

MINIREVIEW

Into the wild: new yeast genomes from natural environments and new tools for their analysis

D. Libkind^{1,*}, D. Peris², F. A. Cubillos^{3,4}, J. L. Steenwyk^{5,†}, D. A. Opulente^{6,7}, Q. K. Langdon⁶, A. Rokas^{5,§} and C. T. Hittinger^{6,7,¶}

¹Centro de Referencia en Levaduras y Tecnología Cervecera (CRELTEC), Instituto Andino Patagónico de Tecnologías Biológicas y Geoambientales (IPATEC) – CONICET/Universidad Nacional del Comahue, Quintral 1250 (8400), Bariloche., Argentina, ²Department of Food Biotechnology, Institute of Agrochemistry and Food Technology-CSIC, Calle Catedrático Dr. D. Agustín Escardino Benloch n°7, 46980 Paterna, Valencia, Spain, ³Millennium Institute for Integrative Biology (iBio). General del Canto 51 (7500574), Santiago, ⁴Universidad de Santiago de Chile, Facultad de Química y Biología, Departamento de Biología. Alameda 3363 (9170002). Estación Central. Santiago, Chile, ⁵Department of Biological Sciences, VU Station B#35-1634, Vanderbilt University, Nashville, TN 37235, USA, ⁶Laboratory of Genetics, Wisconsin Energy Institute, J. F. Crow Institute for the Study of Evolution, Center for Genomic Science Innovation, University of Wisconsin-Madison, 1552 University Avenue, Madison, WI 53726-4084, USA and ⁷DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, 1552 University Avenue, Madison, WI 53726-4084, Madison, WI, USA

*Corresponding author: Centro de Referencia en Levaduras y Tecnología Cervecera (CRELTEC), IPATEC (CONICET-UNComahue), Quintral 1250 (8400), Bariloche, Argentina. Tel: +54-9-294-4623911; E-mail: libkindfd@comahue-conicet.gob.ar

One sentence summary: New genome sequencing technologies together with new bioinformatic tools provide unique insights into the fascinating stories of wild and non-conventional yeast.

Editor: Jean-Marc Daran

¹D. Libkind, <http://orcid.org/0000-0001-9687-6186>

[†]J. L. Steenwyk, <http://orcid.org/0000-0002-8436-595X>

[§]A. Rokas, <http://orcid.org/0000-0002-7248-6551>

[¶]C. T. Hittinger, <http://orcid.org/0000-0001-5088-7461>

ABSTRACT

Genomic studies of yeasts from the wild have increased considerably in the past few years. This revolution has been fueled by advances in high-throughput sequencing technologies and a better understanding of yeast ecology and phylogeography, especially for biotechnologically important species. The present review aims to first introduce new bioinformatic tools available for the generation and analysis of yeast genomes. We also assess the accumulated genomic data of wild isolates of industrially relevant species, such as *Saccharomyces* spp., which provide unique opportunities to further investigate the domestication processes associated with the fermentation industry and opportunistic pathogenesis. The availability of genome sequences of other less conventional yeasts obtained from the wild has also increased substantially, including representatives of the phyla Ascomycota (e.g. *Hanseniaspora*) and Basidiomycota (e.g. *Phaffia*). Here, we review salient examples of both fundamental and applied research that demonstrate the importance of continuing to sequence and analyze genomes of wild yeasts.

Received: 7 December 2019; Accepted: 31 January 2020

© FEMS 2020. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Keywords: *Saccharomyces*; *Hanseniaspora*; *Phaffia*; Biotechnology; Wild yeast

INTRODUCTION

Recent advances in sequencing technologies, the availability of new bioinformatic tools and multiple genomic studies during the past five years have significantly improved our understanding of the evolution, phylogeography, ecology and biotechnology of yeasts. Although most studies have focused on the genus *Saccharomyces*, substantial progress has been achieved with other yeasts of the phylum Ascomycota and, to a lesser extent, with yeasts of the phylum Basidiomycota. Today, approximately one fifth of the 1500 described yeast species have had their genomes fully sequenced (Kurtzman, Fell and Boekhout 2011; Shen et al. 2018). In a few cases, sequences of multiple isolates are available for population genomic studies (Table S1, Supporting Information). Historically, most studies were performed on yeast strains isolated from anthropic environments. In recent years, the number of yeasts from natural environments (wild yeasts) whose genomes have been sequenced has increased rapidly, creating a new opportunity to more fully explore eukaryotic biological mechanisms. This review provides an update on recent advances in the bioinformatics tools available for assembling, annotating and mining yeast genomes from a broad evolutionary range of yeasts (Section 2). Besides the best-studied genus *Saccharomyces* (Section 3), we also include other examples of outstanding interest from the Ascomycota (Section 4.1) and Basidiomycota (Section 4.2), which are rising models of yeast evolution and are becoming important for specific industrial applications.

NEW BIOINFORMATIC TOOLS FOR *de novo* GENOME RECONSTRUCTION AND ANALYSIS OF YEASTS

Access to whole genome sequence data has significantly increased in the past few years. In particular, the number of species of yeasts of the subphylum Saccharomycotina whose genomes have been sequenced has increased at least three-fold (Hittinger et al. 2015; Shen et al. 2018). While these data are more accessible, their analysis can be challenging. Non-conventional yeasts can have ploidy variation, have high heterozygosity, or be natural hybrids. Although there are multiple tools available to explore a broad range of topics in yeast evolution, integrating these tools to answer biological questions can be daunting. Table 1 depicts a description of new bioinformatic tools useful for genomic data processing and their respective references.

Several tools have been developed to quantify ploidy levels and detect hybrids from short-read sequencing data. Both *nQuire* and *sppIDer* are alignment-based approaches developed for detecting ploidy variation and hybridization events, respectively. They are useful to run on raw data prior to genome assembly since these factors create challenges for *de novo* genome assembly programs that affect performance and increase the frequency of assembly errors. Multiple *de novo* genome assembly programs are available that can use short-reads, many of which are available in the wrapper *iWGS*, including the ploidy-aware genome assemble programs *PLATANUS* and *dipSPAdes*, which perform well on highly heterozygous sequences. Additionally, genome assemblies with long-reads can be performed with the wrapper *LRSDAY* (Yue and Liti 2018). Prior to these phylogenetic analyses, the bioinformatic tool *BUSCO* can be used to assess

genome quality and completeness, as well as to curate a robust set of orthologous genes to build phylogenies in programs, such as *RAXML* (Shen et al. 2018). As an alternative to traditional phylogenetic approaches that require aligned sequences, phylogenetic analyzes can be performed prior to genome assembly using *AAF* and *SISRS*. Genome annotations can be performed using *MAKER2* and *YGAP*. *MAKER2* is a wrapper that calls multiple gene annotation tools and makes for multiple sets of gene predictions simultaneously, while *YGAP* is a web-based tool built specifically for yeast genome annotation, especially genomes that are syntenic with the model yeast *Saccharomyces cerevisiae*. Additionally, *HybPiper* can be used to detect candidate genes that are located in hard-to-assemble regions of the genome and does not require genome assembly.

In recent years, DNA reassociation also referred as DNA-DNA hybridization (DDH) has been gradually replaced by high-throughput sequencing, which allows the *in silico* calculation of overall genome related indices (OGRI) (Chun and Rainey 2014). OGRI include any measurements indicating how similar two genome sequences are, but they are only useful for differentiating closely related species (Chun et al. 2018). Examples of OGRI include average nucleotide identity (ANI) and digital DDH (dDDH), which are widely used, and relevant software tools are readily available as web-services and as standalone tools (for a detailed list see Chun et al. 2018; Libkind et al. 2020). Other approaches include the calculation of pairwise similarities (Kr, with the tool *genomediff* of *Genometools*) and genome-wide alignments (*MUMmer*, Marçais et al. 2018). The resulting alignments can be used to obtain syntenic regions, study conservation and assist in ultra-scaffolding.

There are many bioinformatic tools and pipelines available that are not listed here. For example, approaches have been developed to explore gene functions (Pellegrini et al. 1999; Jones et al. 2014), horizontal gene transfers (HGT) (Alexander et al. 2016), species phylogenetic tree inference (Shen et al. 2016) and copy number variation (Steenwyk and Rokas 2017, 2018). Furthermore, new tools are being developed regularly. The availability of these bioinformatic tools, coupled with access to hundreds of genomes, allows us to address a broad range of questions in yeast genomics, evolution and genetics.

GENOMICS IN THE MODEL GENUS *Saccharomyces*

The understanding of the *S. cerevisiae* genome has been driven by the advent of novel-sequencing technologies. Indeed, *S. cerevisiae* was the first eukaryote to be completely sequenced (Goffeau et al. 1996) (Fig. 1). Furthermore, the development of next-generation sequencing (NGS or 2nd generation) technologies and long-read sequencing (3rd generation) technologies, together with bioinformatic tools (see Section 2) (Fig. 2), have enhanced our understanding of yeast genome evolution and led to nearly complete assemblies of the nuclear genomes of four of the eight known *Saccharomyces* species. The new combined data sets allowed the annotation of most eukaryotic genetic elements: centromeres, protein-coding genes, tRNAs, Ty retrotransposable elements, core X' elements, Y' elements and ribosomal RNA genes. The study of the population-scale dynamics of repetitive genomic regions has been relatively underexplored due to the emphasis on short-read (< 300 bp) technologies, such as Illumina sequencing. Regardless, in combination with short-read

Table 1. List of bioinformatics tools.

Software	Description	Input	Output	Notes	Pros	Cons	Reference
AAF	Assembly and alignment-free phylogenetic approach.	Raw-sequencing reads.	Phylogenetic tree.		Works with raw short read data. Alignment- and assembly-free. Can work with low coverage data and has low computational demands.	Does not work with hybrid genomes. Cannot analyze deep nodes. Does not report informative sites.	Fan et al. 2015
BUSCO	Assessment of genome quality and completion. Gene annotation of universal orthologous genes.	Assembled genome.	Summary output of gene counts and locations. Amino acid and protein sequences for complete genes in the genome. Varies depending on tool being used.	Requires a set of genes to search for, which are available through their website.	Easy to use. Provides a set of genes that can be used in downstream analyses (e.g. phylogenomics).	Cannot find novel genes and does not allow for customizable gene sets. Gene sets are always single copy orthologs.	Waterhouse et al. 2018
GenomeTools	Collection of bioinformatic tools. For example, <i>genomediff</i> can be used to calculate pairwise differences between genomes (not gene based). Assembles genes of interest from short reads.	Varies depending on tool being used.		This is a collection of tools.	Has multiple tools, which can be used for whole genome analyses. Easy to use.	Not all tools are useful for the approaches discussed here.	Gremme, Steinbiss and Kurtz 2013
HybPiper	Assembles genes of interest from short reads.	Raw-sequencing reads and sequence of gene of interest.	Assembled contig with gene of interest (e.g. FASTA).	Output requires manual assessment to interpret the results.	Works with raw short read data. Assembles genes in regions that tend to be difficult for <i>de novo</i> assemblers. It can potentially retrieve paralogs.	Negative results do not necessarily mean the gene does not exist. Does not give information about functionality. Downstream steps required for <i>in silico</i> gene functionality assessment.	Johnson et al. 2016

Table 1. Continued

Software	Description	Input	Output	Notes	Pros	Cons	Reference
iWGS	Wrapper which integrates multiple <i>de novo</i> genome assemblers. It allows for customization of which assemblers to include in analyses. It has upstreaming trimming and assembly quality assessments downstream. A package with the EMBOSS software.	Raw-sequencing reads (e.g. FASTQ).	Multiple assembled genomes.	Output requires the manual selection of appropriate assembly for your question. A subset of assemblers can be used on long-read data. See reference for list of assemblers.	Parallelizes genome assembly, with a range of assemblers, including assemblers that are ploidy-aware. Upstream steps, such as trimming, are included. Can simulate genome sequencing experiments	Computationally intensive and can be difficult to set-up initially due to multiple dependencies.	Zhou et al. 2016
jEMBOSS		An assembled genome.	Depends on your analysis. For example GC content and length of scaffolds/chromosomes. Chromosome-level scaffolds.		Useful to detect the mitochondrial scaffold.		Carver and Bleasby 2003
LRSDAY	Genome assembly for Nanopore and PacBio data.	Raw-sequencing reads.			Assembler for yeast long read data. There is a well-detailed extensive step-by-step workflow.	Initial set-up can be difficult due to multiple dependencies. Requires a reference genome for assembly.	Yue and Liti 2018
MAKER2	<i>de novo</i> gene annotation.	Assembled genome.	Integrated gene annotations across multiple platforms.	MAKER2 uses SNAP, Augustus, GenMark. MAKER2 can also predict genes using evidence-based approaches. Prior to final annotation, multiple training rounds must occur.	Runs multiple gene predictors. Integrates multiple tools and lines of evidence for accurate predictions.	Requires multiple dependencies and training for annotations. Difficult to run without prior experience.	Campbell et al. 2014

Table 1. Continued

Software	Description	Input	Output	Notes	Pros	Cons	Reference
MITObim	Mitochondrial baiting and iterative mapping.	A reference mitochondrial genome and NGS reads.	A mitochondrial genome assembly.		It targets mitochondrial reads to increase the chances of recovering a mitochondrial genome assembly. Visualization of large structural variants. Helps during ultrasccaffolding of closely related strains	When reference mitochondrial genome is quite different from your strain, the performance of the pipeline is poor. Small structural variants are much difficult to visualize.	Hahn, Bachmann and Chevreur 2013
MUMmer4	Ultra-fast alignment of long DNA and protein sequences.	Two assembled genomes.	PNG or PDF comparing the structure of both genomes.		Works with raw short-read data, with some upstream steps. Can detect aneuploidies.		Marçais et al. 2018
nQuire	Estimation of ploidy.	Short reads mapped to a reference genome.	Estimates of ploidy up to tetraploid, including plots to visualize the data.		Works with raw short-read data, with some upstream steps. Can detect aneuploidies.	Needs a reference genome. Can only detect up to tetraploids.	Weiß et al. 2018
RAxML	Phylogenetic placement.	Aligned sequences.	Phylogenetic tree.		Statistically robust. Is able to recover deeper nodes. Can generate larger phylogenies. Allows for parallelization.	Computationally intensive and maybe difficult for first-time users. It takes multiple upstream steps, including sequence alignments.	Stamatakis 2014
RepeatMasker	Screens DNA sequences for interspersed repeats and low complexity DNA sequences.	An assembled genome.	A detailed annotation of the repeats that are present in the query sequence.				

Table 1. Continued

Software	Description	Input	Output	Notes	Pros	Cons	Reference
A modified version of the query sequence in which all the annotated repeats have been masked. SISRS	Facilitates the annotation of transposable elements.	Raw sequencing reads.	Some libraries, such as RepBase, require payment.	http://www.repe.atmasker.org/	Works with raw short-read data. It is alignment- and assembly-free.	Does not work with hybrid genomes. It does not report informative sites. Deeper nodes are harder to recover.	Schwartz <i>et al.</i> 2015
splitsTree	Assembly- and alignment-free phylogenetic approach.	Raw sequencing reads.	Phylogenetic tree.		Works to visualize admixture and hybrids. Allows thresholding and gives confidence intervals.	Networks can be difficult to interpret and are not regularly used in phylogenomics.	Huson and Bryant 2006
sppIDer	Phylogenetic network.	Aligned sequences.	Phylogenetic network.		Works with raw short read data. Detects hybrids. Keeps intermediate steps, which can be used in other analyses. Includes statistical analyses.	Needs multiple closely related reference genomes.	Langdon <i>et al.</i> 2018
YGAP	Quick inference of genome composition, hybrid and admixture detection.	Raw-sequencing reads.	Multiple assessment (tables and visualizations) of genome composition.	Output requires manual assessment to interpret the results.	Trained on multiple yeast species. It can support pre- and post-whole genome duplication species. Web-based interface easy to use.	Web-based approach limits high throughput analyses. Does not work well on species highly diverged from <i>S. cerevisiae</i> .	Proux-Wera <i>et al.</i> 2012
	Gene annotation based on synteny.	Assembled genome.	Multiple gene annotation outputs.				

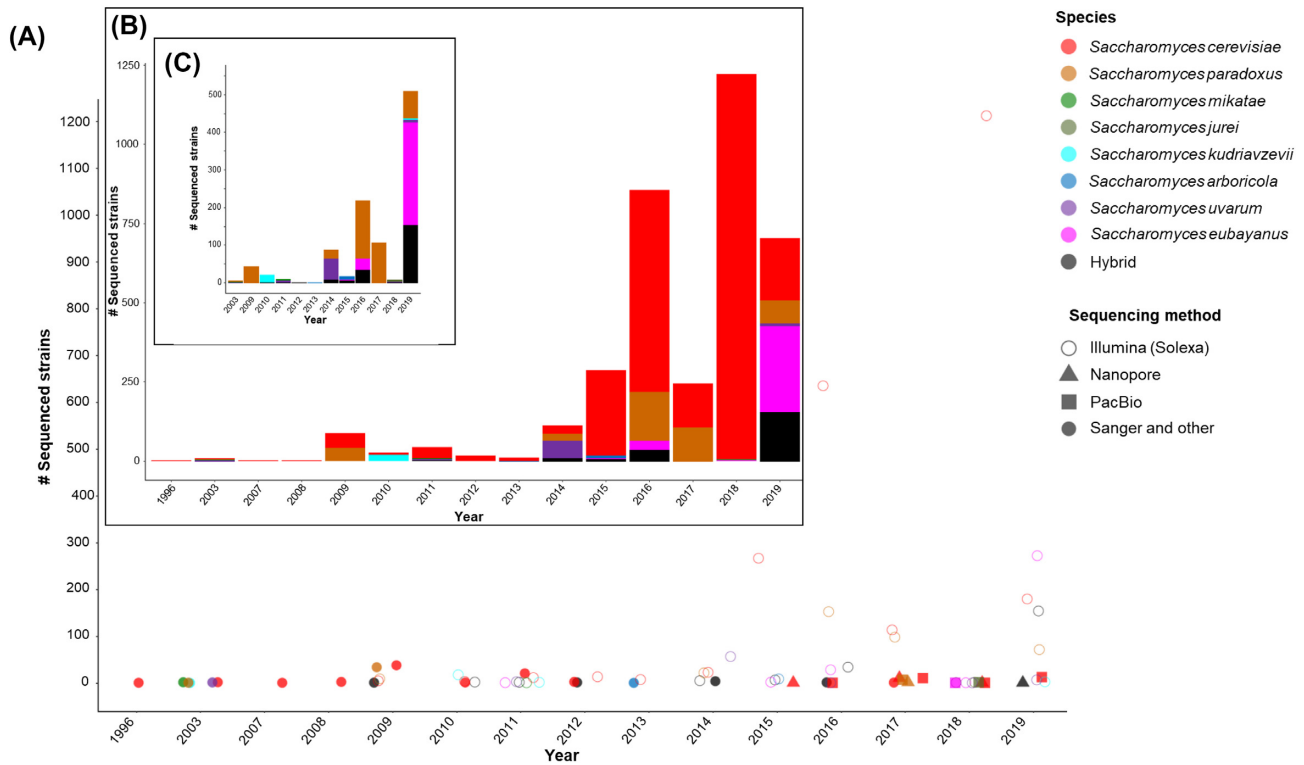


Figure 1. The genomes of more than three thousand *Saccharomyces* strains have been sequenced. At least 3077 unique *Saccharomyces* strains have had their genomes sequenced using various sequencing technologies in the past 23 years (Table S1, Supporting Information). About 71.5% of the sequenced *Saccharomyces* strains belong to *S. cerevisiae*, 11.0% are *S. paradoxus*, 8.5% are *S. eubayanus*, 5.95% are interspecies hybrids, and 2.0% are *Saccharomyces uvarum*. At least 105 *Saccharomyces* strains have been sequenced by more than two studies (Table S1, Supporting Information). Colored circles highlight the total genome sequences published per year per technology (symbol shape) for each *Saccharomyces* species or for interspecies hybrids. Bar plots represent the total number of sequenced strains from each *Saccharomyces* species or interspecies hybrids, including (panel B) and excluding (panel C) *S. cerevisiae* strains. Bar plots are colored according to species.

Illumina	Nanopore	PacBio
<p>Cheap</p> <p>High throughput (Population genomics, GWAS)</p> <p>Low error rate (<1%)</p> <p>Better recovery of binding sites within the telomeric repeats</p> <p>Short reads (150-300 bp)</p> <p>Long run (1-3.5 days)</p> <p>Limited <i>de novo</i> genome assembly</p> <p>Poor coverage of subtelomeric regions</p> <p>No coverage of long repetitive regions (i.e., Ty elements, rDNA cluster, repeated genes)</p> <p>mtDNA usually not assembled or analyzed</p>	<p>The fastest run (1 hour)</p> <p>Mostly complete <i>de novo</i> assembly (Comparative Genomics)</p> <p>Detects structural variation</p> <p>1D² chemistry with low error rate (<4%)</p> <p>Recovery of most subtelomeric regions</p> <p>Recovery of long repetitive regions</p> <p>Study of DNA methylation (5mC)</p> <p>High error rate of 1D chemistry (~13%)</p> <p>Limited coverage of subtelomeric regions with excessive repeats</p> <p>Homomer issues</p> <p>1D² chemistry reduces throughput by half</p> <p>Generates duplicated regions in mtDNA (using standard assembly methods)</p>	<p>Fast (10 hours)</p> <p>Mostly complete <i>de novo</i> assembly (Comparative Genomics)</p> <p>Detects structural variation</p> <p>Long reads in CLR method (30-250 Kb)</p> <p>CC method with low error rate (<1%)</p> <p>Recovery of most subtelomeric regions</p> <p>Recovery of long repetitive regions</p> <p>Study of DNA methylation (4mC and 6mA)</p> <p>Better recovery of mtDNA</p> <p>High error rate of CLR method (~13%)</p> <p>Shorter reads in CC method (25 Kb)</p> <p>Generates duplicated regions in mtDNA</p>

Figure 2. Pros and cons of the genome sequencing methods that are currently used most widely. Pros and cons of technologies used for *de novo* genome assembly and population genomics (Goodwin et al. 2015; Chen et al. 2017; Giordano et al. 2017; Istace et al. 2017). CLR: continuous long read; CC, circular consensus; 4mC, N4-methylcytosine; 5mC, 5-methylcytosine; 6 mA, N6-methyladenine; Kb, kilobase.

data sets, new long-read technologies are beginning to unravel the differences in Ty and other repeat content between different *Saccharomyces* strains and species (Istace et al. 2017; Yue et al. 2017; Czaja et al. 2019), including their contribution to differences in genome size between *S. cerevisiae* and *S. paradoxus* genomes (Yue et al. 2017; Czaja et al. 2019). Assembly of subtelomeric regions has also benefited from combining short-read and long-read data. A recent study comparing the evolutionary

dynamics of subtelomeric genes found that the length of subtelomeric regions to vary greatly (0.13–76 Kb with 0–19 genes) and demonstrated an accelerated rate of evolution in domesticated *S. cerevisiae* strains compared to wild *S. paradoxus* isolates (Yue et al. 2017). Important traits for environmental adaptations and phenotypic diversification can now be detected among subtelomeric structural variants (which can also be important in speciation), the copy number variants can now be quantified and

<u>Wild</u>	<u>Domesticated (<i>S. cerevisiae</i>)</u>
<p>Asian origin Geographic populations Diploids Structural variants between lineages and species CNVs <i>HXT/FLO/PAU/COS/THI/YRF: Saccharomyces</i> <i>GAL80/YML020W/IRNAs: Saccharomyces</i> AQY2: NA <i>Scer</i></p> <p>Aneuploidies Allele variants: <i>GAL: Skud</i> AQYs: NA <i>Scer</i> <i>MAL: Seub</i></p> <p>Low heterozygosity (<0.1%) Positive selection <i>ARO4/DAL3 Skud</i> Non-reference genes Resistance to antifungals</p> <p>Low frequency of admixture <i>Scer</i>: West African x wine <i>Spar</i>: SpC*: 94.2% SpC x 5.8% SpB SpD: 50% SpC* x 50% SpB <i>Suva</i>: SA-A x SA-B <i>Seub</i>: 52% PB1 x 48% PA2</p> <p>Nuclear introgressions <i>Spar</i> → <i>Scer</i> <i>Spar</i> → <i>Sjur</i> <i>Skud</i> → <i>Suva</i>: nitrogen metabolism.</p> <p>Mitochondrial introgressions 2μ plasmid interspecies transfers Evolve by accumulation of SNPs Hybridization Brazil strains: NA <i>Scer</i> x SpB <i>Spar</i></p> <p>HGTs <i>CRG1 Lthe</i> → <i>Smik</i></p>	<p>Multiple domestications Isolation source populations Polyploid: beer, admixed, African palm wine, ADY, Milk Reciprocal translocations <i>SSU1</i>: wine</p> <p>CNVs <i>MAL/HXT/FLO</i>: Lager <i>YRF1/PAU/DUP/ARR</i>: wine <i>CUP1</i>: wine, flor, rum <i>IMAs/MALs/SUC2</i>: bread ADHs: wine <i>DLD3</i>: cheese <i>ENA</i>: olive brine <i>GAL2</i>: milk, cheese ↑Ty elements ↑ORFs B region</p> <p>Aneuploidies Low diacetyl in Lager: ↑chrX, XII Higher flocculation in Lager: ↑chrI Sake, ale beer, admixed, Mantou Ethanol resistance: ↑chrXII, III</p> <p>Heterozygosity High (>0.16%): Beer1, baking, milk, mantou, ADY, Qingkeju, Baijiu Low: wine, sake</p> <p>Positive selection <i>HXT7/YPS6/MTS27</i>: flor <i>IRA2/ATG19/AVT3/BAP2</i>: wine <i>FUR4/SIC1</i>: cheese</p> <p>Allele variants <i>MAL11</i>: beer AQYs: wine, beer, sake <i>FDC1/PAD1</i>: Beer1 and Beer2 (POF), except Wheat beer (POF*) <i>BIO1/BIO6/KHR1/EHL/GNAT</i>: Sake, industrial <i>GAL3/GAL4</i>: Sake <i>GAL/MEL</i>: <i>Scer</i> and hybrids <i>FLO11/ZRT1</i>: flor HXTs: cheese</p> <p>High frequency of admixture Laboratory: wine x lab/clinical Rum: 44% NA x 56% wine Beer: two episodes 60% wine x 40% sake</p> <p>Introgressions <i>Spar</i> → wine, French Guiana, Mexican Agave, Alpechin <i>Sacc</i> → wine/ale beer</p> <p>Mitochondrial introgressions Interspecies hybrids Lager: Ale <i>Scer</i> x HOL <i>Seub</i> Ale beer: Ale <i>Scer</i> x EU <i>Skud</i> Wine: wine <i>Scer</i> x <i>Skud</i> Cider: wine <i>Scer</i> x <i>Suva</i> Muri: English Ale <i>Scer</i> x HOL <i>Suva</i> Olives: wine <i>Scer</i> x EU <i>Spar</i></p> <p>HGTs A region: non-<i>Sacc</i> → wine, flor B region: <i>Zbai</i> → wine, flor C region (<i>FOT1</i>): <i>Tdel</i> → wine, flor: nitrogen-limited</p>
	<u>Pathogenic/Clinical</u>
	<p>Aneuploidies ↑chr IX ↑chr III (heat), ↑chrV (high pH), ↑chrXIII (4-nitroquinoline 1-oxide)</p> <p>Allele variants <i>PDR1/PDR3/PDR5</i> <i>ICT1/YOR1/GRE2/PDR16/YGR035C/YPL088W</i> <i>ERG3/NCS2/MKT1/END3/RHO2</i> <i>SNZs/SNOs</i></p> <p>High mutation rate 46% of isolates heterozygous (0.16-0.65%) Admixture: wine x wine Introgressions <i>Spar</i> → <i>Scer</i></p> <p>Frequent polyploidy Hybrids: <i>Scer</i> x <i>Skud</i> HGTs <i>GNAT: Efae</i></p>

Figure 3. Genomic traits of wild, pathogenic, and domesticated *Saccharomyces* yeasts. Main genomic trait differences inferred from whole genome sequencing studies between wild *Saccharomyces*, domesticated, and clinical *S. cerevisiae* strains (Table S1, Supporting Information). Heterozygosity is represented as the percentage of heterozygous sites in the genome. Arrows (→) indicate the introgression/HGT direction inferred. Arrow (↑) indicates an increase in copy number. ADY, active dry yeast; CNVs, copy number variants; HGT, horizontal gene transfer; LOF, loss of function; POF, phenolic off-flavor; SNPs, single nucleotide polymorphisms; *Sacc*, *Saccharomyces*; *Scer*: *S. cerevisiae*; *Spar*, *S. paradoxus*; *Sjur*, *S. jurei*; *Suva*, *S. uvarum*; *Seub*, *S. eubayanus*; *Efae*, *Enterococcus faecium*; *Tmic*, *Torulospira microellipsoides*; *Zbai*, *Zygosaccharomyces bailii*; *Lthe*, *Lachancea thermotolerans*; PB, Patagonia B; PA, Patagonia A; NA, North America; HOL, Holarctic.

localized (e.g. those observed in the *CUP1* gene and *ARR* cluster) and the presence and absence of metabolic genes can be accurately assessed (McIlwain et al. 2016; Yue et al. 2017; Naseeb et al. 2018; Steenwyk and Rokas 2018).

Mitochondrial genome sequence assemblies have also been missing from most Illumina sequencing studies, except in a handful studies (Baker et al. 2015; Wu, Buljic and Hao 2015; Sulo et al. 2017). In contrast, long-read technologies better capture and facilitate the assembly of mitochondrial genome sequences (Wolters, Chiu and Fiumera 2015; Giordano et al. 2017; Yue et al. 2017). Similarly, despite the fitness disadvantages of possessing the 2μ plasmids (1.5%–3% growth rate disadvantage compared to cured cells) (Mead, Gardner and Oliver 1986), few genome sequencing studies explicitly comment about the recovery of 2μ plasmid sequences (Baker et al. 2015; Strobe et al. 2015; McIlwain et al. 2016; Peter et al. 2018).

Genomic differences among wild, pathogenic and domesticated *Saccharomyces* strains

The importance of *S. cerevisiae* for a multitude of industrial processes, such as making wine, ale beers, biofuels, sake and bread, has greatly influenced genome sequencing efforts, including in other *Saccharomyces* species. Indeed, more 2500 *S. cerevisiae* strains have been sequenced, including many that were independently sequenced by different labs (Fig. 1). These efforts have helped differentiate the genome characteristics of wild, pathogenic/clinical and domesticated *S. cerevisiae* strains

(Fig. 3). However, it has also been necessary to increase isolation efforts of other *Saccharomyces* species to generalize the genomic traits found in wild *S. cerevisiae* strains to other species where all or most known strains are wild. In contrast to wild strains, pathogenic/clinical strains and domesticated strains are both associated with humans. Wild and human-associated strains differ for several genomic characteristics: (i) low heterozygosity in wild isolates, suggesting high-inbreeding rates (Magwene et al. 2011; Wohlbach et al. 2014; Leducq et al. 2016; Peris et al. 2016; Duan et al. 2018; Naseeb et al. 2018; Peter et al. 2018; Nespolo et al. 2019; Langdon et al. 2019b); (ii) fewer admixed strains from the wild, supporting low levels of outcrossing (Liti et al. 2009; Almeida et al. 2014; Leducq et al. 2016; Peris et al. 2016; Eberlein et al. 2019); (iii) the rarity of wild interspecies hybrids [currently only one is known (Barbosa et al. 2016)], suggesting limited opportunities or low fitness for interspecies hybrids in wild environments (Fig. 3); (iv) strong geographic structure of wild *Saccharomyces* populations (Hittinger et al. 2010; Almeida et al. 2014; Gayevskiy 2015; Leducq et al. 2016; Peris et al. 2016; Duan et al. 2018; Peter et al. 2018), which highlights the limited influence of humans on the expansion of wild strains and (v) more copy number variants (CNVs), especially in subtelomeric genes and more aneuploidies in domesticated lineages (Gallone et al. 2016; Gonçalves et al. 2016; Steenwyk and Rokas 2017). In addition, wild strains have evolved mainly by accumulating SNPs, whereas domesticated and clinical samples are more prone to Ty element and gene family expansions (Peter et al. 2018). However, there are common genomic characteristics among wild and

human-associated strains: (i) 75% genes not found in a reference genome are located in subtelomeric regions and are often related to flocculation, nitrogen metabolism, carbon metabolism and stress (Bergström et al. 2014; Steenwyk and Rokas 2017); (ii) subtelomeric regions are hotspots of gene diversity, which influences traits (McIlwain et al. 2016; Yue et al. 2017) and (iii) loss-of-function (LOF) mutations usually occur in non-essential genes and are more frequently found in regions closer to the 3' end of protein-coding sequences (Bergström et al. 2014).

Several HGT events have been described in *Saccharomyces*, including several specific examples that are well supported (Fig. 3) (Hall and Dietrich 2007; Novo et al. 2009; Galeote et al. 2010; League, Slot and Rokas 2012; Marsit et al. 2015; Peter et al. 2018). Nonetheless, caution is warranted for cases built solely using automated BLAST analysis, which can lead to premature conclusions for two main reasons. First, the absence of published genome sequences for most species make the unequivocal identification of donor and recipient species or clades challenging. Second, gene presence and absence variation of a horizontally acquired gene within or between species can mislead the inference of the history of a gene if population or species sampling is insufficient or biased. For example, large gene families found in subtelomeric regions are particularly prone to being identified as involved in HGT events using simple BLAST criteria due to cryptic paralogy. In these cases, the fact that a gene's best BLAST hit is to a distant species may just be due to missing data. For these reasons, we recommend using BLAST-based statistics, such as Alien Index (Alexander et al. 2016; Wisecaver et al. 2016), to identify interesting candidates, followed by explicit gene tree-species tree reconciliation and phylogenetic topology testing to evaluate candidate HGT events (Alexander et al. 2016; Wisecaver et al. 2016; Shen et al. 2018). Furthermore, the identification of HGT events, as well as more accurate identification of donors and recipients, will greatly benefit from the completion of comprehensive whole genome sequencing projects from diverse species, such as the Y1000 + Project (Hittinger et al. 2015; Shen et al. 2018). In summary, a combination of improved genome sampling and formal phylogenetic approaches together provides the best path forward to generating robust inferences about which genes have been horizontally acquired.

Genomic insights into the fascinating phylogeography of the wild lager-brewing yeast ancestor, *Saccharomyces eubayanus*

The yeast species *S. eubayanus* has been isolated exclusively from wild environments; yet, hybridizations between *S. cerevisiae* and *S. eubayanus* were key innovations that enabled cold fermentation and lager brewing (Libkind et al. 2011; Gibson and Liti 2015; Hittinger, Steele and Ryder 2018; Baker and Hittinger 2019; Mertens et al. 2019; Langdon et al. 2019a). Industrial isolates of *S. uvarum*, the sister species of *S. eubayanus*, with genomic contributions from *S. eubayanus* have also been frequently obtained from wine and cider (Almeida et al. 2014; Nguyen and Boekhout 2017; Langdon et al. 2019a), indicating that this species has long been playing a role in shaping many fermented products. Even so, pure strains of *S. eubayanus* have only ever been isolated from the wild. This association with both wild and domesticated environments makes *S. eubayanus* an excellent model where both wild diversity and domestication can be investigated.

Saccharomyces eubayanus was initially discovered in 2011 in Patagonia (Argentina) from locally endemic tree species of the genus *Nothofagus* (Libkind et al. 2011). Since then, it has received

much attention for brewing applications and as a model for understanding the evolution, ecology and population genomics of the genus *Saccharomyces* (Sampaio 2018). Many new globally distributed isolates have been found in different parts of the world since its discovery (Bing et al. 2014; Peris et al. 2014; Rodríguez et al. 2014; Gayevskiy and Goddard 2016; Peris et al. 2016; Eizaguirre et al. 2018) but the abundance and genetic diversity measured by multilocus genetic data is still by far highest in Patagonia (Eizaguirre et al. 2018). Recently, two independent investigations significantly increased the number of *S. eubayanus* American isolates, mainly from Patagonia (Chile and Argentina), and together provide the largest genomic data set for this species with a total of 256 new draft genome sequences (Nespolo et al. 2019; Langdon et al. 2019b). This data set confirms the previously proposed population structure (Peris et al. 2014, 2016; Eizaguirre et al. 2018), where two major populations were detected (Patagonia A/Population A/PA and Patagonia B/Population B/PB), which has been further divided into five subpopulations (PA-1, PA-2, PB-1, PB-2 and PB-3) (Eizaguirre et al. 2018). Other isolates from outside Patagonia belong to PB, either the PB-1 subpopulation that is also found in Patagonia (Gayevskiy and Goddard 2016; Peris et al. 2016), or a Holarctic-specific subpopulation (PB-Holarctic) that includes isolates from Tibet and from North Carolina, USA (Bing et al. 2014; Peris et al. 2016; Brouwers et al. 2019), which represents the closest known wild relatives of the *S. eubayanus* subgenomes of lager-brewing yeasts (Bing et al. 2014; Peris et al. 2016). Furthermore, heterosis was recently demonstrated in a *S. cerevisiae* x Tibetan *S. eubayanus* hybrid, which showed that regulatory cross talk between the two subgenomes is partly responsible for maltotriose and maltose consumption (Brouwers et al. 2019). Multilocus data suggested that two more lineages from China, West China and Sichuan, diverged very early from all other known *S. eubayanus* strains, while Holarctic isolates from China had unusually low sequence diversity (Bing et al. 2014). In this way, *S. eubayanus* can be subdivided into a total of eight non-admixed subpopulations (six likely Patagonian—2 PA, 3 PB and 1 PB-Holarctic and 2 Asian—1 West China and 1 Sichuan) and two admixed lineages (one North American lineage with a broad distribution and South American strain sympatric to the Patagonian lineages) (Langdon et al. 2019b). The global distribution and geographically well-differentiated population structure of *S. eubayanus* is similar to what has been observed for *Saccharomyces* species, such as *S. paradoxus* (Leducq et al. 2014, 2016) and *S. uvarum* (Almeida et al. 2014).

While this species has been easily and repeatedly isolated from South American *Nothofagus* trees (Libkind et al. 2011; Eizaguirre et al. 2018; Nespolo et al. 2019), only a handful of isolates have been recovered from trees in China, New Zealand and North America (Bing et al. 2014; Peris et al. 2014; Gayevskiy and Goddard 2016; 2016). These data suggest that *S. eubayanus* is abundant in Patagonia but sparsely found in North America, Asia and Australasia. Most subpopulations display isolation by distance with genetic diversity that mostly scales with the geographic range of a subpopulation. In Patagonia, one sampling location can harbor more genetic diversity than is found in all of North America (Langdon et al. 2019b). The levels of diversity found within Patagonia is further underscored by the restriction of four subpopulations to this region, suggesting that Patagonia is the origin of *S. eubayanus* diversity or at least the last common ancestor of the PA and PB-Holarctic populations, the latter of which gave rise to lager-brewing hybrids. Different hypotheses and scenarios are discussed in more depth by Langdon et al. (2019b) and Nespolo et al. (2019).

NON-CONVENTIONAL YEASTS WITH NON-CONVENTIONAL GENOMES

Besides the well-studied genus *Saccharomyces*, more than 1500 recognized yeast species are known, which belong either to the Ascomycota or Basidiomycota (Kurtzman, Fell and Boekhout 2011). In this section, we review the interesting stories recently revealed through the use of genome data of two representative genera of both respective phyla, *Hanseniaspora* and *Phaffia*.

The yeasts with the least; the reductive genome evolution of *Hanseniaspora*

A hallmark of evolution in the budding yeast subphylum Saccharomycotina is the loss of traits and their underlying genes (Shen et al. 2018). Arguably, the most dramatic example of reductive evolution observed is the *Hanseniaspora* (Steenwyk et al. 2019), a genus of bipolar budding, apiculate yeasts in the family Saccharomycodaceae. *Hanseniaspora* yeasts can be assigned to two lineages, a faster-evolving one and a slower-evolving one (FEL and SEL, respectively), which differ dramatically in their rates of genome sequence evolution as well as in the extent and types of genes that they have lost (Fig. 4). The types of genes lost can be broadly ascribed to three categories: metabolism, DNA repair, and cell-cycle.

Metabolism-related genes have been lost in both FEL and SEL. Analysis of 45 growth traits across 332 Saccharomycotina yeasts revealed that *Hanseniaspora* species can assimilate fewer carbon substrates compared to most of their relatives (Opulente et al. 2018; Shen et al. 2018) and have lost many of the associated genes and pathways (Steenwyk et al. 2019). Although less pronounced, similar gene and trait losses have been observed in wine strains of *S. cerevisiae* (Gallone et al. 2016; Steenwyk and Rokas 2017) and are thought to be signatures of adaptation to the wine must environment (Steenwyk and Rokas 2018). These gene losses may play a similar role in the ecology of *Hanseniaspora* species, considering their frequent isolation from fruit juices and fermenting musts (Cadez 2006; Kurtzman, Fell and Boekhout 2011), which likely reflects the specialization of *Hanseniaspora* species to sugar-rich environments.

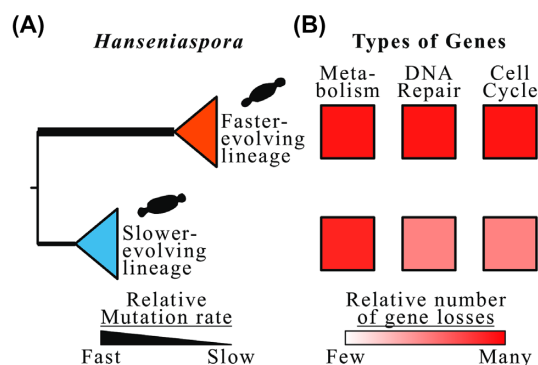


Figure 4. The evolutionary trajectories of *Hanseniaspora* lineages are marked by differential rates of sequence evolution and rates of loss of metabolism, DNA repair and cell-cycle genes. (A), There are two lineages in the budding yeast genus *Hanseniaspora*: the faster-evolving and slower-evolving lineage (FEL and SEL, respectively). The FEL has a long and thicker stem branch indicative of higher rates of sequence evolution or higher mutation rates, whereas the SEL has a much shorter and thinner stem branch indicative of lower rates of sequence evolution or lower mutation rates. (B), Each lineage has lost many genes associated with metabolism, DNA repair and cell-cycle processes; squares with colors toward the red end of the spectrum correspond to greater rates of gene loss, whereas squares on the white end of the spectrum correspond to lower rates of gene loss.

Hanseniaspora species, especially those in the FEL, have lost numerous DNA repair genes spanning multiple pathways and processes (Steenwyk et al. 2019). For example, yeasts in both lineages have lost 14 DNA repair genes, including PHR1, which encodes a photolyase (Sebastian, Kraus and Sancar 1990) and MAG1, which encodes a DNA glycosylase that is part of the base excision repair pathway (Xiao et al. 2001). However, FEL yeasts have lost 33 additional DNA repair genes, which include polymerases (i.e. POL4 and POL32) and numerous telomere-associated genes, such as CDC13 (Lustig 2001). Inactivation or loss of DNA repair genes can cause hypermutator phenotypes, such as those observed in microbial pathogens and in human cancers (Jolivet-Gougeon et al. 2011; Billmyre, Clancey and Heitman 2017; Campbell et al. 2017). In the short-term, hypermutation can facilitate adaptation in maladapted populations by increasing the chance of occurrence of beneficial mutations (e.g. conferring drug resistance); in the long-term, however, hypermutation is not a viable strategy due to the increased accumulation of deleterious mutations (Ram and Hadany 2012). Molecular evolutionary analyses suggest that the stem lineages of FEL and SEL yeasts were hypermutators; interestingly, the increased mutation rates in the two stem lineages reflect the degree of observed DNA repair gene loss in the two lineages. The larger number of gene losses in FEL stem branch is consistent with its higher mutation rate and the smaller number of gene losses in the SEL stem branch is consistent with a lower increase in its mutation rate (Steenwyk et al. 2019). However, the mutation rates of both FEL and SEL crown groups (i.e. every branch after the stem) are similar to those of other yeast lineages, consistent with evolutionary theory's predictions that long-term hypermutation is maladaptive (Ram and Hadany 2012; Steenwyk et al. 2019). Altogether, *Hanseniaspora* have lost DNA repair genes, undergone punctuated sequence evolution, and slowed down their overall mutation rate, despite having a reduced DNA repair gene repertoire. Finally, *Hanseniaspora* yeasts have lost genes associated with key features of the cell cycle, including cell size control, the mitotic spindle checkpoint and DNA-damage-response checkpoint processes, but these losses are more pronounced in the FEL. For example, both lineages have lost WHI5, a negative regulator of the G1/S phase transition in the cell cycle that is critical for cell size control (Jorgensen et al. 2002). Other gene losses are exclusive to the FEL, such as the loss of MAD1 and MAD2, which bind to unattached kinetochores and are required for a functional mitotic spindle checkpoint (Heinrich et al. 2014), as well as RAD9 and MEC3, which function in the DNA-damage-checkpoint pathway and arrest the cell cycle in G2 (Weinert, Kiser and Hartwell 1994). The loss of checkpoint genes is thought to contribute to bipolar budding in both lineages and greater variance in ploidy, as well as strong signatures of mutational burden due to aberrant checkpoint processes in FEL compared to SEL (Steenwyk et al. 2019). These observations suggest landmark features of cell cycle processes are absent in *Hanseniaspora* and warrant future investigations into the functional consequences of these losses.

Phaffia rhodozyma: A colorful genome from the Basidiomycota

The orange-colored yeast *Phaffia rhodozyma* (= *Xanthophylomyces dendrorhous*), an early diverging Agaricomycotina (Basidiomycota), possesses multiple exceptional traits of fundamental and applied interest. The most relevant is the ability to synthesize astaxanthin, a carotenoid pigment with potent

antioxidant activity and of great value for the aquaculture and pharmaceutical industries. Hyperpigmented mutants of *P. rhodozyma* are currently being exploited biotechnologically as a natural source of astaxanthin in aquaculture feed (Rodríguez-Sáiz, De La Fuente and Barredo 2010). These mutants were derived from an initial collection from 1976 from bark exudates of specific tree species (e.g. *Betula* sp.) from the Northern Hemisphere. Today, *P. rhodozyma* is known to have specific niches in association with trees of mountainous regions and a worldwide distribution comprising at least seven different genetic lineages (David-Palma, Libkind and Sampaio 2014). One of these lineages was obtained from Andean Patagonia (Argentina) on *Nothofagus* trees, the same substrates as *S. eubayanus* and *S. uvarum* (Section 3.2) (Libkind et al. 2011), and based on genomic analyzes, Patagonian wild strains were recently proposed as a potential novel variety of *P. rhodozyma* (Bellora et al. 2016). The 19-Mb genome of *P. rhodozyma* CRUB 1149 wild Patagonian isolate was sequenced and assembled, achieving a coverage of 57x. Analysis of its gene structure revealed that the proportion of intron-containing genes and the density of introns per gene in *P. rhodozyma* are the highest hitherto known for fungi, having values more similar to those found in humans than among *Saccharomycotina* where intronless genes predominate. An extended analysis suggested that this trait might be shared with other members of the order Cystofilobasidiales (Bellora et al. 2016).

Genome mining revealed important photoprotection and antioxidant-related genes, as well as genes involved in sexual reproduction. New genomic insight into fungal homothallicism was obtained, including a particular arrangement of the mating-type genes that might explain the self-fertile sexual behavior. All known genes related to the synthesis of astaxanthin were annotated. Interestingly, a hitherto unknown gene cluster potentially responsible for the synthesis of an important UV protective and antioxidant compound (mycosporine-glutaminol-glucoside) (Moliné et al. 2011) was found in the newly sequenced and mycosporinogenic strain. However, this gene cluster was absent in a strain (CBS 6938) that does not accumulate this secondary metabolite, which has potential applications in cosmetics (Colabella and Libkind 2016). Genome mining also revealed an unexpected diversity of catalases and the loss of H₂O₂-sensitive superoxide dismutases in *P. rhodozyma*. Altogether, the *P. rhodozyma* genome is enriched in antioxidant mechanisms, in particular those most effective at coping with H₂O₂, suggesting that the environmental interaction with this reactive species has definitely contributed to shaping the peculiar genome of *P. rhodozyma*.

YEAST BIOTECHNOLOGY GETS WILD WITH GENOMICS

The identification of new yeast strains and novel species could offer valuable innovative opportunities for applied research by taking advantage of traits found by bioprospecting in extreme environments (Pretscher et al. 2018; Cubillos et al. 2019). Newly isolated yeasts are expanding the repertoire of phenotypic diversity, and therefore, the current known variation in physiological and metabolic traits. These yeasts from extreme environments are of considerable interest in biotechnology, owing to diverse advantages, such as: rapid growth rates at extreme temperatures, (Choi, Park and Kim 2017; Yuivar et al. 2017; Cai, Gao and Zhou 2019), extraordinary capacity of fermentation in large-scale cultures (Choi, Park and Kim 2017; Krogerus et al. 2017) and the production of cold-active hydrolytic enzymes (such as lipases, proteases, cellulases and amylases) (Martorell et al.

2019). For example, the cryotolerant yeast *S. eubayanus* exhibits a wide set of relevant traits appropriate for brewing, including efficient biomass production at low temperature and production of high levels of esters and preferred aroma compounds in beer (Libkind et al. 2011; Hebly et al. 2015; Mertens et al. 2015; Alonso-del-Real et al. 2017; Gibson et al. 2017; Krogerus et al. 2017). Similarly, an Antarctic isolate of *Wickerhamomyces anomalus* has been indicated as a high producer and secretor of glucose oxidases, invertases and alkaline phosphatases enzymes at lower temperatures, decreasing the temperature requirement for their production (Schlander et al. 2017; Yuivar et al. 2017). In this context, the availability of new yeasts as biological and genetic resources from the wild immediately opens new avenues, not only for their direct utilization in industrial processes, but also to gather and obtain new genomic data so that their genes can be integrated into complex industrial systems already in use. However, the use and manipulation of these genetic resources are restricted by the limited knowledge in terms of the molecular basis underlying metabolic traits of industrial interest. Mining this genomic and phenotypic diversity provides a great opportunity to pinpoint unique pathways of biotechnological importance, which can then be exported to other systems or improved within the same genetic backgrounds. Recent advances in bioinformatics, quantitative genetics, systems biology and integrative biology, together with the large number of new genome sequencing projects are providing the means to address these challenges (Liti 2015; Peter et al. 2018; Viigand et al. 2018; Cai, Gao and Zhou 2019; Nespolo et al. 2019; Langdon et al. 2019b). Thus, leveraging wild yeast genomes, together with other 'multi-omic' approaches can generate possible targets for biotechnological applications.

Genomics can support predicting biochemical traits in organisms with biotechnological potential, where the combination of comparative genomic and physiological studies can allow key genomic features to be inferred in non-conventional organisms (Riley et al. 2016). Furthermore, efforts to unravel the complexity of yeast genomes have proven successful in providing genome-scale models that can determine their potential metabolic profiles (Loira et al. 2012; Lopes and Rocha 2017). These models can be applied to new yeast genomes to predict an organism's chemical repertoire by reconstructing metabolic pathways and elucidating their biotechnological potential (Wang et al. 2017). Thus far, these approaches have been successfully applied to a subset of strains in model yeasts, such as *Yarrowia lipolytica* (Loira et al. 2012), *S. cerevisiae* (Heavner and Price 2015; Müllleder et al. 2016), and *Komagataella phaffii* (formerly known as *Pichia pastoris*) (Saitua et al. 2017). Their utilization in novel organisms is still in its infancy, but the integration of transcriptional regulatory networks and metabolic networks could guide novel metabolic engineering applications (Shen et al. 2019) to convert new yeasts (strains or species) into potential resources for the production of biofuels and biochemicals.

Biotechnological applications in non-conventional organisms are poised to be enhanced by recent advances in genome-editing techniques, such as CRISPR-Cas9 (Donohoue, Barrangou and May 2018). The utilization of CRISPR-Cas9 requires whole genome sequences so that gRNAs can be designed to specifically target genes of interest. This system is highly effective in *S. cerevisiae* and other *Saccharomyces* species, mostly due to their efficient homology-directed DNA repair machinery (Akhmetov et al. 2018; Kuang et al. 2018; Mertens et al. 2019). For example, novel *S. eubayanus* strains recently isolated from Patagonia (Rodríguez et al. 2014) were successfully engineered for the lower production of phenolic off-flavors (Mertens et al. 2019).

Interestingly, high success rates have also been reported in other non-conventional yeasts, demonstrating the large spectrum of genomes that can be modified using the CRISPR-Cas9 system (Wang et al. 2017; Juergens et al. 2018; Kuang et al. 2018; Cai, Gao and Zhou 2019; Lombardi, Oliveira-Pacheco and Butler 2019; Maroc and Fairhead 2019). For example, CRISPR-Cas9-assisted multiplex genome editing (CMGE) in the thermotolerant methylotrophic yeast *Ogataea polymorpha* allowed for the introduction of all the genes necessary for the biosynthesis of resveratrol, along with the biosynthesis of human serum albumin and cadaverine (Wang et al. 2017). The seemingly universal capacity of the CRISPR-Cas9 genome-editing technique means that many, if not all, yeasts will ultimately be susceptible to being modified using this system. Thus, even newly isolated yeasts and novel species could be used as microbial cell factories, allowing the spectrum of applications and products to be expanded.

CONCLUSIONS

The power of genomics in the study of yeast biology, evolution and biotechnology is highly dependent on the number of genome sequences available, and this factor is currently the main limitation for comprehensive studies. So far, studies have focused mostly on model species or taxa of specific fundamental or applied interest, mainly for ascomycetous yeasts. In contrast, few projects have dealt with basidiomycetous yeast genomes, many of which also likely harbor interesting characteristics. The description of novel species based on complete genome sequences is still not a trend among yeast taxonomists, probably due in part to cost and due in part to the lack of general guidelines for this practice. A review included in this issue represents the first attempt to establish minimal advice for taxonomic descriptions using whole genome sequence data for the formal descriptions of novel yeast species (Libkind et al. 2020 submitted). As this practice becomes more widespread and the genomic database for non-conventional yeasts grows, our ability to answer different biological questions about their history, ecological adaptations and dynamics will increase. Even so, new bioinformatic tools that are more user-friendly and automated will make the power of genomics more accessible to researchers without bioinformatic training. On the technological side, the gradual increase in the use of long-read sequencing technologies will enable the exploration of complete or near-complete genome assemblies, including repeats and telomeres, of non-conventional yeasts.

Here, we provided clear examples of how our understanding of many biological and evolutionary processes has been improved by widening the spectrum of yeasts studied, especially by including non-conventional yeasts from the wild. Emblematic cases from the anthropogenically affected genus *Saccharomyces* were addressed as an example of how genomics helped to cast light into complex microbial domestication processes and to detect genomic signatures of pathogenicity and domestication. This insight would not have been possible if large genomic data sets from wild isolates of *S. cerevisiae* were not available. Similarly, the previously missing wild ancestor of lager-brewing yeasts would have not been found if yeast explorations into pristine and remote environments had not been carried out. Studies in the less known genus *Hanseniaspora*, including both domesticated and wild strains, revealed unexpected evolutionary histories, with surprising and interesting modes of genome evolution. The basidiomycetous yeast *Phaffia rhodozyma* provided an illustrative example of the unique genomic traits that can be found within this understudied phylum. In the future, the large

number of new yeast genomes, along with transcriptomic, proteomic and other multi-omic studies, will rapidly improve our understanding of non-conventional and indeed all organisms at the systems level.

SUPPLEMENTARY DATA

Supplementary data are available at [FEMSYR](https://femsyr.onlinelibrary.wiley.com/doi/10.1111/femsyr.10000) online.

ACKNOWLEDGMENTS

DL has been funded through CONICET (PIP11220130100392CO) and Universidad Nacional del Comahue (B199). Research in AR's lab has been funded through a National Science Foundation grant (DEB-1442113); JLS and AR have also received funding by the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study program. CTH has been funded through the National Science Foundation (DEB-1442148), USDA National Institute of Food and Agriculture (Hatch Project 1020204), and DOE Great Lakes Bioenergy Research Center (DE-SC0018409). CTH is a Pew Scholar in the Biomedical Sciences and a H. I. Romnes Faculty Fellow, supported by the Pew Charitable Trusts and Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation, respectively. DP is a Marie Skłodowska-Curie fellow of the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 747775).

Conflict of interest. None declared.

REFERENCES

- Akhmetov A, Laurent J, Gollihar J et al. Single-step precision genome editing in yeast using CRISPR-Cas9. *Bio-Protocol* 2018;8:e2765.
- Alexander WG, Wisecaver JH, Rokas A et al. Horizontally acquired genes in early-diverging pathogenic fungi enable the use of host nucleosides and nucleotides. *Proc Natl Acad Sci USA* 2016;113:4116–21.
- Almeida P, Gonçalves C, Teixeira S et al. A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. *Nat Commun* 2014;5:4044.
- Alonso-del-Real J, Lairón-Peris M, Barrio E et al. Effect of temperature on the prevalence of *Saccharomyces non cerevisiae* species against a *S. cerevisiae* wine strain in wine fermentation: competition, physiological fitness, and influence in final wine composition. *Front Microbiol* 2017;8:4044.
- Baker ECP, Hittinger CT. Evolution of a novel chimeric maltotriose transporter in *Saccharomyces eubayanus* from parent proteins unable to perform this function. *PLoS Genet* 2019;15:e1007786.
- Baker ECP, Wang B, Bellora N et al. The genome sequence of *Saccharomyces eubayanus* and the domestication of lager-brewing yeasts. *Mol Biol Evol* 2015;32:2818–31.
- Barbosa R, Almeida P, Safar SVB et al. Evidence of natural hybridization in Brazilian wild lineages of *Saccharomyces cerevisiae*. *Genome Biol Evol* 2016;8:317–29.
- Bellora N, Moliné M, David-Palma M et al. Comparative genomics provides new insights into the diversity, physiology, and sexuality of the only industrially exploited tremellomycete: *Phaffia rhodozyma*. *BMC Genomics* 2016;17:901.
- Bergström A, Simpson JT, Salinas F et al. A high-definition view of functional genetic variation from natural yeast genomes. *Mol Biol Evol* 2014;31:872–88.

- Billmyre RB, Clancey SA, Heitman J. Natural mismatch repair mutations mediate phenotypic diversity and drug resistance in *Cryptococcus deuterogattii*. *Elife* 2017;6:e28802.
- Bing J, Han PJ, Liu WQ et al. Evidence for a far east asian origin of lager beer yeast. *Curr Biol* 2014;24:R380–1.
- Brouwers N, Brickwedde A, Vries ARG de et al. Maltotriose consumption by hybrid *Saccharomyces pastorianus* is heterotic and results from regulatory cross-talk between parental sub-genomes. *bioRxiv* 2019;1:679563.
- Cadez N. Phylogenetic placement of *Hanseniaspora-Kloeckera* species using multigene sequence analysis with taxonomic implications: descriptions of *Hanseniaspora pseudoguilliermondii* sp. nov. and *Hanseniaspora occidentalis* var. *citrica* var. nov. *Int J Syst Evol Microbiol* 2006;56:1157–65.
- Cai P, Gao J, Zhou Y. CRISPR-mediated genome editing in non-conventional yeasts for biotechnological applications. *Microb Cell Fact* 2019;18:63.
- Campbell BB, Light N, Fabrizio D et al. Comprehensive analysis of hypermutation in human cancer. *Cell* 2017;171:1042–56 e10.
- Campbell MS, Holt C, Moore B et al. Genome Annotation and Curation Using MAKER and MAKER-P. *Curr Protoc Bioinforma* 2014;48:4.11.1–4.11.39.
- Carver T, Bleasby A. The design of Jemboss: a graphical user interface to EMBOSS. *Bioinformatics* 2003;19:1837–43.
- Chen Q, Lan C, Zhao L et al. Recent advances in sequence assembly: Principles and applications. *Brief Funct Genomics* 2017;16:361–78.
- Choi DH, Park EH, Kim MD. Isolation of thermotolerant yeast *Pichia kudriavzevii* from nuruk. *Food Sci Biotechnol* 2017;26:1357–62.
- Chun J, Oren A, Ventosa A et al. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol Microbiol* 2018;68:461–6.
- Chun J, Rainey FA. Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. *Int J Syst Evol Microbiol* 2014;64:316–24.
- Colabella F, Libkind D. PCR-based method for the rapid identification of astaxanthin-accumulating yeasts (*Phaffia* spp.). *Rev Argent Microbiol* 2016;48:15–20.
- Cubillos FA, Gibson B, Grijalva-Vallejos N et al. Bioprospecting for brewers: Exploiting natural diversity for naturally diverse beers. *Yeast* 2019;36:383–98.
- Czaja W, Bensasson D, Anh HW et al. Evolution of Ty1 copy number control in yeast by horizontal transfer of a gag gene. *bioRxiv* 2019;1:741611.
- David-Palma M, Libkind D, Sampaio JP. Global distribution, diversity hot spots and niche transitions of an astaxanthin-producing eukaryotic microbe. *Mol Ecol* 2014;23:921–32.
- Donohoue PD, Barrangou R, May AP. Advances in industrial biotechnology using CRISPR-Cas systems. *Trends Biotechnol* 2018;36:134–46.
- Duan SF, Han PJ, Wang QM et al. The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. *Nat Commun* 2018;9:2690.
- Eberlein C, Hénault M, Fijarczyk A et al. Hybridization is a recurrent evolutionary stimulus in wild yeast speciation. *Nat Commun* 2019;10:923.
- Eizaguirre JI, Peris D, Rodríguez ME et al. Phylogeography of the wild Lager-brewing ancestor (*Saccharomyces eubayanus*) in Patagonia. *Environ Microbiol* 2018;20:3732–43.
- Fan H, Ives AR, Surget-Groba Y et al. An assembly and alignment-free method of phylogeny reconstruction from next-generation sequencing data. *BMC Genomics* 2015;16:1–18.
- Galeote V, Novo M, Salema-Oom M et al. FSY1, a horizontally transferred gene in the *Saccharomyces cerevisiae* EC1118 wine yeast strain, encodes a high-affinity fructose/H⁺ symporter. *Microbiology* 2010;156:3754–61.
- Gallone B, Steensels J, Prah T et al. Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* 2016;166:1397–410 e16.
- Gayevskiy V, Goddard MR. *Saccharomyces eubayanus* and *Saccharomyces arboricola* reside in North Island native New Zealand forests. *Environ Microbiol* 2016;18:1137–47.
- Gayevskiy V. The origin, diversity and ancestry of *Saccharomyces* yeasts in New Zealand, PhD Thesis; University of Auckland 2015;1994:173.
- Gibson B, Geertman JMA, Hittinger CT et al. New yeasts-new brews: Modern approaches to brewing yeast design and development. *FEMS Yeast Res* 2017;17.
- Gibson B, Liti G. *Saccharomyces pastorianus*: Genomic insights inspiring innovation for industry. *Yeast* 2015;32:17–27.
- Giordano F, Aigrain L, Quail MA et al. De novo yeast genome assemblies from MinION, PacBio and MiSeq platforms. *Sci Rep* 2017;7:3935.
- Goffeau A, Barrell G, Bussey H et al. Life with 6000 genes. *Science* (80-) 1996;274:546–67.
- Gonçalves M, Pontes A, Almeida P et al. Distinct domestication trajectories in top-fermenting beer yeasts and wine yeasts. *Curr Biol* 2016;26:2750–61.
- Goodwin S, Gurtowski J, Ethe-Sayers S et al. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res* 2015;25:1750–6.
- Gremme G, Steinbiss S, Kurtz S. Genome tools: A comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans Comput Biol Bioinforma* 2013;10:645–56.
- Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res* 2013;41:e129.
- Hall C, Dietrich FS. The reacquisition of biotin prototrophy in *Saccharomyces cerevisiae* involved horizontal gene transfer, gene duplication and gene clustering. *Genetics* 2007;177:2293–307.
- Heavner BD, Price ND. Comparative analysis of yeast metabolic network models highlights progress, opportunities for metabolic reconstruction. Ouzounis CA (ed). *PLoS Comput Biol* 2015;11:e1004530.
- Hebly M, Brickwedde A, Bolat I et al. *S. cerevisiae* × *S. eubayanus* interspecific hybrid, the best of both worlds and beyond. *FEMS Yeast Res* 2015;15:fov005.
- Heinrich S, Sewart K, Windecker H et al. Mad1 contribution to spindle assembly checkpoint signalling goes beyond presenting Mad2 at kinetochores. *EMBO Rep* 2014;15:291–8.
- Hittinger CT, Gonçalves P, Sampaio JP et al. Remarkably ancient balanced polymorphisms in a multi-locus gene network. *Nature* 2010;464:54–8.
- Hittinger CT, Rokas A, Bai FY et al. Genomics and the making of yeast biodiversity. *Curr Opin Genet Dev* 2015;35:100–9.
- Hittinger CT, Steele JL, Ryder DS. Diverse yeasts for diverse fermented beverages and foods. *Curr Opin Biotechnol* 2018;49:199–206.
- Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 2006;23:254–67.
- Istace B, Friedrich A, D'Agata L et al. De novo assembly and population genomic survey of natural yeast isolates with

- the Oxford Nanopore MinION sequencer. *Gigascience* 2017;6: giw018.
- Johnson MG, Gardner EM, Liu Y et al. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl Plant Sci* 2016;4:1600016.
- Jolivet-Gougeon A, Kovacs B, Le Gall-David S et al. Bacterial Hypermutation: Clinical implications. *J Med Microbiol* 2011;60:563–73.
- Jones P, Binns D, Chang H-Y et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014;30: 1236–40.
- Jorgensen P, Nishikawa JL, Breitskreutz BJ et al. Systematic identification of pathways that couple cell growth and division in yeast. *Science (80-)* 2002;297:395–400.
- Juergens H, Varela JA, Gorter de Vries AR et al. Genome editing in *Kluyveromyces* and *Ogataea* yeasts using a broad-host-range Cas9/gRNA co-expression plasmid. *FEMS Yeast Res* 2018;18:foy012.
- Krogerus K, Magalhães F, Vidgren V et al. Novel brewing yeast hybrids: creation and application. *Appl Microbiol Biotechnol* 2017;101:65–78.
- Kuang MC, Kominek J, Alexander WG et al. Repeated cis-regulatory tuning of a metabolic bottleneck gene during evolution. Wittkopp P (ed). *Mol Biol Evol* 2018;35:1968–81.
- Kurtzman CP, Fell JW, Boekhout T. *The Yeasts: A Taxonomic Study*, 5th Edition, 2011.
- Langdon QK, Peris D, Baker ECP et al. Fermentation innovation through complex hybridization of wild and domesticated yeasts. *Nat Ecol Evol* 2019a;3:1576–86.
- Langdon QK, Peris D, Eizaguirre JI et al. Genomic diversity and global distribution of *Saccharomyces eubayanus*, the wild ancestor of hybrid lager-brewing yeasts. *bioRxiv* 2019b:709535.
- Langdon QK, Peris D, Kyle B et al. sppIDer: A Species Identification Tool to Investigate Hybrid Genomes with High-Throughput Sequencing. Rosenberg M (ed.). *Mol Biol Evol* 2018;35:2835–49.
- League GP, Slot JC, Rokas A. The ASP3 locus in *Saccharomyces cerevisiae* originated by horizontal gene transfer from *Wickerhamomyces*. *FEMS Yeast Res* 2012;12:859–63.
- Leducq JB, Charron G, Samani P et al. Local climatic adaptation in a widespread microorganism. *Proc R Soc B Biol Sci* 2014;281: 20132472.
- Leducq JB, Nielly-Thibault L, Charron G et al. Speciation driven by hybridization and chromosomal plasticity in a wild yeast. *Nat Microbiol* 2016;1:15003.
- Libkind D, Hittinger CT, Valefio E et al. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc Natl Acad Sci USA* 2011;108:14539–44.
- Libkind D, Čadež N, Opulente D et al. Yeast taxogenomics: description of novel species based on complete genome sequences. *FEMS Yeast Res* 2020 Submitted (this issue).
- Liti G, Carter D, Moses A et al. Population genomics of domestic and wild yeasts. *Nat Preced* 2009;458:337–41.
- Liti G. The fascinating and secret wild life of the budding yeast *S. cerevisiae*. *Elife* 2015;4:e05835.
- Loira N, Dulermo T, Nicaud J-M et al. A genome-scale metabolic model of the lipid-accumulating yeast *Yarrowia lipolytica*. *BMC Syst Biol* 2012;6:35.
- Lombardi L, Oliveira-Pacheco J, Butler G. Plasmid-based CRISPR-Cas9 gene editing in multiple *Candida* species. *mSphere* 2019;4:e00125–19.
- Lopes H, Rocha I. Genome-scale modeling of yeast: chronology, applications and critical perspectives. *FEMS Yeast Res* 2017;17:fox050.
- Lustig AJ. Cdc13 subcomplexes regulate multiple telomere functions. *Nat Struct Biol* 2001;8:297–9.
- Magwene PM, Kayikçi Ö, Granek JA et al. Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 2011;108:1987–92.
- Maroc L, Fairhead C. A new inducible CRISPR-Cas9 system useful for genome editing and study of double-strand break repair in *Candida glabrata*. *Yeast* 2019;36:723–31.
- Marsit S, Mena A, Bigey F et al. Evolutionary advantage conferred by an eukaryote-to-eukaryote gene transfer event in wine yeasts. *Mol Biol Evol* 2015;32:1695–707.
- Martorell MM, Ruberto LAM, de Figueroa LIC et al. Antarctic yeasts as a source of enzymes for biotechnological applications. *Fungi of Antarctica* Springer International Publishing, 2019;285–304.
- Marçais G, Delcher AL, Phillippy AM et al. MUMmer4: A fast and versatile genome alignment system. Darling AE (ed). *PLoS Comput Biol* 2018;14:e1005944.
- McIlwain SJ, Peris D, Sardi M et al. Genome sequence and analysis of a stress-tolerant, wild-derived strain of *Saccharomyces cerevisiae* used in biofuels research. *G3 Genes, Genomes, Genet* 2016;6:1757–66.
- Mead DJ, Gardner DCJ, Oliver SG. The yeast 2 μ plasmid: strategies for the survival of a selfish DNA. *MGG Mol Gen Genet* 1986;205:417–21.
- Mertens S, Gallone B, Steensels J et al. Reducing phenolic off-flavors through CRISPR-based gene editing of the FDC1 gene in *Saccharomyces cerevisiae* x *Saccharomyces eubayanus* hybrid lager beer yeasts. Schacherer J (ed). *PLoS One* 2019;14:e0209124.
- Mertens S, Steensels J, Saelens V et al. A large set of newly created interspecific *Saccharomyces* hybrids increases aromatic diversity in lager beers. *Appl Environ Microbiol* 2015;81:8202–14.
- Moliné M, Arbeloa EM, Flores MR et al. UVB photoprotective role of mycosporines in yeast: photostability and antioxidant activity of mycosporine-glutaminol-glucoside. *Radiat Res* 2011;175:44–50.
- Mülleider M, Calvani E, Alam MT et al. Functional metabolomics describes the yeast biosynthetic regulome. *Cell* 2016;167:553–65 e12.
- Naseeb S, Alsammar H, Burgis T et al. Whole genome sequencing, de novo assembly and phenotypic profiling for the new budding yeast species *Saccharomyces jurei*. *G3 Genes, Genomes, Genet* 2018;8:2967–77.
- Nespolo RF, Villarroel CA, Oporto CI et al. An Out-of-Patagonia dispersal explains most of the worldwide genetic distribution in *Saccharomyces eubayanus*. *bioRxiv* 2019:709253.
- Nguyen H-V, Boekhout T. Characterization of *Saccharomyces uvarum* (Beijerinck, 1898) and related hybrids: assessment of molecular markers that predict the parent and hybrid genomes and a proposal to name yeast hybrids. *FEMS Yeast Res* 2017;17:fox014.
- Novo M, Bigey F, Beyne E et al. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proc Natl Acad Sci U S A* 2009;106:16333–8.
- Opulente DA, Rollinson EJ, Bernick-Roehr C et al. Factors driving metabolic diversity in the budding yeast subphylum. *BMC Biol* 2018;16:26.

- Pellegrini M, Marcotte EM, Thompson MJ et al. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proc Natl Acad Sci USA* 1999;**96**:4285–8.
- Peris D, Langdon QK, Moriarty R V. et al. Complex ancestries of Lager-brewing hybrids were shaped by standing variation in the wild yeast *Saccharomyces eubayanus*. Fay JC (ed). *PLoS Genet* 2016;**12**:e1006155.
- Peris D, Sylvester K, Libkind D et al. Population structure and reticulate evolution of *Saccharomyces eubayanus* and its lager-brewing hybrids. *Mol Ecol* 2014;**23**:2031–45.
- Peter J, De Chiara M, Friedrich A et al. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 2018;**556**:339–44.
- Pretscher J, Fischkal T, Branscheidt S et al. Yeasts from different Habitats and their potential as biocontrol agents. *Fermentation* 2018;**4**:31.
- Proux-Wéra E, Armisen D, Byrne KP et al.. A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. *BMC Bioinformatics* 2012;**13**:1–12.
- Ram Y, Hadany L. The evolution of stress-induced hypermutations in asexual populations. *Evolution (N Y)* 2012;**66**:2315–28.
- Riley R, Haridas S, Wolfe KH et al. Comparative genomics of biotechnologically important yeasts. *Proc Natl Acad Sci USA* 2016;**113**:9882–7.
- Rodríguez-Sáiz M, De La Fuente JL, Barredo JL. *Xanthophyllomyces dendrorhous* for the industrial production of astaxanthin. *Appl Microbiol Biotechnol* 2010;**88**:645–58.
- Rodríguez ME, Pérez-Través L, Sangorrín MP et al. *Saccharomyces eubayanus* and *Saccharomyces uvarum* associated with the fermentation of *Araucaria araucana* seeds in Patagonia. *FEMS Yeast Res* 2014;**14**:948–65.
- Saitua F, Torres P, Pérez-Correa JR et al. Dynamic genome-scale metabolic modeling of the yeast *Pichia pastoris*. *BMC Syst Biol* 2017;**11**:27.
- Sampaio JP. Microbe profile: *Saccharomyces eubayanus*, the missing link to lager beer yeasts. *Microbiol (United Kingdom)* 2018;**164**:1069–71.
- Schlander M, Distler U, Tenzer S et al. Purification and properties of yeast proteases secreted by *Wickerhamomyces anomalus* 227 and *Metschnikovia pulcherrima* 446 during growth in a white grape juice. *Fermentation* 2017;**3**:2.
- Schwartz RS, Harkins KM, Stone AC et al.. A composite genome approach to identify phylogenetically informative data from next-generation sequencing. *BMC Bioinformatics* 2015;**16**:193.
- Sebastian J, Kraus B, Sancar GB. Expression of the yeast PHR1 gene is induced by DNA-damaging agents. *Mol Cell Biol* 1990;**10**:4630–7.
- Shen F, Sun R, Yao J et al. OptRAM: In-silico strain design via integrative regulatory-metabolic network modeling. Ouzounis CA (ed). *PLoS Comput Biol* 2019;**15**:e1006835.
- Shen XX, Opulente DA, Kominek J et al. Tempo and mode of genome evolution in the budding yeast subphylum. *Cell* 2018;**175**:1533–45 e20.
- Shen XX, Zhou X, Kominek J et al. Reconstructing the backbone of the *Saccharomycotina* yeast phylogeny using genome-scale data. *G3 Genes, Genomes, Genet* 2016;**6**:3927–39.
- Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;**30**:1312–3.
- Steenwyk J, Rokas A. Extensive copy number variation in fermentation-related genes among *Saccharomyces cerevisiae* wine strains. *G3 Genes, Genomes, Genet* 2017;**7**:1475–85.
- Steenwyk JL, Opulente DA, Kominek J et al. Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts. *PLoS Biol* 2019;**17**:e3000255.
- Steenwyk JL, Rokas A. Copy number variation in fungi and its implications for wine yeast genetic diversity and adaptation. *Front Microbiol* 2018;**9**:288.
- Strope PK, Kozmin SG, Skelly DA et al. 2 μ plasmid in *Saccharomyces* species and in *Saccharomyces cerevisiae*. *FEMS Yeast Res* 2015;**15**:fov090.
- Sulo P, Szabóová D, Bielik P et al. The evolutionary history of *Saccharomyces* species inferred from completed mitochondrial genomes and revision in the “yeast mitochondrial genetic code.” *DNA Res* 2017;**24**:571–83.
- Viigand K, Põšnograjeva K, Visnapuu T et al. Genome mining of non-conventional yeasts: search and analysis of MAL clusters and proteins. *Genes (Basel)* 2018;**9**:354.
- Wang Z, Danziger SA, Heavner BD et al. Combining inferred regulatory and reconstructed metabolic networks enhances phenotype prediction in yeast. Nielsen J (ed). *PLoS Comput Biol* 2017;**13**:e1005489.
- Waterhouse RM, Seppey M, Simão FA et al. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol Biol Evol* 2018;**35**:543–8.
- Weinert TA, Kiser GL, Hartwell LH. Mitotic checkpoint genes in budding yeast and the dependence of mitosis on DNA replication and repair. *Genes Dev* 1994;**8**:652–65.
- Weiß CL, Pais M, Cano LM et al.. nQuire: A statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinformatics* 2018;**19**:122.
- Wisecaver JH, Alexander WG, King SB et al. Dynamic evolution of nitric oxide detoxifying flavohemoglobins, a family of single-protein metabolic modules in bacteria and eukaryotes. *Mol Biol Evol* 2016;**33**:1979–87.
- Wohlbach DJ, Rovinskiy N, Lewis JA et al. Comparative genomics of *Saccharomyces cerevisiae* natural isolates for bioenergy production. *Genome Biol Evol* 2014;**6**:2557–66.
- Wolters JF, Chiu K, Fiumera HL. Population structure of mitochondrial genomes in *Saccharomyces cerevisiae*. *BMC Genomics* 2015;**16**:451.
- Wu B, Buljic A, Hao W. Extensive horizontal transfer and homologous recombination generate highly chimeric mitochondrial genomes in yeast. *Mol Biol Evol* 2015;**32**:2559–70.
- Xiao W, Chow BL, Hanna M et al. Deletion of the MAG1 DNA glycosylase gene suppresses alkylation-induced killing and mutagenesis in yeast cells lacking AP endonucleases. *Mutat Res - DNA Repair* 2001;**487**:137–47.
- Yue JX, Li J, Aigrain L et al. Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nat Genet* 2017;**49**:913–24.
- Yue JX, Liti G. Long-read sequencing data analysis for yeasts. *Nat Protoc* 2018;**13**:1213–31.
- Yuivar Y, Barahona S, Alcaíno J et al. Biochemical and thermodynamical characterization of glucose oxidase, invertase, and alkaline phosphatase secreted by Antarctic yeasts. *Front Mol Biosci* 2017;**4**:86.
- Zhou X, Peris D, Kominek J et al. In silico Whole Genome Sequencer And Analyzer (iWGS): A computational pipeline to guide the design and analysis of de novo genome sequencing studies. *G3 Genes, Genomes, Genet* 2016;**6**:3655–62.