*Molecular Biology and Evolution*, 2025, **42**, 1–16
https://doi.org/10.1093/molbev/msaf285
Advance access publication 3 November 2025
**Discoveries**

MBE OXFORD

# Stable Hypermutators Revealed by the Genomic Landscape of Genes Involved in Genome Stability Among Yeast Species

**Carla Gonçalves** [1,2,3,4,†], **Jacob L. Steenwyk** [1,2,5,†], **David C. Rinker** [1,2], **Dana A. Opulente** [6,7], **Abigail L. LaBella** [1,2,8], **Marie-Claire Harrison** [1,2], **John F. Wolters** [6], **Xiaofan Zhou** [1,2,9], **Xing-Xing Shen** [1,2,10], **Shay Covo** [11], **Marizeth Groenewald** [12], **Chris Todd Hittinger** [6], **Antonis Rokas** [1,2,*]

[1]Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA
[2]Evolutionary Studies Initiative, Vanderbilt University, Nashville, TN 37235, USA
[3]Associate Laboratory i4HB—Institute for Health and Bioeconomy and UCIBIO—Applied Molecular Biosciences Unit, Department of Life Sciences, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica 2829-516, Portugal
[4]UCIBIO-i4HB, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica 2829-516, Portugal
[5]Howard Hughes Medical Institute and the Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA
[6]Laboratory of Genetics, DOE Great Lakes Bioenergy Research Center, Center for Genomic Science Innovation, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, University of Wisconsin-Madison, Madison, WI 53726, USA
[7]Biology Department, Villanova University, Villanova, PA 19085, USA
[8]Department of Bioinformatics and Genomics, North Carolina Research Center, University of North Carolina at Charlotte, Kannapolis, NC 28223, USA
[9]Guangdong Province Key Laboratory of Microbial Signals and Disease Control, Integrative Microbiology Research Center, South China Agricultural University, Guangzhou 510642, China
[10]College of Agriculture and Biotechnology and Centre for Evolutionary & Organismal Biology, Zhejiang University, Hangzhou 310058, China
[11]Department of Plant Pathology and Microbiology, Hebrew University of Jerusalem, Rehovot 7610001, Israel
[12]Westerdijk Fungal Biodiversity Institute, Utrecht 3584, The Netherlands

[†]Equally contributing authors.
***Corresponding author:** E-mail: antonis.rokas@vanderbilt.edu.
**Associate editor**: Jeffrey Townsend

## Abstract

Mutator phenotypes are short-lived due to the rapid accumulation of deleterious mutations. Yet, recent observations reveal that certain fungi can undergo prolonged accelerated evolution after losing genes involved in DNA repair. Here, we surveyed 1,154 yeast genomes representing nearly all known yeast species of the subphylum Saccharomycotina (phylum Ascomycota) to examine the relationship between reduced gene repertoires broadly associated with genome stability functions (eg DNA repair, cell cycle) and elevated evolutionary rates. We identified 3 distantly related lineages—encompassing 12% of species—that had both the most streamlined sets of genes involved in genome stability (specifically DNA repair) and the highest evolutionary rates in the entire subphylum. Two of these "faster-evolving lineages" (FELs)—a subclade within the order Pichiales and the *Wickerhamiella/Starmerella* (W/S) clade (order Dipodascales)—are described here for the first time, while the third corresponds to a previously documented *Hanseniaspora* FEL. Examination of genome stability gene repertoires revealed a set of genes predominantly absent in these 3 FELs, suggesting a potential role in the observed acceleration of evolutionary rates. In the W/S clade, genomic signatures are consistent with a substantial mutational burden, including pronounced A|T bias and endogenous DNA damage. Interestingly, we found that the W/S clade also contains DNA repair genes possibly acquired through horizontal gene transfer, including a photolyase of bacterial origin. These findings highlight how hypermutators can persist across macroevolutionary timescales, potentially linked to the loss of genes related to genome stability, with horizontal gene transfer as a possible avenue for partial functional compensation.

**Keywords:** gene loss, DNA repair, horizontal gene transfer, rapid evolution, yeast genome evolution, yeast pathogens, macroevolution

## Introduction

Mutations are the raw material of evolution and adaptation. Since mutations are more likely to be deleterious than adaptive (Eyre-Walker and Keightley 2007), cells have evolved numerous interacting mechanisms to ensure high fidelity of genome replication and stability—such as cell cycle checkpoints and DNA damage sensing and repair pathways (Giglia-Mari, et al. 2011; Kreuzer 2013; Steenwyk 2021). Hypermutation can arise from the aberrant function of genes involved in several pathways and functions, such as DNA damage/S-phase

**Received:** March 16, 2025. **Revised:** September 24, 2025. **Accepted:** October 22, 2025

checkpoints, cell cycle, DNA replication, or oxidative stress (Murakami-Sekimata, et al. 2010; Serero, et al. 2014). Impairment specifically in DNA repair pathways has been associated with hypermutation across the tree of life (Bridges 2001; Roberts and Gordenin 2014; Campbell, et al. 2017; Steenwyk, et al. 2019; Phillips, et al. 2021; Gambhir, et al. 2022). For example, defects in the mismatch repair (MMR) system, a highly conserved pathway that corrects mismatches (Fukui 2010), have been shown to lead to diverse mutator phenotypes (Oliver, et al. 2000; Chopra, et al. 2003; Dos Reis, et al. 2019; Gambhir, et al. 2022). In bacteria, a deficiency in the MMR can also result in the relaxation of recombination barriers between species, resulting in higher horizontal gene transfer (HGT) rates between distantly related species (Rayssiguier, et al. 1989; Thomas and Nielsen 2005).

Hypermutators have been mostly observed in fluctuating environments (Swings, et al. 2017; Callens, et al. 2023), such as in the context of human disease (Oliver, et al. 2000; Chopra, et al. 2003; Healey, et al. 2016; Billmyre, et al. 2017; Rhodes, et al. 2017; Gambhir, et al. 2022). In clinical isolates of bacterial and fungal pathogens (Healey, et al. 2016; Billmyre, et al. 2017; Rhodes, et al. 2017; Gambhir, et al. 2022), defects in DNA repair pathways have been associated with increased mutation rates and are thought to facilitate host adaptation. For example, mutations in the MMR gene *MSH2* in *Cryptococcus* (Billmyre, et al. 2017; Rhodes, et al. 2017) and *Nakaseomyces glabratus* (Healey, et al. 2016) opportunistic fungal pathogens are known to give rise to mutator phenotypes, possibly promoting rapid host adaptation and drug resistance. Hypermutator phenotypes have also been implicated in the acceleration of disease progression and the evolution of resistance to treatments in humans (Campbell, et al. 2017; Jiang, et al. 2020).

Fluctuating environments can, therefore, serve as triggers for the emergence and maintenance of hypermutator phenotypes because they enable a broader exploration of genotype-phenotype space and access to a larger pool of potentially advantageous mutations (Shaver, et al. 2002). However, once these beneficial mutations are fixed in the population, and organisms become well adapted to their new environment, compensatory mutations that reduce the mutation rate will be favored (Kimura 2009; Wielgoss, et al. 2013). Genetic constraints such as gene loss may, however, constrain the re-lowering of mutation rates.

In fungi, the DNA repair gene repertoire has been described to be highly variable (Milo, et al. 2019; Shen, et al. 2020), and in a few ancient lineages, loss of DNA repair-related genes has been associated with long-term increased evolutionary rates (Steenwyk, et al. 2019; Phillips, et al. 2021). These lineages include a case of macroevolutionary hypermutation among *Hanseniaspora* yeasts, in which the cell-cycle gene repertoire is also substantially reduced (Steenwyk, et al. 2019). These examples suggest that hypermutator lineages can survive and diversify over macroevolutionary timescales. However, the prevalence of such long-term hypermutator lineages, the extent of their association with the loss of DNA repair genes or other genome stability-related genes, and the mechanisms involved in their long-term survival remain poorly understood.

Here, we used a sequence similarity search approach, partially validated by structural homology and phylogenetic approaches to explore the relationship between repertoires of genes broadly involved in genome stability, including cell cycle and DNA repair-related functions, and evolutionary rates across the entire subphylum Saccharomycotina, an ancient yeast lineage with more than 1,000 known species (Groenewald, et al. 2023; Opulente, et al. 2024). We found that the 3 lineages with the highest rates of sequence evolution (faster-evolving lineages or FELs) in the subphylum have experienced substantial reductions in the number of genes related to genome stability, including substantial reductions of their DNA repair gene repertoires. These lineages encompass a subclade within the order Pichiales (Pichiales subclade), the *Wickerhamiella/Starmerella* (W/S) clade (order Dipodascales), and a previously reported *Hanseniaspora* lineage (order Saccharomycodales) (Steenwyk, et al. 2019). Several genes have been seemingly independently lost in the 3 FELs, suggesting a possible role in accelerating mutation rates. In the W/S clade, we found strong signatures of mutational burden compared to its closest relatives—namely, a pronounced A|T bias and increased frequency of mutations associated with endogenous DNA damage, but no evidence of UV damage. Interestingly, most W/S-clade species harbour either a filamentous fungal-like version or a bacterial version of the gene *PHR1* encoding the photolyase that repairs UV-induced DNA damage (Sebastian, et al. 1990)- likely horizontally acquired and seemingly functional- while most W/S clade closest relatives lack this gene. These results suggest that variation of evolutionary rates through the loss of DNA repair and other genes involved in genome stability may have been common in Saccharomycotina yeast evolution and that other evolutionary mechanisms, such as HGT, might help circumvent constraints imposed by loss.

## Materials and Methods

### Genome Assembly and Annotation

Genomic data were obtained from Opulente et al. (2024). Briefly, reads obtained from paired-end NEBNext Ultra II DNA Illumina runs were used to produce whole genome assemblies using iWGS v101 pipeline. Variation of ploidy and heterozygosity levels was considered. Mitochondrial DNA and putative contamination were subsequently filtered.

For genome annotation, repetitive sequences were identified in each of the genome assemblies and softmasked using RepeatMasker v4.1.2 (Tarailo-Graovac and Chen 2009). Protein-coding genes were annotated using the BRAKER v2.1.6 pipeline (Brůna, et al. 2021) using all Saccharomycetes protein sequences in the OrthoDB v10 as homology evidence and the ab initio gene predictors AUGUSTUS v3.4.0 (Stanke, et al. 2006) and GeneMarkEP + v4.6.1 (Brůna, et al. 2020). The BRAKER pipeline was run in the EP mode to process all protein homology evidence using ProtHint v2.6.0 (Hunter, et al. 2009; Blum, et al. 2021), and the "–fungus" option was turned on to run GeneMark-EP + with the branch point model for fungal genomes. For genes with multiple transcripts, only the longest transcript was retained. The standard genetic code was used to translate the protein annotations.

### Gene Presence and Absence Analysis and Determination of Evolutionary Rates

To assess the presence and absence of genome stability-related genes across the Saccharomycotina subphylum, we first retrieved from UniprotKB all proteins belonging to Saccharomycotina yeasts potentially involved in DNA repair (keywords: dna + repair + saccharomycotina-filtered-reviewed) as of February 2023. Although we searched for all Saccharomycotina DNA repair-related proteins at UniprotKB, most of the retrieved genes

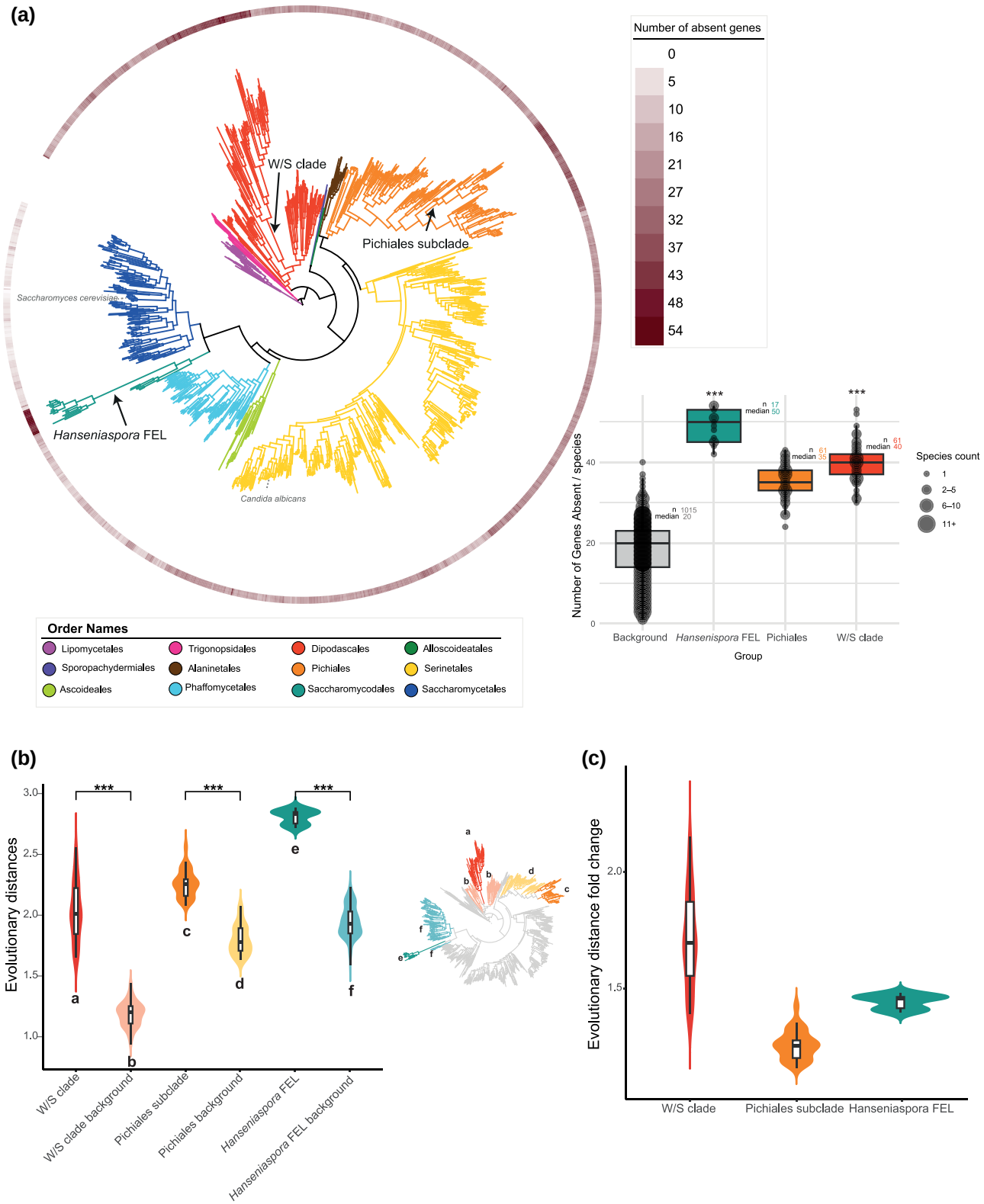


**Fig. 1.** The genome stability gene repertoire is highly variable across the subphylum Saccharomycotina and significantly reduced in three faster-evolving lineages. (a) Distribution of the number of absent genome stability genes (from a total of 393) across the Saccharomycotina species phylogeny, depicting three distantly related lineages that lack the largest number of genome stability genes: W/S clade, Pichiales subclade, and *Hanseniaspora* FEL. The panel on the right shows the number of absent genes across these 3 lineages (foreground species) compared to the number of genes absent in the background species (all others). (b) Evolutionary distances were determined using tip-to-root distances of the Saccharomycotina species phylogeny (Opulente, et al. 2024) as proxy. (c) Fold change difference between tip-to-root distances in the foreground species of each foreground lineage with respect to the average tip-to-root distance in the closest relatives (background). Statistical significance (panels a and (b) was assessed using a pairwise Wilcoxon rank test after testing for normality. *P*-values were adjusted using the Holm correction (*** *P*-value < 0.001). In panel B, statistical significance was assessed between each of the faster-evolving lineages and the background. For reference, the phylogenetic position of *Saccharomyces cerevisiae* and *Candida albicans* are shown in the tree.

belong to *Saccharomyces cerevisiae* (Table S1). These 415 proteins are associated with distinct functions, such as DNA damage response, cell cycle, chromatin, and telomere organization (Table S1); for this reason, we refer to this dataset as the genome stability gene repertoire. Among these proteins, we identified several ohnologs. To remove redundancy, we kept only one ohnolog from each pair, resulting in a total of 393 proteins. For each protein, we built HMMs profiles (Eddy 1998) using an alignment with the top 100 hits retrieved by BLASTp from the NCBI nonredundant (nr) database. Next, we ran Orthofisher (Steenwyk and Rokas 2021) with default parameters (*e*-value cutoff $< e^{-3}$) and with higher levels of stringency (*e*-value cutoff of $1e^{-20}$ and $1e^{-50}$, Figure S1) using the previously constructed HMM profiles against 1,154 proteomes corresponding to at least 1,051 Saccharomycotina species (Opulente, et al. 2024). We also excluded paralogs from all subsequent analyses (Table S1). To assess the distribution pattern of genes specifically involved in DNA repair functions, we focused on those annotated with the Gene Ontology (GO) term GO:0006281—DNA repair (Figure S2).

Absences were confirmed for the common absent genes in the 3 FELs (Fig. 2) and for the list of genes absent in the W/S clade (Table S3) by tBLASTn searches against the respective assemblies using the list of 393 DNA repair-related proteins. Genes were considered absent whenever the *e*-value for the best hit was >0.1. For the hits with *e*-values < 0.1, the identity of the gene was confirmed by BLASTp against the NCBI nr database using *Saccharomyces cerevisiae* as reference.

Evolutionary rates were determined using tip-to-root distances as a proxy. For that, we used the recently published Saccharomycotina phylogeny (Opulente, et al. 2024) and determined the tip-to-root distances using the *distRoot* function (method = patristic) included in the *adephylo* package for R (Jombart, et al. 2010).

## Large Language Model (LLM) Remote Homolog Detection

To search for more diverged homologs without relying on sequence similarity, we used pLM-BLAST, a large language model (LLM) based approach optimized for remote homolog detection (Kaminski, et al. 2023). To prepare to run pLM-BLAST, we first used the pLM-BLAST tool embeddings.py to make protein sequence embeddings of each of the genes, in all the proteomes of the W/S clade (307,448 sequences in total). We then used embeddings.py to make embeddings for representative proteins in the sister clade (*Blastobotrys* group) (45,011 sequences, clustered by mmseq2 (Steinegger and Söding 2017) to represent the 450,863 proteins in the clade). Embeddings were then generated for each query sequence (*AHC1* from *Candida incommunis*, *HNT3* from *Sugiyamaella lignohabitans*, *ISY1* from *Candida incommunis*, *SLX1* from *Sugiyamaella lignohabitans*, and *SLX4* from *Sugiyamaella lignohabitans*). Next, each query embedding was searched using the default pLM-BLAST parameters (alignment cutoff =0.3, cosine cutoff = 90, sigma = 2) against all proteins in the W/S and sister clade, using on 4× NVIDIA RTX A6000 GPUs. Finally, pLM-BLAST hits were sorted by their cosine similarity score, and any queries having scores above 0.5 were considered as possible homologs (Table S5). For hits with scores higher than 0.5, a subsequent reciprocal BLASTp against NCBI nr database (using *Saccharomyces cerevisiae* as reference) was performed. These results were mostly negative, with most hits being fragmentary and likely corresponding to individual domains rather than whole genes. Overall, these remote homolog detection results recapitulate the results of the sequence similarity searches.

## Mutational Signatures

To identify base substitutions and indels (insertions and deletions) in the W/S clade, we followed previously published methodologies (Steenwyk, et al. 2019). First, we select the orthologues to be analyzed by running OrthoFinder v.2.3.8 (Emms and Kelly 2019) using an inflation parameter of 1.5 and DIAMOND v2.0.13.151 (Buchfink, et al. 2015) for protein alignments, on a dataset containing all W/S-clade proteomes (excluding redundancy, ie strains from the same species) as well as proteomes from closest relatives (W/S sister, Outgroups 1 and 2) and outgroups (Table S6). A total of 143 core single-copy orthogroups were obtained and subsequently aligned with MAFFT v7.402 using an iterative refinement method (–localpair) (Katoh and Standley 2014; Vialle, et al. 2018). Next, codon-aware alignments for each amino acid alignment were generated with PAL2NAL v.14 (Suyama, et al. 2006).

Substitution patterns (a proxy for mutational patterns) were examined from the resulting multiple sequence alignments. To do so, substitutions, insertions, and deletions were examined at sites otherwise conserved in the outgroup of closely related taxa (eg, the W/S sister clade is the outgroup to the W/S clade). For substitutions, the nucleotide character for a focal species was compared to the conserved nucleotide in the outgroup taxa at the same position. If the focal species had a nucleotide character that differed from the outgroup taxa, a substitution was determined to have occurred. While doing so, we kept track of the nucleotide character for the focal species and the outgroup taxa as well as the codon position, enabling inference of the directionality of the substitution and positional information. To ensure the number of mutations was comparable for each group, the raw number of substitutions was corrected by the number of conserved sites in the outgroup taxa. The same correction was made for substitutions at the various codon positions. For the AT-bias analysis, a correction was made to account for variation in the conserved number of GC or AT sites. Lastly, a correction was made to normalize for the number of single-copy orthologous genes that could be examined per focal lineage and closely related lineage. To identify insertions and deletions, a sliding window approach with a step size of one nucleotide was used to scan the multiple sequence alignments for positions that had nucleotides in the focal species and gaps in the outgroup taxa (insertions) or vice versa (deletions).

Statistical significance of the differences between the several groups (W/S clade, W/S sister, Outgroup 1 and Outgroup 2) was assessed by pairwise Wilcoxon rank test after testing for normality.

## Phylogenetic Analyses of *PHR1*

Putative Phr1 sequences from W/S-clade species and other Saccharomycotina were retrieved from the Orthofisher run and used in BLASTp searches against the NCBI nr database. Two distinct major lineages were identified as top hits for different W/S-clade species: Pezizomycotina (filamentous fungi) and Bacteria. Putative Phr1 sequences from *Wickerhamiella australiensis* (top hits Bacteria) and *Starmerella bombicola* (top hits Pezizomycotina) were used in a BLASTp search

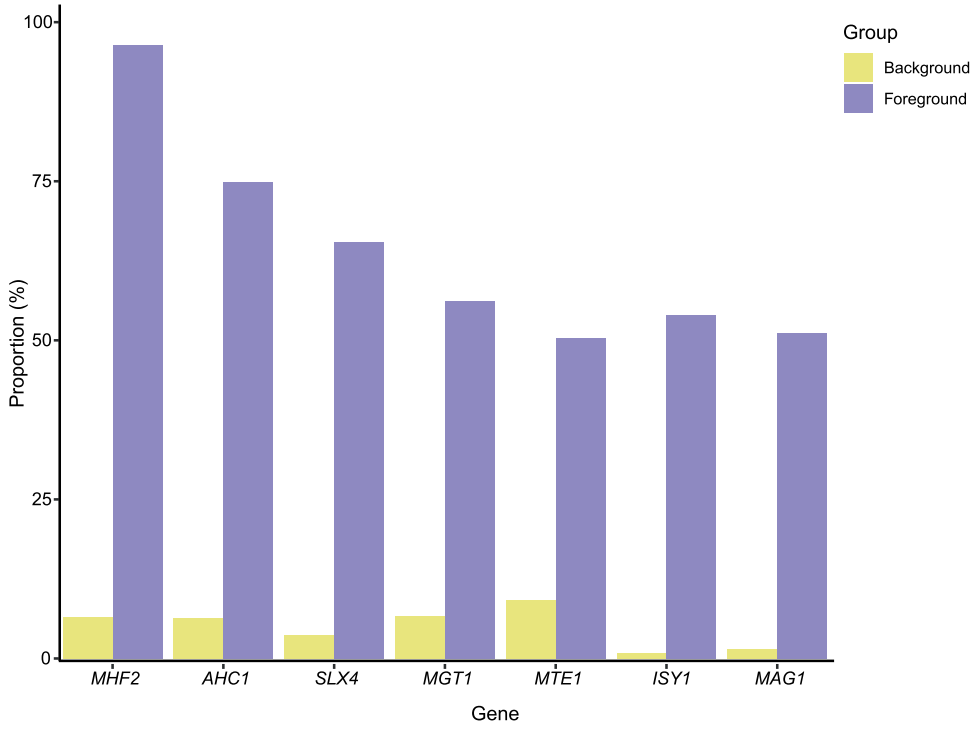


**Fig. 2.** Faster-evolving lineages show overlapping absences of 7 genome stability-related genes. Genes absent in more than 50% of the total foreground strains (W/S clade + *Hanseniaspora* FEL + Pichiales subclade, $N = 139$; shown in purple bars) and in less than 10% in the background species ($N = 1{,}015$; shown in yellow bars) are represented. The Y axis represents the proportion of species in which each gene is absent is represented in foreground and background groups. This proportion was calculated based on the number of proteomes in which each gene was absent out of the total number of proteomes inspected.

against the NCBI RefSeq database (O'Leary, et al. 2016). The top 250 hits from each blast search were retrieved. Redundancy was removed with CD-HIT (keeping sequences with less than 98% identity) (Li and Godzik 2006). The resulting 2,007 sequences and subsequently aligned using MAFFT v7.402 using an iterative refinement method (–localpair) and poorly aligned regions were removed with trimAL (Capella-Gutierrez, et al. 2009) using the "gappyout" option. A Phr1 phylogeny was subsequently constructed with IQ-TREE v2.0.6 (Nguyen, et al. 2015) using an automated method for model selection and 1,000 ultrafast bootstraps (Hoang, et al. 2018). The tree with the highest likelihood score was subsequently chosen from a total of 5 runs (–runs 5). The hypothesis of the horizontal acquisition of Phr1 from both bacteria and filamentous fungi was assessed through topology tests performed in IQ-TREE v2.0.6 (Nguyen, et al. 2015). For that, two constrained topologies were constructed. In the first constrained topology to test HGT from bacteria, W/S-clade sequences clustering with bacteria were considered to be monophyletic with other Saccharomycotina (except for Saccharomycotina sequences that clustered with other fungi). To test for HGT from filamentous fungi, a second constrained topology was constructed by considering "fungal" W/S-clade sequences as monophyletic with other Saccharomycotina (except for Saccharomycotina sequences that clustered with other fungi). Constrained trees were constructed in IQ-TREE v2.0.6 (using the "-g" option). The likelihoods of the constrained and unconstrained tree topologies were subsequently compared in IQ-TREE using the AU test (using "-z" option).

## UV Sensitivity Assays

Single colonies of Phr1-positive and Phr1-negative W/S-clade species (Figure S8) were grown in YPD and incubated for 48 h at 25 °C in total darkness to avoid photoreactivation. A biomass corresponding to an $OD_{600nm}$ ~0.5 was resuspended in 200 µL sterile water. The cell suspensions were serially diluted and spotted onto YPD [1% (w/v) yeast extract, 2% (w/v) bacto peptone, and 2% (w/v) glucose] agar plates. The plates were irradiated with 200 J/m$^2$ using UV-C (254 nm) with or without subsequent 2 h of photoreactivation with UV-A (365 nm) using a Vilber Lourmat VL-6.LC lamp (230 V, 50/60 Hz) as previously described (Milo, et al. 2019; Milo, et al. 2024). All manipulations were carried out in the dark under a safe red light to avoid uncontrolled photoreactivation. The plates were incubated in total darkness for 48 h at 25 °C.

## Search for Additional Bacteria-derived DNA Repair Genes

To investigate whether other DNA repair genes from bacteria were horizontally acquired by W/S-clade species, DNA repair genes from *Escherichia coli* K12 reference strain and *Bacillus subtilis* reference strain were retrieved from UniprotKB (as of February 2023). A local BLASTp search against all W/S-clade proteomes was performed (*e*-value cutoff $1e^{-3}$). The top blast hit for each gene was subsequently retrieved and analyzed through a BLASTp search against the NCBI nr database and whenever the best hit corresponded to a bacterial gene, a phylogenetic tree was reconstructed to confirm the bacterial origin of the gene. Phylogenies were constructed as by retrieving the closest related sequences obtained by BLASTp searches against NCBI nr or UniprotKB databases. Sequences were aligned using MAFFT v7.402 using an iterative refinement method (–localpair) and trees were constructed with IQ-TREE v2.0.6 using an automated method for model selection and 1,000 ultrafast bootstraps.

## Results

### Extensive Variation in the Repertoire of Genome Stability Related Genes Across the Saccharomycotina

Recent investigations of gene family and trait evolution of the Saccharomycotina subphylum have shown that losses of genes and traits are frequent and substantially contribute to yeast diversity (Shen, et al. 2018; Feng, et al. 2025). For example, research in the genus *Hanseniaspora* revealed a substantial loss of cell cycle and DNA repair genes, which seem to be associated with rapid evolution in this lineage (Steenwyk, et al. 2019). However, it remains unclear how DNA repair gene repertoires, and more broadly genes involved in genome stability, and evolutionary rates vary across the entire yeast subphylum. To shed light on this question, we examined the distribution of 393 genes broadly associated with genome stability (henceforth referred to as genome stability genes, Table S1) across 1,154 proteomes representing at least 1,051 species within the subphylum Saccharomycotina (Groenewald, et al. 2023; Opulente, et al. 2024) using Hidden Markov Model sequence similarity searches (Steenwyk, et al. 2019; Phillips, et al. 2021; Steenwyk and Rokas 2021). The list of analyzed genes includes genes involved in distinct pathways that are directly or indirectly involved in DNA repair, as mutation rates are not exclusively impacted by DNA repair genes but also genes involved in cell cycle regulation or DNA replication (Schroeder, et al. 2018; Steenwyk, et al. 2019).

We observed that the repertoire of genome stability genes extensively varied across Saccharomycotina (Fig. 1a, Tables S1 and S2). Some genes were widely conserved, such as *RAD3* encoding a DNA helicase involved in nucleotide excision repair and transcription (Naumovski and Friedberg 1983), which was present in all but one strain. Other genes were poorly conserved, such as *IRC4*, which encodes a protein involved in double-strand break repair and was absent from 98% of the proteomes inspected. Of the 393 genes examined, 209 (53%) were found in all proteomes; in contrast, 105 (26%) were absent in > 10 proteomes (Table S1). Genes considered to be essential in *S. cerevisiae* were generally highly conserved across Saccharomycotina (Table S1), although several exceptions were observed. For instance, *ABF1*, encoding a transcription factor involved in chromatin silencing and remodeling (Rhode, et al. 1989, 1992), was absent in more than 80% of the proteomes examined (Table S1).

Species belonging to Saccharomycetales had the largest genome stability gene repertoires; other lineages generally had smaller genome stability gene repertoires (Table S1, Fig. 1a). The lineage with the smallest repertoire was the *Hanseniaspora* FEL (Fig. 1a, Table S2), which was previously reported to have lost multiple genes involved in DNA repair and cell cycle pathways (Table S2) (Steenwyk, et al. 2019). Two additional lineages were identified as having reduced genome stability gene repertoires compared to their closest relatives and other Saccharomycotina species (Fig. 1a, Figure S1, Table S2). One lineage comprised species belonging to the genera *Pichia*, *Saturnispora*, and *Martiniozyma*, all in the order Pichiales (henceforth referred to as the Pichiales subclade). Within this subclade, the average number of absent genome stability genes (36) exceeded the average for Pichiales as a whole (approximately 19) (Fig. 1b). The second clade includes species belonging to the genera *Wickerhamiella* and *Starmerella* (W/S clade) (Fig. 1a). In the W/S clade, the number of genes absent ranges from 32 to 55, while the number of missing genes ranges from 16 to 29 in close relatives belonging to the genera *Blastobotrys* and *Sugiyamaella*. Notably, the 3 lineages—*Hanseniaspora,* Pichiales subclade, and the W/S clade—are all distantly related, suggesting that the reductions of their genome stability gene repertoires have taken place independently. The observed pattern was consistent across the different sequence homology stringency thresholds employed (Figure S1).

Given that the gene set analyzed spans multiple functional pathways, we next asked whether genes specifically associated with DNA repair functions are also underrepresented in these 3 lineages. To test this, we focused on the 259 genes associated with the GO term GO:0006281-DNA repair (Table S1). The resulting pattern (Figure S2) is similar to that seen when the full set of genes was considered (Fig. 1a), showing that DNA repair repertoires are statistically significantly reduced in the three FELs compared to the background species.

### Independent Losses of Genome Stability-related Functions are Associated With Accelerated Evolutionary Rates

We next determined if evolutionary rate variation is associated with the repertoire of genome stability genes. The species with the highest evolutionary rates belonged to the 3 lineages with the higher proportions of gene absences: *Hanseniaspora* FEL, the Pichiales subclade, and the W/S clade ($P$-value < 0.001, Wilcoxon rank test, Fig. 1a). Evolutionary rates among FELs are significantly higher ($P$-value < 0.001, Wilcoxon rank test) compared to their respective closest relatives (Fig. 1b and c). The differences in evolutionary rates between these clades and their closest relatives (evolutionary rate fold change), were particularly pronounced in the W/S clade and in *Hanseniaspora* FEL (Fig. 1c). Evolutionary rates were relatively uniform within *Hanseniaspora* FEL and in the Pichiales subclade (2.71 to 2.88 substitutions per site in *Hanseniaspora* FEL and 2.08 to 2.58 substitutions per site in the Pichiales subclade) but highly variable in the W/S clade (1.65 to 2.56 substitutions per site; Fig. 1b), which can be partially attributed to the fact that *Starmerella* species exhibit slightly higher evolutionary rates than *Wickerhamiella* species (Figure S3a).

In *Hanseniaspora* FEL, the elevated evolutionary rates were previously found to be concentrated on the stem branch leading to the clade (Steenwyk, et al. 2019). Interestingly, terminal branches appeared to contribute substantially to the observed rate differences in both the W/S clade and the Pichiales subclade (Fig. 1a). We next tested whether there was a correlation between the total number of absent DNA repair genes and evolutionary rate across the yeast phylogeny using a phylogenetically corrected analysis. We found that the correlation is statistically significant ($P$-value $=5.1e^{-12}$, Phylogenetically Independent Contrasts—PIC method, Figure S3b) but weak (adjusted $R$-squared: 0.0397), suggesting that there is no consistent link between the size of the genome stability gene repertoires and altered evolutionary rates.

### Seven DNA Repair-related Genes are Absent From Most Species of the Faster-evolving Lineages

Genes consistently absent across all 3 FELs may represent strong candidates for contributing to accelerated evolutionary rates (Figure S4). To evaluate this, we first selected genes absent in more than 30% of strains of at least 2 foreground lineages (Figure S5) and compared these genes among all foreground strains (*Hanseniaspora* FEL, W/S clade, and

Pichiales subclade, $N = 139$) to the background strains (all others, $N = 1{,}015$) (Figure S5). We found that 17 of the 24 (70%) selected genes are also absent in a high proportion of background species (Figure S5), suggesting that a substantial fraction of the absences is not specific to the foreground clades.

However, seven genes (*MHF2*, *ISY1*, *MAG1*, *MGT1*, *SLX4*, *MTE1*, and *AHC1*) were found to be absent in more than 50% of all foreground strains and only in less than 10% of their background counterparts (Fig. 2, Figure S5). These genes are involved in multiple pathways in *S. cerevisiae*, although apart from *ISY1* and *MTE1*, all are involved in DNA repair or DNA damage response pathways (Table S1). For instance, *MHF2* is part of the MHF histone-fold complex, along with *MHF1*, which is involved in the cellular response to DNA damage stimulus (Yang, et al. 2012). We found that *MFH1* was in more than 50% of the FEL strains lacking *MHF2*, but these proteins are only 90 amino acids long, which might hinder their accurate identification. *MAG1* is involved in the base excision repair (BER) pathway; *MGT1* contributes to repairing DNA alkylation (Xiao, et al. 1991); *AHC1* is a subunit of the Ada histone acetyltransferase complex involved in double strand break repair (Eberharter, et al. 1999); and *SLX4* encodes one of the subunits of the Slx1-Slx4 endonuclease involved in double-strand break repair (Coulon, et al. 2004; Pardo and Aguilera 2012; Gaur, et al. 2015; Covo 2020). In line with this, we found that all predicted W/S proteomes lacking Slx4 also lack Slx1, while *Hanseniaspora* FEL strains that lack Slx4 contain Slx1.

*MTE1* is involved in the maintenance of telomere length but is also involved in double-strand break repair (Yimit, et al. 2016). *ISY1* is part of the NineTeen complex and is involved in the regulation of the fidelity of pre-mRNA splicing (Villa and Guthrie 2005).

After confirming the distribution of these genes through tBLASTn against the genomes and validating the presence of proteins with a patchy distribution in all 3 clades, we found that some lineages almost or completely lack certain genes. For instance, *MAG1* is absent in all *Hanseniaspora* FEL strains inspected, while *MGT1* is absent in 70% of the strains, in line with previous reports (Steenwyk, et al. 2019). *MAG1* and *MGT1* are also inferred to be absent from all strains in the Pichiales subclade, except for *Candida sorboxylosa*, which contains an *MGT1* homolog. A BLASTp search against the NCBI nr database revealed that the *MGT1* homologs present in both *C. sorboxylosa* and 30% of the *Hanseniaspora* FEL strains are likely of bacterial origin, suggesting that they were acquired through HGT (Figure S6). The HGT is further supported by a topology test that rejected the null hypothesis of vertical descent ($P$-value = 0.029, Approximately Unbiased [AU] test). In the W/S clade, all strains lack *AHC1,* and most species lack *ISY1*. *AHC1* is also absent in all Pichiales genomes inspected and *ISY1* is absent in almost all *Hanseniaspora* FEL.

Finally, the independent losses of these genes and respective acceleration of mutation rates likely occurred at different evolutionary time points. Relaxed molecular clock analyses (Opulente et al. 2024) estimate that the onset of accelerated mutation rates occurred approximately 87 million years ago (mya) in the *Hanseniaspora* FEL and 73 mya in the Pichiales subclade, coinciding with their divergence from their closest relatives. In the W/S clade, this acceleration appears to have originated substantially earlier, around 253 mya.

## Reduction in Genome Stability Gene Repertoires is not Associated With Accelerated Evolutionary Rates in W/S-clade Relatives

In all 3 faster-evolving lineages inspected, we observed that high evolutionary rates were associated with a substantial loss of genome stability genes and specifically DNA repair genes. The W/S clade stands out as having the second-highest number of genes absent (after *Hanseniaspora* FEL) and interestingly exhibits significantly higher evolutionary rates when compared to its closest relatives. This lineage is also known for having the highest rates of horizontal gene transfer (HGT) across the yeast subphylum (Gonçalves, et al. 2018, 2020; Shen, et al. 2018; Gonçalves and Gonçalves 2019). Furthermore, the W/S-clade sister lineage, which contains only 3 species [*Candida incommunis*, *Candida bentonensis*, and *Deakozyma indianensis* (Groenewald, et al. 2023; Opulente, et al. 2024)], showed no evidence of acceleration of evolutionary rates (Fig. 3b) even though it too experienced numerous DNA repair gene losses (between 31 and 35 genes were absent, Fig. 3a). We therefore hypothesized that comparing the gene repertoires of the W/S clade and its sister clade could shed light on genes potentially involved in the acceleration of mutation rates.

In the W/S clade, the genes absent among all species range from 30 to 53. In line with the slightly higher evolutionary rates in *Starmerella* compared to *Wickerhamiella* (Figure S3a), we found that the average number of genes absent in *Starmerella* species (42) was slightly higher than in *Wickerhamiella* species (37) (Table S2). Although genes absent are associated with multiple functions, 70% of the genes absent in at least 10% of W/S-clade strains are associated with DNA repair (GO:0006281) or DNA damage response (GO:0006974) functions (Table S1, Table S3). Among these are genes associated with distinct pathways from cell cycle (eg MEC3, *PDS1*, *REC8*), DNA mismatch repair (eg *RPA3*), nonhomologous end joining (eg *XRS2*, *NEJ1*), or base excision repair (eg *TDP1*). Interestingly, *TDP1* is absent in *Starmerella* species but present in almost all *Wickerhamiella* species, suggesting a loss in the most recent common ancestor of the *Starmerella* genus (Table S3, Figure S7).

We also observed that a substantial portion of genes absent in the W/S clade are also absent in the W/S sister clade (Table S3). Specifically, 26 of the 37 genes absent in $\geq 50\%$ of W/S-clade species were also absent in at least 2 of the 3 W/S-sister species (Table S3). Notwithstanding sampling issues associated with the small number of W/S-sister species, these results suggest that loss of DNA repair-related genes occurred both prior to and during the diversification of the W/S clade.

This prompted us to look for genes absent in the W/S clade and present in its closest relatives, which could be strong candidates for being involved in accelerating evolutionary rates. We detected that *AHC1* was absent in all W/S clade species and present in the W/S-sister clade (Table S3). Ahc1 is part of the ADA complex involved in chromatin remodeling and double-strand break repair (Eberharter, et al. 1999; Muñoz-Galván, et al. 2013). Loss-of-function mutations in *AHC1* can cause elevated levels of homologous recombination in *S. cerevisiae* (Wong, et al. 2013). *HNT3* and *ISY1* were also present in W/S sister and absent in most species of the W/S clade. Hnt3 is a DNA 5′-adenylate hydrolase involved in DNA damage response (Daley, et al. 2010; Tumbale, et al. 2014), more specifically in RNA–DNA damage response by

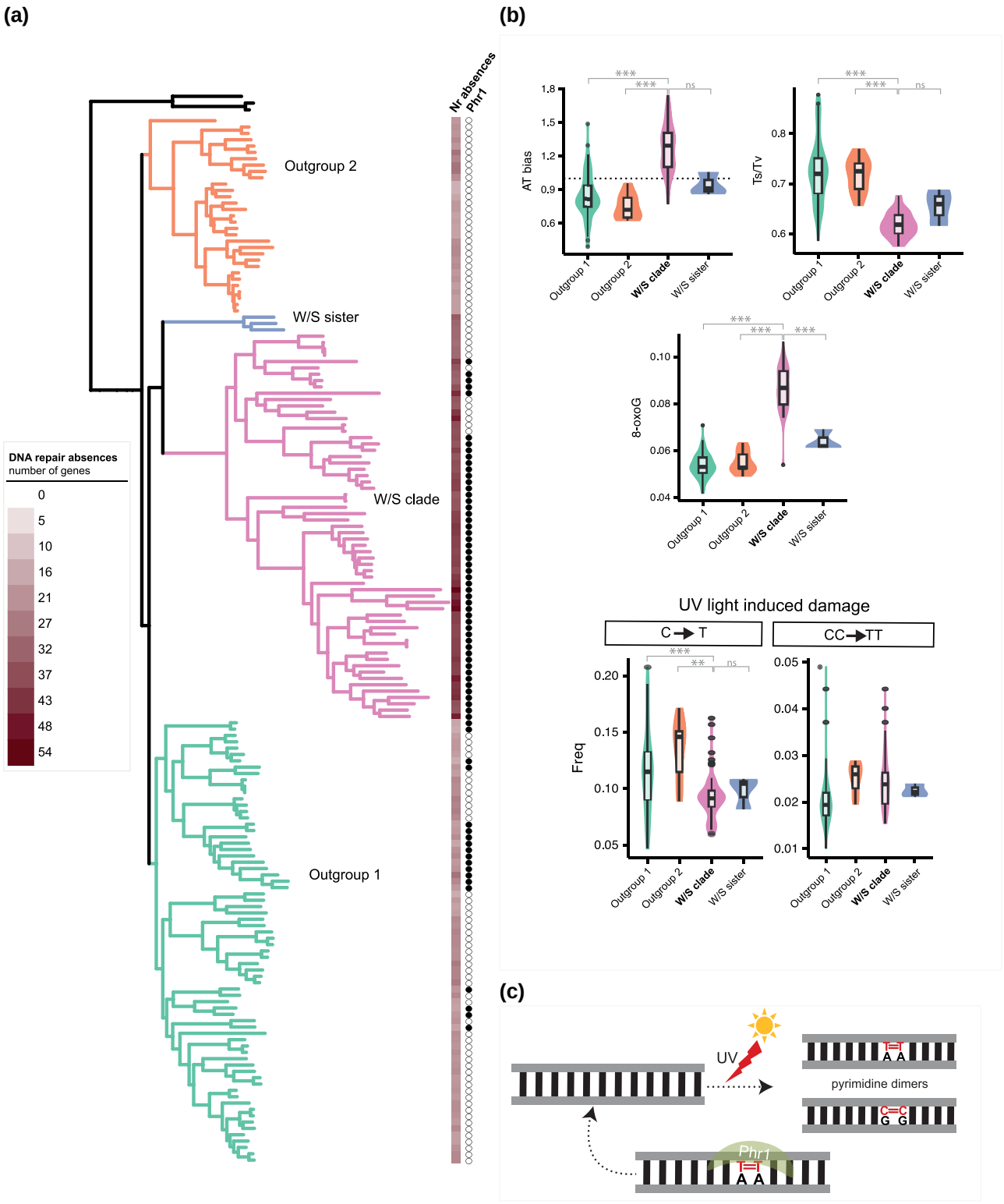


**Fig. 3.** Absence of genome stability genes in the W/S clade is commensurate with mutational burden. (a) Distribution of genome stability gene absences across the W/S clade and closest relatives. Presence/absence of *PHR1*, a gene encoding a DNA photolyase involved in UV damage repair, is shown next to the phylogeny. The Outgroup 1 comprises species from the genera *Blastobotrys*, *Sugyamaella*, *Groenewaldozyma*, *Zygoascus*, *Trichomonascus*, and *Spencermartinsiella* and the Outgroup 2 comprises representative species from the genera *Magnusiomyces*, *Dipodascus*, *Geotrichum*, and *Saprochaete*. (b) Analyses of substitution patterns among codon-based alignments of 143 single-copy orthogroups. Top row: Left) Substitution A|T bias. *Y*-axis is the number of substitutions in the A|T direction divided by the number of substitutions in the G|C direction. Thus, values greater than 1 indicate an A|T bias in substitutions. Right) ts/tv ratios reveal that the W/S clade is approaching a random expectation. Middle row) Mutational signatures associated with 8oxo-G, a common oxidative damage, reveal higher mutational signatures in the W/S clade. Bottom row: Single (T) and double (TT) mutations associated with UV damage. (c) UV damage repair via Phr1 photolyase. UV damage generally results in pyrimidine (cytosine or thymine) dimers. The pruned species phylogeny was obtained from (Opulente, et al. 2024). All pairwise statistical comparisons are shown with ns (no statistical significance), * (*P*-value < 0.05), ** (*P*-value < 0.001), or *** (*P*-value < 0.0001). All statistical analyses were performed using the Wilcoxon test with Holm correction.

repairing adenylated RNA–DNA. On the other hand, Isy1 is involved in mRNA splicing in *S. cerevisiae* (Villa and Guthrie 2005), and its human homolog has been found to be involved in base excision DNA repair (Jaiswal, et al. 2020).

We noted that *CSM2* and *RFA3* were present in the W/S clade's closest relatives and absent in most or all W/S-clade species. However, when we analyzed the sequences found in W/S sister species by a reciprocal BLASTp in the NCBI non-redundant (nr) database, we concluded that these proteins were not orthologs of either *CSM2* or *RFA3*. *RFA3* is an essential gene in *S. cerevisiae* that encodes a DNA-binding subunit of replication protein A complex involved in DNA recombination (Brill and Stillman 1991). The putative Rfa3 sequences in W/S sister returned Hst4, an $NAD^+$-dependent protein deacetylase involved in mitotic DNA replication and genomic stability (Pan, et al. 2006), as top hits, not Rfa3. As for *CSM2*, the top hits were Cdc40 proteins, which in *S. cerevisiae* encode a pre-mRNA splicing factor involved in cell cycle progression and RNA splicing (Dahan and Kupiec 2004). Although these proteins appear to belong to different gene families and perform distinct functions, they share sufficient homology to be detected by our searches. This overestimation of gene presences is likely the result of our use of a permissive e-value cutoff of $1e^{-3}$ in our sequence similarity searches, a decision guided by the need to prioritize confident inference of gene absences.

While our approach for identifying the absence of genes is conservative—employing an e-value cutoff of $1e^{-3}$ alongside tBLASTn searches against genome assemblies to mitigate the impact of annotation issues—the higher evolutionary rates observed in W/S-clade species may cause additional challenges in sequence homology detection. Specifically, some absences could stem from proteins whose sequences are too divergent to be recognized as homologs by sequence similarity search algorithms, a phenomenon previously documented in other highly divergent fungi (Mascarenhas Dos Santos, et al. 2022). To test whether this was an issue, we further employed a protein structural homology-based search that is not reliant on sequence similarity (Kaminski, et al. 2023). We chose 5 proteins that were flagged as absent in most W/S-clade species but present in their closest relatives according to our sequence similarity searches (*AHC1*, *ISY1*, *HNT3*, *SLX1*, and *SLX4*). Employing a protein structure homology algorithm yielded results consistent with the sequence similarity searches: proteins were absent in W/S-clade species and present in the W/S sister clade. These results (Table S5) indicate that our sequence similarity-based results are generally reliable.

## Characterization of Mutational Patterns Reveals a Higher Burden in the W/S Clade

Our findings indicate that most absent genes in the W/S clade are also absent in the W/S-sister clade. However, while evidence of accelerated evolutionary rates was found in the W/S clade, no such evidence was found in its sister clade. Genomic fingerprints of base substitution patterns and indels, which provide insights into the mutational landscape, have previously been used to reveal signatures of mutational burden associated with DNA repair gene loss in *Hanseniaspora* FEL (Steenwyk, et al. 2019). Thus, we analyzed patterns of base substitutions, substitution directionality, indels, and signatures of endogenous and UV-induced damage across W/S-clade species and their closest relatives (Fig. 3 and Figure S8).

We found that the W/S clade exhibited a higher mutational burden than its closest relatives across several mutational signatures examined. For instance, the bias toward A|T substitutions was significantly stronger in the W/S clade than in its relatives ($P$-value < 0.001, Wilcoxon rank test), except when compared to the W/S-sister clade (Fig. 3b). This pattern is consistent with the general A|T bias of spontaneous mutations reported for several organisms (Hershberg and Petrov 2010; Lynch 2010; Liu and Zhang 2021), including the *Hanseniaspora* FEL (Steenwyk, et al. 2019), suggesting deficient DNA repair ability. In line with this, the transition (ts)/transversion (tv) ratio was approximately 0.5 to 0.7 in the W/S clade (Fig. 3b). These values align with the estimated ts/tv ratios attributed to neutral mutations in *S. cerevisiae* (Lynch, et al. 2008; Zhu, et al. 2014). We also found significant mutational load associated with one of the most abundant endogenously damaged bases, 8-oxoguanine (Shibutani, et al. 1991), which causes the transversion mutation of G → T or C → A (De Bont and van Larebeke 2004).

Deletions were also significantly higher in the W/S clade compared to its closest relatives (Figure S8), while no evidence for a higher proportion of insertions was found (Figure S8). We also did not find evidence for mutational burden associated with UV damage, which was evaluated by the number of C → T substitutions at CC sites (or G → A substitutions at GG sites), as well as the less frequent CC → TT (or GG → AA) double substitutions (Fig. 3b). In fact, we found that C → T substitutions in particular might be lower in the W/S clade when compared to closest relatives.

Interestingly, while we did not find evidence of accelerated evolutionary rates in the W/S-sister clade (Fig. 3a), this lineage showed a slight bias toward A|T substitutions (Fig. 3b), suggesting a lower DNA repair efficiency. Consistent with this hypothesis, the ts/tv ratio in this clade was also notably lower than in other closely related clades (Fig. 3b). This finding suggests that, while loss of genome stability genes might not have affected evolutionary rates in this lineage (Fig. 3a), it could have left mutational fingerprints across the genome. However, it is important to note that these considerations are based on only the three species currently described in the W/S sister clade.

## The W/S Clade has Horizontally Acquired Genes Related With UV Damage Repair Functions

*PHR1* encodes a DNA photolyase and is the main gene responsible for directly reversing UV damage in *S. cerevisiae* (Liu, et al. 2011). Considering the absence of UV damage signatures in the W/S clade, we specifically investigated the distribution of the *PHR1* gene within this lineage and its closest relatives. Approximately 84% of species in the W/S clade possess a *PHR1* gene, whereas the gene is substantially less prevalent among their closest relatives (Fig. 3a).

The patchy distribution of *PHR1* spurred us to investigate its evolutionary history in the W/S clade. For that, we reconstructed the phylogenetic history of the gene by first performing a BLASTp search of all putative Phr1 sequences from the W/S clade against the NCBI RefSeq database. For some W/S-clade species, the top hits were from filamentous fungi (subphylum Pezizomycotina); for some *Wickerhamiella* species, the top hits were bacterial proteins. This suggests that HGT event(s) might have contributed to the patchy distribution of Phr1 in the W/S clade. To formally test this, we constructed a phylogenetic tree with all putative orthologs of Phr1 from

Saccharomycotina (using the 1,154-proteome dataset) and the top hits obtained from RefSeq for two distinct W/S clade Phr1 proteins (bacterial and fungal). The phylogeny, represented in Fig. 4a, shows 3 major clades: one of bacterial sequences, one of Saccharomycotina sequences, and a third of Pezizomycotina sequences; the Saccharomycotina and Pezizomycotina sequences were sister clades, as expected from the species phylogeny (Li, et al. 2021). There were 2 exceptions to this pattern. First, some W/S-clade Phr1 proteins cluster within the bacterial clade (within the order Sphingomonadales), and the rest of the W/S-clade Phr1 proteins clustered within the Pezizomycotina clade. The second exception was sequences of *Blastobotrys* species (Outgroup 1), which clustered within the Pezizomycotina clade (together with W/S Phr1 sequences) (Fig. 4a). Constrained topology analyses rejected an alternative topology where bacterial-like W/S sequences clustered with other Saccharomycotina species ($P$-value = 0.001, Approximately Unbiased [AU] test). This result supports the hypothesis that an HGT event occurred, possibly in the most recent common ancestor (MRCA) of the *Wickerhamiella* subclade composed of the bacterial Phr1-harboring species (Fig. 4b). As for the Pezizomycotina-like Phr1 sequences from the W/S clade, constrained topology tests also rejected the alternative hypothesis that clustered these proteins with other Saccharomycotina ($P$-value = 0, AU test). Nevertheless, most Pezizomycotina-like W/S Phr1 sequences formed a monophyletic group, which is sister to the Pezizomycotina. Therefore, we cannot exclude the alternative hypothesis of differential retention of an ancestral paralog.

We next tested whether these putative xenologs are functional by performing UV irradiation assays (Milo, et al. 2019). Damage caused by exposure to ultraviolet light is photorepaired by the activity of Phr1, which is activated by light energy. Therefore, we exposed cells of PHR1 + and *phr*- species to UV-C light (200 to 280 nm range, which causes lesions in DNA) and assessed whether subsequent exposure to UV-A light (320 to 400 nm range, which induces photoreactivation of Phr1) improved UC-V resistance, suggesting that Phr1 is active. We found that in W/S-clade species harboring bacterial and fungal Phr1 versions, cells exposed to UV-A light after UV-C induced damage were more resistant than cells that were not subject to photoreactivation (Figure S9). This is consistent with the UV-A dependent activation of Phr1 (Sancar 1990; Milo, et al. 2019). We also found that W/S-clade species harboring either the fungal or the bacterial version of *PHR1* exhibited a generally stronger response to photoreactivation when compared to W/S-clade species that seemingly lack the gene (Figure S9). We have, however, noted that *phr1*- strains also respond to photoreactivation, suggesting that yet unidentified mechanism(s) of repair responsive to photoreactivation might exist in these species.

It is well established that the W/S clade has acquired numerous foreign genes from both bacteria and filamentous fungi (Gonçalves, et al. 2016, 2018, 2020; Shen, et al. 2018; Gonçalves and Gonçalves 2019; Kominek, et al. 2019; Pontes, et al. 2024). To explore if other DNA repair-related genes might have also been acquired horizontally by W/S-clade species, we retrieved the DNA repair-related genes from *Escherichia coli* and *Bacillus subtilis* (a total of 263 sequences) as of February 2023 from UniProtKB and performed a BLASTp against the W/S proteomes. We next constructed phylogenetic trees for the proteins that showed significant similarity with bacterial proteins based on a second BLAST search against the NCBI nr database. From this analysis, we detected 7 DNA repair-related proteins whose top BLASTp hits were bacterial sequences (Table S4), suggesting that they are of putative bacterial origin. One of these proteins is a homolog of *MGT1*, which appears to have also been involved in HGT events in *Hanseniaspora* FEL and *C. sorboxylosa* (Pichiales subclade) (Figure S6). Phylogenetic reconstruction provided support for the bacterial origin of an endonuclease V, which was likely independently acquired by W/S-clade species (Fig. 4b), possibly once in the MRCA of *Wickerhamiella* and once in the MRCA of *Starmerella* (Figure S10a). Although top BLAST hits for *Wickerhamiella* species were endonuclease V sequences from the bacterial CFB group, the *Starmerella* top hits belonged to the gamma proteobacteria. This topology suggests at least 2 independent acquisitions, but we did not find statistical support for this hypothesis ($P$-value = 0.069, AU test). Endonuclease V is involved in the repair of deaminated DNA bases, which are commonly caused by endogenous and environmental agents (Cao 2013), and is absent in the rest of the Saccharomycotina. However, loss of endonuclease V in *Schizosaccharomyces pombe* (subphylum Taphrinomycotina) induces a strong mutator phenotype (Dalhus, et al. 2009). Another protein for which we found phylogenetic support of HGT was UvrA (Figure S10b), which is partly involved in nucleotide excision repair and might also be involved in UV damage repair in bacteria (Agostini, et al. 1996; Crowley, et al. 2006).

## Discussion

Rapid evolution associated with the loss of DNA repair-related genes is usually observed in microevolutionary contexts, such as during the evolution of human tumours or clinical isolates of microbial pathogens (LeClerc, et al. 1996; Fukui 2010; Healey, et al. 2016; Billmyre, et al. 2017; Campbell, et al. 2017; Gambhir, et al. 2022). A few examples demonstrate how entire fungal lineages can also undergo rapid evolution over macroevolutionary timescales, likely resulting from the loss of DNA repair genes (Steenwyk, et al. 2019; Phillips, et al. 2021). Here, we demonstrate that impairment of DNA repair and possibly other genome stability-related functions is particularly pronounced in the 3 most rapidly evolving lineages of the fungal subphylum Saccharomycotina. One of these lineages, belonging to the genus *Hanseniaspora* (*Hanseniaspora* FEL), was previously identified as having lost numerous genes associated with the cell cycle and DNA repair (Steenwyk, et al. 2019), demonstrating the efficacy of our approach, while the other 2 lineages, the *Wickerhamiella*/*Starmerella* (W/S) clade and a subclade within the Pichiales comprising *Pichia, Saturnispora*, and *Martiniozyma* species, are reported here for the first time.

The number of absent genome stability genes in these 3 FELs is statistically significantly higher than in the rest of the subphylum; we found that this is particularly true for genes specifically involved in DNA repair, which accounted for 65% of all the genes inspected. Furthermore, almost all yeast strains lack at least a portion of the examined genes. This phenomenon can be partially explained by the fact that most of the genes inspected are part of the *S. cerevisiae* gene repertoire, and supports the notion that the DNA repair gene repertoire varies across yeast species (Milo, et al. 2019). It also indicates that many of these genes are not essential and may be partially redundant, which is a critical feature in safeguarding the genome against DNA damage (Darzynkiewicz 2011; Gartner and Engebrecht 2021). Even genes described as essential in

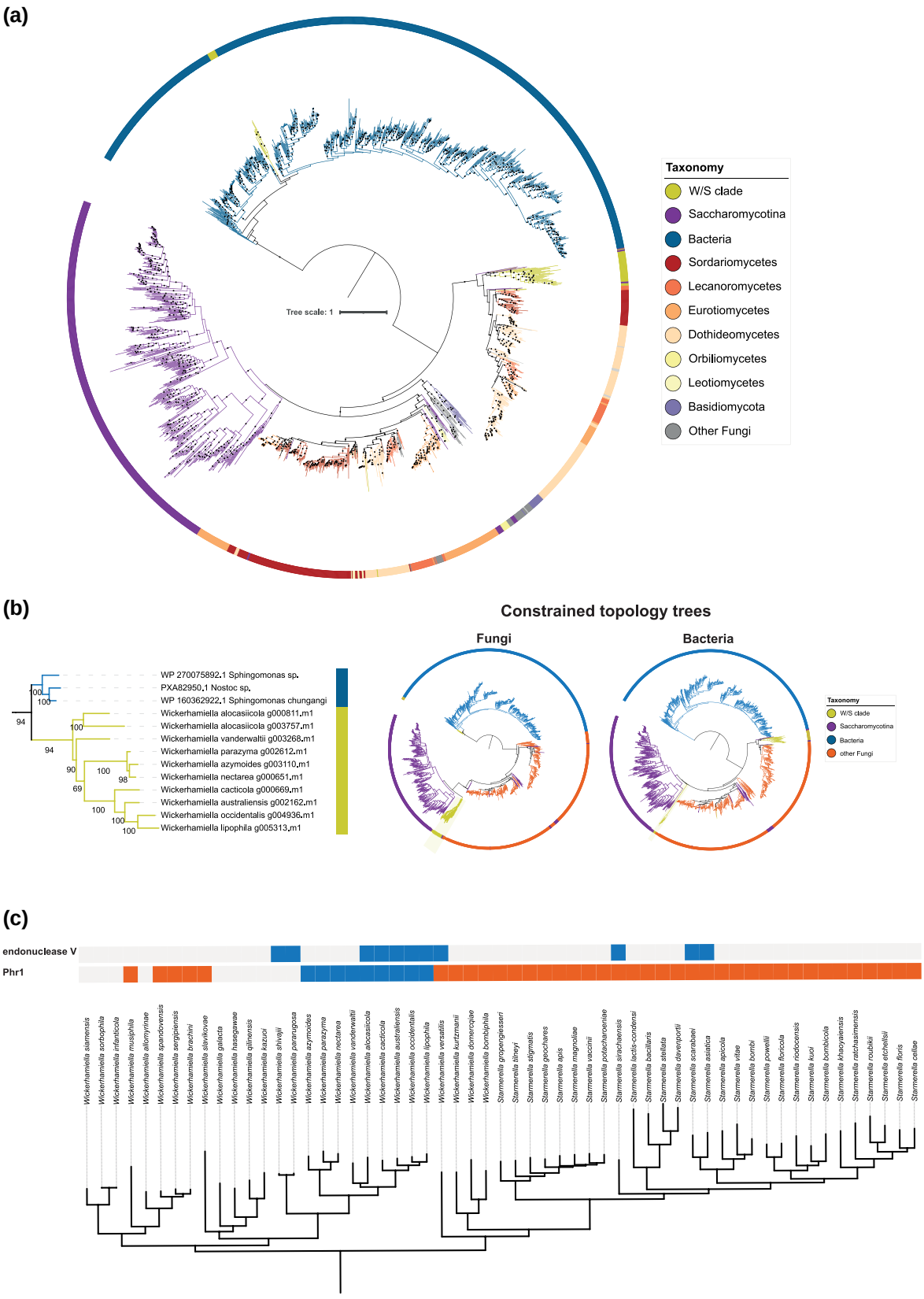


**Fig. 4.** Several W/S-clade species acquired bacterial DNA repair genes through horizontal gene transfer (HGT). **(**a) Phylogenetic tree of sequences with the highest sequence similarity to Phr1 from *Wickerhamiella australiensis* and *Starmerella bombicola*. Branches are colored according to taxonomy (species outside Dikarya or Pezizomycotina species for which no taxonomic class could be determined were categorized as "Other Fungi"). (b) (on the left) Pruned Phr1 phylogeny depicting the HGT event from bacteria (Sphingomonadales) to a W/S subclade. On the right, the resulting trees from the constrained topology analyses are shown: Fungi—constrained topology considering monophyly of the W/S Pezizomycotina-like sequences and other Saccharomycotina (excluding the Saccharomycotina sequences that cluster with other Fungi) and Bacteria-constrained topology considering monophyly of bacteria-like W/S sequences and Saccharomycotina sequences (excluding those clustering with other fungi). (c) Distribution of putative HGT-derived (blue: bacterial, orange: fungal) *PHR1* and endonuclease V genes across the W/S clade. Original tree files and sequence alignments can be found in Figshare (https://figshare.com/s/3d575d6283c970a0c1e4).

*S. cerevisiae* and close relatives, such as the transcription regulator Abf1 (Rhode, et al. 1992; Hernández-Hernández, et al. 2021), might not be essential in the distinct genetic and environmental backgrounds of other species in the Saccharomycotina subphylum. For instance, insects are known to lack several "key" DNA repair genes, suggesting that they may have evolved alternative strategies to deal with DNA-damaging agents (Wyder, et al. 2007; Sekelsky 2017).

Although we found that the lineages with the most reduced gene repertoires were also the fastest evolving, the correlation between the size of genome stability gene repertoires and rates of evolution across the subphylum was weak. One potential explanation for the weak correlation is that high mutation rates can be caused by only a few genes rather than being inversely proportional to the size of the gene repertoire. In line with this hypothesis, we sought to identify genes potentially involved in increasing evolutionary rates by analysing the common absences across the three FELs. We did not find genes exclusively absent in FELs, but several genes (eg *MAG1*, *MGT1*) were absent in most of the faster-evolving species, raising the hypothesis that they might be or have been involved in the acceleration of mutation rates in these genetic backgrounds.

Following this logic, only a few genes may have been responsible for the acceleration of mutation rates, and the others were subsequently lost either due to the accumulation of deleterious mutations that can lead to loss of function or as a compensatory mechanism (Ngo and Lydall 2010). Under this model, acceleration of mutation rates could have preceded massive gene loss (rather than the other way round).

In line with the weak correlation between genome stability, gene repertoire size, and evolutionary rates, we also found that the most significant gene loss event(s) in the W/S clade might have occurred before its diversification, since its sister clade already exhibits reduced genome stability and gene repertoires. We observed that most genes absent in the W/S clade are also absent in species belonging to its sister clade, but we did not find evidence of accelerated evolutionary rates in the latter. However, the W/S-sister clade does display mild signatures of mutational burden, such as A|T bias. This difference could either suggest that the additional gene absences in the W/S clade (eg *AHC1*, *ISY1*) may be involved in the acceleration of the evolutionary rates or that the W/S-sister clade may have compensated for the loss of certain genes through alternative mechanisms.

Interestingly, we found possible mechanisms of such compensation, such as the acquisition of foreign genes associated with UV damage repair. We found that most W/S-clade species encode Phr1 proteins, while most of their closest relatives lack the gene. Some of the W/S-clade Phr1 homologs were likely acquired from bacteria in a single HGT event, while the others are more closely related to Phr1 sequences from filamentous fungi. Whether the fungal-like sequences are the result of additional HGT event(s) from the Pezizomycotina or are the result of a differential retention of an ancient paralog remains unclear.

Notably, we found no evident genomic signatures of UV-induced damage and found that Phr1-containing species are resistant to UV under photoreactivation conditions, suggesting that these proteins might be functional. Additionally, we found that some species also acquired a bacterial endonuclease V, a broad specificity enzyme involved in the repair of deaminated bases that can arise from multiple types of endogenous and exogenous aggressions (Cao 2013). Although these HGT events could have helped W/S clade-species to deal with DNA damage, it is still unclear whether a deceleration of the evolutionary rates has occurred, as observed for *Hanseniaspora* (Steenwyk, et al. 2019). Mutation accumulation experiments will be essential to ascertain whether these species are still evolving faster or have slowed down their mutation rates.

It is important to note that our conservative approach to considering gene absences may generate false positives (false gene presences) by detecting the presence of conserved motifs in distant homologs, as shown for *CSM2* or *RFA3*. Conversely, reduced DNA repair gene repertoire was previously associated with remarkably high levels of sequence divergence in a lineage of intracellular parasites (Microsporidia) (Corradi 2015). However, using a recently developed protein language model for distant homolog detection, DNA repair gene loss in Microsporidia was found to have been less extensive than previously thought (Mascarenhas Dos Santos, et al. 2022). Therefore, we validated some of the gene absences with both tBLASTn searches against the genomes to avoid annotation issues and (in a few cases) with protein language models nonreliant on sequence similarity, but we cannot fully rule out that some genes are fast-evolving and are therefore difficult to identify. However, at least in the W/S clade, loss of genome stability genes seems to have been part of a broader gene loss event (Figure S11), which was also observed in *Hanseniaspora* FEL (Steenwyk, et al. 2019). While *Blastobotrys* and *Sugiyamaella* yeast genomes range between 11 and 25 Mb in size, W/S-clade genomes are around 9 to 11 Mb. Examination of both genome size and gene number suggests that extensive ancient gene losses might have occurred in the MRCA of the W/S and W/S-sister clades, with subsequent gene loss events occurring after the diversification of the W/S clade, including genes involved in genome stability. These ancient losses raise the hypothesis that some of the HGT events of DNA repair genes we uncovered, and others documented in the literature (Gonçalves, et al. 2018, 2022; Gonçalves and Gonçalves 2019; Pontes, et al. 2024), might have been compensatory.

What triggered the ancient losses observed in the 3 FEL lineages remains unclear. Ecological factors may have played a role. For example, many species in the W/S clade are associated with the floral environment (Lachance 2011; de Vega, et al. 2017), particularly with bees and other pollinators. It is plausible that some form of symbiotic relationship with insects existed in W/S ancestors, potentially leading to the loss of redundant functions. Similar patterns of gene loss, especially in DNA repair and metabolic pathways, have been frequently observed in other symbiotic relationships.

The W/S clade is also known for exhibiting the highest number of bacterial genes across the subphylum Saccharomycotina (Gonçalves, et al. 2018, 2020; Shen, et al. 2018; Gonçalves and Gonçalves 2019; Kominek, et al. 2019; Pontes, et al. 2024). Defects in certain DNA repair pathways, such as the MMR system, have been correlated with higher rates of HGT events in bacteria because these mutator strains can recombine more frequently with divergent DNA (Rayssiguier, et al. 1989; Thomas and Nielsen 2005). While we did not find specific MMR losses in the W/S clade, we can speculate whether the periods of high mutation rates might have facilitated the integration of HGT-derived genes into W/S yeast genomes.

In *Hanseniaspora*, for instance, the dysregulation of the cell cycle may offer an advantage in sugar-rich environments where they are usually found, because it allows for a faster generation time (Vegas, et al. 2020). For example, during wine making, *Hanseniaspora* typically outcompetes *S. cerevisiae* in earlier stages of wine fermentation (Albertin, et al. 2016).

In conclusion, while the loss of DNA repair and other genome stability-related functions may be generally detrimental due to the accumulation of deleterious mutations, our results show that entire lineages can maintain elevated mutation rates for prolonged periods of time. In the well-described short-lived hypermutator populations (Billmyre, et al. 2017; Rhodes, et al. 2017; Gambhir, et al. 2022; Callens, et al. 2023; Hall, et al. 2025), high mutation rates are associated with nonsense mutations in DNA repair genes that can be compensated by the emergence of anti-mutator alleles. Over macroevolutionary timescales, genetic constraints stemming from wholesale gene loss may have contributed to the maintenance of high evolutionary rates, while evolutionary mechanisms, such as HGT, may have partially compensated for ancient losses.

## Supplementary material

Supplementary material is available at *Molecular Biology and Evolution* online.

## Acknowledgments

We thank members of the Rokas Lab, Hittinger Lab, Yeast Genomics Lab, and members of the Y1000+ Project (http://y1000plus.org) for helpful discussions. We thank Sérgio Filipe (Molecular Microbiology & Genomics Lab, NOVA-FCT) for kindly providing access to the UV lamp. This work was performed using resources contained within the Advanced Computing Center for Research and Education at Vanderbilt University in Nashville, TN.

## Funding

National Science Foundation from the USA Grant DEB-2110403 (CTH). National Science Foundation from the USA Grant DEB-2110404 (AR). DOE Great Lakes Bioenergy Research Center, funded by BER Office of Science Grant DE-SC0018409 (CTH). USDA National Institute of Food and Agriculture Hatch Project 1020204 (CTH). USDA National Institute of Food and Agriculture Hatch Project 7005101 (CTH). H. I. Romnes Faculty Fellow, supported by the Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation (CTH). National Institutes of Health/National Institute of Allergy and Infectious Diseases Grant R01 AI153356 (AR). Burroughs Wellcome Fund (AR). Research supported by the National Key R&D Program of China Grant 2022YFD1401600 (XXS). National Science Foundation for Distinguished Young Scholars of Zhejiang Province Grant LR23C140001 (XXS). Fundamental Research Funds for the Central Universities Grant 226-2023-00021 (XXS). National Institutes of Health Grant T32 HG002760-16 (JFW). National Science Foundation Grant Postdoctoral Research Fellowship in Biology 1907278 (JFW). Fundação para a Ciência e a Tecnologia Grant PTDC/BIA-EVL/0604/2021 (CG). Fundação para a Ciência e a Tecnologia Grant UIDP/04378/2020 (CG). Fundação para a Ciência e a Tecnologia Grant UIDB/04378/2020 (CG). Fundação para a Ciência e a Tecnologia Grant LA/P/0140/2020 (CG). JLS is a Howard Hughes Medical Institute Awardee of the Life Sciences Research Foundation.

## Conflict of Interest

JLS is a scientific consultant to FutureHouse Inc. JLS is an advisor for ForensisGroup Inc. During this project, JLS was a scientific advisor for WittGen Biotechnologies and a Bioinformatics Visiting Scholar at MantleBio Inc. AR is a scientific consultant for LifeMine Therapeutics, Inc.

## Data Availability

All input data for all analyses and raw results files can be found in Figshare: https://figshare.com/s/3d575d6283c970a0c1e4. Genomic data used in this work is publicly available and can be accessed via the Figshare repository from Opulente et al. (2024): https://plus.figshare.com/articles/dataset/Genomes_and_Annotations_of_1_154_budding_yeasts/22802147.

## References

Agostini HJ, Carroll JD, Minton KW. Identification and characterization of uvrA, a DNA repair gene of *Deinococcus radiodurans*. *J Bacteriol*. 1996:178:6759–6765. https://doi.org/10.1128/jb.178.23.6759-6765.1996.

Albertin W *et al.* *Hanseniaspora uvarum* from winemaking environments show spatial and temporal genetic clustering. *Front Microbiol*. 2016:6:1569. https://doi.org/10.3389/fmicb.2015.01569.

Billmyre RB, Clancey SA, Heitman J. Natural mismatch repair mutations mediate phenotypic diversity and drug resistance in *Cryptococcus deuterogattii*. *eLife*. 2017:6:e28802. https://doi.org/10.7554/eLife.28802.

Blum M *et al.* The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res*. 2021:49:D344–d354. https://doi.org/10.1093/nar/gkaa977.

Bridges BA. Hypermutation in bacteria and other cellular systems. *Philos Trans R Soc Lond B Biol Sci*. 2001:356:29–39. https://doi.org/10.1098/rstb.2000.0745.

Brill SJ, Stillman B. Replication factor-A from *Saccharomyces cerevisiae* is encoded by three essential genes coordinately expressed at S phase. *Genes Dev*. 1991:5:1589–1600. https://doi.org/10.1101/gad.5.9.1589.

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform*. 2021:3:lqaa108. https://doi.org/10.1093/nargab/lqaa108.

Brůna T, Lomsadze A, Borodovsky M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom Bioinform*. 2020:2:lqaa026. https://doi.org/10.1093/nargab/lqaa026.

Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015:12:59–60. https://doi.org/10.1038/nmeth.3176.

Callens M *et al.* Hypermutator emergence in experimental *Escherichia coli* populations is stress-type dependent. *Evol Lett*. 2023:7:252–261. https://doi.org/10.1093/evlett/qrad019.

Campbell BB *et al.* Comprehensive analysis of hypermutation in human cancer. *Cell*. 2017:171:1042–1056.e10. https://doi.org/10.1016/j.cell.2017.09.048.

Cao W. Endonuclease V: an unusual enzyme for repair of DNA deamination. *Cell Mol Life Sci*. 2013:70:3145–3156. https://doi.org/10.1007/s00018-012-1222-z.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009:25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Chopra I, O'Neill AJ, Miller K. The role of mutators in the emergence of antibiotic-resistant bacteria. *Drug Resist Updat*. 2003:6:137–145. https://doi.org/10.1016/S1368-7646(03)00041-4.

Corradi N. Microsporidia: eukaryotic intracellular parasites shaped by gene loss and horizontal gene transfers. *Annu Rev Microbiol*. 2015:69:167–183. https://doi.org/10.1146/annurev-micro-091014-104136.

Coulon S *et al.* Slx1-slx4 are subunits of a structure-specific endonuclease that maintains ribosomal DNA in fission yeast. *Mol Biol Cell*. 2004:15:71–80. https://doi.org/10.1091/mbc.e03-08-0586.

Covo S. Genomic instability in fungal plant pathogens. *Genes (Basel)*. 2020:11:421. https://doi.org/10.3390/genes11040421.

Crowley DJ *et al.* The uvrA, uvrB and uvrC genes are required for repair of ultraviolet light induced DNA photoproducts in *Halobacterium* sp. NRC-1. *Saline Syst*. 2006:2:11. https://doi.org/10.1186/1746-1448-2-11.

Dahan O, Kupiec M. The *Saccharomyces cerevisiae* gene *CDC40/PRP17* controls cell cycle progression through splicing of the *ANC1* gene. *Nucleic Acids Res*. 2004:32:2529–2540. https://doi.org/10.1093/nar/gkh574.

Daley JM, Wilson TE, Ramotar D. Genetic interactions between *HNT3*/aprataxin and *RAD27/FEN1* suggest parallel pathways for 5′ end processing during base excision repair. *DNA Repair (Amst)*. 2010:9:690–699. https://doi.org/10.1016/j.dnarep.2010.03.006.

Dalhus B *et al.* Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nat Struct Mol Biol*. 2009:16:138–143. https://doi.org/10.1038/nsmb.1538.

Darzynkiewicz Z. Redundancy in response to DNA damage: the key to protection of genome integrity. *Cell Cycle*. 2011:10:3425. https://doi.org/10.4161/cc.10.20.17525.

De Bont R, van Larebeke N. Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis*. 2004:19:169–185. https://doi.org/10.1093/mutage/geh025.

de Vega C *et al.* Flowers as a reservoir of yeast diversity: description of *Wickerhamiella nectarea* f.a. sp. nov., and *Wickerhamiella natalensis* f.a. sp. nov. from South African flowers and pollinators, and transfer of related *Candida* species to the genus *Wickerhamiella* as new combinations. *FEMS Yeast Res*. 2017:17:1–11. https://doi.org/10.1093/femsyr/fox054.

Dos Reis TF *et al.* The *Aspergillus fumigatus* mismatch repair *MSH2* homolog is important for virulence and azole resistance. *mSphere*. 2019:4:e00416–e00419. https://doi.org/10.1128/mSphere.00416-19.

Eberharter A *et al.* The ADA complex is a distinct histone acetyltransferase complex in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 1999:19:6621–6631. https://doi.org/10.1128/MCB.19.10.6621.

Eddy SR. Profile hidden markov models. *Bioinformatics*. 1998:14:755–763. https://doi.org/10.1093/bioinformatics/14.9.755.

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019:20:238. https://doi.org/10.1186/s13059-019-1832-y.

Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. *Nat Rev Genet*. 2007:8:610–618.

Feng B *et al.* Unique trajectory of gene family evolution from genomic analysis of nearly all known species in an ancient yeast lineage. *Mol Syst Biol*. 2025:21:1066–1089. https://doi.org/10.1038/s44320-025-00118-0.

Fukui K. DNA mismatch repair in eukaryotes and bacteria. *J Nucleic Acids*. 2010:2010:260512. https://doi.org/10.4061/2010/260512.

Gambhir N, Harris SD, Everhart SE. Evolutionary significance of fungal hypermutators: lessons learned from clinical strains and implications for fungal plant pathogens. *mSphere*. 2022:7:e0008722. https://doi.org/10.1128/msphere.00087-22.

Gartner A, Engebrecht J. DNA repair, recombination, and damage signaling. *Genetics*. 2021:220:iyab178. https://doi.org/10.1093/genetics/iyab178.

Gaur V *et al.* Structural and mechanistic analysis of the Slx1-slx4 endonuclease. *Cell Rep*. 2015:10:1467–1476. https://doi.org/10.1016/j.celrep.2015.02.019.

Giglia-Mari G, Zotter A, Vermeulen W. DNA damage response. *Cold Spring Harb Perspect Biol*. 2011:3:a000745. https://doi.org/10.1101/cshperspect.a000745.

Gonçalves C *et al.* Evidence for loss and reacquisition of alcoholic fermentation in a fructophilic yeast lineage. *eLife*. 2018:7:e33034. https://doi.org/10.7554/eLife.33034.

Gonçalves C, Coelho MA, Salema-Oom M, Gonçalves P. Stepwise functional evolution in a fungal sugar transporter family. *Mol Biol Evol*. 2016:33:352–366. https://doi.org/10.1093/molbev/msv220.

Gonçalves C, Gonçalves P. Multilayered horizontal operon transfers from bacteria reconstruct a thiamine salvage pathway in yeasts. *Proc Natl Acad Sci U S A*. 2019:116:22219–22228. https://doi.org/10.1073/pnas.1909844116.

Gonçalves C, Marques M, Gonçalves P. Contrasting strategies for sucrose utilization in a floral yeast clade. *mSphere*. 2022:7:e0003522. https://doi.org/10.1128/msphere.00035-22.

Gonçalves P, Gonçalves C, Brito PH, Sampaio JP. The *Wickerhamiella/Starmerella* clade—A treasure trove for the study of the evolution of yeast metabolism. *Yeast*. 2020:37:313–320. https://doi.org/10.1002/yea.3463.

Groenewald M *et al.* A genome-informed higher rank classification of the biotechnologically important fungal subphylum *Saccharomycotina*. *Stud Mycol*. 2023:105:1–22. https://doi.org/10.3114/sim.2023.105.01.

Hall KM *et al.* Mutational signature analysis predicts bacterial hypermutation and multidrug resistance. *Nat Commun*. 2025:16:19. https://doi.org/10.1038/s41467-024-55206-w.

Healey KR *et al.* Prevalent mutator genotype identified in fungal pathogen *Candida glabrata* promotes multi-drug resistance. *Nat Commun*. 2016:7:11128. https://doi.org/10.1038/ncomms11128.

Hernández-Hernández G *et al.* Abf1 is an essential protein that participates in cell cycle progression and subtelomeric silencing in *Candida glabrata*. *J Fungi (Basel)*. 2021:7:1005. https://doi.org/10.3390/jof7121005.

Hershberg R, Petrov DA. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet*. 2010:6:e1001115. https://doi.org/10.1371/journal.pgen.1001115.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018:35:518–522. https://doi.org/10.1093/molbev/msx281.

Hunter S *et al.* InterPro: the integrative protein signature database. *Nucleic Acids Res*. 2009:37:D211–D215. https://doi.org/10.1093/nar/gkn785.

Jaiswal AS *et al.* The splicing component *ISY1* regulates *APE1* in base excision repair. *DNA Repair (Amst)*. 2020:86:102769. https://doi.org/10.1016/j.dnarep.2019.102769.

Jiang M *et al.* Alterations of DNA damage repair in cancer: from mechanisms to applications. *Ann Transl Med*. 2020:8:1685. https://doi.org/10.21037/atm-20-2920.

Jombart T, Balloux F, Dray S. *Adephylo*: new tools for investigating the phylogenetic signal in biological traits. *Bioinformatics*. 2010:26:1907–1909. https://doi.org/10.1093/bioinformatics/btq292.

Kaminski K, Ludwiczak J, Pawlicki K, Alva V, Dunin-Horkawicz S. pLM-BLAST: distant homology detection based on direct comparison of sequence representations from protein language models. *Bioinformatics*. 2023:39:btad579. https://doi.org/10.1093/bioinformatics/btad579.

Katoh K, Standley DM. MAFFT: iterative refinement and additional methods. *Methods Mol Biol*. 2014:1079:131–146. https://doi.org/10.1007/978-1-62703-646-7_8.

Kimura M. On the evolutionary adjustment of spontaneous mutation rates. *Genet Res*. 2009:9:23–34. https://doi.org/10.1017/S0016672300010284.

Kominek J *et al.* Eukaryotic acquisition of a bacterial operon. *Cell*. 2019:176:1356–1366.e10. https://doi.org/10.1016/j.cell.2019.01.034.

Kreuzer KN. DNA damage responses in prokaryotes: regulating gene expression, modulating growth patterns, and manipulating replication forks. *Cold Spring Harb Perspect Biol*. 2013:5:a012674. https://doi.org/10.1101/cshperspect.a012674.

Lachance MA. Starmerella Rosa & lachance (1998). In: Kurtzman CF, Fell JW, Boekhout T, editors. *The yeasts: a taxonomic study*. Elsevier; 2011. p. 811–815.

LeClerc JE, Li B, Payne WL, Cebula TA. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science*. 1996:274:1208–1211. https://doi.org/10.1126/science.274.5290.1208.

Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006:22:1658–1659. https://doi.org/10.1093/bioinformatics/btl158.

Li Y *et al*. A genome-scale phylogeny of the kingdom fungi. *Curr Biol*. 2021:31:1653–1665.e5. https://doi.org/10.1016/j.cub.2021.01.074.

Liu H, Zhang J. The rate and molecular spectrum of mutation are selectively maintained in yeast. *Nat Commun*. 2021:12:4044. https://doi.org/10.1038/s41467-021-24364-6.

Liu Z *et al*. Dynamics and mechanism of cyclobutane pyrimidine dimer repair by DNA photolyase. *Proc Natl Acad Sci U S A*. 2011:108:14831–14836. https://doi.org/10.1073/pnas.1110927108.

Lynch M *et al*. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci U S A*. 2008:105:9272–9277. https://doi.org/10.1073/pnas.0803466105.

Lynch M. Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci U S A*. 2010:107:961–968. https://doi.org/10.1073/pnas.0912629107.

Mascarenhas Dos Santos AC, Julian AT, Pombert JF. The Rad9-rad1-hus1 DNA repair clamp is found in microsporidia. *Genome Biol Evol*. 2022:14:evac053. https://doi.org/10.1093/gbe/evac053.

Milo S, Harari-Misgav R, Hazkani-Covo E, Covo S. Limited DNA repair gene repertoire in ascomycete yeast revealed by comparative genomics. *Genome Biol Evol*. 2019:11:3409–3423. https://doi.org/10.1093/gbe/evz242.

Milo S, Namawejje R, Krispin R, Covo S. Dynamic responses of *Fusarium mangiferae* to ultra-violet radiation. *Fungal Biol*. 2024:128:1714–1723. https://doi.org/10.1016/j.funbio.2024.02.007.

Muñoz-Galván S, Jimeno S, Rothstein R, Aguilera A. Histone H3K56 acetylation, Rad52, and non-DNA repair factors control double-strand break repair choice with the sister chromatid. *PLoS Genet*. 2013:9:e1003237. https://doi.org/10.1371/journal.pgen.1003237.

Murakami-Sekimata A, Huang D, Piening BD, Bangur C, Paulovich AG. The *Saccharomyces cerevisiae RAD9*, *RAD17* and *RAD24* genes are required for suppression of mutagenic post-replicative repair during chronic DNA damage. *DNA Repair (Amst)*. 2010:9:824–834. https://doi.org/10.1016/j.dnarep.2010.04.007.

Naumovski L, Friedberg EC. A DNA repair gene required for the incision of damaged DNA is essential for viability in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*. 1983:80:4818–4821. https://doi.org/10.1073/pnas.80.15.4818.

Ngo H-P, Lydall D. Survival and growth of yeast without telomere capping by Cdc13 in the absence of Sgs1, Exo1, and Rad9. *PLoS Genet*. 2010:6:e1001072. https://doi.org/10.1371/journal.pgen.1001072.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015:32:268–274. https://doi.org/10.1093/molbev/msu300.

O'Leary NA *et al*. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016:44:D733–D745. https://doi.org/10.1093/nar/gkv1189.

Oliver A, Cantón R, Campo P, Baquero F, Blázquez J. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science*. 2000:288:1251–1253. https://doi.org/10.1126/science.288.5469.1251.

Opulente DA *et al*. Genomic factors shape carbon and nitrogen metabolic niche breadth across Saccharomycotina yeasts. *Science*. 2024:384:eadj4503. https://doi.org/10.1126/science.adj4503.

Pan X *et al*. A DNA integrity network in the yeast *Saccharomyces cerevisiae*. *Cell*. 2006:124:1069–1081. https://doi.org/10.1016/j.cell.2005.12.036.

Pardo B, Aguilera A. Complex chromosomal rearrangements mediated by break-induced replication involve structure-selective endonucleases. *PLoS Genet*. 2012:8:e1002979. https://doi.org/10.1371/journal.pgen.1002979.

Phillips MA, Steenwyk JL, Shen X-X, Rokas A. Examination of gene loss in the DNA mismatch repair pathway and its mutational consequences in a fungal phylum. *Genome Biol Evol*. 2021:13:evab219. https://doi.org/10.1093/gbe/evab219.

Pontes A *et al*. Extensive remodeling of sugar metabolism through gene loss and horizontal gene transfer in a eukaryotic lineage. *BMC Biol*. 2024:22:128. https://doi.org/10.1186/s12915-024-01929-7.

Rayssiguier C, Thaler DS, Radman M. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature*. 1989:342:396–401. https://doi.org/10.1038/342396a0.

Rhode PR, Elsasser S, Campbell JL. Role of multifunctional autonomously replicating sequence binding factor 1 in the initiation of DNA replication and transcriptional control in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 1992:12:1064–1077. https://doi.org/10.1128/mcb.12.3.1064-1077.1992.

Rhode PR, Sweder KS, Oegema KF, Campbell JL. The gene encoding ARS-binding factor I is essential for the viability of yeast. *Genes Dev*. 1989:3:1926–1939. https://doi.org/10.1101/gad.3.12a.1926.

Rhodes J *et al*. A population genomics approach to assessing the genetic basis of within-host microevolution underlying recurrent cryptococcal meningitis infection. *G3 (Bethesda)*. 2017:7:1165–1176. https://doi.org/10.1534/g3.116.037499.

Roberts SA, Gordenin DA. Hypermutation in human cancer genomes: footprints and mechanisms. *Nat Rev Cancer*. 2014:14:786–800. https://doi.org/10.1038/nrc3816.

Sancar GB. DNA photolyases: physical properties, action mechanism, and roles in dark repair. *Mutat Res*. 1990:236:147–160. https://doi.org/10.1016/0921-8777(90)90002-M.

Schroeder JW, Yeesin P, Simmons LA, Wang JD. Sources of spontaneous mutagenesis in bacteria. *Crit Rev Biochem Mol Biol*. 2018:53:29–48. https://doi.org/10.1080/10409238.2017.1394262.

Sebastian J, Kraus B, Sancar GB. Expression of the yeast *PHR1* gene is induced by DNA-damaging agents. *Mol Cell Biol*. 1990:10:4630–4637. https://doi.org/10.1128/mcb.10.9.4630-4637.1990.

Sekelsky J. DNA repair in *Drosophila*: mutagens, models, and missing genes. *Genetics*. 2017:205:471–490. https://doi.org/10.1534/genetics.116.186759.

Serero A, Jubin C, Loeillet S, Legoix-Né P, Nicolas AG. Mutational landscape of yeast mutator strains. *Proc Natl Acad Sci U S A*. 2014:111:1897–1902. https://doi.org/10.1073/pnas.1314423111.

Shaver AC *et al*. Fitness evolution and the rise of mutator alleles in experimental *Escherichia coli* populations. *Genetics*. 2002:162:557–566. https://doi.org/10.1093/genetics/162.2.557.

Shen XX *et al*. Tempo and mode of genome evolution in the budding yeast subphylum. *Cell*. 2018:175:1533–1545.e20. https://doi.org/10.1016/j.cell.2018.10.023.

Shen X-X *et al*. Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota. *Sci Adv*. 2020:6:eabd0079. https://doi.org/10.1126/sciadv.abd0079.

Shibutani S, Takeshita M, Grollman AP. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature*. 1991:349:431–434. https://doi.org/10.1038/349431a0.

Stanke M *et al*. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res*. 2006:34:W435–W439. https://doi.org/10.1093/nar/gkl200.

Steenwyk JL *et al*. Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts. *PLoS Biol*. 2019:17:e3000255. https://doi.org/10.1371/journal.pbio.3000255.

Steenwyk JL. Evolutionary divergence in DNA damage responses among fungi. *mBio*. 2021:12:e03348-20. https://doi.org/10.1128/mBio.03348-20.

Steenwyk JL, Rokas A. Orthofisher: a broadly applicable tool for automated gene identification and retrieval. *G3 (Bethesda)*. 2021:11:jkab250. https://doi.org/10.1093/g3journal/jkab250.

Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol*. 2017:35:1026–1028. https://doi.org/10.1038/nbt.3988.

Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 2006:34:W609–W612. https://doi.org/10.1093/nar/gkl315.

Swings T *et al*. Adaptive tuning of mutation rates allows fast response to lethal stress in Escherichia coli. *eLife*. 2017:6:e22939. https://doi.org/10.7554/eLife.22939.

Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*. 2009:4:4.10.1–14.10.14. https://doi.org/10.1002/0471250953.bi0410s25.

Thomas CM, Nielsen KM. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol*. 2005:3:711–721. https://doi.org/10.1038/nrmicro1234.

Tumbale P, Williams JS, Schellenberg MJ, Kunkel TA, Williams RS. Aprataxin resolves adenylated RNA-DNA junctions to maintain genome integrity. *Nature*. 2014:506:111–115. https://doi.org/10.1038/nature12824.

Vegas C, Zavaleta AI, Canales PE, Esteve-Zarzoso B. Yeasts associated with various Amazonian native fruits. *Pol J Microbiol*. 2020:69:251–261. https://doi.org/10.33073/pjm-2020-027.

Vialle RA, Tamuri AU, Goldman N. Alignment modulates ancestral sequence reconstruction accuracy. *Mol Biol Evol*. 2018:35:1783–1797. https://doi.org/10.1093/molbev/msy055.

Villa T, Guthrie C. The Isy1p component of the NineTeen complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing. *Genes Dev*. 2005:19:1894–1904. https://doi.org/10.1101/gad.1336305.

Wielgoss S *et al*. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc Natl Acad Sci U S A*. 2013:110:222–227. https://doi.org/10.1073/pnas.1219574110.

Wong ED, Karra K, Hitz BC, Hong EL, Cherry JM. The YeastGenome app: the *Saccharomyces* genome database at your fingertips. *Database (Oxford)*. 2013:2013:bat004. https://doi.org/10.1093/database/bat004.

Wyder S, Kriventseva EV, Schröder R, Kadowaki T, Zdobnov EM. Quantification of ortholog losses in insects and vertebrates. *Genome Biol*. 2007:8:R242. https://doi.org/10.1186/gb-2007-8-11-r242.

Xiao W, Derfler B, Chen J, Samson L. Primary sequence and biological functions of a *Saccharomyces cerevisiae* O6-methylguanine/O4-methylthymine DNA repair methyltransferase gene. *EMBO J*. 1991:10:2179–2186. https://doi.org/10.1002/j.1460-2075.1991.tb07753.x.

Yang H *et al*. *Saccharomyces cerevisiae* MHF Complex structurally resembles the histones (H3-H4)2 heterotetramer and functions as a heterotetramer. *Structure*. 2012:20:364–370. https://doi.org/10.1016/j.str.2011.12.012.

Yimit A *et al*. *MTE1* functions with *MPH1* in double-strand break repair. *Genetics*. 2016:203:147–157. https://doi.org/10.1534/genetics.115.185454.

Zhu YO, Siegal ML, Hall DW, Petrov DA. Precise estimates of mutation rate and spectrum in yeast. *Proc Natl Acad Sci U S A*. 2014:111:E2310–E2318. https://doi.org/10.1073/pnas.1323011111.