obtained through the secondary screening, whole genome analysis was performed using DNBSEQ-G400 (MGISEQ-2000RS) short-read sequencing. The genomic DNA (gDNA) reads of each strain were mapped using BWA (Li 2013) to the previously reported *C. utilis* reference genome sequence (Yasukawa et al. 2022). In all nine strains, which include the wild type, three TAQing2.0-treated mutants, and five NTG-treated mutants, more than 99.6% of the reads could be mapped to the reference genome sequence by Yasukawa et al. (2022).

In Yasukawa et al., the *C. utilis* reference genome sequence for the wild-type was assembled with MaSuRCA (Zimin et al. 2013) and Falcon (Chin et al. 2016) using a combination of PacBio Sequel long-read and Illumina short-read sequencing. For PacBio long-read sequencing, assembly was performed using Canu (Koren et al. 2017). Unmappable reads of the mutants in this study and the wild-type strain (0.4% of the total reads) were re-mapped using BWA to all the contigs from the above-mentioned assemblies. As a result, 71.0%–92.6% of these previously unmapped reads became mappable (Table 2), suggesting that these newly mapped reads were derived from unique allele sequences other than the reference genome sequence. The remaining unmapped reads was further mapped using the *C. utilis* (NBRC0988) genome sequence (Tomita et al. 2012) and the genome sequence of *Cyberlindnera jadinii* NRRL Y-1542, considered an ancestor of *C. utilis* (Rupp et al. 2015). Ultimately, 99.90%–99.98% of all reads from both the wild-type and the mutant strains were mapped to *C. utilis*-related reference sequences.

For the remaining unmapped reads, we further attempted to map them to foreign DNA sequences derived from the *Escherichia coli* and vector sequences used in constructing the TaqI expression vector. However, the reads mapping to these foreign genomes were less than 4.3×10^{-4}% in both the TAQing2.0-treated and the NTG-treated mutants (Table 2). They had no significant read depth greater than 4× to the *E. coli* or TaqI vector sequences for any of the mutant strain reads while the average depth of *C. utilis* chromosomes were between 1,400×–3,400×.

Additionally, we explored whether there were any common unmapped reads between the wild-type and the acid-resistant mutants. First, we prepared a custom BLAST database using the remaining wild-type unmapped reads. Then, we performed local BLAST searches (Camacho et al. 2009) of the remaining unmapped reads from the TAQing2.0-treated and NTG-treated acid-resistant mutants against this custom BLAST database of wild-type reads. As a result, 12.7%–35.0% and 26.4%–37.7% of the unmapped reads of the TAQing2.0-treated and the NTG-treated had at least one hit against the wild-type custom BLAST database, respectively (Table 3). These results show that 99.971%–99.983% of the reads from TAQing2.0-treated mutants and 99.927%–99.982% of the reads from NTG-treated mutants can be mapped to *C. utilis*-related genome sequences (Table 3). Since the Quality scores of the final remaining unmapped reads were significantly lower than those of the mappable reads to the genome reference in all groups (Figure S3), it is suggested that the final remaining sequences may contain sequencing-derived specific errors. This demonstrates that the likelihood of foreign DNA contamination in TAQing2.0-treated mutants is negligible, comparable to that in NTG-treated mutants.

2.3 | Analysis of Genome Rearrangements in the Acid-Resistant Mutants

Next, we analyzed the genome rearrangements in the acid-resistant mutants. First, changes in chromosome lengths of

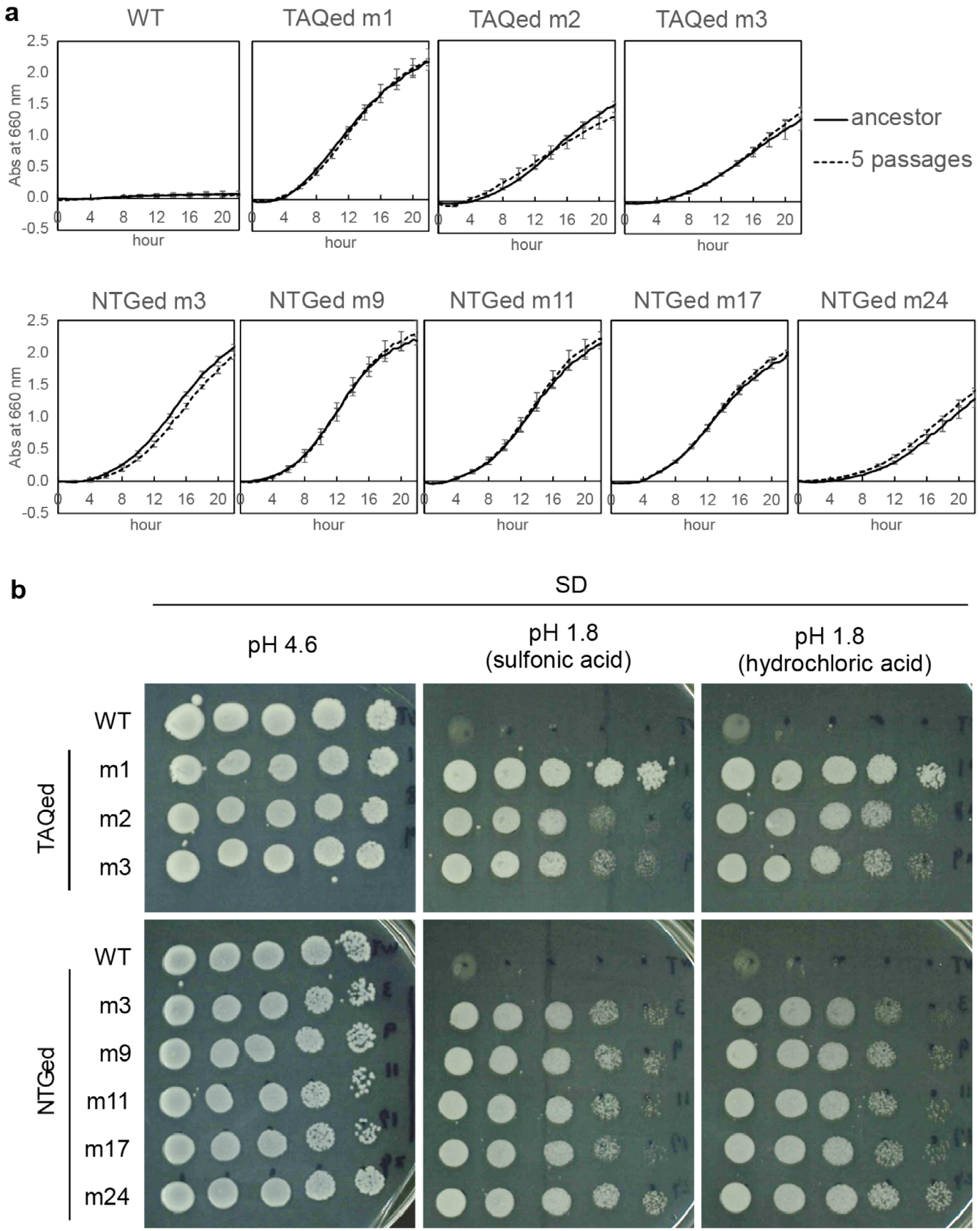

FIGURE 1 | Schematic diagram of TAQing2.0 technology and selection of acid-tolerant mutants using TAQing2.0 or NTG mutagenesis. (a) Comparison of the growth curves between ancestor and 5-passage. Strain was cultured in SD (pH 2.0) at 30°C. Error bars represent standard deviation from three independent experiments. (b) A spot test for observing the growth of CuWT and acid-tolerant mutants under non-stress or acidic environment. Cells were spotted onto SD (pH 4.6) or SD (pH 1.8), adjusted with sulfonic acid or hydrophobic acid. Two independent experiments were performed and a representative image is shown.

TABLE 2 | Number and ratio of reads that were mapped and remaining unmapped to *C. utilis* related genomic sequence and the *E. coli* and vector sequences used in constructing the TaqI expression vector.

			TAQed						NTGed									
	WT		m1		m2		m3		m3		m9		m11		m17		m24	
Num. of total reads	1.4523×10^8		1.3083×10^8		1.5229×10^8		6.6524×10^7		6.7558×10^7		6.3405×10^7		7.3500×10^7		7.5352×10^7		7.2646×10^7	
Mapped to *C. utilis* refs.	1.4520×10^8		1.3081×10^8		1.5226×10^8		6.6494×10^7		6.7529×10^7		6.3362×10^7		7.3427×10^7		7.5304×10^7		7.2601×10^7	
Remaining unmapped	36,294		27,178		29,176		29,914		29,398		42,428		72,794		48,076		44,818	
Mappable to *E. coli* or vector	14		6		12		79		150		225		318		194		181	
Remaining unmapped	36,280		27,172		29,164		29,835		29,248		42,203		72,476		47,882		44,637	
Read ratio/ total reads (%)	mapped	remaining	mapped	remaining	mapped	remaining	mapped	remaining	mapped	remaining	mapped	remaining	mapped	remaining	mapped	remaining	mapped	remaining
C. utilis genome	99.729	0.2713	99.729	0.2709	99.728	0.2724	99.704	0.2959	99.716	0.2842	99.689	0.3106	99.655	0.3446	99.697	0.3026	99.696	0.3038
C. utilis contigs	0.2456	0.0258	0.2494	0.0216	0.2524	0.0200	0.2500	0.0460	0.2402	0.0442	0.2428	0.0678	0.2449	0.0998	0.2382	0.0645	0.2414	0.0624
NBRC0988	0.0005	0.0252	0.0006	0.0211	0.0006	0.0194	0.0008	0.0452	0.0004	0.0437	0.0007	0.0671	0.0004	0.0994	0.0004	0.0641	0.0006	0.0619
C. jadinii	0.0002	0.0250	0.0003	0.0208	0.0003	0.0192	0.0002	0.0450	0.0002	0.0435	0.0002	0.0669	0.0003	0.0990	0.0003	0.0638	0.0002	0.0617

TABLE 3 | BLAST result of mutants' remaining unmapped reads and ratio of final unmappable and identified reads of TAQed and NTG mutants.

	TAQed			NTGed				
	m1	**m2**	**m3**	**m3**	**m9**	**m11**	**m17**	**m24**
Remaining unmapped (BLAST query)	27,172	29,164	29,835	29,248	42,203	72,476	47,882	44,637
Num. of reads with more than 1 hit	3524	3716	10,452	10,142	15,912	19,184	17,296	13,464
Read with hit/query reads (%)	13.0	12.7	35.0	34.7	37.7	26.5	36.1	30.2
Total reads	1.308×10^8	1.523×10^8	6.652×10^7	6.756×10^7	6.340×10^7	7.350×10^7	7.535×10^7	7.265×10^7
Num. of unknown reads	23,648	25,448	19,383	19,106	26,291	53,292	30,586	31,173
Unmappable read/total reads (%)	0.018	0.017	0.029	0.028	0.041	0.073	0.041	0.043
Mapped read/total reads (%)	99.982	99.983	99.971	99.972	99.959	99.927	99.959	99.957

the mutants were analyzed by pulsed-field gel electrophoresis (PFGE). Extra bands corresponding to newly emerging shorter or longer chromosomes were observed in the NTG-treated mutants m11 and m17, as well as in the TAQing2.0-treated mutants m2 and m3 (Figure S4, open arrowheads), demonstrating that large-scale chromosome rearrangements occurred in these mutants.

We then analyzed the gDNA sequencing data to identify the sites of point mutations and rearrangements in the TAQing2.0-treated and NTG-treated mutants. We observed single nucleotide variations (SNVs), insertions/deletions (InDels), homologous recombinations (HRs), break-induced replications (BIRs), TLs, aneuploidy, and large Dels (Figure 2a–c). TLs generally exhibit discontinuous changes in sequence coverage and are often observed as instances where different chromosomal sequences are linked through the TaqI recognition sequence "TCGA" at the breakpoints (Muramoto et al. 2018; Yasukawa et al. 2022). Indeed, in the TAQing2.0-treated strain m3, two inter-chromosomal TLs were observed. One TL occurred between the left arm of chromosome I (ChrI) near the rDNA repeat region and the left arm of chromosome VI (ChrVI), where the TaqI recognition sequence "TCGA" was present at the breakpoint (Figures 2b and 3a). The connection between these chromosomes was confirmed by PCR amplification using primers around the breakpoints on ChrI and ChrVI (Figure 3b). Another TL was observed on ChrV (Figure 3c). CNVs were found on ChrV, which were observed as ×4 coverage in the left arm and ×2 coverage in the right arm. One of the major allele chromosomes lacks its right arm, and the minor allele has an aneuploidy which also lacks its right arm. These right ends were attached as a TL (Figure 2b). This TL occurred at the TaqI recognition sequence, and the connection of both chromosomes was confirmed by PCR amplification (Figure 3d).

Furthermore, large Dels involving extensive coverage changes were detected in TAQing2.0-treated m2, m3, NTG-treated m11, and m17, notably a large Del on the left arm of chromosome IV (ChrIV) (Figure 2a–c). These deletions were verified by PCR amplification using primers near the chromosome breakpoints and telomeric sequence primers (Figure 3e). TAQing2.0-treated m2 and m3 mutants harbored large Dels that occurred on different alleles of the right arm of chromosome II (ChrII) (Figures 3e and S5).

Next, we analyzed point mutations and HR events. Since wild-type *C. utilis* is triploid, heterozygous SNVs generally show allele distributions at ratios of 1/3 or 2/3 throughout the genome (Figure S6a). Point mutations were defined as newly generated SNVs without recombination templates (Novel SNVs; Figure S6b). DSBs often induce copy-paste type HR such as gene conversion (GC) using homologous sequences as templates, which can overwrite some SNVs and alter local SNV ratios. Chromosomal regions exhibiting these local SNV ratio changes were defined as GC-type HR sites. For example, when a sequential local SNV ratio changed from 2:1 to 1:2, one of the two major alleles was seemingly repaired by GC using the minor allele as a template (Figure S6c). Especially when a sequential local SNV ratio changed from 2:1 to 3:0, it can be regarded as LOH-type HR whose minor allele was repaired using major alleles as a template (Figure S6d). Note that even a single SNV that had a repair template was provisionally regarded as HR. In regions with altered coverage due to genome rearrangements (e.g., regions with decreased coverage due to Dels or CNVs), we used different thresholds for the identification of SNVs, taking into account the number of alleles at the loci (ex. 1/2, 3/4 as altered allele ratio; Figures S6e, S7 and S8a,b). We also set different thresholds for the short ChrII fragment which has a ratio of 2:2 allele separation with telomeres at both ends observed even in the wild-type.

Through these analyses, we found that the TAQing2.0-treated mutants had no InDels and only 1–4 newly occurring SNVs per strain. In contrast, NTG-treated mutants exhibited a small number of InDels and 503-1,402 new SNVs per strain (Table 4,

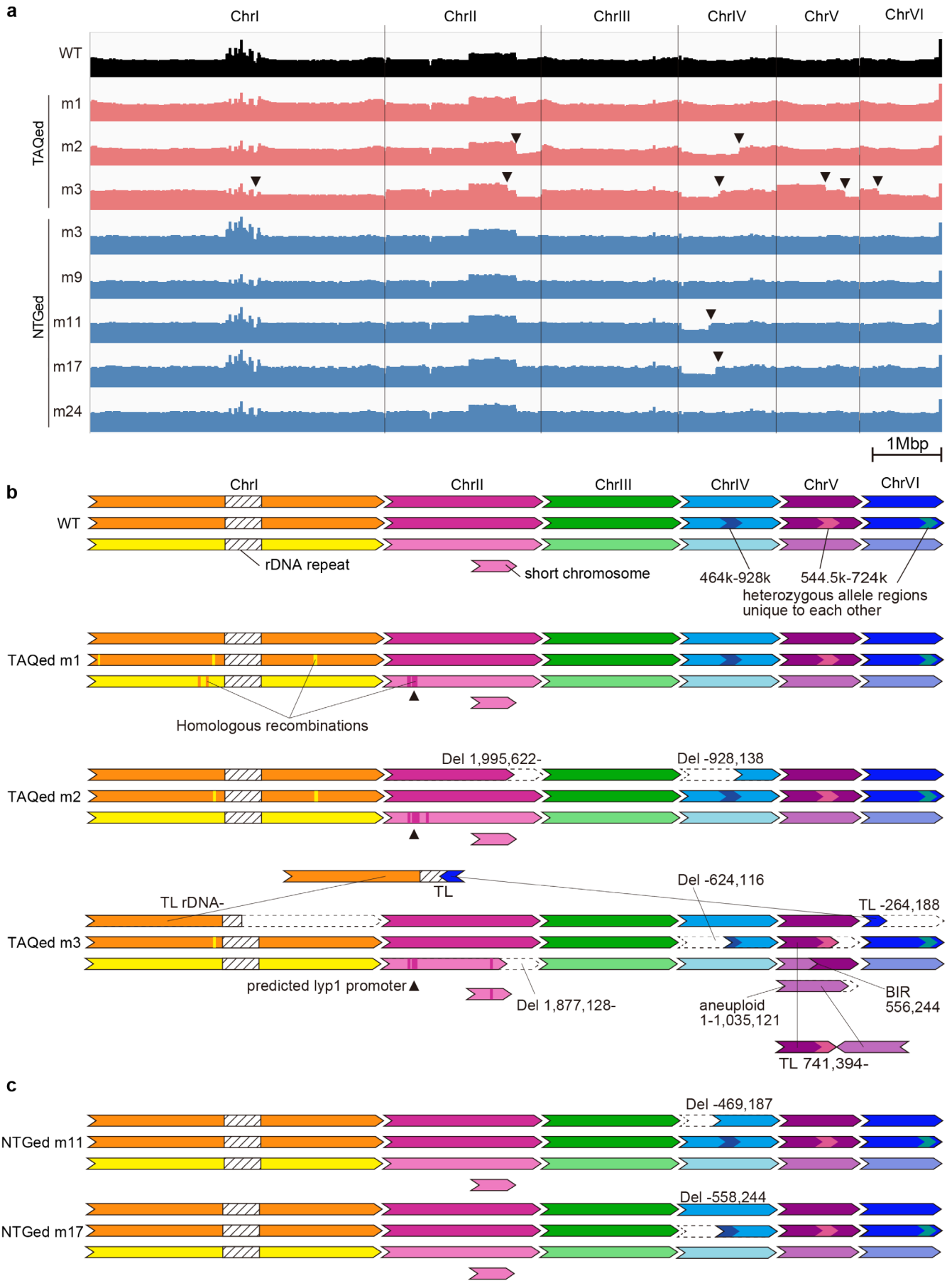

FIGURE 2 | Legend on next page.

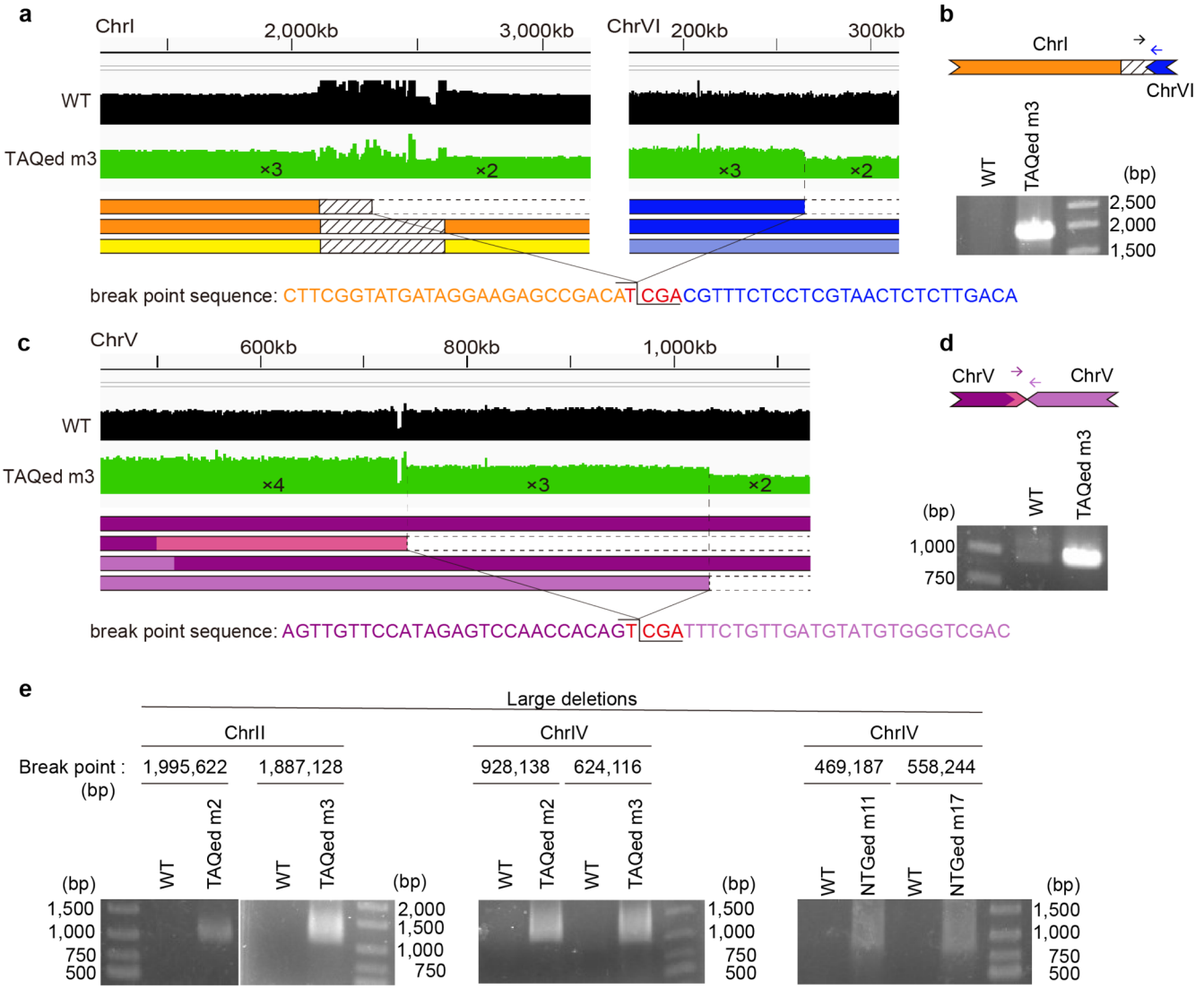

FIGURE 3 | Translocations and large deletions in the mutants. (a) Breakpoint sequence of inter-chromosomal translocated chromosomes between ChrI and Chr VI in the TAQed m3 strain. (b) Local PCR amplification check of (a). Arrows indicate primers for PCR. (c) Breakpoint sequences of inter-chromosomal translocated chromosomes within ChrV in the TAQed m3 strain. (d) Local PCR amplification check of (c). Arrows indicate primers for PCR. (e) Local PCR amplification check of large deletions in TAQed m2, m3, NTGed m11 and 17.

Data S1). Notably, a single SNV was commonly detected in all three acid-resistant TAQing2.0-treated mutants, whereas 11 SNVs and one InDel were commonly found across the five NTG-treated mutants (Table S2). Additionally, we detected 5–7 HR events in TAQing2.0-treated mutants, while NTG-treated mutants exhibited 73–393 HR events. Three HR events were consistently detected in all three TAQing2.0-treated mutants, two of which occurred discontinuously at very close positions on ChrII around 199,988 and 204,000 bp. In contrast, 24 common HR events were observed among the NTG-treated mutants (Table S3).

Since NTG induces alkylation damage, its mutation spectrum predominantly features G/C to A/T transitions (Harper and

FIGURE 2 | Genome rearrangements in TAQed mutants. (a) Overview of the whole genome sequencing for WT (black), three TAQing2.0-treated mutants (red) and five NTG-treated mutants (blue). Aligned short read data were visualized by Interactive Genome Viewer (IGV). Black arrowheads indicate the chromosome locus with coverage changes caused by translocations or deletions. (b) Schematic diagrams of rearranged chromosomes in TAQing2.0-treated mutants. WT Cu is a triploid yeast and two alleles (major alleles) tend to have similar SNV and InDel patterns, which were shown in the same color. Since ChrIV, V and ChrVI have partial heterozygous allele regions which enable distinction of two major alleles, these regions are marked in different colors. Homologous recombination (HR) and large-scale rearrangements such as translocations (TL), large deletions (Del) aneuploid and break induced replication (BIR) of TAQing2.0-treated mutants are shown. Black arrowhead shows predicted lyp1 promoter locus. The origin of multiple mutations on the major alleles is indistinguishable, so they are tentatively marked on one of the chromosomes. (c) Schematic diagrams of large deletions in ChrIV in NTG-treated mutants m11 and m17.

TABLE 4 | Small mutations and large rearrangements.

	TAQed			NTGed				
	m1	**m2**	**m3**	**m3**	**m9**	**m11**	**m17**	**m24**
SNVs	3	4	1	1333	503	799	984	1402
InDels	0	0	0	1	1	1	2	1
HR	7	5	5	123	119	73	116	393
TLs	0	0	2	0	0	0	0	0
Dels	0	2	2	0	0	1	1	0
aneuploid	0	0	1	0	0	0	0	0
BIR	0	0	1	0	0	0	0	0

Note: SNVs, InDels, homologous recombination (HR), translocations (TLs), large deletions (Dels), aneuploid, and break-induced replication (BIR) were counted in each strain. Dels are defined as the loss of multiple kilobases to megabases of DNA and detected by chromosome-arm wide coverage change, while InDels are small insertions or deletions ranging from a single nucleotide to a few hundred base pairs to be detected by GATK.

Lee 2011; Loechler et al. 1984; Ohnishi et al. 2008). In mutants treated with NTG, novel SNVs exhibited a strong bias toward G/C to A/T transitions, whereas mutants treated with TAQing2.0 showed no such bias (Figure S9a,b). However, in both TAQing2.0 and NTG-treated mutants, the mutation spectrum in HR regions displayed no such strong bias and resembled the spectrum of spontaneous SNVs (Figure S9c). Furthermore, it has been reported that NTG treatment can activate HRs in *Saccharomyces cerevisiae* (Buettner and Parks 1970; Zimmermann and Schwaier 1967). Thus, even if a single SNV is mutated, it is highly likely to be repaired through the GC pathway, provided a homologous repair template is available. Therefore, classifying SNVs with a repair template as HR-related was deemed appropriate.

Using a similar approach, we explored HR regions with continuous changes in the SNV allele ratio across entire chromosomal arms, defining these regions as BIR segments. As mentioned above, TAQing2.0-treated m3 strain had an aneuploidy of minor ChrV allele that was translocated to one of the major ChrV alleles, so that the copy number of its left arm was four (Figure 2b). The major-to-minor allele ratio of heterozygous SNV was 2:2 on the left side of the boundary at 556,244 bp, which shifted to 3:1 to the right of this boundary (Figures 2b and 3c). This suggests that BIR occurred in one of the minor allele chromosomes using the major allele chromosome as a replication template.

These results indicate that while TAQing2.0-treated mutants exhibit various large-scale chromosome rearrangements, they have very few point mutations. In contrast, NTG-treated mutants predominantly exhibit point mutations along with HR events.

2.4 | Identification of Genes Responsible for Low pH Resistance

We previously reported that the original TAQing method has the advantage of efficient mapping of genes responsible for the altered phenotypes by identifying common genetic alterations in the TAQing-generated mutants (Yone et al. 2022). In this study, we employed the mapping method to determine the genes underlying the acid resistance phenotype generated by TAQing2.0-treated and NTG-treated mutants. We hypothesized that regions with common SNVs, InDels, or HR might harbor candidate responsible genes. In the three TAQing2.0-treated mutants, common alteration sites were identified as follows: (1) a single SNV at 1,422,278 bp on ChrIV, (2) an HR at ChrI 320,039, and (3) two HRs on ChrII around 199,988 and 204,000 bp (Figure 4a). On the other hand, NTG-treated mutants did not exhibit these common SNVs or HR regions (Figure 4b), though they had numerous other common mutations of SNVs and HRs within NTG-treated mutants (Tables S2 and S3).

For the SNV at 1,422,278 bp on ChrIV, the nearby sequence was subjected to a BLAST search against the genome database of the closely related fungi *C. jadinii* NRRL Y-1542 (Rupp et al. 2015). The search revealed no protein-coding genes within 3.7 kbp upstream and 6.7 kbp downstream of the common SNV. As we examined RNA expression of the common SNV locus and two closest gene loci in this region by reverse transcription-PCR (RT-PCR), we could not detect any changes in expression patterns between mutants and wild-type (Figure S10a). At the hetero HR on ChrI:320,039, which was in the homologous sequence of the MoaE (molybdopterin synthase) protein-coding sequence (CDS) in *C. jadinii*, a 1:2 heterozygous allele ratio was observed where the wild-type genome showed a 2:1 allele ratio. The RNA expression change was not significant between the mutants and the wild-type at the locus (Figure S10b). We therefore concluded that this region does not include the loci responsible for low pH resistance.

For the region with the two HRs around 199,985 and 204,000 bp on ChrII, we searched 2,000 bp upstream and downstream of these two HRs in the *C. jadinii* genome database and identified *CuPEX13* (44.5% identical to a peroxisomal importer complex component gene *ScPEX13* related to selective autophagy called pexophagy and human Zellweger syndrome), a predicted aldo/keto reductase gene (44.1% identical to *ScYPR127W*), and *CuLYP1* (36.3% identical to lysine permease gene *ScLYP1*) (Figures 4c and S10c,d). We then conducted RT-PCR experiments using wild-type and the three TAQing2.0-treated mutants cultured in SD medium at pH 4.6 and pH 2.0 to examine the RNA transcription

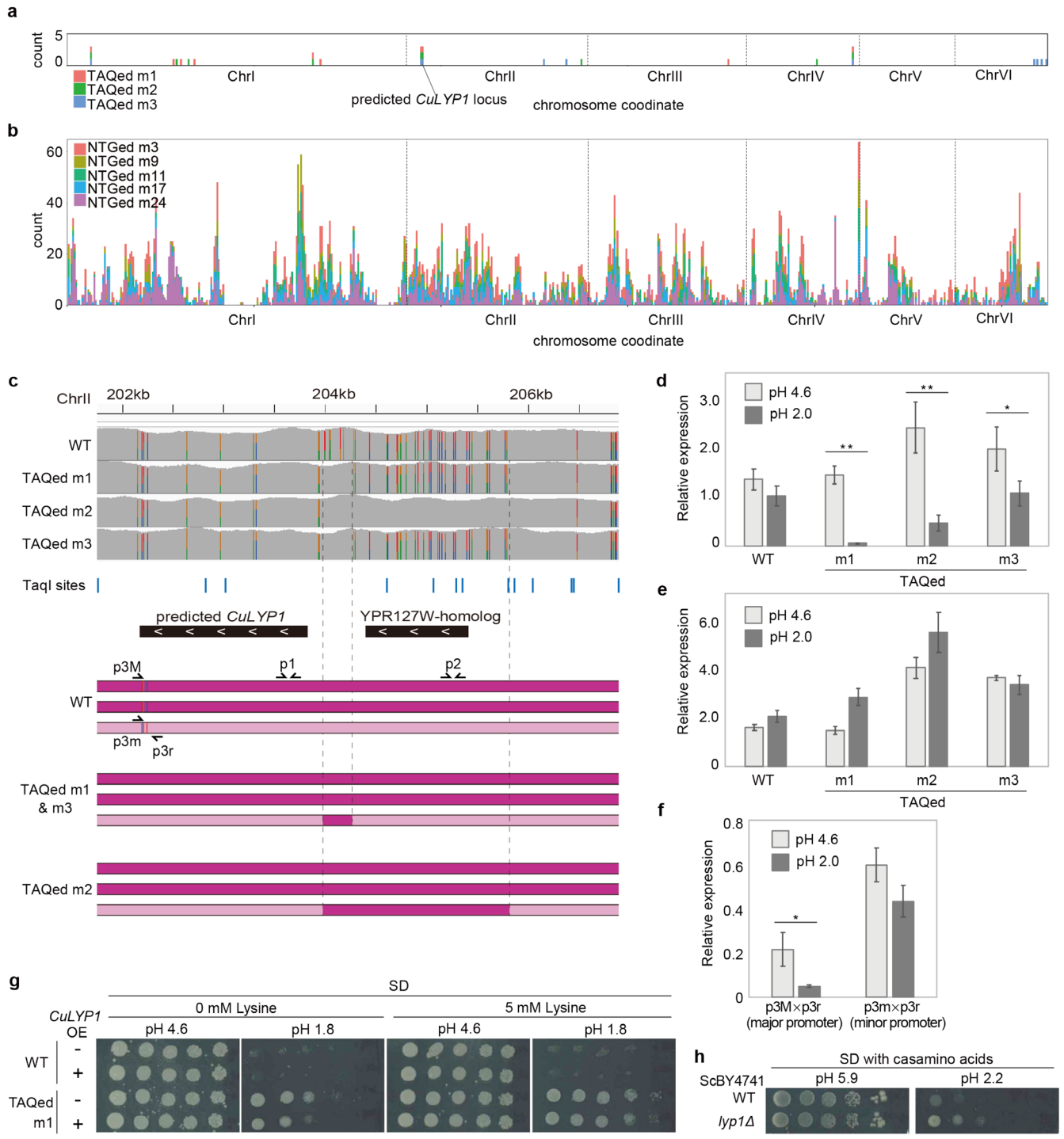

FIGURE 4 | Systematic search for genes responsible for low-pH resistance. (a, b) Stacked bar graph of the SNV and HR counts par 20 kb in (a) TAQing2.0-treated and (b) NTG-treated mutants. (c) IGV visualization of SNP patterns by DNA sequencing of WT and TAQing2.0-treated mutants and the schematic image of alleles with loss of heterozygosity at the *CuLYP1* promoter locus which were commonly detected in the TAQing2.0-treated mutants. TaqI recognition sites were marked with blue lines. Two major alleles and one minor allele of the triploid chromosomes were shown in dark magenta and light magenta, respectively. (d, e) Relative gene expression in WT, TAQed m1, m2 and m3 in pH 4.6 and pH 2.0 SD media at (d) *CuLYP1*, and at (e) a predicted aldo/keto reductase gene, a *YPR127W* homolog. The expression levels under each condition were normalized to that of *act1* as internal control. Data are presented as mean ± SEM ($n = 3$). Two asterisk indicates $p < 0.05$ for a two-tailed test. (f) Allele specific *CuLYP1* expression of WT in pH 4.6 and pH 2.0 SD media. The forward primer p3M detects RNAs from the major alleles (the dark magenta chromosome), while p3m detects RNAs from the minor allele (the light magenta chromosome) and the reverse primer p3r is used in common as indicated in (c). The expression levels under each condition were normalized to that of *act1*. Data are presented as mean ± SEM ($n = 3$). One asterisk indicates $p < 0.05$ for a one-tailed test. (g) *C. utilis* wild-type and TAQed m1 spot assay when *CuLYP1* is overexpressed with or without 5 mM Lysine. (h) Spot assay of a *lyp1*Δ mutant in *S. cerevisiae*.

levels of *CuPEX13*, the predicted aldo/keto reductase gene, and *CuLYP1*. We observed all TAQing2.0-treated mutants showed a marked decrease in the expression of *CuLYP1* at pH 2.0 compared to that at pH 4.6, whereas *CuPEX13* and the aldo/keto reductase gene (*YPR127W* homolog) did not exhibit a clear common difference (Figures 4d,e and S10c). The expression of the homozygotic intergenic region (the upstream of *CuLYP1* and downstream of the *ScYRP127W* homolog) was quite low in all samples, although TAQed m2 and m3 showed slightly higher expression at pH 4.6 (Figure S10d). These observations seem consistent with the location of the common SNVs around *CuLYP1* in the putative transcriptional promoter for *CuLYP1*. It should also be noted that the SNVs in *CuPEX13* were silent SNVs (T225 → A and G228 → A) and resulted in synonymous codon substitutions (Gly75 and Gly76). Therefore, we conclude that the promoter mutation at *CuLYP1* can be responsible for the low pH resistance of TAQing2.0-treated mutants. We also found a SNV of *CuLYP1* CDS in the NTG-treated acid-resistant mutant m17 (the 1079th C to T in NTG which caused amino acid substitution of G360E), but no mutation within the promoter region in the NTG-treated mutants (Figure S10d).

To compare the RNA expression levels of the major and minor promoters under different pH conditions, we performed RT-qPCR using allele-specific primers at pH 4.6 and 2.0 in WT. As a result, the major promoter exhibited significantly lower activity at pH 2.0 compared to pH 4.6, with RNA production decreasing to approximately 20% of the level observed at pH 4.6 (Figure 4f). Similarly, the minor promoter also showed reduced activity at pH 2.0, but to a lesser extent, with RNA levels decreasing to about 70% of those at pH 4.6 (Figure 4f). Importantly, the major allele promoter exhibits a more pronounced decrease in expression when exposed to lower pH conditions, suggesting that its activity is more sensitive to acidic environments compared to the minor allele promoter. Our results indicate the pH-dependent regulatory mechanism of the major and minor *CuLYP1* promoters, emphasizing the greater susceptibility of the major allele *CuLYP1* promoter to transcriptional suppression under highly acidic conditions, which is consistent with the strong acid-resistant phenotype of the TAQing treated LOH strains without the minor *CuLYP1* promoter.

These analyses suggest that the decreased expression of *CuLYP1* under low pH conditions could confer acid tolerance to the mutants. To test these possibilities, we first overexpressed *CuLYP1* under low pH conditions in TAQing2.0-treated acid-resistant mutants. We constructed an overexpression vector by placing *CuLYP1* under the control of the Cu glyceraldehyde-3-phosphate dehydrogenase gene (*CuTHD3*) promoter. This cassette was introduced into the wild-type (CuWT) and the TAQed m1 strain. Spot assay results indicated a partial decrease in low pH resistance of the *CuLYP1* overexpressed strain (TAQed m1 *CuLYP1* OE) compared to TAQed m1 (Figure 4g, 0 mM Lysine). Since Lyp1 is a lysine permease, we tested the effect of lysine addition to the medium and found that the acid tolerance of both TAQed m1 and TAQed m1 *CuLYP1* OE improved (Figure 4g, 5 mM Lysine).

Next, we studied the effects of *LYP1* gene disruption. Due to the difficulty of gene disruption in triploid *C. utilis*, we obtained a *S. cerevisiae* strain with deletion of the orthologous gene *ScLYP1* and tested its low pH tolerance. We confirmed that the *S. cerevisiae LYP1* deletion strain exhibited slightly enhanced acid resistance compared to the wild type (Figure 4h).

Although we cannot eliminate the possibility of synergistic effects of other unique mutations in each TAQ2.0-treated mutant, these results suggest that *CuLYP1* is a strong candidate gene responsible for the improved acid resistance of all three TAQing2.0-treated mutants.

3 | Discussion

Nonconventional yeast, which often means "non-*Saccharomyces*" yeasts, have various abilities that are absent in *S. cerevisiae*, and they are well studied from an evolutionary perspective of fungi (Aoki et al. 2017; Shen et al. 2018). From the industrial point of view, hosts that can achieve purposes are selected by nonconventional yeasts. *Candida utilis*, also known as *Cyberlindnera jadinii* and recognized as Generally Regarded As Safe (GRAS) by the United States (U.S.) Food and Drug Administration (FDA), is a crabtree-negative yeast and presents high respiratory activity, and its ability to synthesize various compounds; it is often used in food and pharmaceutical industries to produce food additives, supplements, and organic acids (Sousa-Silva et al. 2021). *Ogataea polymorpha* (either known as *Hansenula polymorpha* or *Pichia angusta*) is one of the most thermotolerant yeasts identified to date, with growth up to a maximal temperature of 50°C (Guerra et al. 2005). Some nonconventional yeasts such as *Kazachstania bulderi*, *Pichia occidentalis*, and *Issatchenkia orientalis* are reported to have low pH tolerance and a high ability to produce organic acids (Balarezo-Cisneros et al. 2023; Pyne et al. 2023; Wu et al. 2023), although native *C. utilis* does not show strong low pH tolerance. We succeeded in gaining *C. utilis* strains that can grow in highly acidic conditions by conventional mutagenesis and our recent technology TAQing2.0 and assessed their ability and safety.

This study demonstrates that TAQing2.0 is an efficient technology capable of producing mutant strains with favorable traits, with minimal risk of foreign DNA incorporation. Therefore, TAQed strains are expected to have a lower barrier to social acceptance than genetically modified organisms (GMOs), particularly in fields with a strong demand for natural products, such as food and cosmetics. Additionally, TAQing2.0 is likely to have economic advantages regarding containment measures required by GMO regulations. Another advantage of TAQing2.0 compared to conventional mutagenesis is the readiness of identification of genes responsible for altered phenotypes. This is mainly owing to the fewer point mutation rate of TAQing2.0-treated mutants.

Exploring unmapped reads in multicellular organisms can sometimes reveal meaningful biological information such as infection or contamination (Chen and Li 2020; Laine et al. 2019; Neumann et al. 2023). In these studies, mappable reads are around 70%–90% (Sangiovanni et al. 2019), but our wild-type *C. utilis* sequence mapping rate was much higher, more than 99.7%, since contamination risks are much lower when sequencing pure cultured microorganisms. But still, there were 0.018%

of unmappable reads. It is reported that gDNA sequences by NGS systems can contain systematic error reads. Especially, sequencers are known to have substitution biases (Stoler and Nekrutenko 2021). The quality scores of the final unmapped reads in the wild-type and mutants were very low (Figure S2), suggesting that these reads were unreliable for searching their origins.

The direct protein transfection technology by using recombinant Cas9 expressed in *E. coli* and guide RNA complexes is the most obvious approach to achieve transgene-free genome editing, and recently it has attracted much attention in the field of genetically modified plants (Gong et al. 2021). The reason why is not only the possibility that this technique could avoid various barriers accompanied by GMO regulation, but also the contribution to reduced off-target effect (Tang et al. 2018; Vakulskas and Behlke 2019; Wolter and Puchta 2017). Here, Cas9 ribonucleoproteins are often manufactured by the *E. coli* expression system, and commercially available products have a purity standard of more than 90% or 95%. It is, however, notable that whether residual DNA from the host cell is determined as regulatory specs will depend on a case-by-case basis. Thus, the safety risk has not previously been assessed that foreign DNA mixed in reagents or raw materials could be transferred into the cells and unconsciously be integrated into their genome. In this study, we performed the TAQing2.0 system by introducing *E. coli* expression system-derived recombinant TaqI into *C. utilis* cells. Nevertheless, neither the *E. coli* genome nor the integration of the TaqI expression plasmid into the *C. utilis* genome was confirmed (Table 2). This result could become a role model for the rigorous safety evaluation of gene-edited plants, foods, and feeds via direct protein transfection.

The whole genome sequencing analysis also revealed the precise mutations such as SNVs, GCs, and LOHs by NTG treatment as a representative of conventional methods in triploid yeast. We succeeded in identifying SNVs and gene conversion events accurately by NTG treatment, which forms dose-dependent alkyl adducts in the genome. Normally, O- and N-methylated bases in DNA, e.g., O6-methylguanine, which is known to induce G to A mutations, are repaired by the DSB repair pathway or the single-strand gap repair pathway (Loechler et al. 1984; Wyatt and Pittman 2006). Actually, in bacterial NTG mutagenesis, whole genome sequencing revealed that their mutations are mainly SNVs, especially G/C to A/T transitions (Harper and Lee 2011; Ohnishi et al. 2008). But even in haploid bacteria, HR-related DNA repair mechanisms function to remove adductive DNA modifications (Nowosielska et al. 2006). Although there are few whole genome studies about HR frequency by NTG treatment in yeast, it is reported that the HR pathways are activated in *S. cerevisiae* (Buettner and Parks 1970; Zimmermann and Schwaier 1967). Since *C. utilis* is a triploid, it is natural that the increased frequency of DNA repair via HR pathways induced the GC- and LOH-type mutations (Table 4). On the other hand, TAQing2.0 system induced DSBs with two-base overhanging sticky ends by the restriction enzyme, which caused more frequent HR and non-homologous end-joining events.

In this study, we could identify a new unexpected candidate for low pH resistance genes. One possible explanation for the enhanced acid tolerance in TAQing2.0-treated strains was the decreased expression of the *CuLYP1* gene under acidic conditions. CuLyp1 protein is a lysine permease, and its direct relationship with low pH tolerance has not been explicitly demonstrated. However, Viridiana Olin-Sandoval and colleagues have reported that high extracellular lysine concentrations confer oxidative stress resistance to cells (Olin-Sandoval et al. 2019). Therefore, it is conceivable that the suppression of *LYP1* expression under low pH conditions activates the oxidative stress response, indirectly reducing acid-induced cellular damage.

Moreover, in the TAQed m1 *CuLYP1* OE strain, where overexpression of *CuLYP1* resulted in reduced acid tolerance, the addition of excessive lysine from the extracellular environment improved growth to the same level as TAQed m1. This suggests that supplying lysine in amounts exceeding the capacity of the overexpressed *CuLYP1* might have enhanced the acid stress response mechanisms, thereby improving acid tolerance. Since TAQed m1 showed the strongest pH-resistant phenotype within the three TAQing2.0 treated mutants (Figure 1a,b), it is consistent that *CuLYP1* RNA expression was the most repressed in this mutant (Figure 4d). TAQed m1 had minimal SNVs and HRs without Dels or TLs, while TAQed m2 and m3 had a few large rearrangements. These large rearrangements might become indirect causes for repression of *CuLYP1* downregulation or slower growth (Figure 1a, Table S1).

WT harbored heterozygous *CuLYP1* promoters with different pH sensitivities (Figure 4f), while low pH resistant TAQing2.0-treated strains had only acid-sensitive *CuLYP1* promoters. Although the acid-resistant strain TAQed m1 had little trade-off for growth in pH4.6 (Figures 1a and S2), the coexistence of heterozygous promoters with different sensitivities in WT may have been advantageous for various environmental stresses. In this study, TAQed m1 showed the strongest resistance with limited mutations within the three TAQing2.0-treated mutants. Moreover, TAQed m1 did not show a trade-off between low pH resistance and growth, since its growth rate in non-acidic culture was almost the same as that of WT. However, the other strains tended to proliferate significantly slower, so we cannot rule out the possibility of a trade-off between the ability of acid tolerance and proliferation. Within each mutation treatment group, the strains with large deletions tended to grow more slowly than the other strains. Among them, TAQed m2 and m3 with large deletions on both chromosomes II and IV showed the lowest growth in the non-acidic environment. Therefore, for the mutants with large deletions, it might be difficult to simply conclude the relationship of a trade-off growth.

On the other hand, it seems that the TAQed mutants, m1–m3, had genetically limited diversity as compared with the five NTGed mutants (Figure 4). For this point, we speculate two possible reasons: First, the limited numbers of mutation sites, as TCGA sequences exist approximately once every 256 bases; and second, the nature of selection pressure, low pH, as resistance to low pH could be achieved with only a few changes. However, this does not imply that TAQing2.0 consistently generates similar genotypic rearrangements. In our former report, two TAQing2.0-treated mutants from one experiment with aggregation phenotype did not necessarily have similar genotypes; in detail, one had translocation and large deletions, while another had aneuploidy (Yasukawa et al. 2022). It is probably because

these two mutants were selected not through qualitative selection pressure, but via the quantitative trait of relatively small colony size.

Finally, while NTG treatment is acknowledged for its practicality and versatility as conventional mutagenesis, the not less efficiency of mutant production using TAQing2.0 than that of NTG mutagenesis (Table 1) demonstrates that TAQing2.0 is also a practical method for breeding. Besides, the phenotypic stability of the TAQed mutants from the first selection was much higher than that of NTG-treated mutants, suggesting that TAQing2.0 can save effort and time in generating favorable strains with stable phenotypes, whereas NTG treatment often induces an excessive number of point mutations as off-targets which require subsequent backcrossing. In contrast, there are two disadvantages in the TAQing2.0 method; one is that mutated sites by TAQing2.0 cannot be identified unless whole-genome sequencing is performed, and another is that their mutations are limited to the restriction enzyme sites. However, it is reported that the extended TAQing usage of other restriction enzymes from TaqI changes the way mutations are introduced, which is especially effective for plant genome rearrangement (Tanaka et al. 2020).

Besides, the fewer the number of mutations introduced into the gDNA, the less unintended secondary metabolites will be produced. There are some arguments in the several new genomic techniques (NGTs) that the notifier might provide appropriate information about the bred strain to enable the competent authorities to assess a risk to human health or the environment (Voigt 2023). Among the three TAQed mutants, thus, TAQed m1 will be the best strain with high industrial availability and low acquisition costs, including economic advantages on containment measures. But all the TAQing2.0-treated strains successfully gained effective and stable phenotypes with small numbers of genomic changes which can reduce the time and effort required to identify the unexpected causative genes. Therefore, the TAQing2.0 system would be a powerful and versatile tool without foreign DNA induction in the context of the creation of non-GMO products.

4 | Experimental Procedures

4.1 | Strains and Cultures

Candida utilis NBRC0988 (Cu) and *Saccharomyces cerevisiae* BY4741 (*MAT*a *his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0*) (Sc) were used in this study. The *LYP1*-deletant, *lyp1*Δ, with a genetic background of BY4741, was purchased from the Yeast MATa Knock Out Strain Collection (Horizon Discovery Ltd., Cambridge, UK), and its deletion was verified by yeast colony PCR using the primer pairs listed in Table S4. TAQed Cu mutants (TAQed m1, m2, and m3), NTG-treated Cu mutants (NTGed m3, m9, m11, m17, and m24), and *CuLyp1* OE mutants were generated as described below. SD medium (pH 4.6) containing 20 g/L glucose, 1.7 g/L Difco Yeast Nitrogen Base without ammonium sulfate and amino acids (Becton Dickinson and Company, BD, Maryland, USA), and 5 g/L ammonium sulfate was used, and both SD-agar (pH 1.8) and SD (pH 2.0) media were prepared by adding sulfuric acid or hydrochloric acid to SD (pH 4.6) after autoclave treatment. The ammonium sulfate in the SD-agar (pH 4.6) was replaced with 1.0 g/L monosodium glutamate (SD/MSG-agar medium). Cu and Sc cells were cultured at 30°C with agitation at 180 rpm using a BioShaker BR-43FL (Taitec Corp., Saitama, Japan).

4.2 | TAQing2.0 Technology

TAQing2.0 technology was performed as previously described (Yasukawa et al. 2022). Briefly, a recombinant TaqI expression system was constructed in *E. coli* Rosetta (DE3) pLysS (Novagen, California, USA) mediated by pTaqI (Muramoto et al. 2018), where the TaqI open reading frame sequence amplified from the *Thermus thermophilus* HB8 genome is cloned into pET-15b (Novagen), and a nuclear localization signal (PKKKRKV) is fused to the N-terminus of TaqI. After *E. coli* cells were cultivated in LB medium with 1% glucose (catabolite repression) at 30°C, the expression of the recombinant 6× His-tagged NLS-TaqI (rTaqI) was induced by 0.1 mM isopropyl *β*-D-1thiogalactopyranoside, and then purified from the sonicated extract of the cells by heat treatment (65°C) and affinity chromatography.

(TALON Metal Affinity Resin, TaKaRa Bio), and filtration (Amicon Ultra-15 Centrifugal Filter Units 10 kDa cutoff, Merck Millipore Ltd., County Cork, Ireland). rTaqI was introduced into alive Cu cells in a nucleic acid-free manner using a cell penetrating peptide, Xfect Protein Transfection Reagent (TaKaRa Bio Inc. Shiga, Japan). Cu WT cells (5.0×10^7 cells) grown at early-log phase in YPD at 30°C were suspended in 10 mM salt-free MES (pH 6.0), mixed with the rTaqI/Xfect complex in the Xfect reaction buffer at 3°C for 1 h, and then transferred to fresh YPD for recovery step, followed by incubation at 30°C for 30 min. Subsequently, cells were warmed at 38°C for 90 min to temporarily activate rTaqI, and after being centrifuged to eliminate their supernatants, cells were suspended with distilled water (TAQed mutant library). For calculation of the library size, a small aliquot was cultured on YPD-agar plates at 30°C for 2 days and colonies that formed were counted.

4.3 | NTG-Treatment Mutagenesis

Cu WT cells (5.5×10^8 cells) grown in the early-log phase in YPD at 30°C were harvested on a 1.0 μm cellulose ester filter and suspended at RT for 10 min with distilled water containing 0.3 mg/mL of freshly dissolved 1-methyl-3-nitro-1-nitrosoguanidine (NTG) (D. Moore 1969; Lim et al. 2012). Immediately, the cells on the filter were washed with distilled water and cultured in fresh YPD at 30°C for 90 min for the recovery step, followed by the substitution of their medium with distilled water (NTGed mutant library). For the calculation of the library size, a small aliquot was cultured on YPD-agar plates at 30°C for 2 days, and colonies were counted.

4.4 | Isolation of Acid-Tolerant Mutant

To isolate mutants with tolerance to acid, TAQed and NTGed mutant libraries described above, and Cu WT cells fully grown

in YPD and suspended with distilled water were each spread on SD-agar plates (pH 1.8) with incubation at 30°C for 6 days. As the first screening, mutants with larger colony sizes than those of WT were selected and re-streaked on SD agar plates (pH 1.8) to form single colonies (ancestor). Next, as the second screening, we selected those mutants that stably inherited an acid-tolerant trait as follows: Each ancestor was grown overnight in SD (pH 4.6) at 30°C without selection, and this procedure was repeated over 5-rounds of passage resulting in the acquisition of 5 passages. Growth curves between the ancestor and 5 passages in the specific strain were compared by adopting the recording apparatus described below, and the mutant that showed no significant difference was finally selected.

4.5 | Drawing Growth Curve and Spot Assay

To plot the time course of growth, overnight cultures inoculated with Cu WT, ancestor, or 5 passage were added to 4 mL SD (pH 2.0) at an OD_{600} of 0.1 by using a Spectrophotometer U-5100 (Hitachi Ltd., Tokyo, Japan), and then cultured at 30°C with agitation at 40 rpm using a compact rocking incubator TVS062CA (Advantec Co. Ltd., Osaka, Japan). The OD_{660} scores were recorded every 30 min. Spot assay experiments were performed as described below. Cu WT and mutants were grown in SD (pH 4.6) at 30°C overnight, diluted to an OD_{600} of 2.5 (Spectrophotometer U-5100) with distilled water, and then ten-fold serial dilutions were spotted on the SD-agar (pH 4.6 or 2.0) plates, followed by incubation at 30°C for 3–4 days.

4.6 | Pulsed-Field Gel Electrophoresis (PFGE)

Sample preparation was performed according to the manufacturer's instructions attached to the CHEF Genomic DNA Plug Kits. (Bio-Rad Laboratories Inc., Hercules, California, USA) and our previous report (3). Cu cells grown in SD (pH 4.6) until their OD_{600} of 1 were harvested (about 6.0×10^7 cells), washed twice with ice-cold 50 mM EDTA solution (pH 7.5), and then suspended in the same buffer. The cells were treated with Zymolyase-100T (Seikagaku Biobusiness Corp., Tokyo, Japan) at 50°C, embedded in 2% CleanCut Agarose, and then transferred into plug molds, followed by gelling at 4°C for 10–15 min.

The plugs were immersed in Lyticase buffer (10 mM Tris (pH 7.2), 50 mM EDTA) mixed with Zymolyase and incubated at 37°C for 2 h. Then, after removal of washing water, the plugs were filled with Proteinase K Reaction Buffer mixed with Proteinase K (Qiagen, Hilden, Germany), digested at 50°C overnight, washed four times with 1 × wash buffer at RT for 1 h each time with gentle agitation, and stored in 1 × wash buffer at 4°C. Agarose gels, certified Mega agarose #1613108, were prepared at a concentration of 1% in 1 × TAE buffer. Electrophoresis was performed in 1 × TAE buffer at 14°C for 48 h with a voltage of 3 V/cm, a reorientation angle of 106°, and a switching time of 500 s using the CHEF Mapper XA System (Bio-Rad). CHEF DNA size marker #170–3667 (Bio-Rad) was applied. The gels were stained with GelRedTM (Wako, Osaka, Japan) and scanned using the ChemiDoc MP Imaging System (Bio-Rad).

4.7 | Genomic DNA Preparation and Whole-Genome Sequencing

Cu genomic DNA (gDNA) was extracted from WT or mutant cells using the genomic DNA buffer set (Qiagen, Hildem, Germany) and Genomic-tip 100/G (Qiagen) according to the manufacturer's procedure. Quality check of the purified genomic DNA, library construction, and sequencing were performed by GeneBay Inc., (Kanagawa, Japan). Purified gDNA was sheared and fragmented by using Covaris Focused-ultrasonicator ME220 (Covaris Inc., Vienna, Austria). MGIEasy PCR-Free DNA Library Prep Set (MGI Tech Co. Ltd., Shenzhen, China) was used to construct the library, and whole genome sequencing with 200-bp paired-end reads was performed on the DNBSEQ-G400 (MGISEQ-2000RS) platform (MGI Tech).

4.8 | RNA Isolation and qRT-PCR Analysis

For the qRT-PCR experiment, three biological replicates ($n = 3$) in each sample were used. CuWT and TAQed m1, m2, and m3 mutants were cultured in SD (pH 4.6) or SD (pH 2.0) at 30°C until log phase. Then, the cultures were centrifuged at 8000 × g for 3 min at RT, the supernatant was removed and the pellets were frozen in liquid nitrogen without washing cells and immediately stored at −80°C until use. Total RNA was extracted with the hot acid-phenol method (Collart and Oliviero 1993), purified using the RNeasy Mini Kit (Qiagen) according to the manufacturer's protocols, and dissolved in RNase-free distilled water as previously described (Yasukawa et al. 2024). RT-PCR was performed using the PrimeScript RT Reagent Kit with gDNA eraser (Takara Bio, Japan) and Thunderbird SYBR qPCR Mix (Toyobo, Osaka, Japan) as previously described (Oda et al. 2022).

4.9 | Construction of *CuLYP1* Plasmid and *CuLYP1* OE Mutant

All primers used in this study were listed in Table S4. Expression plasmid, p*CuLYP1*, was constructed as follows: 5'-UTR (-) from 204,300 to −204,831 bp in ChrII of *CuLYP1* (adenine of the initiation codon ATG is numbered as +1), KanMX4 gene, promoter (+) of glyceraldehyde-3-phosphate dehydrogenase from 2,221,073 to 2,221,803 bp in ChrII, and ORF of *LYP1* homolog in Cu (DNA segment (+) from 202,178 to 203,830 bp in ChrII) were joined in this order into pBluescript II SK (+) by using the InFusion system (TaKaRa Bio, Kusatsu, Japan). Integrity of the cloned inserts was checked by DNA sequencing analysis. To generate strains overexpressing *CuLYP1*, p*CuLYP1* was linearized by the restriction enzyme *BamHI* and *EcoRI*, introduced into the Cu WT or TAQed m1 through double-crossover recombination using the lithium acetate method (Gietz and Woods 2002; Gietz and Schiestl 2007), and then transformants, *CuLyp1* OE (OE; overexpressed), were selected on the SD/MSG-agar plates with 120 μg/mL G418.

4.10 | Detection of Foreign DNA Sequences

We previously reported the wild-type *C. utilis* complete genome sequence using the PacBio Sequel and Illumina MiSeq in 2022

(Yasukawa et al. 2022). The long and short reads were assembled by MaSuRCA (Zimin et al. 2013) and Falcon (Chin et al. 2016). The long reads were also assembled using the assembler Canu (v1.6) (Koren et al. 2017). To detect TAQing2.0-treated strain-specific foreign DNA introduction, gDNA sequence reads from MGI-seq (200 bp ×2 paired-end) of 9 strains, the wild-type, three TAQing2.0-treated strains, and five NTG-treated strains were mapped using BWA-MEM (v0.7.17) with default mapping parameters to the *C. utilis* reference genome sequence (Yasukawa et al. 2022). Their unmapped reads were extracted using Samtools (v1.15.1) (Li 2011) and they were mapped to the contig sequences from three assemblers, MaSuRCA, Falcon, and Canu. Then their unmapped reads were extracted using Samtools again and mapped to the previously reported contig sequences of *C. utilis* (NBRC0988) (Tomita et al. 2012) and the sequence of a closely related species, *C. jadinii* (NRRL Y-1542) (Rupp et al. 2015).

Next, the remaining unmapped reads were extracted again and mapped to the *E. coli* BL21 (DE3) chromosome sequence and the pET-15b vector (Novagen) with a TaqI-coding DNA fragment amplified from the *T. thermophilus* HB8 sequence. The overlapping of the mapped regions between TAQing2.0 and NTG-treated strains was compared using Bedtools (v2.27.1) (Quinlan and Hall 2010) to detect uniquely induced foreign sequences. Then the remaining unmapped reads were extracted using Samtools again. We prepared an nBLAST database from remaining unmappable reads of wild-type using local Nucleotide-Nucleotide BLAST (v2.5.0+) (Camacho et al. 2009). The unmappable reads from TAQing2.0 and NTG-treated strains were nBLAST-searched with default parameters against the custom database. The quality score of the remaining unmappable reads and reads mapped to the *C. utilis* reference was checked by FastQC.

4.11 | Large Rearrangements and SNV Analysis

First, TLs, Dels, and aneuploid were detected by genome-wide coverage change. Chimera reads mapped to different loci were detected to decide the breakpoints for TLs and Dels. Dels were determined when their deficient end sequences were attached to telomeric repeat sequences. The breakpoint sequences of TLs and Dels were locally checked by PCR using primers listed in Table S4. Although most *C. utilis* chromosomes are triploid with two major alleles and a minor allele, some TLs, Dels, and aneuploid were within an altered heterozygous allele pattern whose alleles are unique to each other (Figure S4 as an example), so that even the two major alleles were distinguishable.

C. utilis variant calling in all the strains was performed using the GATK4.2.6.1 (Van der Auwera and O'Connor 2020) and Bcftools (ver 1.3.1) (Li et al. 2009). HaplotypeCaller from GATK and mpileup Bcftools detected similar SNVs (97.5%). InDels and a single nucleotide with 1:1:1 SNV separation were detected better by GATK with standard filters. Bcftools detected SNVs in a 2:1 heteroallelic triploid locus more accurately. SNVs and InDels from both tools were merged, and mutant-specific SNVs and InDels were marked by Bcftools. Mutations were classified into novel SNVs, HRs, and BIRs according to their SNV separation ratio (Coverage proportion mapped to the reference vs the alternative reference) and consecutive patterns (Figure S5 and Data S1). Since *C. utilis* is mostly triploid, it required custom recalibration of SNV separation ratio. Base composition at each SNV was recalculated by Bcftools. The threshold to define 2:1 or 1:2 of SNV in triploid loci were set as 0.4 and 0.55 by plotting the frequency of reference versus alternative reference mapping proportion, as shown in Figure S6. The threshold of SNV separation ratio for novel SNVs and LOHs was defined as 0.05 (Figure S7a,b). A novel SNV was defined as an SNV without recombination template. HR was defined as other SNVs with template alleles. A group of sequential changes of SNV separation rate was regarded as one HR. Especially when the SNV ratio change covers a chromosome arm, it was defined as a BIR.

Author Contributions

Arisa H. Oda: data curation, formal analysis, visualization, writing – review and editing, funding acquisition, validation. **Taishi Yasukawa:** conceptualization, investigation, methodology, visualization, writing – review and editing. **Miki Tamura:** investigation. **Ayumu Sano:** investigation. **Naohisa Masuo:** project administration, supervision. **Kunihiro Ohta:** project administration, supervision, writing – original draft, funding acquisition.

Acknowledgments

The authors have nothing to report.

Conflicts of Interest

Taishi Yasukawa, Ayumu Sano, and Naohisa Masuo are employees of Mitsubishi Corporation Life Sciences Limited (MCLS). Experimental materials and mutants were provided by MCLS, but the authors from MCLS were not involved in the genome data analysis. All authors do not have a share in Mitsubishi Corporation Life Sciences Limited.

Data Availability Statement

All data are available in the main text or the Supporting Information. Sequencing data from DNBSEQ-G400RS are deposited in the DDBJ Sequence Read Archive database under accession number DRA017141.

References

Adelberg, E. A., M. Mandel, and G. C. Ching Chen. 1965. "Optimal Conditions for Mutagenesis by N-Methyl-N′-Nitro-N-Nitrosoguanidine in *Escherichia Coli* K12." *Biochemical and Biophysical Research Communications* 18, no. 5: 788–795. https://doi.org/10.1016/0006-291X(65)90855-7.

Aoki, K., K. Furuya, and H. Niki. 2017. "Schizosaccharomyces Japonicus: A Distinct Dimorphic Yeast Among the Fission Yeasts." *Cold Spring Harbor Protocols* 2017, no. 12: pdb.top082651. https://doi.org/10.1101/pdb.top082651.

Balarezo-Cisneros, L. N., S. Timouma, A. Hanak, A. Currin, F. Valle, and D. Delneri. 2023. "High Quality De Novo Genome Assembly of the Non-Conventional Yeast Kazachstania Bulderi Describes a Potential Low pH Production Host for Biorefineries." *Communications Biology* 6, no. 1: 918. https://doi.org/10.1038/s42003-023-05285-0.

Bekatorou, A., C. Psarianos, and A. A. Koutinas. 2006. "Production of Food Grade Yeasts." *Food Technology and Biotechnology* 44, no. 3: 407–415.

Buettner, M. J., and L. W. Parks. 1970. "Nitrosoguanidine-Induced Gene Conversion During Growth and Amino Acid Starvation in *Saccharomyces Cerevisiae*." *Mutation Research, Fundamental and Molecular Mechanisms of Mutagenesis* 10, no. 1: 73–76. https://doi.org/10.1016/0027-5107(70)90148-X.

Camacho, C., G. Coulouris, V. Avagyan, et al. 2009. "BLAST+: Architecture and Applications." *BMC Bioinformatics* 10: 421. https://doi.org/10.1186/1471-2105-10-421.

Chen, X., and D. Li. 2020. "Sequencing Facility and DNA Source Associated Patterns of Virus-Mappable Reads in Whole-Genome Sequencing Data." *Genomics* 113, no. 1 Pt 2: 1189–1198. https://doi.org/10.1016/j.ygeno.2020.12.004.

Chin, C.-S., P. Peluso, F. J. Sedlazeck, et al. 2016. "Phased Diploid Genome Assembly With Single-Molecule Real-Time Sequencing." *Nature Methods* 13, no. 12: 1050–1054. https://doi.org/10.1038/nmeth.4035.

Collart, M. A., and S. Oliviero. 1993. "Preparation of Yeast RNA." *Current Protocols in Molecular Biology* 23, no. 1: 13.12.1–13.12.5. https://doi.org/10.1002/0471142727.mb1312s23.

Gietz, R. D., and R. H. Schiestl. 2007. "High-Efficiency Yeast Transformation Using the LiAc/SS Carrier DNA/PEG Method." *Nature Protocols* 2, no. 1: 31–34. https://doi.org/10.1038/nprot.2007.13.

Gietz, R. D., and R. A. Woods. 2002. "Transformation of Yeast by Lithium Acetate/Single-Stranded Carrier DNA/Polyethylene Glycol Method." In *Methods in Enzymology*, edited by C. Guthrie and G. R. Fink, vol. 350, 87–96. Academic Press. https://doi.org/10.1016/S0076-6879(02)50957-5.

Gong, Z., M. Cheng, and J. R. Botella. 2021. "Non-GM Genome Editing Approaches in Crops." *Frontiers in Genome Editing* 3: 817279. https://doi.org/10.3389/fgeed.2021.817279.

Guerra, E., P. P. Chye, E. Berardi, and P. W. Piper. 2005. "Hypoxia Abolishes Transience of the Heat-Shock Response in the Methylotrophic Yeast Hansenula Polymorpha." *Microbiology* 151, no. Pt 3: 805–811. https://doi.org/10.1099/mic.0.27272-0.

Harper, M., and C. J. Lee. 2011. "Genome-Wide Analysis of Mutagenesis Bias and Context Sensitivity of N-Methyl-N′-Nitro-N-Nitrosoguanidine (NTG)." *Mutation Research* 731, no. 1–2: 64–67. https://doi.org/10.1016/j.mrfmmm.2011.10.011.

Koren, S., B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, and A. M. Phillippy. 2017. "Canu: Scalable and Accurate Long-Read Assembly via Adaptive k-Mer Weighting and Repeat Separation." *Genome Research* 27, no. 5: 722–736. https://doi.org/10.1101/gr.215087.116.

Kuspa, A., and W. F. Loomis. 1992. "Tagging Developmental Genes in Dictyostelium by Restriction Enzyme-Mediated Integration of Plasmid DNA." *Proceedings of the National Academy of Sciences of the United States of America* 89: 8803–8807. https://doi.org/10.1073/pnas.89.18.8803.

Laine, V. N., T. I. Gossmann, K. Van Oers, M. E. Visser, and M. A. M. Groenen. 2019. "Exploring the Unmapped DNA and RNA Reads in a Songbird Genome." *BMC Genomics* 20, no. 1: 19. https://doi.org/10.1186/s12864-018-5378-2.

Li, H. 2011. "A Statistical Framework for SNP Calling, Mutation Discovery, Association Mapping and Population Genetical Parameter Estimation From Sequencing Data." *Bioinformatics* 27, no. 21: 2987–2993. https://doi.org/10.1093/bioinformatics/btr509.

Li, H. 2013. "Aligning Sequence Reads, Clone Sequences and Assembly Contigs With BWA-MEM," ArXiv. https://doi.org/10.48550/arXiv.1303.3997.

Li, H., B. Handsaker, A. Wysoker, et al. 2009. "Genome Project Data Processings: The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Lim, J., Y. Jang, Y. Lim, and K. Kim. 2012. "Construction of a Thermotolerant *Saccharomyces Cerevisiae* Strain for Bioethanol Production With Reduced Fermentation Time and Saccharifying Enzyme Dose." *Journal of Microbiology and Biotechnology* 22: 1401–1405. https://doi.org/10.4014/jmb.1203.03069.

Loechler, E. L., C. L. Green, and J. M. Essigmann. 1984. "In Vivo Mutagenesis by O6-Methylguanine Built Into a Unique Site in a Viral Genome." *Proceedings of the National Academy of Sciences of the United States of America* 81, no. 20: 6271–6275. https://doi.org/10.1073/pnas.81.20.6271.

Moore, D. 1969. "The Mutagenic Action of N-Methyl-N'-Nitro-N-Nitrosoguanidine on Coprinus Lagopus." *Microbiology* 55, no. 1: 121–125. https://doi.org/10.1099/00221287-55-1-121.

Muramoto, N., A. Oda, H. Tanaka, et al. 2018. "Phenotypic Diversification by Enhanced Genome Restructuring After Induction of Multiple DNA Double-Strand Breaks." *Nature Communications* 9, no. 1: 1995. https://doi.org/10.1038/s41467-018-04256-y.

Neumann, G. B., P. Korkuć, M. Reißmann, et al. 2023. "Unmapped Short Reads From Whole-Genome Sequencing Indicate Potential Infectious Pathogens in German Black Pied Cattle." *Veterinary Research* 54, no. 1: 95. https://doi.org/10.1186/s13567-023-01227-0.

Nowosielska, A., S. A. Smith, B. P. Engelward, and M. G. Marinus. 2006. "Homologous Recombination Prevents Methylation-Induced Toxicity in *Escherichia Coli*." *Nucleic Acids Research* 34, no. 8: 2258–2268. https://doi.org/10.1093/nar/gkl222.

Oda, A. H., M. Tamura, K. Kaneko, K. Ohta, and T. S. Hatakeyama. 2022. "Autotoxin-Mediated Latecomer Killing in Yeast Communities." *PLoS Biology* 20, no. 11: e3001844. https://doi.org/10.1371/journal.pbio.3001844.

Ohnishi, J., H. Mizoguchi, S. Takeno, and M. Ikeda. 2008. "Characterization of Mutations Induced by N-Methyl-N′-Nitro-N-Nitrosoguanidine in an Industrial *Corynebacterium Glutamicum* Strain." *Mutation Research, Genetic Toxicology and Environmental Mutagenesis* 649, no. 1: 239–244. https://doi.org/10.1016/j.mrgentox.2007.10.003.

Olin-Sandoval, V., J. S. L. Yu, L. Miller-Fleming, et al. 2019. "Lysine Harvesting Is an Antioxidant Strategy and Triggers Underground Polyamine Metabolism." *Nature* 572, no. 7768: 249–253. https://doi.org/10.1038/s41586-019-1442-6.

Pyne, M. E., J. A. Bagley, L. Narcross, et al. 2023. "Screening Non-Conventional Yeasts for Acid Tolerance and Engineering Pichia Occidentalis for Production of Muconic Acid." *Nature Communications* 14, no. 1: 5294. https://doi.org/10.1038/s41467-023-41064-5.

Quinlan, A. R., and I. M. Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics* 26, no. 6: 841–842. https://doi.org/10.1093/bioinformatics/btq033.

Rupp, O., K. Brinkrolf, C. Buerth, et al. 2015. "The Structure of the Cyberlindnera Jadinii Genome and Its Relation to *Candida Utilis* Analyzed by the Occurrence of Single Nucleotide Polymorphisms." *Journal of Biotechnology* 211: 20–30. https://doi.org/10.1016/j.jbiotec.2015.06.423.

Sangiovanni, M., I. Granata, A. S. Thind, and M. R. Guarracino. 2019. "From Trash to Treasure: Detecting Unexpected Contamination in Unmapped NGS Data." *BMC Bioinformatics* 20, no. Suppl 4: 168. https://doi.org/10.1186/s12859-019-2684-x.

Shen, X. X., D. A. Opulente, J. Kominek, et al. 2018. "Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum." *Cell* 175, no. 6: 1533–1545. https://doi.org/10.1016/j.cell.2018.10.023.

Sousa-Silva, M., D. Vieira, P. Soares, M. Casal, and I. Soares-Silva. 2021. "Expanding the Knowledge on the Skillful Yeast Cyberlindnera Jadinii." *Journal of Fungi* 7, no. 1: 36. https://doi.org/10.3390/jof7010036.

Stoler, N., and A. Nekrutenko. 2021. "Sequencing Error Profiles of Illumina Sequencing Instruments." *NAR Genomics and Bioinformatics* 3, no. 1: lqab019. https://doi.org/10.1093/nargab/lqab019.

Tanaka, H., N. Muramoto, H. Sugimoto, A. H. Oda, and K. Ohta. 2020. "Extended TAQing System for Large-Scale Plant Genome Reorganization." *Plant Journal: For Cell and Molecular Biology* 103, no. 6: 2139–2150. https://doi.org/10.1111/tpj.14888.

Tang, X., G. Liu, J. Zhou, et al. 2018. "A Large-Scale Whole-Genome Sequencing Analysis Reveals Highly Specific Genome Editing by Both Cas9 and Cpf1 (Cas12a) Nucleases in Rice." *Genome Biology* 19, no. 1: 84. https://doi.org/10.1186/s13059-018-1458-5.

Tomita, Y., K. Ikeo, H. Tamakawa, T. Gojobori, and S. Ikushima. 2012. "Genome and Transcriptome Analysis of the Food-Yeast *Candida Utilis*." *PLoS One* 7, no. 5: e37226. https://doi.org/10.1371/journal.pone.0037226.

Vakulskas, C. A., and M. A. Behlke. 2019. "Evaluation and Reduction of CRISPR off-Target Cleavage Events." *Nucleic Acid Therapeutics* 29, no. 4: 167–174. https://doi.org/10.1089/nat.2019.0790.

Van der Auwera, G. A., and B. O'Connor. 2020. *Genomics in the Cloud*. O'Reilly Media, Inc.

Voigt, B. 2023. "EU Regulation of Gene-Edited Plants-A Reform Proposal." *Frontiers in Genome Editing* 5: 1119442. https://doi.org/10.3389/fgeed.2023.1119442.

Wolter, F., and H. Puchta. 2017. "Knocking out Consumer Concerns and Regulator's Rules: Efficient Use of CRISPR/Cas Ribonucleoprotein Complexes for Genome Editing in Cereals." *Genome Biology* 18, no. 1: 43. https://doi.org/10.1186/s13059-017-1179-1.

Wu, Z.-Y., W. Sun, Y. Shen, et al. 2023. "Metabolic Engineering of Low-pH-Tolerant Non-Model Yeast, Issatchenkia Orientalis, for Production of Citramalate." *Metabolic Engineering Communications* 16: e00220. https://doi.org/10.1016/j.mec.2023.e00220.

Wyatt, M. D., and D. L. Pittman. 2006. "Methylating Agents and DNA Repair Responses: Methylated Bases and Sources of Strand Breaks." *Chemical Research in Toxicology* 19, no. 12: 1580–1594. https://doi.org/10.1021/tx060164e.

Yasukawa, T., R. Iwama, Y. Yamasaki, N. Masuo, and Y. Noda. 2024. "Yeast Rim11 Kinase Responds to Glutathione-Induced Stress by Regulating the Transcription of Phospholipid Biosynthetic Genes." *Molecular Biology of the Cell* 35, no. 1: ar8. https://doi.org/10.1091/mbc.E23-03-0116.

Yasukawa, T., A. H. Oda, T. Nakamura, et al. 2022. "TAQing2.0 for Genome Reorganization of Asexual Industrial Yeasts by Direct Protein Transfection." *Communications Biology* 5, no. 1: 144. https://doi.org/10.1038/s42003-022-03093-6.

Yone, H., H. Kono, H. Hirai, and K. Ohta. 2022. "Gene Mapping Methodology Powered by Induced Genome Rearrangements." *Scientific Reports* 12, no. 1: 16658. https://doi.org/10.1038/s41598-022-20999-7.

Zimin, A. V., G. Marçais, D. Puiu, M. Roberts, S. L. Salzberg, and J. A. Yorke. 2013. "The MaSuRCA Genome Assembler." *Bioinformatics* 29, no. 21: 2669–2677. https://doi.org/10.1093/bioinformatics/btt476.

Zimmermann, F. K., and R. Schwaier. 1967. "Induction of Mitotic Gene Conversion With Nitrous Acid, 1-Methyl-3-Nitro-1-Nitrosoguanidine and Other Alkylating Agents in *Saccharomyces Cerevisiae*." *Molecular and General Genetics MGG* 100, no. 1: 63–76. https://doi.org/10.1007/BF00425776.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.