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

Libkind, D.[1]; Peris, D.[2]; Cubillos, F.A.[3,4].; Steenwyk, J.L.[5]; Opulente, D.A.[6,7]; Langdon, Q.K.[6]; Bellora, N. [1]; Rokas, A. [5]; Hittinger, C.T[6,7].

Affiliations:

[1]Centro de Referencia en Levaduras y Tecnología Cervecera (CRELEC), Instituto Andino Patagónico de Tecnologías Biológicas y Geoambientales (IPATEC) – CONICET / Universidad Nacional del Comahue, Bariloche, Argentina.

[2]Department of Food Biotechnology, Institute of Agrochemistry and Food Technology, CSIC, Valencia, Spain

[3]Millennium Institute for Integrative Biology (iBio)

[4]Universidad de Santiago de Chile, Facultad de Química y Biología, Departamento de Biología, Santiago, Chile.

[5]Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA

[6]Laboratory of Genetics, Wisconsin Energy Institute, J. F. Crow Institute for the Study of Evolution, Genome Center of Wisconsin, University of Wisconsin-Madison, Madison, WI, USA

[7]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI, USA

Abstract

Genomic studies of yeasts from the wild have increased considerably in the last 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 bioinformatics tools available for the generation and analysis of yeast genomes, as well as 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.

1. Introduction.

Recent advances in sequencing technologies, the availability of new bioinformatics tools, and multiple genomic studies during the last 5 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 1,500 described yeast species have had their genomes fully sequenced (Kurtzman et al. 2011; Shen et al., 2018). In a few cases, sequences of multiple isolates are available for population genomic studies (See Suppl Table 1). 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.

2. 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 budding yeasts of the subphylum Saccharomycotina whose genomes have been sequenced has increased at least 3-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 bioinformatics 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 et al 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 analyses 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–DNA hybridization (DDH) has been gradually replaced by high-throughput sequencing, which allows the *in silico* calculation of overall genome related indices (OGRI) (Chun & 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., submitted for same issue). Other approaches

include the calculation of pairwise similarities (Kr, with the tool *genomediff* of *Genometools*) and genome-wide alignments (*MUMmer*, (Marcais 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 (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.

TABLE 1: List of bioinformatics tools

3. 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) (Figure 1). Furthermore, the development of the next-generation sequencing (NGS or 2nd generation) technologies and the long-read sequencing (3rd generation) technologies, together with bioinformatic tools (see Section 2) (Figure 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 datasets 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 datasets, 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 being detected among subtelomeric structural variants (which can also be important in speciation), the quantification and localization of copy number variants (e.g. those observed in the *CUP1* gene and *ARR* cluster), and presence and absence of metabolic genes (McIlwain et al 2016, Yue et al 2017, Naseeb et al 2018; Steenwyk & 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 et al 2015, Sulo et al 2017). In contrast, long-read technologies better capture and facilitate the assembly of mitochondrial genome sequences (Wolters et al 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 et al 1986), few genome sequencing studies explicitly comment about the recovery of 2µ plasmid sequences (Baker et al 2015, Strope et al 2015, McIlwain et al 2016, Peter et al 2018).

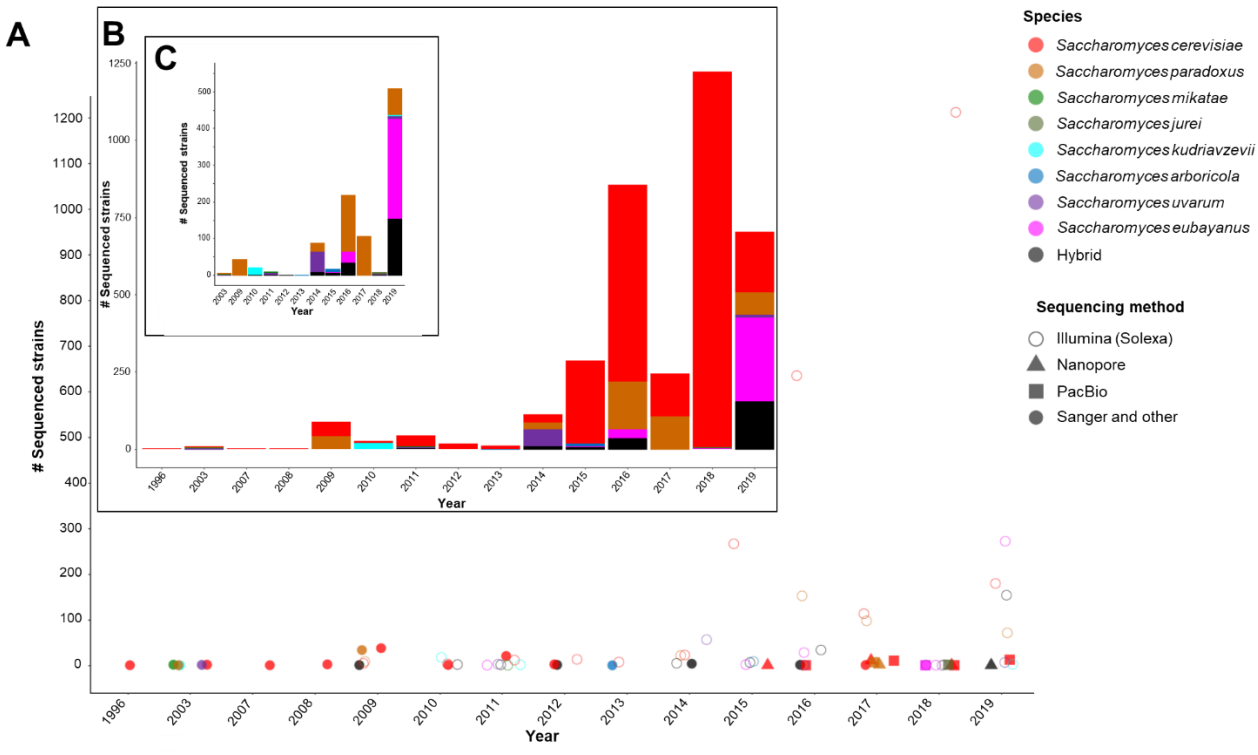

Figure 1. The genomes of more than three thousand *Saccharomyces* strains have been sequenced. At least 3,077 unique *Saccharomyces* strains have had their genomes sequenced using various

sequencing technologies in the last twenty-three years (Supplementary Table 1). 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 (Supplementary Table 1). 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.

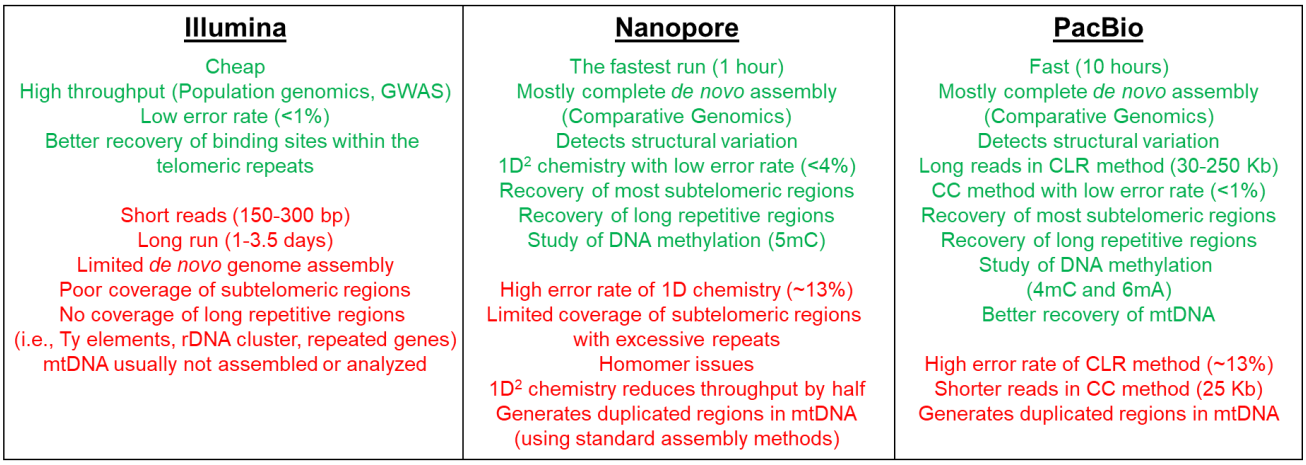

Figure 2. Pros and cons of the genome sequencing methods that are currently used most widely. Pros and cons of technologies used for for *de novo* genome assembly and population genomics (Goodwin et al 2015, Chen et al 2017, Istace et al 2017, Giordano et al 2017). CLR: continuous long read; CC, circular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

3.1 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 2,500 *S. cerevisiae* strains have been sequenced, including many that were independently sequenced by different labs (Figure 1). These efforts have helped differentiate the genome characteristics of wild, pathogenic/clinical, and domesticated *S. cerevisiae* strains (Figure 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 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, Naseeb et al 2018, Peter et al 2018, Duan et al 2018; Langdon et al 2019a, Nespolo et al 2019); ii) fewer admixed strains from the wild, supporting low

levels of outcrossing (Liti et al 2009, Almeida et al 2014, Leducq et al 2016, Eberlein et al 2019, Peris et al 2016); 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 (Figure 3); iv) strong geographic structure of wild *Saccharomyces* populations (Hittinger et al 2010, Almeida et al 2014, Gayevskiy et al 2015, Leducq et al 2016, Peris et al 2016, Peter et al 2018, Duan 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 (Goncalvez et al., 2016; Gallone et al., 2016; Steenwyk & 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 genomes are located in subtelomeric regions and are often related to flocculation, nitrogen metabolism, carbon metabolism, and stress (Bergstrom et al 2014; Steenwyk & 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 frequent in regions closer to the 3' end of protein-coding sequences (Bergstrom et al 2014).

Several horizontal gene transfer (HGT) events have been described in *Saccharomyces,* including several specific examples that are well supported (Figure 3) (Peter et al 2018; Marsit et al., 2015; Novo et al., 2009; Galeote et al., 2010; League et al., 2012). 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. And second, gene presence and absence variation of a horizontally acquired gene within or between species can mislead 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 missing data. For these reasons, we recommend using BLAST-based statistics, such as Alien Index (Wisecaver et al., 2016; Alexander 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, 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.

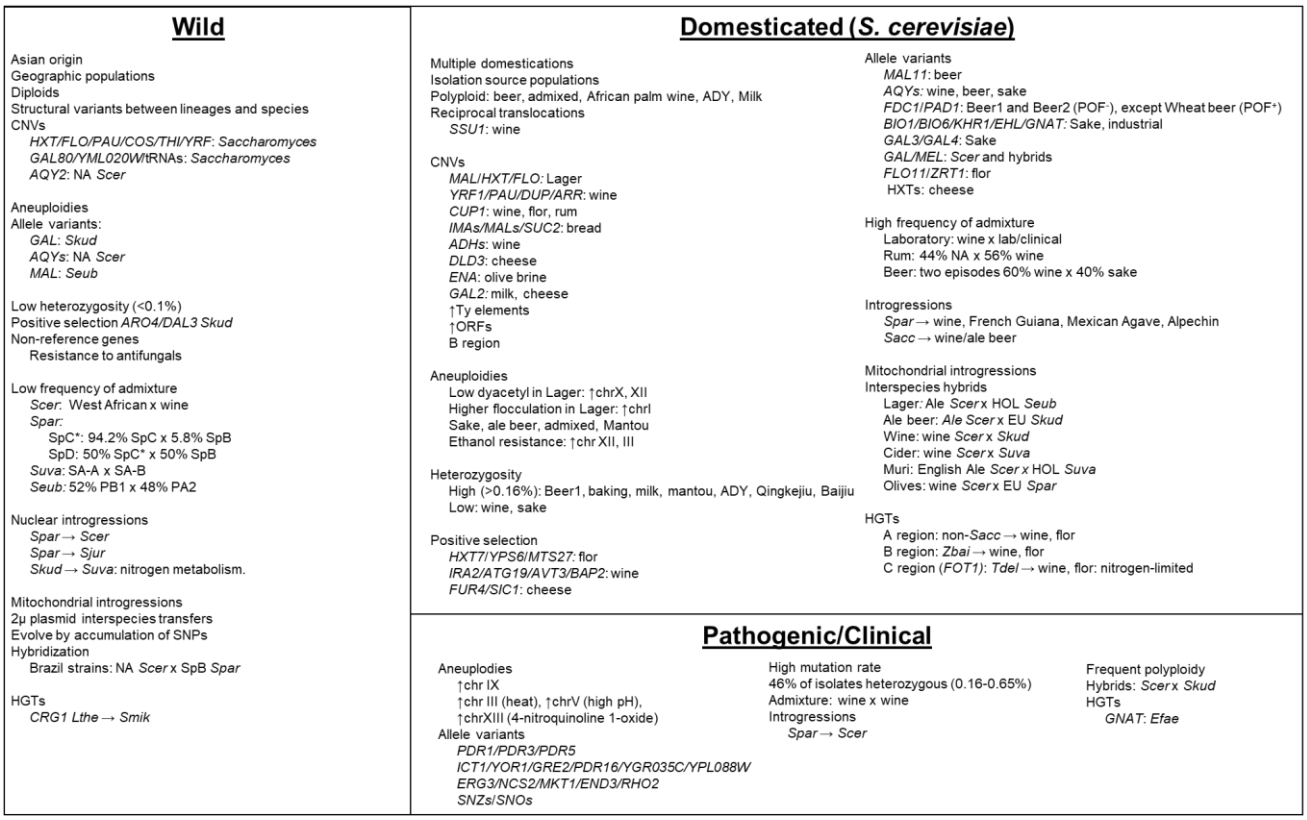

Figure 3. Genomic traits of wild, pathogenic, and domesticated *Saccharomyces* yeasts. Main genomic trait differences inferred from whole genome sequencing studies between wild *Saccharomyces*, domesticated, and clinical *S. cerevisiae* strains (Supplementary Table 1). Heterozygosity is represented as 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*, *Torulospora microellipsoides; Zbai*, *Zygosaccharomyces bailii; Lthe*, *Lachance thermotolerans;* PB, Patagonia B; PA, Patagonia A; NA, North America; HOL, Holarctic.

3.2 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 et al. 2018; Baker et al. 2019, Langdon et al 2019b, Gallone et al., 2019). 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 2019b), 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.

S. 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 (Peris et al. 2014; Bing 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 dataset for this species with a total of 256 new draft genome sequences (Langdon et al, 2019a; Nespolo et al., 2019). This dataset 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 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 Himalayan *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 (5 likely Patagonian, 3 from Patagonia B and 2 from Patagonia A, and 1 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 2019a). 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; Gayevskiy and Goddard 2016; Peris et al. 2016; Langdon et al., 2019a). 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., 2019a). 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., (2019a) and Nespolo et al., (2019).

4. 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 et al., 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*.

4.1 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 (Figure 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 *Hanseniaspora* yeast ecology, considering their frequent isolation from fruit juices and fermenting musts (Cadez 2006; Kurtzman et al. 2011), which likely reflects the specialization of *Hanseniaspora* species to sugar-rich environments.

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 et al. 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 et al. 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* yeasts 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 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 et al. 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.

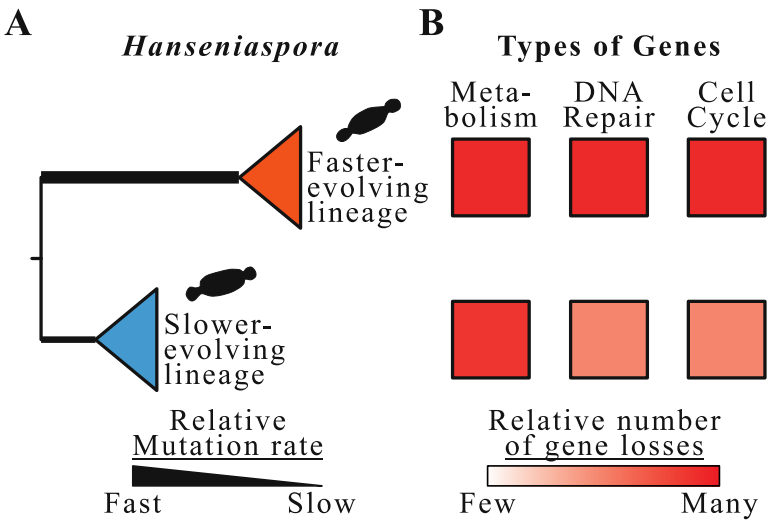

Figure 4: The evolutionary trajectories of *Hanseniaspora* lineages are marked by differential rates of sequence evolution and rates of loss of metabolism, DNA repair, and cell-cycle genes. (A) There are two lineages in the budding yeast genus *Hanseniaspora*: the faster-evolving and slower-evolving 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.

4.2 *Phaffia rhodozyma*: A colorful genome from the Basidiomycota

The orange-colored yeast *Phaffia rhodozyma* (= *Xanthophyllomyces 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 et al. 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 (Palma et al., 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 analyses, 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 homothallism 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) (Moline et al., 2011) was found in the newly sequenced and mycosporinogenic strain. However, this gene cluster was absent in the strain CBS 6938, which was shown not to accumulate this secondary metabolite, which has potential applications in cosmetics (Colabella et al., 2016). Genome mining also revealed an unexpected diversity of catalases and the loss of H_2O_2-sensitive superoxide dismutases in *P. rhodozyma*. Altogether, the *P. rhodozyma* genome is enriched in antioxidant mechanisms, in particular those most effective at coping with H_2O_2, suggesting that the environmental interaction with this reactive species has definitely contributed to shaping the peculiar genome of *P. rhodozyma*.

5. 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 *et al.*, 2017,

Yuivar *et al.*, 2017, Cai *et al.*, 2019), extraordinary capacity of fermentation in large-scale cultures (Choi *et al.*, 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, comprising 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, the Antarctic yeast *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 *et al.*, 2019, Langdon *et al.*, 2019a, Nespolo *et al.*, 2019). 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 & 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 & Price, 2015, Mülleder *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 *et al.*, 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 (Rodriguez *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 *et al.*, 2019, Lombardi *et al.*, 2019, Maroc & 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 be 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.

6. 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., submitted to this issue). 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 automatable 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 datasets 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.

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 Sklodowska-Curie fellow of the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 747775).

References

Akhmetov A, Laurent JM, Gollihar J, Gardner EC, Garge RK, Ellington AD, Kachroo AH & Marcotte EM (2018) Single-step Precision Genome Editing in Yeast Using CRISPR-Cas9. Bio Protoc 8: e2765.

Alexander WG, Doering DT & Hittinger CT (2014) High-efficiency genome editing and allele replacement in prototrophic and wild strains of Saccharomyces. Genetics 198: 859-866.

Almeida P, Gonçalves C, Teixeira S et al. 2014. A Gondwanan imprint on global diversity and domestication of wine and cider yeast Saccharomyces uvarum. Nat Commun. 5:4044

Alonso-Del-Real J, Lairon-Peris M, Barrio E & Querol A (2017) 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 8: 150.

Baker EP, Hittinger CT (2019) Evolution of a novel chimeric maltotriose transporter in *Saccharomyces eubayanus* from parent proteins unable to perform this function. PLoS Genet 15(4): e1007786. https://doi.org/10.1371/journal.pgen.1007786

Barbosa R, Almeida P, Safar SVB et al. 2016. Evidence of natural hybridization in Brazilian wild lineages of Saccharomyces cerevisiae. Genome Biol Evol. 8:317-329.

Bellora N, Moliné M, David-Palma M, Coelho MA, Hittinger CT, Sampaio JP, Libkind D. 2016. Comparative genomics provides new insights into the diversity, physiology, and sexuality of the only industrially exploited tremellomycete: Phaffia rhodozyma. BMC Genomics. 17:901.

503 Bergström A, Simpson JT, Salinas F et al. 2014. A high-definition view of functional genetic
504 variation from natural yeast genomes. Mol Biol Evol. 31:872-888.

505 Billmyre RB, Clancey SA, Heitman J. 2017. Natural mismatch repair mutations mediate phenotypic
506 diversity and drug resistance in Cryptococcus deuterogattii. Elife 6.

507 Bing J, Han P-J, Liu W-Q, Wang Q-M, Bai F-Y. 2014. Evidence for a Far East Asian origin of lager
508 beer yeast. Curr Biol 24: R380-1. http://www.ncbi.nlm.nih.gov/pubmed/24845661.

509 Borneman AR, Forgan AH, Kolouchova R, Fraser JA, Schmidt SA. 2016. Whole genome
510 comparison reveals high levels of inbreeding and strain redundancy across the spectrum of
511 commercial wine strains of Saccharomyces cerevisiae. G3. 6:957-971.

512 Brouwers N, Brickwedde A, Gorter de Vries AR, van den Broek M, Weening SM, van den Eijnden
513 L, Diderich JA, Bai F-Y, Pronk JT, Daran J-MG. 2019. Maltotriose consumption by hybrid
514 Saccharomyces pastorianus is heterotic and results from regulatory cross-
515 talk between parental sub-genomes. bioRxiv 679563.
516 http://biorxiv.org/content/early/2019/06/28/679563.abstract.

517 Cadez N. 2006. Phylogenetic placement of Hanseniaspora-Kloeckera species using multigene
518 sequence analysis with taxonomic implications: descriptions of Hanseniaspora
519 pseudoguilliermondii sp. nov. and Hanseniaspora occidentalis var. citrica var. nov. Int. J. Syst.
520 Evol. Microbiol. 56:1157–1165.

521 Cai P, Gao J & Zhou Y (2019) CRISPR-mediated genome editing in non-conventional yeasts for
522 biotechnological applications. Microbial Cell Factories 18: 63.

523 Campbell BB, Light N, Fabrizio D, Zatzman M, Fuligni F, de Borja R, Davidson S, Edwards M,
524 Elvin JA, Hodel KP, et al. 2017. Comprehensive Analysis of Hypermutation in Human Cancer. Cell
525 171:1042–1056.e10.

526 Chen Q, Lan C, Zhao L, Wang J, Chen B, Chen YPP. 2017. Recent advances in sequence assembly:
527 principles and applications. Brief Funct Genomic. 16:361-378.

528 Choi D-H, Park E-H & Kim M-D (2017) Isolation of thermotolerant yeast Pichia kudriavzevii from
529 nuruk. Food Sci Biotechnol 26: 1357-1362.

Colabella, F. & Libkind, D. 2016. PCR based method for the rapid identification of astaxanthin-accumulating yeast isolates of the genus Phaffia. Revista Argentina de Microbiología. 48(1):15-20. 10.1016/j.ram.2015.10.006

Cubillos FA, Gibson B, Grijalva-Vallejos N, Krogerus K & Nikulin J (2019) Bioprospecting for brewers: Exploiting natural diversity for naturally diverse beers. Yeast. Volume36, Issue6 June 2019 Pages 383-398

David-Palma, M.; Libkind, D.; Sampaio, J.P. 2014. Global distribution, diversity hotspots and niche transitions of an astaxanthin-producing eukaryotic microbe. Molecular Ecology. 23, 921–932.

Donohoue PD, Barrangou R & May AP (2018) Advances in Industrial Biotechnology Using CRISPR-Cas Systems. Trends in Biotechnology 36: 134-146.

Duan SF, Han PJ, Wang QM, Liu WQ, Shi JY, Li K, Zhang XL, Bai FY. 2018. The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. Nature Communications. 9:2690

Eberlein C, Hénault M, Fijarczyk A, Charron G, Bouvier M, Kohn LM, Anderson JB, Landry CR. 2019. Hybridization is a recurrent evolutionary stimulus in wild yeast speciation. Nature Communications. 10:923

Eizaguirre J.I., Peris, D.; Rodríguez, ME,; Lopes, C.; de Los Ríos, P.; Hittinger, CT; Libkind, D. 2018. Phylogeography of the wild Lager-brewing ancestor (Saccharomyces eubayanus) in Patagonia. Environ Microbiol. 20(10):3732-3743 doi: 10.1111/1462-2920.14375.

Fay J, Liu P, Ong G, Dunham M, Cromie G, Jeffrey E, Ludlow C, Dudley A. 2019. A polyploid admixed origin of beer yeasts derived from European and Asian wine populations. PLOS Biology. 17:e3000147

Galeote V, Novo M, Salema-Oom M, Brion C, Valério E, Gonçalves P, Dequin S. FSY1, a horizontally transferred gene in the Saccharomyces cerevisiae EC1118 wine yeast strain, encodes a high-affinity fructose/H+ symporter. Microbiology. 2010 Dec;156(Pt 12):3754-61. doi: 10.1099/mic.0.041673-0. Epub 2010 Aug 12.

Gallone B, Steensels J, Prahl T et al. 2016. Domestication and divergence of Saccharomyces cerevisiae beer yeasts. Cell. 166:1397-1410.

Gallone B, Steensels J, Prahl T, Soriaga L, Saels V, Herrera-Malaver B, Merlevede A, Roncoroni M, Voordeckers K, Miraglia L, et al. 2016. Domestication and Divergence of Saccharomyces cerevisiae Beer Yeasts. Cell [Internet] 166:1397–1410.e16. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0092867416310716

Gayevskiy V, Goddard MR. 2015. Saccharomyces eubayanus and Saccharomyces arboricola reside in North Island native New Zealand forests. Environ Microbiol. 18:1137-1147.

Gayevskiy V, Goddard MR. 2016. Saccharomyces eubayanus and Saccharomyces arboricola reside in North Island native New Zealand forests. Environ Microbiol 18: 1137–1147.

Gibson B, Geertman JA, Hittinger CT, Krogerus K, Libkind D, Louis EJ, Magalhaes F & Sampaio JP (2017) New yeasts-new brews: modern approaches to brewing yeast design and development. FEMS Yeast Res 17.

Gibson B, Liti G. 2015. Saccharomyces pastorianus: genomic insights inspiring innovation for industry. Yeast 32: 17–27. https://onlinelibrary.wiley.com/doi/full/10.1002/yea.3033 (Accessed February 6, 2019).

Giordano F, Aigrain L, Quail MA et al. 2017. De novo yeast genome assemblies from MinION, PacBio and MiSeq platforms. Scientific Reports. 7:3935

Goffeau A, Barrell BG, Bussey H et al. 1996. Life with 6000 Genes. Science. 274:546

Gonçalves M, Pontes A, Almeida P, Barbosa R, Serra M, Libkind D, Hutzler M, Gonçalves P, Sampaio JP. 2016. Distinct domestication trajectories in Top-Fermenting beer yeasts and wine yeasts. Curr Biol. 26:1-12.

Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz MC, McCombie WR. 2015. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. Genome Res. 25:1750-1756.

Heavner BD & Price ND (2015) Comparative Analysis of Yeast Metabolic Network Models Highlights Progress, Opportunities for Metabolic Reconstruction. PLOS Computational Biology 11: e1004530.

584 Hebly M, Brickwedde A, Bolat I, Driessen MR, de Hulster EA, van den Broek M, Pronk JT,
585 Geertman JM, Daran JM & Daran-Lapujade P (2015) S. cerevisiae x S. eubayanus interspecific
586 hybrid, the best of both worlds and beyond. FEMS Yeast Res 15.

587 Heinrich S, Sewart K, Windecker H, Langegger M, Schmidt N, Hustedt N, Hauf S. 2014. Mad1
588 contribution to spindle assembly checkpoint signalling goes beyond presenting Mad2 at
589 kinetochores. EMBO Rep. 15:291–298.

590 Hittinger CT, Rokas A, Bai FY, Boekhout T, Gonçalves P, Jeffries TW, Kominek J, Lachance
591 MA, Libkind D, Rosa CA, Sampaio JP, Kurtzman CP. Genomics and the making of yeast
592 biodiversity. Curr Opin Genet Dev. 2015 Dec;35:100-9.

593 Hittinger CT, Gonçalves P, Sampaio JP, Dover J, Johnston M, Rokas A. 2010. Remarkably ancient
594 balanced polymorphisms in a multi-locus gene network. Nature. 464:54-58.

595 Hittinger CT, Steele JL, Ryder DS. 2018. Diverse yeasts for diverse fermented beverages and foods.
596 Curr Opin Biotechnol 49: 199–206. http://dx.doi.org/10.1016/j.copbio.2017.10.004 (Accessed
597 February 6, 2019).

598 Istace B, Friedrich A, Agata L et al. 2017. de novo assembly and population genomic survey of
599 natural yeast isolates with the Oxford Nanopore MinION sequencer. GigaScience. 6:1-13.

600 Jenjaroenpun P, Wongsurawat T, Pereira R, Patumcharoenpol P, Ussery DW, Nielsen J, Nookaew I.
601 2018. Complete genomic and transcriptional landscape analysis using third-generation sequencing:
602 a case study of Saccharomyces cerevisiae CEN.PK113-7D. Nucl Acids Res. gky014-gky014.

603 Jolivet-Gougeon A, Kovacs B, Le Gall-David S, Le Bars H, Bousarghin L, Bonnaure-Mallet M,
604 Lobel B, Guille F, Soussy C-J, Tenke P. 2011. Bacterial hypermutation: clinical implications. J.
605 Med. Microbiol. 60:563–573.

606 Jorgensen P. 2002. Systematic Identification of Pathways That Couple Cell Growth and Division in
607 Yeast. Science (80-.). 297:395–400.

608 Juergens H, Varela JA, Gorter de Vries AR, et al. (2018) Genome editing in Kluyveromyces and
609 Ogataea yeasts using a broad-host-range Cas9/gRNA co-expression plasmid. FEMS Yeast Research
610 18.

611 Krogerus K, Magalhaes F, Vidgren V & Gibson B (2017) Novel brewing yeast hybrids: creation and
612 application. Appl Microbiol Biotechnol 101: 65-78.

613 Krogerus K, Seppanen-Laakso T, Castillo S & Gibson B (2017) Inheritance of brewing-relevant
614 phenotypes in constructed Saccharomyces cerevisiae x Saccharomyces eubayanus hybrids. Microb
615 Cell Fact 16: 66.

616 Kuang MC, Kominek J, Alexander WG, Cheng J-F, Wrobel RL & Hittinger CT (2018) Repeated
617 Cis-Regulatory Tuning of a Metabolic Bottleneck Gene during Evolution. Molecular biology and
618 evolution 35: 1968-1981.

619 Kurtzman CP, Fell JW, Boekhout T. 2011. The Yeasts: A Taxonomic Study, 5th Edition.

620 Kurtzman, C.; J.W. Fell Teun Boekhout The Yeasts. A Taxonomic Study 5th Edition
621 **ISBN:** 9780444521491, Elsevier Science, 2011. pp 2354

622 Langdon QK, Peris D, Eizaguirre JI, et al. (2019a) Genomic diversity and global distribution of
623 Saccharomyces eubayanus, the wild ancestor of hybrid lager-brewing yeasts. bioRxiv
624 709535.

625 Langdon QK, Peris D, Baker EP, Opulente DA, Nguyen HV, Bond U, Gonçalves P, Sampaio JP,
626 Libkind D, Hittinger CT. 2019b. Fermentation innovation through complex hybridization of wild
627 and domesticated yeasts. Nat Ecol Evol. 2019 Nov;3(11):1576-1586. doi: 10.1038/s41559-019-
628 0998-8

629 League GP, Slot JC, Rokas A. The ASP3 locus in Saccharomyces cerevisiae originated by horizontal
630 gene transfer from *Wickerhamomyces*. FEMS Yeast Res. 2012 Nov;12(7):859-63.

631 Leducq J-B, Charron G, Samani P, Dubé AK, Sylvester K, James B, Almeida P, Sampaio JP,
632 Hittinger CT, Bell G, et al. 2014. Local climatic adaptation in a widespread microorganism. Proc
633 Biol Sci 281: 20132472. http://www.ncbi.nlm.nih.gov/pubmed/24403328.

634 Leducq JB, Nielly-Thibault L, Charron G et al. 2016. Speciation driven by hybridization and
635 chromosomal plasticity in a wild yeast. Nature Microbiology. 1:15003

Leducq J-B, Nielly-Thibault L, Charron G, Eberlein C, Verta J-P, Samani P, Sylvester K, Hittinger CT, Bell G, Landry CR. 2016. Speciation driven by hybridization and chromosomal plasticity in a wild yeast. Nat Microbiol 1: 1–10.

Legras JL, Galeote V, Bigey F et al. 2018. Adaptation of S. cerevisiae to fermented food environments reveals remarkable genome plasticity and the footprints of domestication. Mol Biol Evol. msy066-msy066.

Libkind D, Hittinger CT, Valerio E, Goncalves C, Dover J, Johnston M, Goncalves P & Sampaio JP (2011) Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. Proc Natl Acad Sci U S A 108: 14539-14544.

Liti G (2015) The fascinating and secret wild life of the budding yeast S. cerevisiae. Elife 4.

Liti G, Carter DM, Moses AM et al. 2009. Population genomics of domestic and wild yeasts. Nature. 458:337-341.

Loira N, Dulermo T, Nicaud J-M & Sherman DJ (2012) A genome-scale metabolic model of the lipid-accumulating yeast Yarrowia lipolytica. BMC Systems Biology 6: 35.

Lombardi L, Oliveira-Pacheco J & Butler G (2019) Plasmid-Based CRISPR-Cas9 Gene Editing in Multiple Candida Species. mSphere 4: e00125-00119.

Lopes H & Rocha I (2017) Genome-scale modeling of yeast: chronology, applications and critical perspectives. FEMS yeast research 17: fox050.

Lustig AJ. 2001. Cdc13 subcomplexes regulate multiple telomere functions. Nat. Struct. Biol. 8:297–299.

Maite Novo, Frédéric Bigey, Emmanuelle Beyne, Virginie Galeote, Frédérick Gavory, Sandrine Mallet, Brigitte Cambon, Jean-Luc Legras, Patrick Wincker, Serge Casaregola, and Sylvie Dequin. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast Saccharomyces cerevisiae EC1118. PNAS September 22, 2009 106 (38) 16333-16338.

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. PLoS computational biology. 2018 Jan 26;14(1):e1005944.

Maroc L & Fairhead C (2019) A new inducible CRISPR-Cas9 system useful for genome editing and study of DSB repair in Candida glabrata. Yeast.

664 Martorell MM, Ruberto LAM, de Figueroa LIC & Mac Cormack WP (2019) Antarctic Yeasts as a
665 Source of Enzymes for Biotechnological Applications. Fungi of Antarctica: Diversity, Ecology and
666 Biotechnological Applications,(Rosa LH, ed.) p.^pp. 285-304. Springer International Publishing,
667 Cham.

668 McIlwain SJ, Peris D, Sardi M et al. 2016. Genome sequence and analysis of a stress-tolerant, wild-
669 derived strain of Saccharomyces cerevisiae used in biofuels research. G3. 6:1757-1766.

670 Mead DJ, Gadner DCJ, Olivere SG 1986 The yeast 2□ plasmid: strategies for the survival of a selfish
671 DNA. Molecular General Genetics 205:417-421

672 Mertens S, Gallone B, Steensels J, Herrera-Malaver B, Cortebeek J, Nolmans R, Saels V, Vyas VK
673 & Verstrepen KJ (2019) Reducing phenolic off-flavors through CRISPR-based gene editing of the
674 FDC1 gene in Saccharomyces cerevisiae x Saccharomyces eubayanus hybrid lager beer yeasts.
675 PLoS One 14: e0209124.

676 Mertens S, Steensels J, Saels V, De Rouck G, Aerts G & Verstrepen KJ (2015) A large set of newly
677 created interspecific Saccharomyces hybrids increases aromatic diversity in lager beers. Appl
678 Environ Microbiol 81: 8202-8214.

679 Moliné, M.; Arbeloa, E. M.; Regina Flores, M.; Libkind, D.; M.C.; Farías, M.E.; Bertolotti, S.G.;
680 Churio, M.S. & van Broock, M. 2011. UV-B photoprotective role of mycosporines in yeasts:
681 photostability and antioxidant activity of mycosporine-glutaminol-glucoside. Radiation Research.
682 175 (1), 44-50.

683 Mülleder M, Calvani E, Alam MT, Wang RK, Eckerstorfer F, Zelezniak A & Ralser M (2016)
684 Functional Metabolomics Describes the Yeast Biosynthetic Regulome. Cell 167: 553-565.e512.

685 Naseeb S, Alsammar H, Burgis T, Donaldson I, Knyazev N, Knight C, Delneri D. 2018. Whole
686 genome sequencing, de novo assembly and phenotypic profiling for the new budding yeast species
687 Saccharomyces jurei. G3. 8:2967-2977.

688 Nespolo RF, Villarroel CA, Oporto CI, et al. (2019) An Out-of-Patagonia dispersal explains most of
689 the worldwide genetic distribution in Saccharomyces eubayanus. bioRxiv 709253.

690 Nguyen HV, Boekhout T. 2017. Characterization of Saccharomyces uvarum (Beijerinck, 1898) and
691 related hybrids: Assessment of molecular markers that predict the parent and hybrid genomes and a
692 proposal to name yeast hybrids. FEMS Yeast Res 17: 1–19.

693 Olsen RA, Bunikis I, Tiukova I, Holmberg K, Lötstedt B, Pettersson OV, Passoth V, Käller M, Vezzi
694 F. De novo assembly of Dekkera bruxellensis: a multi technology approach using short and long-
695 read sequencing and optical mapping. GigaScience. 2015 Dec;4(1):56.

696 Opulente DA, Rollinson EJ, Bernick-Roehr C, Hulfachor AB, Rokas A, Kurtzman CP, Hittinger CT.
697 Factors driving metabolic diversity in the budding yeast subphylum. BMC Biol. 2018 Mar
698 2;16(1):26. doi: 10.1186/s12915-018-0498-3.

699 Peris D, Langdon Q, Moriarty RV, Sylvester K, Bontrager M, Charron G, Leducq J-B, Landry CR,
700 Libkind D, Hittinger CT. 2016. Complex origins of lager-brewing hybrids were shaped by standing
701 variation in the wild yeast Saccharomyces eubayanus. Plos Genetics. 12(7):e1006155.

702 Peris, D.; Sylvester, K.; Libkind, D.; Gonçalves, P.; Sampaio, J.P.; Alexander, W.; Hittinger, C. 2014
703 Population Structure and Reticulate Evolution of Saccharomyces eubayanus and Its Lager-Brewing
704 Hybrids. Molecular Ecology. 23, 2031-2045.

705 Peter J, De Chiara M, Friedrich A et al. 2018. Genome evolution across 1,011 Saccharomyces
706 cerevisiae isolates. Nature. 556:339-344.

707 Press MO, Hall AN, Morton EA, Queitsch C. Substitutions are boring: Some arguments about
708 parallel mutations and high mutation rates. Trends in Genetics. 2019 Feb 20.

709 Pretscher J, Fischkal T, Branscheidt S, Jäger L, Kahl S, Schlander M, Thines E & Claus H (2018)
710 Yeasts from Different Habitats and Their Potential as Biocontrol Agents. Fermentation 4: 31.

711 Ram Y, Hadany L. 2012. The evolution of stress-induced hypermutation in asexual populations.
712 Evolution (N. Y). 66:2315–2328.

713 Riley R, Haridas S, Wolfe KH, et al. (2016) Comparative genomics of biotechnologically important
714 yeasts. Proceedings of the National Academy of Sciences 113: 9882-9887.

715 Rodríguez ME, Pérez-Través L, Sangorrín MP, Barrio E, Lopes CA. 2014. Saccharomyces
716 eubayanus and Saccharomyces uvarum associated with the fermentation of Araucaria araucana
717 seeds in Patagonia. FEMS Yeast Res 14: 948–965.

718 Rodríguez-Sáiz M1, de la Fuente JL, Barredo JL. Xanthophyllomyces dendrorhous for the industrial
719 production of astaxanthin. Appl Microbiol Biotechnol. 2010 Oct;88(3):645-58. doi:
720 10.1007/s00253-010-2814-x. Epub 2010 Aug 14.

721 Saitua F, Torres P, Pérez-Correa JR & Agosin E (2017) Dynamic genome-scale metabolic modeling
722 of the yeast Pichia pastoris. BMC Systems Biology 11: 27.

723 Sampaio JP. 2018. Microbe profile: Saccharomyces eubayanus, the missing link to lager beer yeasts.
724 Microbiol (United Kingdom) 164: 1069–1071.
725 http://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.000677 (Accessed May
726 22, 2019).

727 Schlander M, Distler U, Tenzer S, Thines E & Claus H (2017) Purification and Properties of Yeast
728 Proteases Secreted by Wickerhamomyces anomalus 227 and Metschnikovia pulcherrima 446 during
729 Growth in a White Grape Juice. Fermentation 3: 2.

730 Sebastian J, Kraus B, Sancar GB. 1990. Expression of the yeast PHR1 gene is induced by DNA-
731 damaging agents. Mol. Cell. Biol. 10:4630–4637.

732 Shen F, Sun R, Yao J, Li J, Liu Q, Price ND, Liu C & Wang Z (2019) OptRAM: In-silico strain
733 design via integrative regulatory-metabolic network modeling. PLOS Computational Biology 15:
734 e1006835.

735 Shen XX, Zhou X, Kominek J, Kurtzman CP, Hittinger CT, Rokas A. Reconstructing the Backbone
736 of the Saccharomycotina Yeast Phylogeny Using Genome-Scale Data. G3 (Bethesda). 2016 Dec
737 7;6(12):3927-3939. doi: 10.1534/g3.116.034744.

738 Shen X-X, Opulente DA, Kominek J, Zhou X, Steenwyk JL, Buh K V., Haase MAB, Wisecaver JH,
739 Wang M, Doering DT, et al. 2018. Tempo and Mode of Genome Evolution in the Budding Yeast
740 Subphylum. Cell 175:1533–1545.e20.

741 Shen, X-X.; D.A. Opulente, J. Kominek, X. Zhou, J. Steenwyk, K. V. Buh, M.A. B. Haase, J. H.
742 Wisecaver, M. Wang, J. T. Boudouris, R. M. Schneider, Q. K. Langdon, M. Ohkuma, R. Endoh, M.
743 Takashima, R. Manabe, N. Čadež, D. Libkind, C. A. Rosa, J. DeVirgilio, A. Beth Hulfachor, M.
744 Groenewald, C. P. Kurtzman, C. T. Hittinger & A. Rokas. 2018. The tempo and mode of genome

evolution across the budding yeast subphylum. Cell. 175(6):1533-1545.e20. doi: 10.1016/j.cell.2018.10.023.

Souhir Marsit, Adriana Mena,Frédéric Bigey,François-Xavier Sauvage,Arnaud Couloux, Julie Guy, Jean-Luc Legras, Eladio Barrio, Sylvie Dequin, and Virginie Galeote. Evolutionary Advantage Conferred by an Eukaryote-to-Eukaryote Gene Transfer Event in Wine YeastsMol Biol Evol. 2015 Jul; 32(7): 1695–1707.

Steenwyk J, Rokas A. 2017. Extensive Copy Number Variation in Fermentation-Related Genes Among Saccharomyces cerevisiae Wine Strains. G3 Genes, Genomes, Genet. [Internet] 7. Available from: http://www.g3journal.org/content/7/5/1475#ref-25

Steenwyk JL, Opulente DA, Kominek J, Shen X-X, Zhou X, Labella AL, Bradley NP, Eichman BF, Čadež N, Libkind D, et al. 2019. Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts.Kamoun S, editor. PLOS Biol. 17:e3000255.

Steenwyk JL, Rokas A. 2018. Copy Number Variation in Fungi and Its Implications for Wine Yeast Genetic Diversity and Adaptation. Front. Microbiol. [Internet] 9. Available from: http://journal.frontiersin.org/article/10.3389/fmicb.2018.00288/full

Strope PK, Kozmin SG, Skelly DA, Magwene PM, Dietrich FS, McCusker JH. 2015. 2µ plasmid in Saccharomyces species and in Saccharomyces cerevisiae. FEMS Yeast Res. 15: fov090Sulo P, Szabóóvá D, Bielik P, Poláková S, Šoltys K, Jatzová K, Szemes Tomáš. 2017. The evolutionary history of Saccharomyces species inferred from completed mitochondrial genomes and revision in the 'yeast mitochondrial genetic code'. DNA Res. 24:571-583.

Viigand K, Põšnograjeva K, Visnapuu T & Alamäe T (2018) Genome Mining of Non-Conventional Yeasts: Search and Analysis of MAL Clusters and Proteins. Genes (Basel) 9: 354.

Wang Z, Danziger SA, Heavner BD, et al. (2017) Combining inferred regulatory and reconstructed metabolic networks enhances phenotype prediction in yeast. PLoS computational biology 13: e1005489-e1005489.

Weinert TA, Kiser GL, Hartwell LH. 1994. Mitotic checkpoint genes in budding yeast and the dependence of mitosis on DNA replication and repair. Genes Dev. 8:652–665.

772 Wohlbach DJ, Rovinskiy N, Lewis JA et al. 2014. Comparative genomics of Saccharomyces
773 cerevisiae natural isolates for bioenergy production. Genome Biol Evol. 6:2557-2566.

774 Wolters JF, Chiu K, Fiumera HL. 2015. Population structure of mitochondrial genomes in
775 Saccharomyces cerevisiae. BMC Genomics. 16:451

776 Xiao W, Chow BL, Hanna M, Doetsch PW. 2001. Deletion of the MAG1 DNA glycosylase gene
777 suppresses alkylation-induced killing and mutagenesis in yeast cells lacking AP endonucleases.
778 Mutat. Res. - DNA Repair.

779 Yue JX, Li J, Aigrain L et al. 2017. Contrasting evolutionary genome dynamics between
780 domesticated and wild yeasts. Nat Genet. 49:913-924.

781 Yue JX, Liti G. 2018. Long-read sequencing data analysis for yeasts. Nature Protocols. 13:1213

782 Yuivar Y, Barahona S, Alcaíno J, Cifuentes V & Baeza M (2017) Biochemical and
783 Thermodynamical Characterization of Glucose Oxidase, Invertase, and Alkaline Phosphatase
784 Secreted by Antarctic Yeasts. Front Mol Biosci 4: 86-86

Campbell MS, Holt C, Moore B, Yandell M. 2014. Genome Annotation and Curation Using MAKER and MAKER-P. Current Protocols in Bioinformatics. 48:4-4.

Carver T, Bleasby A. 2003. The design of Jemboss: a graphical user interface to EMBOSS. Bioinformatics. 19:1837-1843.

Fan H, Ives AR, Surget-Groba Y, Cannon CH. 2015. An assembly and alignment-free method of phylogeny reconstruction from next-generation sequencing data. BMC Genomics. 16:1-18.

Gremme G, Steinbiss S, Kurtz S. 2013. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 10:645-656.

Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads-a baiting and iterative mapping approach. Nucl Acids Res. 41:e129-e129.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 23:254-267.

Johnson MG, Gardner EM, Liu Y, Medina R, Goffinet B, Shaw AJ, Zerega NJC, Wickett NJ. 2016. HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Appl Plant Sci. 4:1600016

Langdon QK, Peris D, Kyle B, Hittinger CT. 2018. sppIDer: a species identification tool to investigate hybrid genomes with high-throughput sequencing. Mol Biol Evol. 35:2835-2849.

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: A fast and versatile genome alignment system. PLoS Computational Biology. 14:e1005944

Proux-Wera E, Armisen D, Byrne K, Wolfe K. 2012. A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. BMC Bioinformatics. 13:237

Schwartz RS, Harkins KM, Stone AC, Cartwright RA. 2015. A composite genome approach to identify phylogenetically informative data from next-generation sequencing. BMC Bioinformatics. 16:193

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312-1313.

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 35:543-548.

Weiß CL, Pais M, Cano LM, Kamoun S, Burbano HA. 2018. nQuire: a statistical framework for ploidy estimation using next generation sequencing. BMC Bioinformatics. 19:122

Yue JX, Liti G. 2018. Long-read sequencing data analysis for yeasts. Nature Protocols. 13:1213

Zhou X, Peris D, Kominek J, Kurtzman CP, Hittinger CT, Rokas A. 2016. in silico Whole Genome Sequencer & Analyzer (iWGS): a computational pipeline to guide the design and analysis of de novo genome sequencing studies. G3 (Bethesda). 6:3655-3670.

785