

RESEARCH ARTICLE



How to characterize a strain? Clonal heterogeneity in industrial *Saccharomyces* influences both phenotypes and heterogeneity in phenotypes

Hanna Viktória Rác^{1,2} | Fezan Mukhtar¹ | Alexandra Imre^{1,3} |
Zoltán Rádai⁴ | Andreas Károly Gombert⁵ | Tamás Rátónyi⁶ | János Nagy⁶ |
István Pócsi¹ | Walter P. Pfliegler¹

¹Department of Molecular Biotechnology and Microbiology, University of Debrecen, Debrecen, Hungary

²Doctoral School of Nutrition and Food Sciences, University of Debrecen, Debrecen, Hungary

³Kálmán Laki Doctoral School of Biomedical and Clinical Sciences, University of Debrecen, Debrecen, Hungary

⁴MTA-ÖK Lendület Seed Ecology Research Group, Institute of Ecology and Botany, Centre for Ecological Research, Vácrátót, Hungary

⁵School of Food Engineering, Universidade Estadual de Campinas, Campinas, SP, Brazil

⁶Institute of Land Use, Technology and Regional Development, University of Debrecen, Debrecen, Hungary

Correspondence

Walter P. Pfliegler, Department of Molecular Biotechnology and Microbiology, University of Debrecen, Egyetem tér 1, Debrecen 4032, Hungary.

Email: pfliegler.valter@science.unideb.hu

Funding information

Ministry of Innovation and Technology, Hungary, Grant/Award Numbers: GINOP-2.2.1-15-2016-00001, TKP2020-IKA-04; Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Grant/Award Number: 2018/04962-5; Hungarian Academy of Sciences; Hungarian-American Fulbright Commission; Ministry of Innovation and Technology, Grant/Award Number: NKFIH-1150-6/2019

Abstract

Populations of microbes are constantly evolving heterogeneity that selection acts upon, yet heterogeneity is nontrivial to assess methodologically. The necessary practice of isolating single-cell colonies and thus subclone lineages for establishing, transferring, and using a strain results in single-cell bottlenecks with a generally neglected effect on the characteristics of the strain itself. Here, we present evidence that various subclone lineages for industrial yeasts sequenced for recent genomic studies show considerable differences, ranging from loss of heterozygosity to aneuploidies. Subsequently, we assessed whether phenotypic heterogeneity is also observable in industrial yeast, by individually testing subclone lineages obtained from products. Phenotyping of industrial yeast samples and their newly isolated subclones showed that single-cell bottlenecks during isolation can indeed considerably influence the observable phenotype. Next, we decoupled fitness distributions on the level of individual cells from clonal interference by plating single-cell colonies and quantifying colony area distributions. We describe and apply an approach using statistical modeling to compare the heterogeneity in phenotypes across samples and subclone lineages. One strain was further used to show how individual subclonal lineages are remarkably different not just in phenotype but also in the level of heterogeneity in phenotype. With these observations, we call attention to the fact that choosing an initial clonal lineage from an industrial yeast strain may vastly influence downstream performances and observations on karyotype, on phenotype, and also on heterogeneity.

Take Away

- Industrial yeast strains are more heterogeneous than often assumed.
- Isolating and examining a subclone lineage alters observable genotype.
- Subclones may display altered phenotypes along with very variable degree of heterogeneity.
- Yeast products are in fact evolved, heterogeneous populations.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. Yeast published by John Wiley & Sons Ltd.

KEYWORDS

aneuploidy, domestication, fermented foods and beverages, genomic structural variation, phenotypic variability, *Saccharomyces boulardii*

1 | INTRODUCTION

Yeasts have played an important role in human societies since ancient times. Their biochemical versatility, tolerance to a wide range of stress factors, and the ease of applying traditional and later molecular strain improvement strategies have only increased their roles in many agricultural and industrial fields (Barbosa et al., 2018; Gallone et al., 2016; Gonçalves et al., 2016; Peter et al., 2018; Steensels et al., 2019). This industrial applicability is most pronounced in the species *Saccharomyces cerevisiae* that has become unsurmountable in the production of leavened bread, alcoholic beverages, bioethanol, and modern biotechnology while also being widely utilized in fields like bioprotection or food and feed supplements (Legras et al., 2007; Peter et al., 2018). The species is not merely utilized for industrial fermentations but may be part of the human microbiome or be used as a probiotic (under the taxonomically obsolete name *Saccharomyces boulardii*), whereas in some cases, it also has been reported as an opportunistic human pathogen (Pérez-Torrado & Querol, 2015; Peter et al., 2018; Zhu et al., 2016). Importantly, colonizing and infectious isolates are often derived from commercial probiotic or baking strains or are known to be members of the wine yeast clade (Imre et al., 2019; Pérez-Torrado & Querol, 2015; Peter et al., 2018; Pflieger et al., 2017; Zhu et al., 2016).

S. cerevisiae is known to be a genetically diverse species with dozens of globally distributed or locally endemic phylogenetic clades, many of which show hallmarks of domestication. A number of clades have become adapted to the production of fermented beverages or foods, and these are regarded as prime examples of microbe domestication that quite often led to the existence of polyploid and/or aneuploid lineages (Duan et al., 2018; Gallone et al., 2016; Peter et al., 2018; Steensels et al., 2019; Strobe et al., 2015). The most widespread yeast-fermented product worldwide that uses a different yeast “species” is lager beer, where fermentation is carried out by domesticated hybrids of *S. cerevisiae* and *Saccharomyces eubayanus* (known as *Saccharomyces pastorianus* and *Saccharomyces carlsbergensis* as well), whereas hybrids of other combinations have also been found in several fermentation industries (Gallone et al., 2019; Langdon et al., 2019; Morard et al., 2020; Salazar et al., 2019).

The key factors in *S. cerevisiae* becoming so ubiquitous in human-made environments are improved fermentation characteristics, including the utilization of various sugars and production of aroma components (Pontes et al., 2020; Steensels et al., 2019), stress tolerance, and elevated adaptability (Peter et al., 2018; Tattini et al., 2019; Yue et al., 2017). The species is not merely capable of coping with various stress factors found under industrial circumstances, but it also very quickly adapts to changing environments, a

trait of utmost importance in the fluctuating environments of various alcoholic fermentations. *Saccharomyces* species are sexual yeasts, able to utilize meiotic recombination to enhance genetic variability to facilitate adaptation (McDonald et al., 2016; Mortimer, 2000). However, during most industrial processes, yeasts reproduce mitotically. These clonal populations, however, retain their ability to generate novel variants for selection to act upon, in the form of mutations and genome structure variations (GSVs). The latter include ploidy changes, aneuploidies/chromosome copy number variations (CCNVs), loss of heterozygosity (LOH), gross chromosomal rearrangements (GCRs), and mitotic crossing overs (Peter et al., 2018; van den Broek et al., 2015; Zhu et al., 2016). These phenomena can alter their industrial performance and may happen very rapidly (Gorter de Vries et al., 2020; Kadowaki et al., 2017; Large et al., 2020; Morard et al., 2019, 2020; Zhang et al., 2016). Along with point mutations, these GSV events result in clonal populations gradually accumulating differences in various traits, leading to clonal heterogeneity, clonal interference (competition among isogenic asexual lineages), and hence the emergence of so-called subclonal lineages, reminiscent of the experimental evolution setups conducted with laboratory strains (Blundell et al., 2018; Lang et al., 2011; Large et al., 2020; Payen et al., 2016). These evolving and competing subclones ultimately determine the fitness and the performance of the industrial strains during technological applications. Clonal interference may later be alleviated by sexual reproduction, but only if the yeasts survive the technological processes and are able to recolonize the fermentation environment, as happens in traditional wineries (Magwene, 2014; Mortimer, 2000). Most modern technological protocols, however, completely remove the applied yeast populations, either immediately or after a limited number of repitchings (Large et al., 2020), and new fermentations are carried out with fresh inocula from established propagation companies, for example, starter cultures in wineries (Ciani et al., 2016) or beer yeast starters (Large et al., 2020).

In spite of the considerations above, a yeast strain is in general treated as a uniform entity, both in studies aiming at assessing the diversity and characteristics of the species and in the commercialization and handling of industrial starter yeasts. These strains, upon transfer from one lab to another or even before each experimental round in the same lab, are conventionally spread on agar media to isolate genuine single-cell colonies void of any potential contaminants. Single-cell colonies are conventionally considered to be genetically identical (e.g., Eyler, 2013), and even in experimental evolution setups, heterogeneity is only considered after the start of the experiment (for a review on ale and lager experimental evolution, see Gibson et al., 2020). In the present study, we aimed to investigate whether the wide-spread (and necessary) process of

TABLE 1 Strains used in this study, with accession numbers for whole genome sequencing data

Strain name used in this study	Other names	Origin of cultures in this study	Genomes analyzed in this study and their origins				Reference
			BioSample	SRA experiment	SRA run	Coverage (calculated for haploid <i>S. cerevisiae</i> genome)	
ADY_Baker	Commercial name	Commercial vendor, Hungary, manufactured in Germany, producer undisclosed by vendor	SAMN15579325 (subclone 1)	SRX8770000	SRR12264806	43.0	This study
			SAMN15579326 (subclone 2)	SRX8770001	SRR12264805	51.5	This study
Ale	Commercial name	Commercial vendor, Hungary, manufactured in France by a subsidiary of a French company	SAMEA3895632 (subclone 1, sequenced by Peter et al.)	ERX1380630	ERR1309393	210.3	Peter et al. (2018)
			SAMN10973883 (subclone 2, sequenced by Langdon et al.)	SRX6781686	SRR10047300	33.8	Langdon et al. (2019)
			SAMN10375233 (subclone 3, sequenced by Fay et al.)	SRX4993536	SRR8173067	14.9	Fay et al. (2019)
Bioethanol	PE-2; NCYC 3233, JAY270 (JAY270 is a pure culture isolate derived from a PE2 commercial stock)	NCYC (National Collection of Yeast Cultures)	SAMN04965971 (JAY270 subclone)	SRX2038376	SRR4047520	102.7	Rodrigues-Prause et al. (2018)
		Commercial vendor, Brazil, manufactured in Brazil	SAMN15559291 (product subclone 1)	SRX8748432	SRR12240130	42.3	This study
		Commercial vendor, Brazil, manufactured in Brazil	SAMN15559292 (product subclone 2)	SRX8748433	SRR12240129	30.1	This study
		Commercial vendor, Brazil, manufactured in Brazil	SAMN15559293 (product subclone 3)	SRX8748434	SRR12240128	27.4	This study
Lager	Weihenstephan 34/70	Commercial vendor, Hungary, manufactured in France by a subsidiary of a French company	SAMN03174146 (isolate A1)	SRX758144	SRR1649183	71.0	van den Broek et al. (2015)
			SAMN03174146 (isolate A2)	SRX758149	SRR1649191	87.3	van den Broek et al. (2015)
			SAMN03174146 (isolate A1 + B11)	SRX758145	SRR1649190	48.2	van den Broek et al. (2015)
			SAMD00035489 (isolate sequenced by Okuno et al.)	DRX036594	DRR040651	430.9	Okuno et al. (2016), also used in Langdon et al. (2019)
			SAMN10375239 (isolate sequenced by Fay et al.)	SRX4993613	SRR8172990	23.9	Fay et al. (2019)

(Continues)

TABLE 1 (Continued)

Strain name used in this study	Other names	Origin of cultures in this study	Genomes analyzed in this study and their origins				Reference
			BioSample	SRA experiment	SRA run	Coverage (calculated for haploid <i>S. cerevisiae</i> genome)	
Probiotic	Commercial name	Commercial vendor, Hungary, manufactured in Hungary with license from a Canadian company	SAMN11634143	SRX5874542	SRR9099591	256.8	Offei et al. (2019)
Wine	Commercial name	Commercial vendor (Hungary), manufactured in Switzerland by a subsidiary of a Canadian company	SAMN04286169	SRX1457336	SRR2967887	20.2	Borneman et al. (2016)

isolating single-cell colonies (subclones) from commercial yeast products and from strains in collections has a hitherto neglected effect on the observable genotype and phenotype of these strains. In particular, as industrial yeasts are propagated en masse (under relatively stressful conditions) by companies producing and packaging them for dozens of generations (Large et al., 2020; Qiu et al., 2019), we hypothesized that standing genetic variation and clonal heterogeneity stemming from mutations and genome structure variations may already be present in commercial products and may have considerable effects on the phenotypes of industrial yeast. We also assumed that such a diversity in subclone lineages may confer plasticity to the industrial yeast population as a whole, manifesting in clonal phenotypic heterogeneity. Heterogeneity may presumably cause unpredictable biases to genotypic and phenotypic studies involving yeast lineages that need to be isolated from products, whether for basic research, for strain improvement, or to study opportunistic infections by yeasts (e.g., Large et al., 2020; Pfliegler et al., 2017).

To observe and compare heterogeneity, we used wine, ale, lager, probiotic, and bread yeasts of various ploidies to study how heterogeneous these yeasts are when colony phenotypes and stress tolerance are considered. Using a baker's yeast sample, we assessed how even a single population bottleneck, namely, the first instance of single-colony isolation, may considerably influence the observed phenotypic characteristics and, surprisingly, also the observable clonal heterogeneity of a given strain. Additionally, we discuss that recent genomic studies of yeasts occasionally investigated very different subclone lineages from a single strain, corroborating the widespread nature of heterogeneity in industrial lineages.

2 | MATERIALS AND METHODS

2.1 | Strains and (sub)culturing

We obtained commercially available ale, bakery, bioethanol, lager, probiotic, and wine yeasts (Table 1) and precultured the products in YPD medium (VWR, Radnor, PA, USA) for 30 min at room temperature. The bioethanol PE-2 strain was also obtained from NCYC, and the supplier's protocol was followed for reviving. These cultures were immediately used to isolate 12 subclone lineages from each yeast, which were designated with letters from “a” to “l.”

2.2 | Whole-genome sequencing and genomic analysis

Whole-genome analysis involved previously sequenced genomes downloaded from NCBI SRA. For the lager and ale strain, multiple lineages have been sequenced in recent studies, and these were all used individually in our genomics pipeline (Table 1) after giving them individual identifiers (e.g., subclone 1 to subclone 5). In the case of the ADY_Baker yeast, two subclones from a product commercially obtained in Hungary were newly sequenced at the core facility of the University of Debrecen: one typical colony, and one smaller, rough phenotype colony (named subclone 1 and subclone 2, respectively). These lineages were subcultured only once (multiple single-cell bottlenecks were avoided) and were saved as stocks at -70°C . Genomic DNA was isolated from the lineages after 24 h growth of the cultures following inoculation in the form of a streak on YPD agar from stocks stored at

–70°C. DNA isolation followed Hanna and Xiao (2006). Library preparation was performed using tagmentation with the Nextera DNA Flex Library Prep kit (Illumina, San Diego, CA, USA) according to the manufacturer's protocol; sequencing was performed using 150 bp paired-end reads on an Illumina NextSeq 500 system, with approximately 50× coverage of the nuclear genome. Altogether, three subclones of the bioethanol strain (named subclone 2 to 4), obtained from a commercial product in Brazil in active dry yeast form, containing yeast PE-2, were isolated and sequenced at the Bauer Core, Harvard University, Cambridge, MA, using Illumina NextSeq 150 bp paired-end reads. Genomic DNA for these samples was extracted using an in-house protocol; library preparation was carried out using an adapted tagmentation and Nextera kit from Illumina (Baym et al., 2015). Raw reads were deposited to NCBI SRA under BioProject PRJNA646688.

2.2.1 | Mapping

We only used single runs and single experiments for each SRA genome to avoid any effect of clonal heterogeneity in biosamples with multiple available experiments. Newly generated FASTQ sequencing files along with those obtained from SRA were trimmed and filtered using fastp (<https://github.com/OpenGene/fastp>, RRID:SCR_016962) (Chen et al., 2018) and mapped to the S288C reference genome (R64.2.1.) downloaded from the SGD database (yeastgenome.org) and the reference genomes of the other *Saccharomyces* species (Baker et al., 2015; Naseeb et al., 2018; Scannell et al., 2011) concatenated to it after corrections based on Kim et al. (2017) where applicable. After establishing the nonhybrid nature of *S. cerevisiae* yeasts, another round of mapping solely to the S288c reference genome was applied, whereas lager yeasts' sensu stricto mapping was used downstream. Finally, a next round of mapping was done for lager yeasts onto the *S. eubayanus* reference genome for the purpose of mitochondrial variant calling. Mapping was performed using BWA 0.7.17 (<http://bio-bwa.sourceforge.net/>, RRID:SCR_010910) (Li & Durbin, 2009). Sorted BAM files were obtained using Samtools 1.7 (<http://htslib.org/>, RRID:SCR_002105) (Li et al., 2009), and Picard-tools 2.23.8 (<http://broadinstitute.github.io/picard/>, RRID:SCR_006525) was used to mark duplicated reads.

2.2.2 | Coverage mapping

For the sensu stricto mapping, we used BEDTools 2.30.0 (<https://github.com/arq5x/bedtools2>, RRID:SCR_006646) (Quinlan & Hall, 2010) to calculate the median coverage of chromosomes in 10,000 base windows sliding every 5000 bases and to calculate the median coverage of the whole chromosome. Plots generated from these data were corrected for ploidy (as described below) and were used to check for major introgressions from other species. The software Y_{MAP} (<http://lovelace.cs.umn.edu/Ymap/>) was used to generate coverage plots with chromosome end and GC content bias correction (Abbey et al., 2014) for the S288c mapping, and in the case of the hybrid W34/70 lager genome, we created a hybrid reference in Y_{MAP} to be

able to represent the strain with the same method as well. Chromosomal rearrangements manifesting in sudden changes of apparent copy numbers within a chromosome were identified on Y_{MAP} plots. Only rearrangements of at least ~100,000 base length were identified.

2.2.3 | Variant calling

Using BAM files, local realignment around indels and joint variant calling and filtering for the six samples were performed with GATK 4.1.9.0 (<https://software.broadinstitute.org/gatk/>, RRID:SCR_001876) (Poplin et al., 2018; Van der Auwera et al., 2013) with regions annotated in the SGD database as simple repeats, centromeric regions, telomeric regions, or LTRs excluded. First, genomic VCF files were obtained, joint calling was applied, and in the resulting VCF files, only SNPs or only INDELS were selected. SNPs were filtered according to the parameters used by Fay et al. (2019): QD < 5.0; QUAL < 30.0; SOR > 3.0; FS > 60.0; MQ < 40.0; MQRankSum < –12.5; ReadPosRankSum < –8.0. INDELS were filtered according to the parameters QD < 5.0; QUAL < 30.0; FS > 60.0; ReadPosRankSum < –20.0. INDELS were then left-aligned. For the final VCF files, INDELS and SNPs were merged and filtered, and nonvariant sites were removed. Called VCF files were uploaded to FigShare (doi: 10.6084/m9.figshare.14216648).

2.2.4 | Phylogenomic network analysis

The SNP VCF files were used to produce genotype matrices using vcf2phylip (<https://github.com/edgarmortiz/vcf2phylip>) (Ortiz, 2019) that contained heterozygous IUPAC codes. This matrix containing heterozygosities was used in SplitsTree SplitsTree4 V4.17.0 (<https://software-ab.informatik.uni-tuebingen.de/download/splitstree4/welcome.html>, RRID:SCR_014734) (Huson & Bryant, 2006) to create Neighbor Nets with an uncorrected P distance and averaging heterozygous positions. This was done for the nuclear and mitochondrial genomes separately. In the case of lager yeasts, the two parental subgenomes were analyzed individually.

2.2.5 | Allele ratio and chromosome copy number analysis

Variants in the individual strains were selected and exported to a csv file using the query option of SAMtools/BCFtools 1.10.2 (<http://samtools.sourceforge.net/mpileup.shtml>, RRID:SCR_005227). Allele frequency plots were obtained from these. Allele frequencies were used to estimate ploidy following Zhu et al. (2016), with the assumptions that diploids have allele ratios of approximately 1:0 or 1:1, triploids of 1:0, 1:2, and 2:1, tetraploids of 1:0, 1:3, 1:1, or 3:1, and so forth. These ploidy and CCNV results obtained from allele ratios were compared with coverage plots and coverage ratios generated by the

software Y_{MAP} using chromosome end and GC content bias correction (Abbey et al., 2014). The results were compared with previous literature on the given strains' ploidies where available, for ale (Fay et al., 2019; Langdon et al., 2019; Peter et al., 2018), bioethanol (Rodrigues-Prause et al., 2018), and lager strains (Fay et al., 2019; Okuno et al., 2016; van den Broek et al., 2015). Subsequently, allele ratio plots were used to identify major events of allele ratio changes (e.g., from approximately 3:1 to 1:1 in a tetraploid genome, or 1:1 to 1:0 in any genome) spanning at least $\sim 100,000$ bases from the plots. Only events differing between subclones were identified; this was helped by superimposing two plots with different colors. Of these allele ratio changes, any event that led to homozygosity (1:0 ratios) were considered LOH events. For this LOH analysis, too, only inter-subclone differences were scored, leaving out runs of homozygosity that were identical among all subclones of a given yeast strain.

2.2.6 | Identity-by-state analysis

Identity-by-state (IBS) analysis was carried out on ale, baker's, bioethanol, and lager strains and their subclones. We used SNP data as described above with the R package "SNPRelate" 1.24.0 (<https://bioconductor.org/packages/release/bioc/html/SNPRelate.html>) (Zheng et al., 2012), using the function "snpgdsIBSNum" with default settings. This function calculates the minor allele frequency and missing rate for each SNP over all the samples and creates three n -by- n matrices that contain the number of SNPs sharing 0, 1, or 2 IBS. Results were visualized with the R package "heatmapply" (<https://cran.r-project.org/web/packages/heatmapply/index.html>) (Galili et al., 2017) with default settings.

2.2.7 | Pangenome ORF copy number analysis

We performed a mapping to the complete pangenome open reading frame (ORF) collection generated by Peter et al. (2018) for all *S. cerevisiae* strains (as the pangenome of lager yeasts is not yet described, these were omitted) following the parameters in that reference (compared with mapping described above, reads with unmapped pairs were not discarded). BEDTools was used to calculate the median coverage for each ORF. After optimization, the subsequent analysis differed from the reference and was performed as follows. S288c core genes: medians of the median coverages of the upstream and downstream 10 ORFs along with the ORF in question (altogether 21 ORFs) was calculated for each ORF (for the first and last 10 ORFs of each S288c chromosome, medians of the median coverages of the first and last 21 genes of the chromosome, respectively, were used). Then the ratio of the ORF to the 21 local ORFs was calculated and used as an estimation of gene copy number corrected for haploid genome. Gene copy number variations (CNVs) below 0.25 (tetraploid genomes) or below 0.4 (diploid genomes) were set to 0 (note that fraction gene copy numbers are not unrealistic when the CNVs are corrected for haploid genome). Pangenome ORFs: because the loci of pangenome

ORFs are not known, the median coverages of these were compared with median coverages of high coverage 21-ORF bins of each genome that were obtained in the previous step. High coverage ORF bins were defined as the lowest values of the top 10% of 21-ORF median coverage bins. In the case of pangenome ORFs, only copy numbers of at least one were kept in the final analysis. Finally, gene CNVs in the S288c and the extended pangenome were compared across subclones: presence/absence variations of ORFs were assessed, along with at least twofold variations in copy numbers. Plots were generated from these data, whereas calculated copy numbers and comparisons are uploaded to FigShare in a table format (doi: 10.6084/m9.figshare.14216663). mtDNA, 2 μ plasmid, and Ty transposon copy numbers were also calculated for each genome using the pangenome mapping, from the median of their respective ORF median coverages. These values were compared with the median of chromosomal median coverages (obtained from the mapping to the S288c genome). Their ratio was used as the copy numbers of these features calculated for the haploid genome.

2.3 | Multiplex PCR

We performed our recently developed interdelta and microsatellite fingerprinting multiplex PCR method to rule out that the subclones obtained from products are contaminations that do not correspond to the actual strain. Briefly, we combined interdelta, microsatellite (YLR177w, YOR267c), and, as a control, ITS 1-4 primer pairs in a single PCR reaction (Imre et al., 2019). Then, after gel electrophoresis, we compared the strains to the derived subclones to identify band patterns that could indicate the presence of isolates other than the original strain.

2.4 | Colony morphology and petite test

Heterogeneity in colony morphologies (colony phenotype switch) and frequency of petite mitochondrial mutants in packed products were assessed by plating samples directly after the first preculturing (as described above) onto YPD agar plates (for colony morphologies) and onto GlyYP (glycerol yeast extract peptone) + 0.1% glucose agar plates with cell densities of approximately 200 per plate (after cell counting in a hemocytometer). Vented plastic plates with 90-mm diameter and 14-mm height were used. Plates were incubated for 10 days at 30°C (with agar surface facing down) and were visually scored for various phenotypes on YPD (rough, wrinkled, sectorized, stalk-like, and very small colonies) and for potential petite mutants on GlyYP. Presumed petites were transferred to YPD and after overnight culturing were inoculated onto GlyYP plates without glucose. Subclones unable to grow on glucose-free GlyYP were scored as petites. Finally, YPD colonies were washed under tap water to determine the frequency of invasivity into agar. At least 1000 subclone colonies were counted for each strain, and for each assay, raw data were uploaded to FigShare (doi: 10.6084/m9.figshare.12673256).

2.5 | Spot-plate assays

Tolerance to various stress factors with a focus on industrially relevant stresses (Gibson et al., 2007; Qiu et al., 2019) and growth on rich and minimal media were assessed using the spot-plate method for all strains and all subclones of the strains. The following stress media based on SD (synthetic defined, 2% glucose, 0.67% yeast nitrogen base without amino acids) were used: ethanol and high sugar osmotic stress (bioethanol, ale, lager, wine yeasts), NaCl and high sugar osmotic stress (ADY_Baker), and salt and oxidative stress (H_2O_2) media for the probiotic yeast. In preliminary experiments, we determined the optimal concentrations of stressors that may enable differentiating between subclone lineages (Table S2). Vented plastic plates with 90-mm diameter and 14-mm height were used. Samples grown overnight (30°C) on YPD plates were washed in ddH₂O, prepared in equal cell concentrations after cell counting with a hemocytometer, and spotted in 10- μ l drops in a series of approximately 50,000, 5000, 500, 50, and five cells to the various plates. The samples originating from the initial isolations were briefly stored at 4°C; single-cell bottle-necks were avoided as described above. Plates were incubated at 30°C for 2 days before photographing them using a DSLR camera. Growth was evaluated visually, and plate photographs were uploaded to FigShare (doi: 10.6084/m9.figshare.12673253).

2.6 | Clonal heterogeneity test

Clonal heterogeneity under stress was assessed by using the same stress conditions as in the spot plate assays, supplemented with assays on SD and YPD media. Vented plastic plates with 90-mm diameter and 14-mm height were used. To test the effect of using various culture dishes, we also used round culture dishes of 50-mm diameter and 14.7-mm height, 140-mm diameter and 20-mm height (Sarstedt, Nümbrecht, Germany), and rectangular 128 mm \times 86 mm OmniTray (Nunc, Roskilde, Denmark) plates.

Freshly grown cells were counted in a hemocytometer and spread to land about 200 cells per plate. For test experiments, we used both 2 and 18 h precultures in liquid YPD (30°C, 180 rpm rotation) in liquid YPD (30°C, 180 rpm rotation). For test experiments, plates were incubated at 30°C for 3 days. For subsequent experiments, overnight precultures were used and plates were incubated at 30°C for 2 days (YPD and SD media) or for 2, 3, 4, and 6 days (stress media) as colonies reached sizes that were visible but not yet close to each other. For each condition, three replicate plates were used for each sample. Photographs were taken of the plates with a DSLR camera in high resolution (14 megapixels) and with constant lighting and previously adjusted white balance and with uniform black background. Using these photographs, data on colony area was gathered by using the Fiji software package CountPHICS (<https://www.fuw.edu.pl/~bbrzozow/FizMed/countPHICS.html>) (Brzozowska et al., 2019) with circularity set to 0.8 (this helped to exclude colonies grown too close and directly bordering others). Area of interest was specified in a manner to avoid colonies near the edge of the plate. The smallest colonies on

the plates were measured using ImageJ (<https://imagej.net/>, RRID: SCR_003070) (Rueden et al., 2017), and minimum colony size was specified accordingly in COUNTPHICS. Pixel to millimeter ratios were measured, and area calculations were randomly verified by manual measurement in ImageJ for altogether 10 colonies. To test the reliability of measurements, we also mixed a high and a low heterogeneity yeast strain directly before plating in 1:1 ratio to test whether the resulting observable heterogeneity was in between that of the two mixed yeasts in pure culture. Plate photographs and all colony area values were uploaded to FigShare (doi: 10.6084/m9.figshare.14216624 and doi: 10.6084/m9.figshare.12673256 for plate optimization and for tests, respectively).

2.7 | Statistical analysis of clonal heterogeneity data

Analyses were done in the R environment for statistical computing (R Core Team, 2020). Prior to analyses, colony size data were square root-transformed to bring value distributions closer to Gaussian; also, following square root transformation, data were rescaled by carrying out z score transformation (i.e., subtracting variable mean from all values, then dividing by standard deviation) to aid model fitting in later analyses. We used the linear regression modeling of the Bayesian approach, utilizing the R package “MCMCglmm” (<https://cran.r-project.org/web/packages/MCMCglmm/index.html>) (Hadfield, 2010), because it allows for flexible model specifications and estimates are less sensitive to group size differences than ordinary least squares methods. First, to test how heterogeneity was dependent on growth conditions across the different strains, we fitted a model with the rescaled colony area measurements as response variable and specific grouping variable accounting for both strain and growth condition (i.e., practically controlling for strain, condition, and the interaction of these, without including empty factor levels, i.e., untested strain-condition pairs) as fixed predictor. Model specification was done in a way so that group-level residual variances could be estimated. Because measurements originated from Petri dish repeats and colonies within Petri dishes were of common origin, we included repetition ID nested within strain as random effect to control for nonindependence in the data. In the results, we assessed growth condition-related differences in group heterogeneities by contrasting posterior distributions of residual variance estimates. Statistical significance was established by using 95% highest posterior density (HPD) intervals (analogous to confidence intervals in frequentist modeling): for contrast estimates where the 95% HPD interval did not cross zero, the difference between the contrasted groups is considered to be statistically significant.

Second, when testing how subclones differ in colony area heterogeneity from their original sample under salt stress, two separate models were fitted, using data from 4 and 6 days of incubation. This separate analysis for 4 and 6 days data were necessary because of the nonindependence in the data due to the temporal correlation between the measures carried out at Days 4 and 6. Because we did

not want to test the effect of time (4 vs. 6 days), and measurements of Day 6 inherently depend on (are correlated with) measurements on Day 4, using separate models was preferable to more complicated model specifications. We note here that it is possible that a single or multiple subclone lineages within a single colony may appear and quickly invade a sector in a colony, resulting in an asymmetrically growing sectored colony. In such a case, Day 6 measurements would not inherently depend on Day 4 measurements. However, sectored colonies were found to be very rare even after 10 days of incubation; thus, their hypothetical effects can be ignored here. In these models, square root- and z score transformed colony size was the response variable, and strain was a fixed predictor. Because we wanted to compare heterogeneities of subclone lineages with that of the commercial ADY_Baker product, in the models, residual variances of groups were estimated for strains separately. Similarly to the above described model, repetition ID nested within sample was used as random effect. In the results, we report posterior distributions of contrast parameters for residual variance estimates compared between subclones and the initial commercial sample.

For all models, weakly informative proper priors were specified; for random effect variances, parameter expanded priors were used to aid the mixing of the Markov chains for random effect variances. During model fitting, sampling of the posterior distributions was run for 105,000 iterations, from which the first 5000 were discarded as “burn-in,” and from the Markov chain Monte Carlo (MCMC) process, only every 50th samples were retained (called thinning interval), yielding a nominal sample size for parameter estimate posterior distributions equal to 2000. Model diagnostics included visual checking of MCMC chains for trends in the chain trajectories (plotting MCMC samples in the order of iterations) and calculation of autocorrelation in the MCMC chains at lag of the thinning interval (MCMC chains were considered to be mixing well if absolute value of estimated autocorrelation coefficient was lower than 0.1).

2.8 | Data availability

The data underlying this article are available in NCBI SRA and can be accessed under BioProject ID PRJNA646688 (raw genome sequencing reads) and in FigShare at doi: 10.6084/m9.figshare.14216624, doi: 10.6084/m9.figshare.12673256, and doi: 10.6084/m9.figshare.12673250 (raw phenotyping results, including all plate images); doi: 10.6084/m9.figshare.14216648 (called SNP VCF files) and doi: 10.6084/m9.figshare.14216663 (pangenome ORF data).

3 | RESULTS

3.1 | Industrial yeast samples and ploidy

In this study, we obtained probiotic, ale, lager, wine, bioethanol, and baking (active dry) yeasts from commercial vendors. Five of these samples belong to strains with sequenced genomes and known ploidy, whereas the ADY_Baking yeast was sequenced and analyzed in this study for the first time. The genomes of the probiotic, bioethanol, and wine yeast were euploid. The ADY_Baker was euploid tetraploid or aneuploid diploid, depending on the subclone lineage (two of which were sequenced). The ale and lager yeast showed previously described extensive aneuploidies; however, these were not identical when different studies were compared and genomes from these were reanalyzed for aneuploidies (Figure 1, Table S1).

Strains with multiple sequenced sublineages showed variable amounts of genetic distance when the nuclear genomes were compared and visualized in the form of neighbor nets (Figure S1a) (largest difference was among the two subclones of the ADY_Baking yeast), but the mitochondrial genomes were almost identical for all subclones of the individual strains (Figure S1b). All *S. cerevisiae* yeast genomes lacked interspecific hybrid chromosomes and only contained short

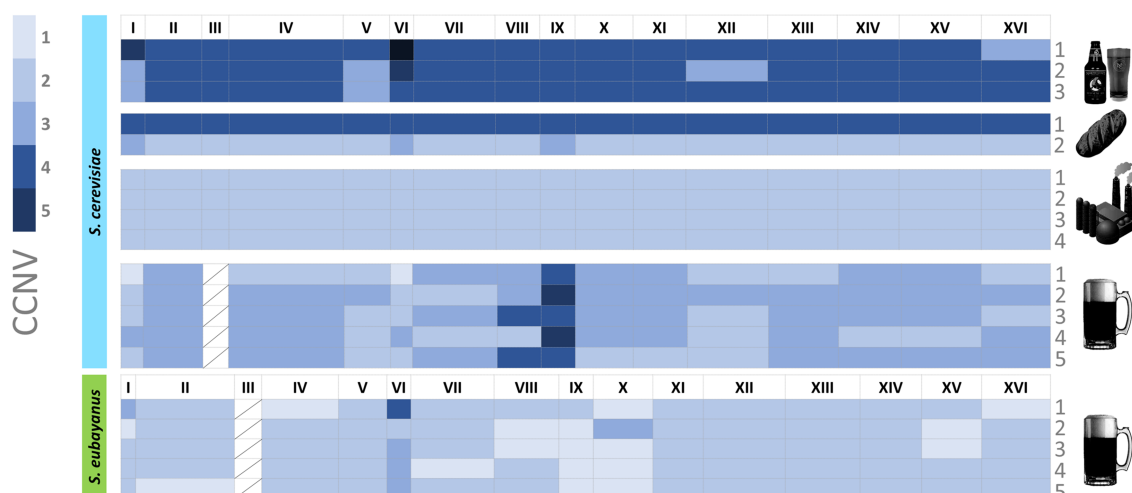


FIGURE 1 Observed chromosome copy numbers of the yeasts with multiple subclone lineages subjected to genome analysis. Commercial yeast strains are represented by pictograms of their applications (ale, baker, bioethanol, lager); numbers given to subclones (as in Table 1) are indicated. Note that the rearrangements affecting the third chromosomes of Lager yeasts prevent the assignment of a simple copy number (see also Figure S3) [Colour figure can be viewed at wileyonlinelibrary.com]

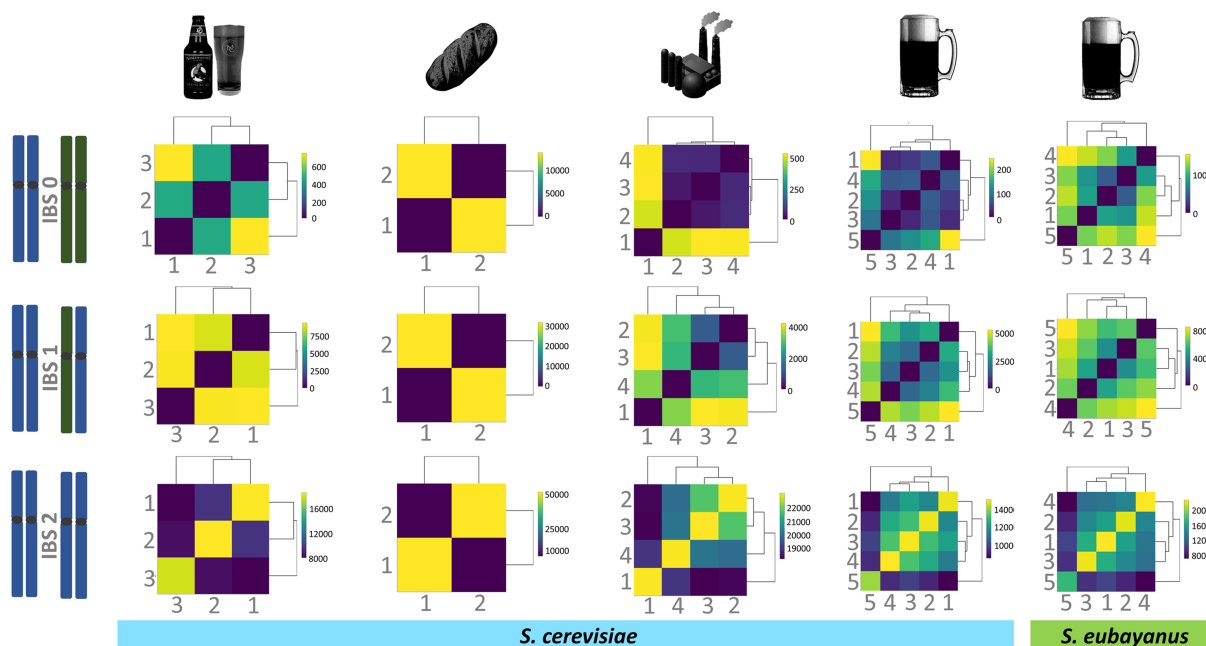


FIGURE 2 Identity-by-state analysis. SNPs of yeasts with multiple sequenced subclones were analyzed for IBS0 (no shared alleles), IBS1 (one shared out of two alleles), and IBS2 (all alleles shared) states (IBS states are illustrated with chromosome cartoons on the left). Commercial yeast strains are represented by pictograms of their applications (ale, baker, bioethanol, lager); numbers given to subclones (as in Table 1) are indicated. Numbers of each IBS state are color coded in the pairwise comparisons. Note that scales are different for all panels [Colour figure can be viewed at wileyonlinelibrary.com]

elements of likely introgressions; thus, in subsequent analyses, mapping to the S288c reference genome was used for these (Figure S2). Identity-by-state analysis showed the highest number of SNPs to be in IBS1 or IBS2 states (Figure 2). The sequenced sublineages showed various conspicuous intrachromosomal changes in coverage pointing to GCR events (Figures S3 and S5) that often differed between subclones, especially in the case of the lager yeast, along with runs of homozygosity (ROH) and various allele ratio changes (Figures S4 and S5). Some of the ROH regions were different among the subclones; we scored these as LOH events. Figure S5 summarizes the amount of GCR, non-LOH and LOH allele ratio event identified between subclone lineages of the strains, ranging from a single identified event to several dozens. Of particular interest are the bioethanol strains that showed a high variation in the number of LOH events identified (ranging from 1 to 12 across subclone comparisons) and the ADY_Baking yeasts, with subclones differing by four rearrangements, 32 allele ratio changes, of these 13 were LOH events. GCR events were most common in the lager yeasts. Furthermore, we calculated copy numbers of all ORFs of the subclone lineages (all S288c ORFs and various other ORFs described for the species' pangenome), and this showed that many have different copy numbers (Figures 3a and S6). These inter-subclone CNVs mostly affected the S288c ORFs, with 63 to 117 showing inter-subclone CNVs in the case of the ale yeast, 27 in the ADY_Baker yeast, and 17 to 89 in the bioethanol yeast. In almost all cases, CNV differences were less than 10-fold when subclone lineages were compared. Pangenome ORFs more commonly showed presence/absence variation (Figure 3). Ty transposon, mtDNA, and 2μ

plasmid CNVs were also observed (Figure 3b–d). The CNV of transposons was generally low, except for the comparison of ale subclone 3 with the other ale yeasts. mtDNA CNV was highest among the ale yeasts and plasmid CNV among the bioethanol yeasts.

As the dozens of novel subclone lineages generated during this study were not sequenced, genome data were only used to confirm the existence of karyotypically different subclone lineages of strains analyzed in recent literature and in our study, without an in-depth analysis of GSV phenomena's phenotypic effects.

3.2 | Heterogeneity of colony phenotypes in commercial yeast products

We determined heterogeneity in colony morphology, invasivity, and petite frequency in the industrial yeast strains directly, without initially subculturing the actual product. We found remarkably variable colony phenotypes (Figures 4 and S7) and, at the same time, large variability in the fraction of atypical colonies, ranging from 0.78% of variable morphologies (bioethanol) to as much as 27.36% in the ale yeasts. In the case of the ale strain, wrinkled and conspicuously small colonies were the most prevalent. Stalk-like colonies (Figure 4: st) were observed, although with negligible frequencies, in four of the six strains. Proportion of invasivity among the colonies ranged between 0.31% (bioethanol) and 34.62% (ADY_Baking), and various types of invasive growth could be observed among the samples. Especially in the case of the ADY_Baking and the probiotic yeast, different invasive

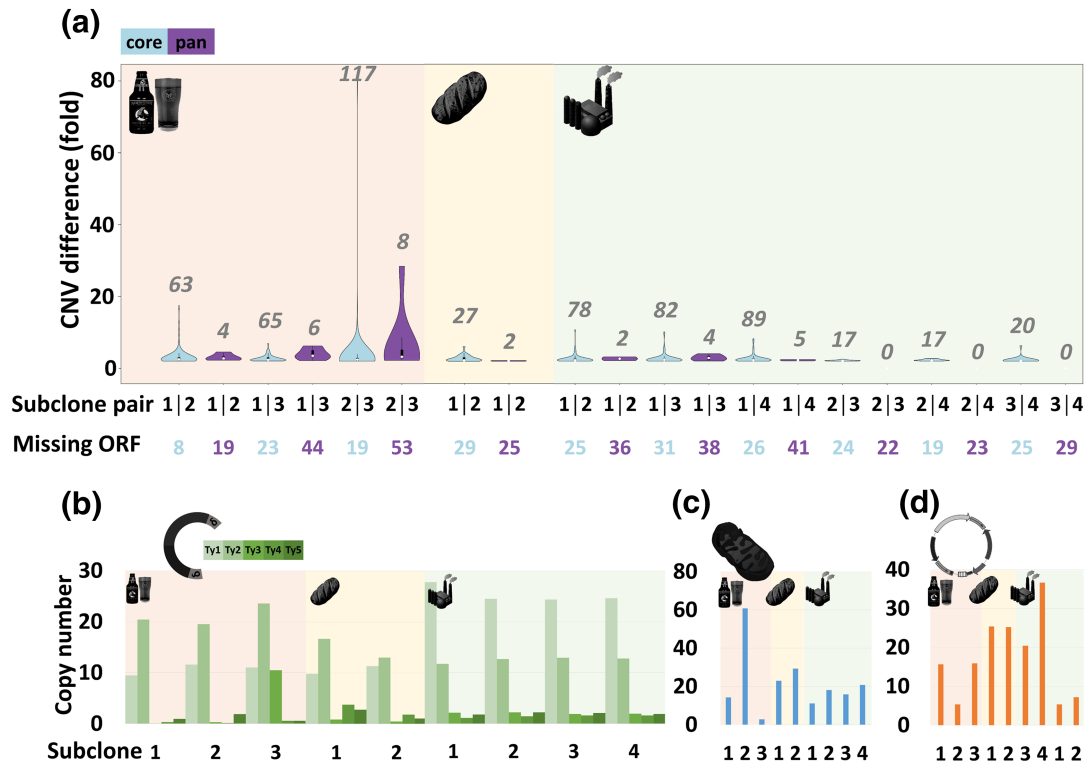


FIGURE 3 Copy number analysis. (a) Numbers of ORFs with at least twofold copy number variations with the distributions of fold differences (vertical axis) in copy numbers. Pairwise subclone comparisons are on the horizontal axis, and below each comparison, the numbers of ORFs missing from one of the compared subclones is indicated. All comparisons are separated for S288c core ORFs (light blue) and pangenome ORFs (purple). (b) Copy numbers of Ty1–5 retrotransposons in each subclone, calculated for haploid genome. (c) Copy numbers of mitochondrial DNA in each subclone, calculated for haploid genome. (d) Copy numbers of 2μ plasmids in each subclone, calculated for haploid genome. Commercial yeast strains are represented by pictograms of their applications (ale, baker, bioethanol); numbers given to subclones (as in Table 1) are indicated [Colour figure can be viewed at wileyonlinelibrary.com]

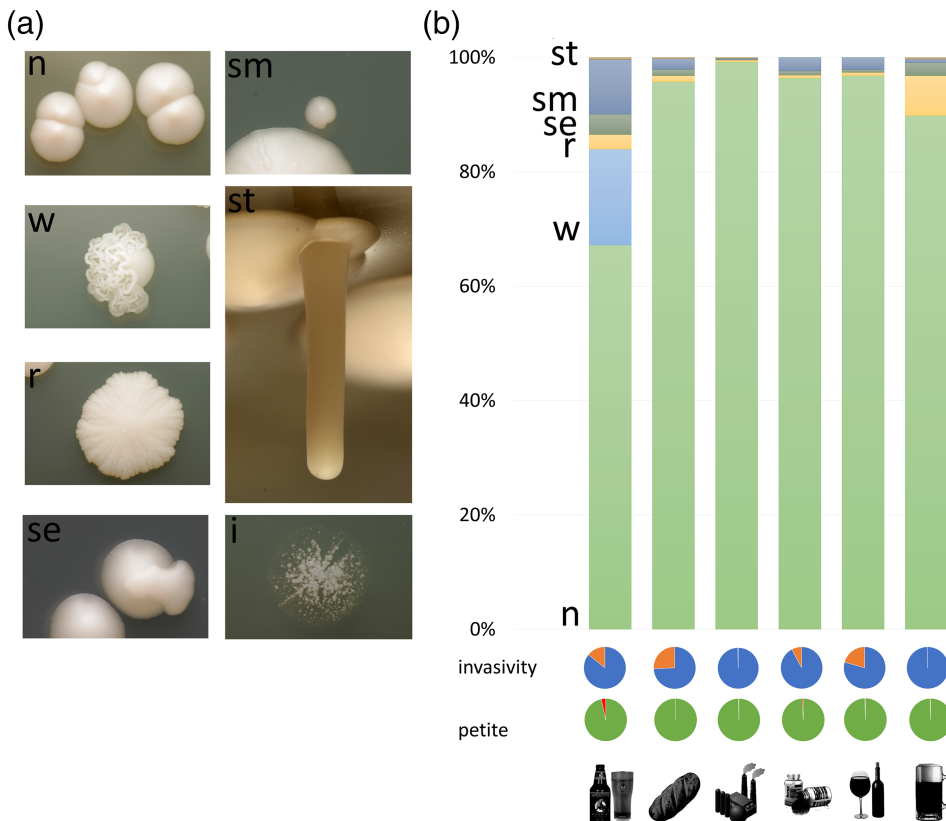


FIGURE 4 Clonal heterogeneity: colony morphologies. (a) Examples of colony morphologies; n, normal; sm, small; w, wrinkled; st, stalk-like; r, rough; se, sectorial; i, invasive. (b) Stacked bar charts representing the distribution of colony morphology types in rich medium for each commercial yeast product (ale, baker, bioethanol, probiotic, wine, lager) identified by pictograms of their applications. Percentage of invasive (on rich medium, in orange) and petite mutants (in red) are shown in pie charts below the main stacked bars [Colour figure can be viewed at wileyonlinelibrary.com]

phenotypes co-occurred. In the case of the bioethanol strain, rough morphology and invasiveness always co-occurred; in other strains, such a clear link was not observed between these traits. Frequency of petites reached more than 1% only in the case of the ale yeast.

3.3 | Heterogeneity of typical colonies and influence on stress tolerance

After observing heterogeneous colony phenotypes and considerable differences in the frequencies of abnormal colony phenotypes, we isolated 12 subclone colonies from each industrial strain that showed the entire, circular, smooth-surface colony phenotypes with the assumption that such regular colonies are the ones most likely to be chosen upon isolation and establishment of a pure lineage in laboratories working with yeasts, whereas very small or highly unusual colonies are consistently avoided. We avoided subculturing (single-cell bottlenecks) and prolonged culturing of these subclone lineages and characterized them within weeks of isolation phenotypically, using the colonies saved at 4°C on YPD plates. Thus, we avoided preparing stocks and reviving yeasts from stocks as it has genotypic and phenotypic consequences on yeast populations with standing genetic variation (Wing et al., 2020). These individual lineages were subjected to multiplex fingerprinting PCR. All subclone lineages showed fingerprinting patterns that were identical or, in the case of the ADY_Baking and ale yeasts, identical except for the occasional loss of a single band (out of 12 bands). All strains showed clearly different patterns from each other, thus, contamination or cross-contamination of the samples could be excluded, and subclone lineages were proven to be derived from the actual strain (Figure S8).

Spot-plate tests revealed differences in stress tolerance among these subclones lineages established from regular colonies. Visible differences in growth under various stress conditions were observed for half of the strains with the spot-plate method, namely, for the probiotic, ale, and ADY_Baker yeasts (Table S2). In all of these cases, a minority of subclone lineages (one to three subclones depending on strain and condition) showed impaired growth under stress when compared with other subclones or to the original sample that was not subjected to single-cell bottlenecks.

Following spot plate phenotyping, we determined whether we can reliably assess clonal phenotypic heterogeneity using single-cell derived colony area as a proxy to fitness. We chose two yeasts that showed low and high variance in colony areas during the experiments originally designed to assess heterogeneity in colony morphologies, as described above (bioethanol and ADY_Baking, respectively). These were plated onto various plate types with rich medium to give single-cell derived colonies (Figures S9–S11). We plated samples from the yeast products after a very short incubation to revive cells (2 h) and also after overnight (18 h) incubation. Colony growth was followed and later evaluated (with colonies touching each other or dish walls automatically discarded, as described in Section 2). This test showed that (1) overnight precultures somewhat diminished diversity in colony sizes for both yeasts, in the case of the ADY_Baker yeast, mostly

by depleting cells that could have formed very small colonies after plating, (2) plate types can affect observable heterogeneity but mostly when too small plates and long incubation times are used, but the more heterogeneous yeast was the ADY_Baker yeast in all test plates, (3) mixing low and high heterogeneity yeasts before plating results in an observable heterogeneity that is in between those of the unmixed cultures (Figure S12). Under stress condition, we were also able to obtain a strongly bimodal distribution of colony sizes when mixing the lager and ADY_Baker yeast (Figure S11) that showed different tolerance of osmotic stress on spot plates.

Subsequently, we settled for 90-mm diameter plates and shorter incubation times for the rich medium and longer ones for the stress media. Clonal heterogeneity of the six strains during growth on rich and minimal medium and under stress was thus evaluated. Clonal heterogeneity in the form of variable colony sizes in a single sample from a single strain was prevalent in most samples (Figures 5, S13, and S14). Based on the posterior distributions of residual variance parameters, the strains showed variable levels of heterogeneity in different conditions, which was also apparent from the estimated contrast parameters comparing group-level residual variances between groups (Figure S15). Group-level residual variances of the measurements estimated with MCMC-generalized linear mixed model (GLMM) was used to interpret heterogeneity (in this context, higher variation around the group mean corresponds to higher heterogeneity in the measured phenotype, i.e., the colony areas). Figure S15 shows that clonal heterogeneity under various growth conditions differs significantly in most cases. That is, when heterogeneities under different conditions were compared, the strains showed significant differences in all (ale, lager) or all but one (ADY_Baking, bioethanol, probiotic, wine) of those comparisons. Thus, for every strain, the level of observable heterogeneity was greatly dependent on the stress/nonstress condition applied, and the ADY_Baking strain showed the highest differences across conditions. When we compared group heterogeneities (i.e., posterior distributions of group residual variances) between strain pairs, separately for each stress/nonstress condition (Figure S16), similarly, the ADY_Baking strain was the yeast that displayed significantly higher measures than others in the highest number of cases. For example, in minimal medium, its heterogeneity was significantly higher than that of the ale, lager, and bioethanol strain and statistically not different from that of the probiotic and wine yeast. In rich medium, its heterogeneity was significantly higher in all but one pairwise comparison (compared with the probiotic yeast, its difference was not significant). Under stress conditions (where fewer pairwise comparisons were made due to different stress conditions applied), the wine and the ADY_Baking strains' heterogeneities were notable. The former showed significantly lower heterogeneity in three out of four pairwise comparisons, whereas the latter showed significantly higher heterogeneity in the same number of comparisons.

Based on these results, the ADY_Baking yeast and its 12 isolated subclones were subsequently chosen to further compare how clonal heterogeneity can influence not merely phenotypes but the level of diversity in cell populations derived from subclones. As described above, this strain showed considerable differences in subclones' spot

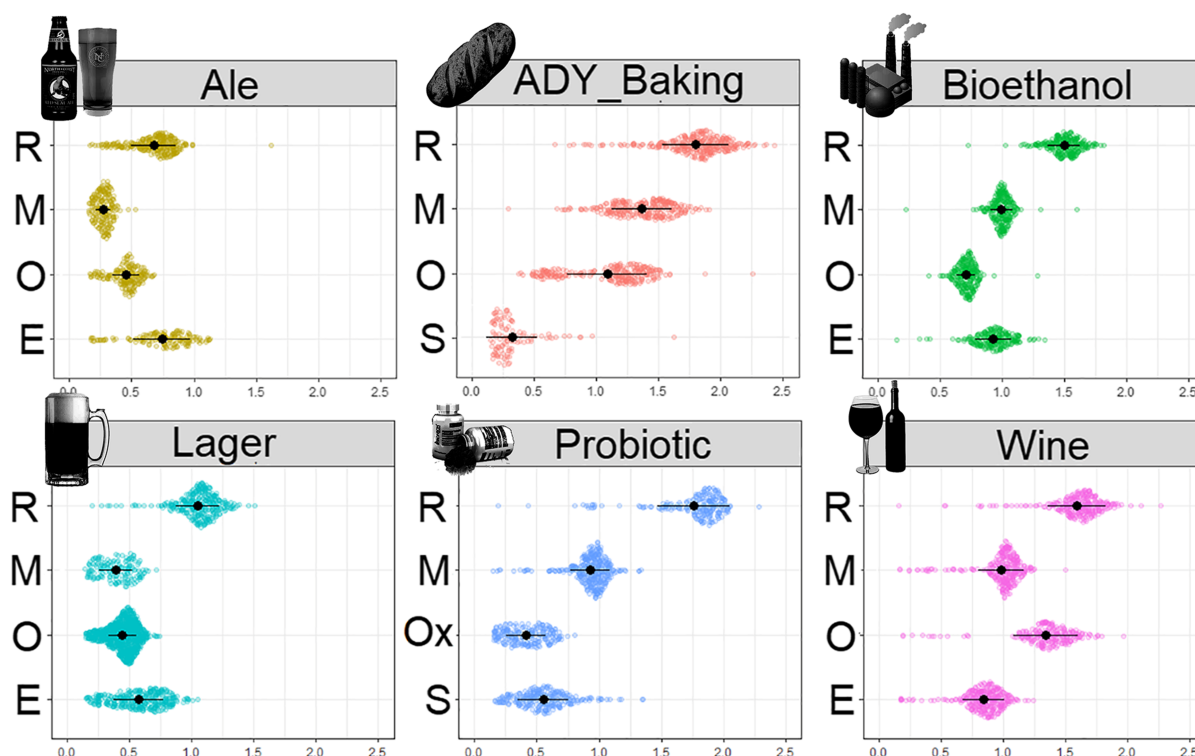


FIGURE 5 Clonal heterogeneity: colony area (square-root transformed) distributions under various conditions for the six commercial yeasts (identified by pictograms of their applications). R: rich; M: minimal; E: ethanol stress; O: osmotic stress; S: salt stress; Ox: oxidative stress medium. Black dots represent group means; black horizontal lines represent standard deviations [Colour figure can be viewed at wileyonlinelibrary.com]

plate tests (Table S2, Figure 6a,b). The level of heterogeneity in the case of subclones and in the original sample under the salt stress condition was compared at two different time points (4 and 6 days) after inoculation, in the following manner. First, growth on minimal SD medium was confirmed to be identical for the subclone lineages using the spot plate method, and then the distributions of colony areas were compared under salt stress (Figures 6a,b, S17, and S18). In most cases, heterogeneity was significantly different between the original commercial sample (which generally showed weaker stress tolerance manifesting in generally smaller colonies but significantly higher heterogeneity) and each of its subclones, except for subclone B (Day 4) and subclones B and J (Day 6) when mean phenotypes (without residuals) were considered (Figure S19). Regarding residual variances in colony size distributions, all subclones showed significantly lower heterogeneity compared with the initial commercial sample at both time points except for subclones B, D, and J in the case of Day 4 measurements.

4 | DISCUSSION

Clonal heterogeneity is a familiar phenomenon for anyone working with culturable microbes. Single-cell isolates from microbial cultures are routinely obtained for various purposes, for example, for subsequent physiological studies, genetic characterization/modification, metabolic engineering, or even for industrial stock propagation, with the advantage of leveraging a simple visual check for eventual

contamination with other microbial species. Differences in morphology or size among the grown colonies are often observable to the naked eye. Yet, the underlying causes and, perhaps more importantly, the consequences of single-cell bottlenecks (the isolation of a given single-cell colony before an experiment) are mostly neglected: the start of evolutionary processes are assumed to coincide with the start of the respective experiments, whereas apparently, diversification and selection starts much sooner.

Studies on the emergence of de novo mutations, genome structure variations, and clonal interference in industrial *Saccharomyces* strains (Bellon et al., 2018; Gibson et al., 2020; Gorter de Vries et al., 2019; Lairón-Peris et al., 2020; Mangado et al., 2018; Sampaio et al., 2019; Voordeckers et al., 2015; Zhang et al., 2016) have led to increased understanding on their adaptation. In comparison, relatively few yeast studies have been devoted to the importance of clonal heterogeneity in adaptation (Bódi et al., 2017; Holland et al., 2014; Vázquez-García et al., 2017) or to understand how epigenetics, gene expression noise, metabolic state, unequal cell division, chronological or replicative age differences, or prions cause yeast populations to be heterogeneous (Ackermann, 2015; Adamczyk et al., 2016; Cerulus et al., 2016; Duveau et al., 2018; Halfmann et al., 2012; Hewitt et al., 2016; Nadal-Ribelles et al., 2019; Yvert et al., 2013). The latter study areas, to our knowledge, exclusively focus on lab strains and on phenomena constantly generating heterogeneity on time scales (Sumner & Avery, 2002; Wheals & Lord, 1992) shorter than those discussed in our study.

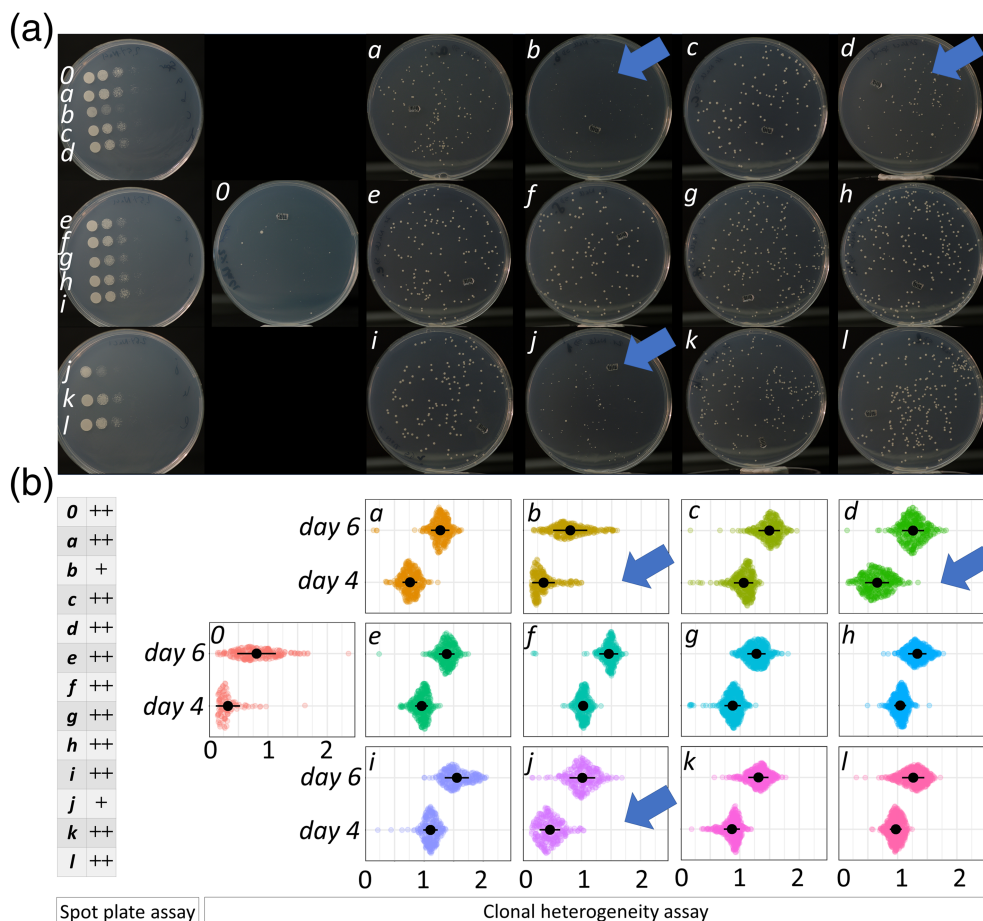


FIGURE 6 Clonal heterogeneity both in growth and in the heterogeneity of growth under stress for the baking yeast. (a) Spot plate images obtained for salt stress condition (left) showing variation among subclones. The ADY_Baking sample is identified by “0” and its 12 subclones (named a-l) are identified by their letters. Individual plate images (right) show clonal heterogeneity visually after 4 days of incubation under salt stress. Highly heterogeneous subclones (that are statistically not different in heterogeneity from the original ADY_Baking sample) are identified by blue arrows. (b) Left: evaluation of spot plate results. Right: Colony area (square-root transformed, on x axis) distributions after 4 and 6 days under salt stress for the ADY_Baking sample and its 12 subclones, obtained from the plate test illustrated in the top of the figure. Black dots represent group means; black horizontal lines represent standard deviations. Highly heterogeneous subclones (where residual variances were not significantly different from that of the ADY_Baking sample) are identified by blue arrows [Colour figure can be viewed at wileyonlinelibrary.com]

Among the factors mentioned above, de novo mutations and GSVs can result in heritable differences among subclone lineages (whereas other above mentioned mechanisms cause constant cell-to-cell heterogeneity without genetic heritability in the strict sense). However, studies comparing *Saccharomyces* strains rarely address the “founder effect” of using a subclone lineage of a strain (due to methodological constraints) to characterize the strain itself. Only a few studies have focused on heterogeneous subclone lineages as well as cryptic variation of the PE-2 bioethanol strain or its derivative JAY270 (Reis et al., 2014; Rodrigues-Prause et al., 2018; Sampaio et al., 2019) and those of the lager W34/70 strain (Bolat et al., 2008; van den Broek et al., 2015), as well as on lineages on the VIN7 commercial hybrid wine starter yeast (Morard et al., 2020), whereas in most other cases, strains are used interchangeably with subclone lineages in literature. In fact, the commonly used and well-known tetraploid Ale strain has been sequenced and analyzed by three recent studies. All of these studies found different karyotypes due to apparent genotypic heterogeneity of

the actual subclone lineages studied by each. This genotypic heterogeneity did not only affect the karyotypes, but allele ratios, LOHs as well (Table S1, Figures 1 and S5). In the case of our tetraploid ADY_Baking active dry yeast, we could identify excessive karyotype heterogeneity (along with allele ratio changes partially attributable to gene conversions, LOH events, and chromosomal rearrangements) within a single batch of yeasts, which may either be caused by meiotic or mitotic processes. Bioethanol yeasts did not show variable karyotypes but often showed LOHs, whereas karyotype, GCR, and allele ratio changes/LOHs were, as expected from literature, common among the lager yeasts. Karyotype changes and LOH events are important as they are known to be adaptive (Gilchrist & Stelkens, 2019) and may even influence cell and colony morphology (note the highly heterogeneous colony morphology for the ale strain observed here) (Tan et al., 2013) and stress adaptations not only in industrial strains (e.g., Kadowaki et al., 2017; Morard et al., 2019; 2020) but also in pathogenic *Saccharomyces* (Raghavan et al., 2019).

It must be noted that in the case of pathogenic yeast species, the existence of genotypically different subclone lineages of strains is more often taken into account in the context of comparability among labs (Abbey et al., 2014; Franzot et al., 1998) or in the context of heteroresistance to antimycotics (Stone et al., 2019). As clinical *Saccharomyces* isolates are regularly derived from commercial (baking and probiotic) yeasts (Imre et al., 2019; Pfliegler et al., 2017), the clonal heterogeneity inside yeast products should be taken more often into account, when the goal is to understand how stress resistance of industrial yeasts translates into colonizing and pathogenic potential. For example, in a recent study, we compared commercial and clinical yeasts but did not test multiple subclone lineages of a given strain (Pfliegler et al., 2017), a fact that may have influenced our observations due to founder effects.

Based on the facts that genotypic heterogeneity is widespread, especially in tetraploid and hybrid lineages, and that the relatively long (Large et al., 2020) industrial yeast cultivations may already be considered a stressful selective environment (Qiu et al., 2019), here we designed experiments to quantify and compare heterogeneity in industrial, commercially obtained yeasts. It must be noted that the PE-2 bioethanol strain was obtained from a culture collection for the phenotyping test (while its sequenced subclones originated from a commercial product) and thus did not go through extensive culturing before packaging in the form of a yeast product. We found an immense heterogeneity in several cases when colony morphologies and invasivity were assessed inside single batches of one strain (Figures 4 and S8). Besides rough, wrinkled, and very small colonies, two other observable types are especially interesting. Sectored colonies are themselves naturally arising illustrations of clonal heterogeneity and interference (when lineages inside the colony compete for space as the colony grows), and the fact that in merely 10 days of incubation sectored colonies were as common as ~2% and ~3% in the lager and ale strains, respectively, shows that the emergence of heterogeneous subclone lineages is more of a rule than an exception. The second remarkable colony phenotype was the stalk-like growth previously described and linked to craters in the agar surface by two studies with *Saccharomyces* (Engelberg et al., 1998; Scherz et al., 2001).

After assessing heterogeneity of single-cell colonies in our strains, we assumed that in routine microbiological workflow, unusual colonies are usually avoided when a pure lineage is to be established. Thus, we obtained 12 subclone lineages that did not show altered morphologies and subsequently showed that even these seemingly uniform lineages can be heterogeneous in their fitness under various stresses (Table S2). Subsequently, the simple plating method used by us decoupled fitness from clonal interference by isolating cells to form hundreds of distant colonies, enabling the simultaneous study of high and very low fitness subclone lineages at a given timepoint within a strain or within a subclone lineage. By applying MCMC-GLMM statistic modeling to such single-cell colony measurements, we showed that each strain is different in the level of heterogeneity, whereas a single strain may also display different levels of heterogeneity depending on the condition (Figures S15 and S16). Finally, we also showed that subclone lineages do not only differ in their phenotypes but can also be significantly different in their potential to generate clonal heterogeneity (Figure S17). Although we have not determined the relative contributions of genetic, epigenetic, or cell age factors affecting heterogeneity, our experimental design of phenotyping (evaluated after days of growth on agar media) plausibly strongly suppressed all but the heritable genetic factors among the intrinsic factors of variation. Some extrinsic factors of variation, for example, microvariations in the agar medium or physical location of colonies on different parts of the plate likely remained in our setup (although colonies on plate edges were discarded from measurements), but these presumably affected all yeasts similarly and would not result in differences in heterogeneity. Yet, various plate types all enabled the differentiation of high and low heterogeneity yeasts, underscoring the applicability of our method. Additionally, the separate growth of colonies on agar media eliminated clonal interference on the test media, enabling the observation of very low fitness lineages emerging from a given strain or a given subclone.

Interestingly, when cells are propagated to be used in an industrial process, these initial propagation conditions can be rather different from the conditions under which the process per se is carried

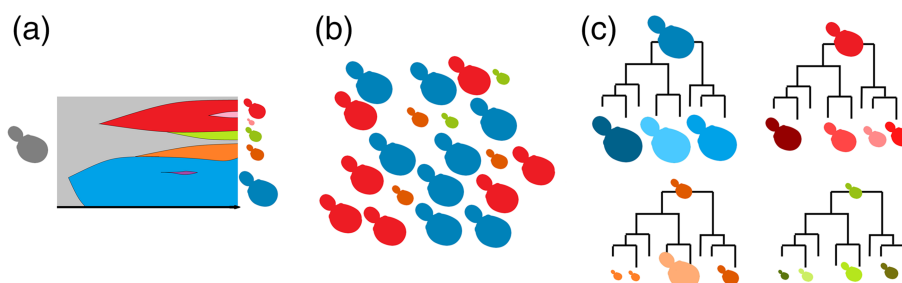


FIGURE 7 Proposed model of the origin and consequences of clonal heterogeneity in commercial *Saccharomyces* yeasts. (a) A stock culture (gray) is used as an inoculum to initiate production, leading to large amounts of cells grown over the course of days (x axis). Subclone lineages (in red, pink, green, orange, and blue) emerge due to genome instability and mutations and compete under (clonal interference) and are selected by stress factors, resulting in changing frequencies of the lineages, as represented on the y axis. (b) A final product that contains a heterogeneous yeast population with various frequencies of subclone lineages (clonal heterogeneity). (c) Subsequent experimental results with single-cell derived subclone lineages lead to a founder effect in the form of different phenotypes and different variability. Colored yeast figure sizes refer to variable fitness; frequency of colors refers to frequency of various lineages in yeast products [Colour figure can be viewed at wileyonlinelibrary.com]

out. Whereas the initial propagation steps have the aim of increasing the microbial population, the process has the aim of generating as much of the product as possible (best TRY compromise; T = titer, R = rate, Y = yield). Thus, the selective pressure during the propagation step might not only be rather different but even somehow unfavorable from the selective pressure during the process itself; that is, propagation might select subclones that may not be the best ones for the process.

In conclusion, our experimental setup shows that (1) clonal heterogeneity is widespread in various clades of commercial yeasts as a presumed consequence of microevolution during industrial cell propagation, (2) this heterogeneity affects observable colony morphologies, invasivity, and stress tolerance, and (3) heterogeneity in subsequent generations of a yeast culture is also greatly dependent on which subclone an experiment is based on (summarized in Figure 7). The surprisingly complex heterogeneity of industrial strains should be taken into account in phenotyping and genotyping studies, as well as in strain improvement strategies, as retaining only a single subclone may lead to the loss of interesting phenotypes and diversity potentially important in performance.

ACKNOWLEDGMENTS

This work was supported by the project “Establishing a scale-independent complex precision consultancy system (GINOP-2.2.1-15-2016-00001)” of the Ministry of Innovation and Technology, Hungary. We acknowledge support from the Thematic Excellence Programme (TKP2020-IKA-04) of the Ministry of Innovation and Technology in Hungary. H.V.R. was supported by the ÚNKP-20-3-I-DE-159, A.I. by the ÚNKP-19-3-I-234 and ÚNKP-20-3-II-DE-103, W.P.P. by the ÚNKP-20-5-DE-121 New National Excellence Program of the Ministry of Innovation and Technology from the source of the National Research, Development and Innovation Fund. W.P.P. was supported by a Fulbright Research Award from the Hungarian-American Fulbright Commission. This work was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences for W.P.P. A.K.G. acknowledges funding from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), São Paulo, Brazil, grant number 2018/04962-5. We are deeply grateful for Michael M. Desai, Harvard University, for enabling the genomic sequencing of the three PE-2 clones and agreeing with sharing the corresponding data. We are very grateful for the three anonymous reviewers for their helpful comments and suggestions during the review process.

CONFLICT OF INTEREST

No conflict of interest declared.

AUTHOR CONTRIBUTION

Conceptualization: WPP; Formal analysis: HVR, AI, FM, ZR, WPP; Investigation: HVR, AI, FM, ZR, AKG, WPP; Resources: IP, WPP, TR, JN; Funding acquisition: JN, TR; Visualization: HVR, ZR, WPP; Supervision: WPP, IP; Writing: WPP, AKG, IP.

ORCID

Hanna Viktória Rác  <https://orcid.org/0000-0003-0130-7445>

Alexandra Imre  <https://orcid.org/0000-0001-5544-0327>

Zoltán Rádai  <https://orcid.org/0000-0001-7011-5055>

Andreas Károly Gombert  <https://orcid.org/0000-0001-9557-3773>

István Pócsi  <https://orcid.org/0000-0003-2692-6453>

Walter P. Pflieger  <https://orcid.org/0000-0001-6723-4416>

REFERENCES

- Abbey, D. A., Funt, J., Lurie-Weinberger, M. N., Thompson, D. A., Regev, A., Myers, C. L., & Berman, J. (2014). YMAP: A pipeline for visualization of copy number variation and loss of heterozygosity in eukaryotic pathogens. *Genome Medicine*, 6(11), 100. <https://doi.org/10.1186/s13073-014-0100-8>
- Ackermann, M. (2015, August 6). A functional perspective on phenotypic heterogeneity in microorganisms. *Nature Reviews Microbiology*, 13, 497–508. <https://doi.org/10.1038/nrmicro3491>
- Adamczyk, J., Deregwiska, A., Panek, A., Golec, E., Lewinska, A., & Wnuk, M. (2016). Affected chromosome homeostasis and genomic instability of clonal yeast cultures. *Current Genetics*, 62(2), 405–418. <https://doi.org/10.1007/s00294-015-0537-3>
- Baker, E., Wang, B., Bellora, N., Peris, D., Hulfachor, A. B., Koshalek, J. A., Adams, M., Libkind, D., & Hittinger, C. T. (2015). The genome sequence of *Saccharomyces eubayanus* and the domestication of lager-brewing yeasts. *Molecular Biology and Evolution*, 32(11), 2818–2831. <https://doi.org/10.1093/molbev/msv168>
- Barbosa, R., Pontes, A., Santos, R. O., Montandon, G. G., De Ponzzes-Gomes, C. M., Morais, P. B., Goncalves, P., Rosa, C. A., & Sampaio, J. P. (2018). Multiple rounds of artificial selection promote microbe secondary domestication—The case of cachaça yeasts. *Genome Biology and Evolution*, 10(8), 1939–1955. <https://doi.org/10.1093/gbe/evy132>
- Baym, M., Kryazhimskiy, S., Lieberman, T. D., Chung, H., Desai, M. M., & Kishony, R. K. (2015). Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One*, 10(5), e0128036. <https://doi.org/10.1371/journal.pone.0128036>
- Bellon, J. R., Ford, C. M., Borneman, A. R., & Chambers, P. J. (2018). A novel approach to isolating improved industrial interspecific wine yeasts using chromosomal mutations as potential markers for increased fitness. *Frontiers in Microbiology*, 9, 1442. <https://doi.org/10.3389/fmicb.2018.01442>
- Blundell, J. R., Schwartz, K., Francois, D., Fisher, D. S., Sherlock, G., & Levy, S. F. (2018). The dynamics of adaptive genetic diversity during the early stages of clonal evolution. *Nature Ecology & Evolution*, 3, 293–301. <https://doi.org/10.1038/s41559-018-0758-1>
- Bódi, Z., Farkas, Z., Nevozhay, D., Kalapis, D., Lázár, V., Csörgő, B., Nyerges, Á., Szamecz, B., Fekete, G., Papp, B., Araújo, H., Oliveira, J. L., Moura, G., Santos, M. A. S., Székely Jr, T., Balázs, G., & Pál, C. (2017). Phenotypic heterogeneity promotes adaptive evolution. *PLoS Biology*, 15(5), 1–26. <https://doi.org/10.1371/journal.pbio.2000644>
- Bolat, I., Walsh, M. C., & Turtoi, M. (2008). Isolation and characterization of two new lager yeast strains from the WS34/70 population. *Roumanian Biotechnological Letters*, 13(6), 62–73.
- Borneman, A. R., Forgan, A. H., Kolouchova, R., Fraser, J. A., & Schmidt, S. A. (2016). Whole genome comparison reveals high levels of inbreeding and strain redundancy across the spectrum of commercial wine strains of *Saccharomyces cerevisiae*. *G3: Genes, Genomes, Genetics*, 6(4), 957–971. <https://doi.org/10.1534/g3.115.025692>
- Brzozowska, B., Galecki, M., Tartas, A., Ginter, J., Kaźmierczak, U., & Lundholm, L. (2019). Freeware tool for analysing numbers and sizes of cell colonies. *Radiation and Environmental Biophysics*, 58(1), 109–117. <https://doi.org/10.1007/s00411-018-00772-z>

- Cerulus, B., New, A. M., Pougach, K., & Verstrepen, K. J. (2016). Noise and epigenetic inheritance of single-cell division times influence population fitness. *Current Biology*, 26(9), 1138–1147. <https://doi.org/10.1016/j.cub.2016.03.010>
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Ciani, M., Capece, A., Comitini, F., Canonico, L., Siesto, G., & Romano, P. (2016). Yeast interactions in inoculated wine fermentation. *Frontiers in Microbiology*, 7, 555. <https://doi.org/10.3389/fmicb.2016.00555>
- Duan, S.-F., Han, P.-J., Wang, Q.-M., Liu, W.-Q., Shi, J.-Y., Li, K., Zhang, X. L., & Bai, F.-Y. (2018). The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. *Nature Communications*, 9(1), 2690. <https://doi.org/10.1038/s41467-018-05106-7>
- Duveau, F., Hodgins-Davis, A., Metzger, B. P., Yang, B., Tryban, S., Walker, E. A., Lybrook, T., & Wittkopp, P. J. (2018). Fitness effects of altering gene expression noise in *Saccharomyces cerevisiae*. *eLife*, 7, e37272. <https://doi.org/10.7554/eLife.37272>
- Engelberg, D., Mimran, A., Martinetto, H., Otto, J., Simchen, G., Karin, M., & Fink, G. R. (1998). Multicellular stalk-like structures in *Saccharomyces cerevisiae*. *Journal of Bacteriology*, 180(15), 3992–3996. <https://doi.org/10.1128/jb.180.15.3992-3996.1998>
- Eyler, E. (2013). Pouring agar plates and streaking or spreading to isolate individual colonies. *Methods in Enzymology*, 533, 3–14. <https://doi.org/10.1016/B978-0-12-420067-8.00001-5>
- Fay, J. C., Liu, P., Ong, G. T., Dunham, M. J., Cromie, G. A., Jeffery, E. W., Ludlow, C. L., & Dudley, A. M. (2019). A polyploid admixed origin of beer yeasts derived from European and Asian wine populations. *PLoS Biology*, 17(3), e3000147. <https://doi.org/10.1371/journal.pbio.3000147>
- Franzot, S. P., Mukherjee, J., Chorniak, R., Chen, L. C., Hamdan, J. S., & Casadevall, A. (1998). Microevolution of a standard strain of *Cryptococcus neoformans* resulting in differences in virulence and other phenotypes. *Infection and Immunity*, 66(1), 89–97. <https://doi.org/10.1128/IAI.66.1.89-97.1998>
- Galili, T., O'Callaghan, A., Sidi, J., & Sievert, C. (2017). heatmaply: An R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics*, 34(9), 1600–1602. <https://doi.org/10.1093/bioinformatics/btx657>
- Gallone, B., Steensels, J., Mertens, S., Dzialo, M. C., Gordon, J. L., Wauters, R., Theßeling, F. A., Bellinazzo, F., Saels, V., Herrera-Malaver, B., Prah, T., White, C., Hutzler, M., Meußdoerffer, F., Malcorps, P., Souffriau, B., Daenen, L., Baele, G., Maere, S., & Verstrepen, K. J. (2019). Interspecific hybridization facilitates niche adaptation in beer yeast. *Nature Ecology and Evolution*, 3(11), 1562–1575. <https://doi.org/10.1038/s41559-019-0997-9>
- Gallone, B., Steensels, J., Prah, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede, A., Roncoroni, M., Voordeckers, K., Miraglia, L., Teiling, C., Steffy, B., Taylor, M., Schwartz, A., Richardson, T., White, C., Baele, G., Maere, S., & Verstrepen, K. J. (2016). Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell*, 166(6), 1397–1410. <https://doi.org/10.1016/j.cell.2016.08.020>
- Gibson, B., Dahabieh, M., Krogerus, K., Jouhten, P., Magalhães, F., Pereira, R., Siewers, V., & Vidgren, V. (2020). Adaptive laboratory evolution of ale and lager yeasts for improved brewing efficiency and beer quality. *Annual Review of Food Science and Technology*, 11(1), 23–44. <https://doi.org/10.1146/annurev-food-032519-051715>
- Gibson, B., Lawrence, S. J., Leclaire, J. P. R., Powell, C. D., & Smart, K. A. (2007). Yeast responses to stresses associated with industrial brewery handling. *FEMS Microbiology Reviews*, 31, 535–569. <https://doi.org/10.1111/j.1574-6976.2007.00076.x>
- Gilchrist, C., & Stelkens, R. (2019, June 13). Aneuploidy in yeast: Segregation error or adaptation mechanism? *Yeast*, 36, 525–539. <https://doi.org/10.1002/yea.3427>
- Gonçalves, M., Pontes, A., Almeida, P., Barbosa, R., Serra, M., Libkind, D., Hutzler, M., Gonçalves, P., & Sampaio, J. P. (2016). Distinct domestication trajectories in top-fermenting beer yeasts and wine yeasts. *Current Biology*, 26(20), 2750–2761. <https://doi.org/10.1016/j.cub.2016.08.040>
- Gorter de Vries, A. R., Knibbe, E., van Roosmalen, R., van den Broek, M., de la Torre Cortés, P., O'Herne, S. F., Vijverberg, P. A., El Masoudi, A., Brouwers, N., Pronk, J. T., & Daran, J. M. G. (2020). Improving industrially relevant phenotypic traits by engineering chromosome copy number in *Saccharomyces pastorianus*. *Frontiers in Genetics*, 11, 518. <https://doi.org/10.3389/fgene.2020.00518>
- Gorter de Vries, A. R., Voskamp, M. A., Van Aalst, A. C. A., Kristensen, L. H., Jansen, L., Van Den Broek, M., Salazar, A. N., Brouwers, N., Abeel, T., Pronk, J. T., & Daran, J. M. G. (2019). Laboratory evolution of a *Saccharomyces cerevisiae* × *S. eubayanus* hybrid under simulated lager-brewing conditions. *Frontiers in Genetics*, 10, 242. <https://doi.org/10.3389/fgene.2019.00242>
- Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, 33(2), 1–22. <https://doi.org/10.18637/jss.v033.i02>
- Halfmann, R., Jarosz, D. F., Jones, S. K., Chang, A., Lancaster, A. K., & Lindquist, S. (2012). Prions are a common mechanism for phenotypic inheritance in wild yeasts. *Nature*, 482(7385), 363–368. <https://doi.org/10.1038/nature10875>
- Hanna, M., & Xiao, W. (2006). Isolation of nucleic acids. In W. Xiao (Ed.), *Yeast protocols* (pp. 15–20). Totowa, NJ: Humana Press. <https://doi.org/10.1385/1-59259-958-3:015>
- Hewitt, S. K., Foster, D. S., Dyer, P. S., & Avery, S. V. (2016). Phenotypic heterogeneity in fungi: Importance and methodology. *Fungal Biology Reviews*, 30(4), 176–184. <https://doi.org/10.1016/j.fbr.2016.09.002>
- Holland, S. L., Reader, T., Dyer, P. S., & Avery, S. V. (2014). Phenotypic heterogeneity is a selected trait in natural yeast populations subject to environmental stress. *Environmental Microbiology*, 16(6), 1729–1740. <https://doi.org/10.1111/1462-2920.12243>
- Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2), 254–267. <https://doi.org/10.1093/molbev/msj030>
- Imre, A., Rácz, H. V., Antunovics, Z., Rádai, Z., Kovács, R., Lopandic, K., Pócsi, I., & Pflieger, W. P. (2019). A new, rapid multiplex PCR method identifies frequent probiotic origin among clinical *Saccharomyces* isolates. *Microbiological Research*, 227, 126298. <https://doi.org/10.1016/j.micres.2019.126298>
- Kadowaki, M., Fujimaru, Y., Taguchi, S., Ferdouse, J., Sawada, K., Kimura, Y., Terasawa, Y., Agrimi, G., Anai, T., Noguchi, H., Toyoda, A., Fujiyama, A., Akao, T., & Kitagaki, H. (2017). Chromosomal aneuploidy improves the brewing characteristics of sake yeast. *Applied and Environmental Microbiology*, 83(24), e01620-17. <https://doi.org/10.1128/AEM.01620-17>
- Kim, S., Liachko, I., Brickner, D. G., Cook, K., Noble, W. S., Brickner, J. H., Shendure, J., & Dunham, M. J. (2017). The dynamic three-dimensional organization of the diploid yeast genome. *Elife*, 6, e23623. <https://doi.org/10.7554/eLife.23623>
- Lairón-Peris, M., Pérez-Través, L., Muñoz-Calvo, S., Guillamón, J. M., Heras, J. M., Barrio, E., & Querol, A. (2020). Differential contribution of the parental genomes to a *S. cerevisiae* × *S. uvarum* hybrid, inferred by phenomic, genomic, and transcriptomic analyses, at different industrial stress conditions. *Frontiers in Bioengineering and Biotechnology*, 8, 129. <https://doi.org/10.3389/fbioe.2020.00129>
- Lang, G. I., Botstein, D., & Desai, M. M. (2011). Genetic variation and the fate of beneficial mutations in asexual populations. *Genetics*, 188(3), 647–661. <https://doi.org/10.1534/genetics.111.128942>
- Langdon, Q. K., Peris, D., Baker, E. P., Opulente, D. A., Nguyen, H.-V., Bond, U., Gonçalves, P., Sampaio, J. P., Libkind, D., & Hittinger, C. T. (2019). Fermentation innovation through complex hybridization of wild and domesticated yeasts. *Nature Ecology & Evolution*, 3(11), 1576–1586. <https://doi.org/10.1038/s41559-019-0998-8>

- Large, C. R. L., Hanson, N. A., Tsouris, A., Saada, O. A., Koonthongkaew, J., Toyokawa, Y., Schmidlin, T., Moreno-Habel, D. A., McConnelllogue, H., Preiss, R., Takagi, H., Schacherer, J., & Dunham, M. J. (2020). Genomic stability and adaptation of beer brewing yeasts during serial repitching in the brewery. *bioRxiv*. Preprint. <https://doi.org/10.1101/2020.06.26.166157>
- Legras, J.-L., Merdinoglu, D., Cornuet, J.-M., & Karst, F. (2007). Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Molecular Ecology*, 16(10), 2091–2102. <https://doi.org/10.1111/j.1365-294X.2007.03266.x>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Magwene, P. M. (2014). Revisiting Mortimer's genome renewal hypothesis: Heterozygosity, homothallism, and the potential for adaptation in yeast. *Advances in Experimental Medicine and Biology*, 781, 37–48. https://doi.org/10.1007/978-94-007-7347-9_3
- Mangado, A., Morales, P., Gonzalez, R., & Tronchoni, J. (2018). Evolution of a yeast with industrial background under winemaking conditions leads to diploidization and chromosomal copy number variation. *Frontiers in Microbiology*, 9, 1816. <https://doi.org/10.3389/fmicb.2018.01816>
- McDonald, M. J., Rice, D. P., & Desai, M. M. (2016). Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature*, 531(7593), 233–236. <https://doi.org/10.1038/nature17143>
- Morard, M., Ibáñez, C., Adam, A. C., Querol, A., Barrio, E., & Toft, C. (2020). Genomic instability in an interspecific hybrid of the genus *Saccharomyces*: A matter of adaptability. *Microbial Genomics*, 6(10), mgen000448. <https://doi.org/10.1099/mgen.0.000448>
- Morard, M., Macías, L. G., Adam, A. C., Lairón-Peris, M., Pérez-Torrado, R., Toft, C., & Barrio, E. (2019). Aneuploidy and ethanol tolerance in *Saccharomyces cerevisiae*. *Frontiers in Genetics*, 10, 82. <https://doi.org/10.3389/fgene.2019.00082>
- Mortimer, R. K. (2000). Evolution and variation of the yeast (*Saccharomyces*) genome. *Genome Research*, 10(4), 403–409. <https://doi.org/10.1101/gr.10.4.403>
- Nadal-Ribelles, M., Islam, S., Wei, W., Latorre, P., Nguyen, M., de Nadal, E., Posas, F., & Steinmetz, L. M. (2019). Sensitive high-throughput single-cell RNA-seq reveals within-clonal transcript correlations in yeast populations. *Nature Microbiology*, 4(4), 683–692. <https://doi.org/10.1038/s41564-018-0346-9>
- Naseeb, S., Alsammar, H., Burgis, T., Donaldson, I., Knyazev, N., Knight, C., & Delneri, D. (2018). Whole genome sequencing, de novo assembly and phenotypic profiling for the new budding yeast species *Saccharomyces jurei*. *G3 (Bethesda)*, 8(9), 2967–2977. <https://doi.org/10.1534/g3.118.200476>
- Offei, B., Vandecruys, P., De Graeve, S., Foulquié-Moreno, M. R., & Thevelein, J. M. (2019). Unique genetic basis of the distinct antibiotic potency of high acetic acid production in the probiotic yeast *Saccharomyces cerevisiae* var. *boulardii*. *Genome Research*, 29(9), 1478–1494. <https://doi.org/10.1101/gr.243147.118>
- Okuno, M., Kajitani, R., Ryusui, R., Morimoto, H., Kodama, Y., & Itoh, T. (2016). Next-generation sequencing analysis of lager brewing yeast strains reveals the evolutionary history of interspecies hybridization. *DNA Research*, 23(1), dsV037. <https://doi.org/10.1093/dnares/dsv037>
- Ortiz, E.M. (2019). vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. <https://doi.org/10.5281/zenodo.2540861>
- Payen, C., Sunshine, A. B., Ong, G. T., Pogachar, J. L., Zhao, W., & Dunham, M. J. (2016). High-throughput identification of adaptive mutations in experimentally evolved yeast populations. *PLoS Genetics*, 12(10), e1006339. <https://doi.org/10.1371/journal.pgen.1006339>
- Pérez-Torrado, R., & Querol, A. (2015). Opportunistic strains of *Saccharomyces cerevisiae*: A potential risk sold in food products. *Frontiers in Microbiology*, 6, 1522. <https://doi.org/10.3389/fmicb.2015.01522>
- Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergstrom, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J. M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., ... Schacherer, J. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*, 556(7701), 339–344. <https://doi.org/10.1038/s41586-018-0030-5>
- Pflieger, W. P., Boros, E., Pázmándi, K., Jakab, Á., Zsuga, I., Kovács, R., Urbán, E., Antunovics, Z., Bácsi, A., Sipiczki, M., Majoros, L., & Pócsi, I. (2017). Commercial strain-derived clinical *Saccharomyces cerevisiae* can evolve new phenotypes without higher pathogenicity. *Molecular Nutrition and Food Research*, 61(11), 1601099. <https://doi.org/10.1002/mnfr.201601099>
- Pontes, A., Hutzler, M., Brito, P. H., & Sampaio, J. P. (2020). Revisiting the taxonomic synonyms and populations of *Saccharomyces cerevisiae*—Phylogeny, phenotypes, ecology and domestication. *Microorganisms*, 8(6), 903. <https://doi.org/10.3390/microorganisms8060903>
- Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur, D. G., & Banks, E. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*. Preprint. <https://doi.org/10.1101/201178>
- Qiu, X., Zhang, J., Zhou, J., Fang, Z., Zhu, Z., Li, J., & Du, G. (2019). Stress tolerance phenotype of industrial yeast: Industrial cases, cellular changes, and improvement strategies. *Applied Microbiology and Biotechnology*, 103(16), 6449–6462. <https://doi.org/10.1007/s00253-019-09993-8>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- R Core Team. (2020). R: A language and environment for statistical computing. <https://www.r-project.org/>
- Raghavan, V., Aquadro, C. F., & Alani, E. (2019, September). Baker's yeast clinical isolates provide a model for how pathogenic yeasts adapt to stress. *Trends in Genetics*, 35, 804–817. <https://doi.org/10.1016/j.tig.2019.08.002>
- Reis, V. R., Bassi, A. P. G., da Silva, J. C. G., & Ceccato-Antonini, S. R. (2014). Characteristics of *Saccharomyces cerevisiae* yeasts exhibiting rough colonies and pseudohyphal morphology with respect to alcoholic fermentation. *Brazilian Journal of Microbiology*, 44(4), 1121–1131. <https://doi.org/10.1590/S1517-83822014005000020>
- Rodrigues-Prause, A., Sampaio, N. M. V., Gurol, T. M., Aguirre, G. M., Sedam, H. N. C., Chapman, M. J., Malc, E. P., Ajith, V. P., Chakraborty, P., Tizei, P. A., Pereira, G. A. G., Mieczkowski, P. A., Nishant, K. T., & Argueso, J. L. (2018). A case study of genomic instability in an industrial strain of *Saccharomyces cerevisiae*. *G3: Genes, Genomes, Genetics*, 8(11), 3703–3713. <https://doi.org/10.1534/g3.118.200446>
- Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., & Eliceiri, K. W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*, 18(1), 529. <https://doi.org/10.1186/s12859-017-1934-z>
- Salazar, A. N., Gorter de Vries, A. R., van den Broek, M., Brouwers, N., de la Torre Cortés, P., Kuijpers, N. G., Daran, J.M.G., Abeel, T. (2019). Nanopore sequencing and comparative genome analysis confirm lager-brewing yeasts originated from a single hybridization. Unpublished, 603480. <https://doi.org/10.1101/603480>
- Sampaio, N. M. V., Watson, R. A., & Argueso, J. L. (2019). Controlled reduction of genomic heterozygosity in an industrial yeast strain reveals wide cryptic phenotypic variation. *Frontiers in Genetics*, 10, 782. <https://doi.org/10.3389/fgene.2019.00782>

- Scannell, D. R., Zill, O. A., Rokas, A., Payen, C., Dunham, M. J., Eisen, M. B., Rine, J., Johnston, M., & Hittinger, C. T. (2011). The awesome power of yeast evolutionary genetics: New genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. *G3: Genes, Genomes, Genetics*, 1(1), 11–25. <https://doi.org/10.1534/g3.111.000273>
- Scherz, R., Shinder, V., & Engelberg, D. (2001). Anatomical analysis of *Saccharomyces cerevisiae* stalk-like structures reveals spatial organization and cell specialization. *Journal of Bacteriology*, 183(18), 5402–5413. <https://doi.org/10.1128/JB.183.18.5402-5413.2001>
- Steensels, J., Gallone, B., Voordeckers, K., & Verstrepen, K. J. (2019). Domestication of industrial microbes. *Current Biology*, 29, R381–R393. <https://doi.org/10.1016/j.cub.2019.04.025>
- Stone, N. R. H., Rhodes, J., Fisher, M. C., Mfinanga, S., Kivuyo, S., Rugemalila, J., Segal, E. S., Needleman, L., Molloy, S. F., Kwon-Chung, J., Harrison, T. S., Hope, W., Berman, J., & Bicanic, T. (2019). Dynamic ploidy changes drive fluconazole resistance in human cryptococcal meningitis. *Journal of Clinical Investigation*, 129, 999–1014. <https://doi.org/10.1172/JCI124516>
- Strope, P. K., Skelly, D. A., Kozmin, S. G., Mahadevan, G., Stone, E. A., Magwene, P. M., Dietrich, F. S., & McCusker, J. H. (2015). The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Research*, 25(5), 762–774. <https://doi.org/10.1101/gr.185538.114>
- Sumner, E. R., & Avery, S. V. (2002). Phenotypic heterogeneity: Differential stress resistance among individual cells of the yeast *Saccharomyces cerevisiae*. *Microbiology*, 148(2), 345–351. <https://doi.org/10.1099/00221287-148-2-345>
- Tan, Z., Hays, M., Cromie, G. A., Jeffery, E. W., Scott, A. C., Ahyong, V., Sirr, A., Skupin, A., & Dudley, A. M. (2013). Aneuploidy underlies a multicellular phenotypic switch. *Proceedings of the National Academy of Sciences*, 110(30), 12367–12372. <https://doi.org/10.1073/pnas.1301047110>
- Tattini, L., Tellini, N., Mozzachiodi, S., D'Angiolo, M., Loeillet, S., Nicolas, A., & Liti, G. (2019). Accurate tracking of the mutational landscape of diploid hybrid genomes. *Molecular Biology and Evolution*, 36(12), 2861–2877. <https://doi.org/10.1093/molbev/msz177>
- van den Broek, M., Bolat, I., Nijkamp, J. F., Ramos, E., Luttkik, M. A. H., Koopman, F., Geertman, J. M., de Ridder, D., Pronk, J. T., & Daran, J. M. (2015). Chromosomal copy number variation in *Saccharomyces pastorianus* is evidence for extensive genome dynamics in industrial lager brewing strains. *Applied and Environmental Microbiology*, 81(18), 6253–6267. <https://doi.org/10.1128/AEM.01263-15>
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 43(1), 11.10.1–11.10.33. <https://doi.org/10.1002/0471250953.bi1110s43>
- Vázquez-García, I., Salinas, F., Li, J., Fischer, A., Barré, B., Hallin, J., Bergström, A., Alonso-Perez, E., Warringer, J., Mustonen, V., & Liti, G. (2017). Clonal heterogeneity influences the fate of new adaptive mutations. *Cell Reports*, 21(3), 732–744. <https://doi.org/10.1016/j.celrep.2017.09.046>
- Voordeckers, K., Kominek, J., Das, A., Espinosa-Cantú, A., De Maeyer, D., Arslan, A., Van Pee, M., van der Zande, E., Meert, W., Yang, Y., Zhu, B., Marchal, K., DeLuna, A., Van Noort, V., Jelier, R., & Verstrepen, K. J. (2015). Adaptation to high ethanol reveals complex evolutionary pathways. *PLoS Genetics*, 11(11), e1005635. <https://doi.org/10.1371/journal.pgen.1005635>
- Wheals, A. E., & Lord, P. G. (1992). Clonal heterogeneity in specific growth rate of *Saccharomyces cerevisiae* cells. *Cell Proliferation*, 25(3), 217–223. <https://doi.org/10.1111/j.1365-2184.1992.tb01396.x>
- Wing, K. M., Phillips, M. A., Baker, A. R., & Burke, M. K. (2020). Consequences of cryopreservation in diverse natural isolates of *Saccharomyces cerevisiae*. *Genome Biology and Evolution*, 12, 1302–1312. <https://doi.org/10.1093/gbe/evaa121>
- Yue, J.-X., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., Bergström, A., Coupland, P., Warringer, J., Lagomarsino, M. C., Fischer, G., Durbin, R., & Liti, G. (2017). Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nature Genetics*, 49(6), 913–924. <https://doi.org/10.1038/ng.3847>
- Yvert, G., Ohnuki, S., Nogami, S., Imanaga, Y., Fehrmann, S., Schacherer, J., & Ohya, Y. (2013). Single-cell phenomics reveals intra-species variation of phenotypic noise in yeast. *BMC Systems Biology*, 7(1), 54. <https://doi.org/10.1186/1752-0509-7-54>
- 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 Research*, 36(2), 1–12. <https://doi.org/10.1093/femsyr/fov118>
- Zheng, X., Levine, D., Shen, J., Gogarten, S., Laurie, C., & Weir, B. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28(24), 3326–3328. <https://doi.org/10.1093/bioinformatics/bts606>
- Zhu, Y. O., Sherlock, G., & Petrov, D. A. (2016). Whole genome analysis of 132 clinical *Saccharomyces cerevisiae* strains reveals extensive ploidy variation. *G3: Genes, Genomes, Genetics*, 6(8), 2421–2434. <https://doi.org/10.1534/g3.116.029397>

SUPPORTING INFORMATION

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

How to cite this article: RÁCZ HV, Mukhtar F, Imre A, et al.

How to characterize a strain? Clonal heterogeneity in industrial *Saccharomyces* influences both phenotypes and heterogeneity in phenotypes. *Yeast*. 2021;38:453–470.

<https://doi.org/10.1002/yea.3562>