

# Genome Diversity and Evolution in the Budding Yeasts (*Saccharomycotina*)

Bernard A. Dujon<sup>\*,†,1</sup> and Edward J. Louis<sup>\*,§</sup>

<sup>\*</sup>Department Genomes and Genetics, Institut Pasteur, Centre National de la Recherche Scientifique UMR3525, 75724-CEDEX15 Paris, France,

<sup>†</sup>University Pierre and Marie Curie UFR927, 75005 Paris, France, <sup>‡</sup>Centre for Genetic Architecture of Complex Traits, and <sup>§</sup>Department of Genetics, University of Leicester, LE1 7RH, United Kingdom

ORCID ID: 0000-0003-1157-3608 (E.J.L.)

**ABSTRACT** Considerable progress in our understanding of yeast genomes and their evolution has been made over the last decade with the sequencing, analysis, and comparisons of numerous species, strains, or isolates of diverse origins. The role played by yeasts in natural environments as well as in artificial manufactures, combined with the importance of some species as model experimental systems sustained this effort. At the same time, their enormous evolutionary diversity (there are yeast species in every subphylum of Dikarya) sparked curiosity but necessitated further efforts to obtain appropriate reference genomes. Today, yeast genomes have been very informative about basic mechanisms of evolution, speciation, hybridization, domestication, as well as about the molecular machineries underlying them. They are also irreplaceable to investigate in detail the complex relationship between genotypes and phenotypes with both theoretical and practical implications. This review examines these questions at two distinct levels offered by the broad evolutionary range of yeasts: inside the best-studied *Saccharomyces* species complex, and across the entire and diversified subphylum of *Saccharomycotina*. While obviously revealing evolutionary histories at different scales, data converge to a remarkably coherent picture in which one can estimate the relative importance of intrinsic genome dynamics, including gene birth and loss, vs. horizontal genetic accidents in the making of populations. The facility with which novel yeast genomes can now be studied, combined with the already numerous available reference genomes, offer privileged perspectives to further examine these fundamental biological questions using yeasts both as eukaryotic models and as fungi of practical importance.

**KEYWORDS** diversity; evolution; genomes; population; yeasts; YeastBook

## TABLE OF CONTENTS

Abstract	717
Introduction	718
What Did We Learn from Comparative and Population Genomics of the <i>Saccharomyces</i> Species Complex?	719
<i>The monophyletic <i>Saccharomyces</i> (formerly <i>S. sensu stricto</i>) clade</i>	719
<i>Chromosome evolution within the clade</i>	720
<i>Gene losses and differentiation of function</i>	720
<i>Population genomics</i>	721
<i>Clean lineages vs. mosaics</i>	721

*Continued*

Copyright © 2017 Dujon and Louis

doi: <https://doi.org/10.1534/genetics.116.199216>

Manuscript received December 14, 2016; accepted for publication April 3, 2017

Available freely online through the author-supported open access option.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

<sup>1</sup>Corresponding author: Department Genomes and Genetics, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris, France. E-mail: [bdujon@pasteur.fr](mailto:bdujon@pasteur.fr)

CONTENTS, *continued*

<i>Genetic vs. phenotypic diversity</i>	722
<i>Domestication or harnessing existing potential</i>	722
<i>Subtelomeres and the pan-genome</i>	723
<i>Horizontal gene transfers</i>	723
<i>Hybrids, introgressions, and reticulate evolution</i>	724
<i>Reproductive isolation and speciation</i>	724
GCRs	724
Bateson–Dobzhansky–Müller incompatibilities	725
Sequence divergence and mismatch repair	725
<i>Remaining issues raised by Saccharomyces</i>	725
What Did We Learn from Comparative Genomics of Other Saccharomycotina?	726
<i>The rise of comparative yeast genomics: a brief history</i>	726
<i>Four major genomic architectures within the Saccharomycotina</i>	728
<i>Origin of the distinctive features of Saccharomycotina genomes</i>	729
Introns	729
Centromeres	731
Mating cassettes and mating-type switching	735
tRNA decoding and alternative genetic codes	736
RNAi machinery	737
mtDNA	738
<i>Tentative reconstruction of the evolution of Saccharomycotina genomes</i>	738
<i>Molecular basis of the evolutionary dynamics of yeast genomes</i>	739
<i>Intrinsic genome dynamics: sequence divergence and chromosomal rearrangements</i>	739
<i>Intrinsic genome dynamics: loss, duplication, and neofunctionalization of genes</i>	740
<i>Horizontal exchanges and interspecies hybridization</i>	742
Perspectives	743

It has now been 20 years since the genome of the S288C laboratory strain of *Saccharomyces cerevisiae* was sequenced (Goffeau *et al.* 1996, 1997). At this time, it was the first and only eukaryotic genome entirely sequenced, and probably only few of the many participants to this scientific milestone had a clear idea of the considerable developments that would follow and their impact on almost every field of biological sciences and their applications. Today, full genome sequences are available for less than a 10th of all described yeast species (>1500 in total, see Kurtzman *et al.* 2011) and, for several of them, sequences of multiple isolates are available for comparisons. These figures are rapidly increasing, and create a novel situation in which our knowledge about basic biological mechanisms, deduced from favored experimental models such as *S. cerevisiae* and *Schizosaccharomyces pombe*, can be usefully complemented with observations of the natural diversity of a group of organisms with an extremely long evolutionary history.

It is now clear that yeasts do not represent a monophyletic group of primitive, unicellular eukaryotes but have repeatedly emerged during evolution from distinct lineages of Ascomycota or Basidiomycota containing more complex forms of fungi (Nagy *et al.* 2014). If budding yeasts, the Saccharomycotina subphylum of Ascomycota to which *S. cerevisiae* belongs, rep-

resent the most successful monophyletic group of yeasts by the total number of species described (almost two thirds of all known yeasts), other yeast species are distributed between the three subphyla of Basidiomycota (representing altogether about a third of all known yeasts) and the Taphrinomycotina subphylum of Ascomycota (~3% of total) (Kurtzman *et al.* 2011). A few unicellular or dimorphic fungi in which the unicellular form (yeast) is restricted to specific environmental conditions also exist in the Pezizomycotina subphylum of Ascomycota, otherwise comprised of filamentous fungi, as exemplified by the black yeast *Hortaea werneckii* (Lenassi *et al.* 2013) and the temperature-dependent *Talaromyces marneffeii* (Woo *et al.* 2011). This unequal taxonomic distribution is further biased in presently available genomic sequences (reviewed in Dujon 2015a). If numerous data are available for the Saccharomycotina genomes, much more fragmented information exists for the other lineages despite their broader evolutionary diversity. For this reason, and also because the *YeastBook* series is devoted to *Saccharomyces*, the present review will focus on the budding yeasts, with only occasional mention of the other lineages, as appropriate for discussion. Information can be found on the genomes of other yeasts in recent publications such as Rhind *et al.* (2011) and Farrer *et al.* (2015).

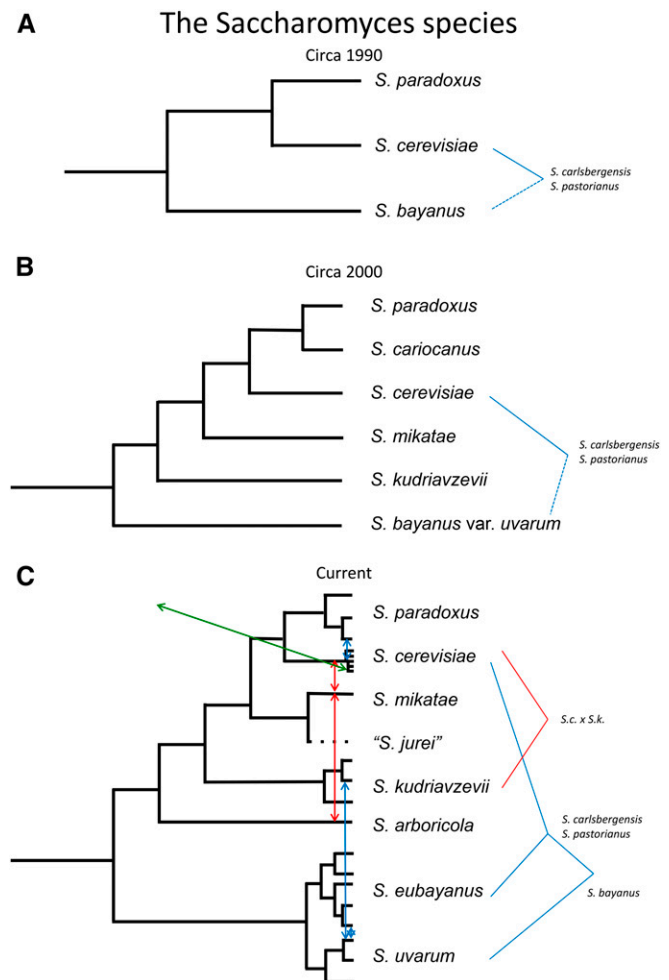
The review is made of two parts. The first one focuses on *Saccharomyces*, the best-studied genus of yeasts containing, in addition to well-characterized laboratory strains, wild isolates of several species spread on all continents as well as domesticated strains and numerous hybrids. The second part explores the entire subphylum of Saccharomycotina at the genome level, trying to identify the major events that shaped its diversity across a long evolutionary range. Lessons learned at both scales remarkably converge, providing a coherent picture of the major mechanisms underlying the dynamics of eukaryotic genomes in evolving populations.

## What Did We Learn from Comparative and Population Genomics of the *Saccharomyces* Species Complex?

### *The monophyletic Saccharomyces (formerly S. sensu stricto) clade*

Over the decades, *Saccharomyces* species were defined by various physiological criteria, the biological species definition, and by DNA–DNA reassociation. Species appeared and disappeared in the literature until the time of the start of the yeast genome-sequencing project. Both the biological species definition (Naumov 1987) and DNA–DNA reassociation studies (Vaughan Martini and Martini 1987) resulted in the community settling on three species and the hybrid used in lager production (*S. cerevisiae*, *S. paradoxus*, *S. bayanus*, and *S. pastorianus/carlsbergensis*) (Figure 1A). The isolation and identification of new strains and perhaps new species was hampered by the lack of knowledge of the ecology and natural history of *Saccharomyces* yeasts. A survey of culture collections where budding yeasts had been isolated from various sources, using the biological species definition (interspecies sterility and intraspecies fertility), along with some molecular analysis [pulsed-field gels and ribosomal DNA (rDNA) sequences], did result in the identification of new species and hybrids such that by the turn of the millennium we had doubled their number (Figure 1B). Three new species with apparently limited ranges were described: *S. mikatae* and *S. kudriavzevii* from the Far East, and *S. cariocanus* from Brazil (Naumov *et al.* 1995a,b, 2000). *S. bayanus* appeared to contain both hybrids and a *bona fide* species called *S. bayanus var. uvarum* (Nguyen *et al.* 2011). Other hybrids were found, in particular a hybrid between *S. cerevisiae* and *S. kudriavzevii* used in some European wine production (Gonzalez *et al.* 2006; Lopandic *et al.* 2007).

Of the original species, *S. cerevisiae* was found in many niches, mostly associated with fermentation activities, while *S. paradoxus* was found mostly associated with oak trees around the world. *S. bayanus* was mostly found in fermentation activities as was the hybrid *S. pastorianus*. The new species were found in natural environments, associated with oak trees or insects, and had much more limited (or so was thought at the time of their discovery) geographic distributions. As most of the species could be isolated in the Far East,



**Figure 1** Evolution of the phylogeny of the *Saccharomyces* (formerly *S. sensu stricto*) group since the first genome sequence. (A) By the mid-1980s to mid-1990s, the use of DNA–DNA reassociation (Vaughan Martini and Martini 1987; Vaughan Martini 1989) and the biological species definition (Naumov 1987) led to the consolidation of the *Saccharomyces* yeasts into three species and one hybrid used in lager fermentation. This hybrid was between *S. cerevisiae* and something close to *S. bayanus*, but not *S. bayanus* itself (Casey and Pedersen 1988; Hansen and Kiehlbrandt 1994; Hansen *et al.* 1994). (B) By the late 1990s, the use of the biological species definition, along with electrophoretic karyotyping and presence/absence of specific repeated sequences, on isolates in various culture collections resulted in the discovery of three new species, *S. cariocanus*, *S. mikatae*, and *S. kudriavzevii*, and the refinement of *S. bayanus var. uvarum* as a species while *S. bayanus* itself appears to be a hybrid. (C) In recent years, whole genome sequencing along with genetic analysis has resulted in the current view of the group. One species (*S. cariocanus*) has disappeared based on phylogeny (see text) and three others have been discovered (*S. arboricola*; *S. eubayanus*, the other parent in the lager hybrids; and *S. jurei*). There are many examples of HGT, red arrows, as well as introgressions, blue arrows (see text). Perhaps the most interesting is the HGT of genes that provide useful traits in wine fermentation, green arrow, which distinguishes the wine yeast from wild European yeasts (see text and Fig. 2).

it was speculated that the *Saccharomyces* clade may have originated there (Naumov 1987). This has been supported by more recent surveys of diversity, where most genetic diversity for several of the species is found in the Far East as well (Wang *et al.* 2012). The genomic analysis of hybrids

presented a conundrum to yeast researchers. *S. pastorianus* is a hybrid between *S. cerevisiae* and another species close to but not *S. uvarum* (Casey and Pedersen 1988; Hansen and Kielland-Brandt 1994; Hansen *et al.* 1994). The European *S. cerevisiae* × *S. kudriavzevii* hybrids presented another conundrum, namely that no European population of *S. kudriavzevii* was known, bringing into question where the hybridization took place (Hittinger *et al.* 2004). Improved sampling techniques (Sniegowski *et al.* 2002) and sampling at different temperatures (Sampaio and Goncalves 2008) along with a wider effort in environmental sampling has led to the identification both of the sister species to *S. uvarum*, called *S. eubayanus* (Libkind *et al.* 2011), and of a European population called *S. kudriavzevii* (Sampaio and Goncalves 2008). *S. eubayanus* was first identified in Patagonia which led to a further conundrum as to how the lager hybrid was formed, given that it was likely to be in use prior to Columbus' arrival in the Americas. Further sampling across the globe has revealed populations of *S. eubayanus* in China, Tibet (Bing *et al.* 2014), New Zealand (Gayevskiy and Goddard 2016), as well as in North America (Peris *et al.* 2014), providing a way out of this historical problem. In addition to these new species and populations, other species have been discovered. *S. arboricola* was recently found in China (Wang and Bai 2008) and subsequently in New Zealand (Gayevskiy and Goddard 2016), and very recently a novel species has been found in France, provisionally called "*S. jurei*" [MycoBank registration number MB819910 (<http://www.mycobank.org>); Naseeb *et al.* 2017], which is a sister to *S. mikatae*. We now have three more species and have lost one (see species discussion below on *S. cariocanus* and the concepts of biological species vs. phylogenetic species), and have several distinct populations of many of the species. Our current understanding of the clade is shown in Figure 1C and consists of *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. jurei*, *S. kudriavzevii*, *S. arboricola*, *S. uvarum*, and *S. eubayanus*, along with numerous hybrids involving many of the species.

As sampling improves and sequencing becomes less expensive, we have increased our knowledge of *S. cerevisiae* and its relatives and have revealed evolutionary processes underlying the evolution of this clade. More species are being found and new populations of species previously thought to have limited geographic distributions are being discovered. This may not supplant the Far East as the origin of the clade but does lead to the conclusion that perhaps we can find *Saccharomyces* yeasts everywhere if we look hard enough. The new field of environmental DNA (Martin and Rygiewicz 2005; Thomsen and Willerslev 2015) may shed light on the true distribution of *Saccharomyces* yeasts.

### **Chromosome evolution within the clade**

As sequencing was still expensive and time consuming at the time of the first genome being completed, comparative analysis was done via physical- and limited-sequence comparisons. This more limited approach was still valuable in terms of determining phylogenetic relationships and evolutionary processes. Individual representatives of most of the species

known at the time were sequenced [*S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, and *S. uvarum* (Cliften *et al.* 2003; Kellis *et al.* 2003)], which allowed a more detailed comparison of genomes. Since then there has been a more complete, high quality assembly of these genomes (Scannell *et al.* 2011). The *Saccharomyces* clade shares a great deal of gene content and synteny, having settled down to a similar chromosome karyotype after more rapid chromosomal rearrangements and deletions following the whole-genome duplication (WGD) (Scannell *et al.* 2006). Other post-WGD species, including *Candida glabrata*, *Naumovozyma castellii*, and *Vanderwaltozyma polyspora*, have very different karyotypes having diverged from the *Saccharomyces* clade soon after the WGD (Scannell *et al.* 2006, 2007a).

Early electrophoretic karyotype analysis of some of the species (Naumov *et al.* 1992) was followed by a more thorough physical analysis of the six known species in 2000 (Fischer *et al.* 2000). Using probes designed for each chromosome arm and each centromere region, their gross chromosomal structure and breakpoints of large rearrangements between species was described. This analysis was able to map the breakpoints of several translocations relative to *S. cerevisiae* in the clade. Most of these were coincident with Ty elements or Ty LTRs. Two of the species were collinear with *S. cerevisiae* and these were not correlated with the phylogenetic relationships. This indicated that the *S. cerevisiae* chromosome configuration was likely the ancestral configuration. These rearrangements were therefore not the driver of reproductive isolation leading to speciation across the whole clade. The translocations that were mapped were nonrandomly distributed with respect to phylogenetic branch lengths and indeed appeared to have occurred in bursts, perhaps due to Ty activity. In particular, the *S. cariocanus* karyotype had four reciprocal translocations relative to the very closely related *S. paradoxus*, which is collinear with *S. cerevisiae*, the next closest relative. The idea of bursts of Ty activity has been supported by more recent population genomic sequencing (Liti *et al.* 2009a) where the Ty content of the two *S. cariocanus* isolates, with four reciprocal translocations mapped, have the highest number of Ty elements, second only to S288C; whereas most *S. paradoxus* isolates, to which *S. cariocanus* phylogenetically belongs, have lower numbers of Ty elements than the majority of *S. cerevisiae* strains. Other examples of large numbers of gross chromosomal rearrangements (GCRs), such as the Malaysian *S. cerevisiae* lineage (Liti *et al.* 2009a; Cubillos *et al.* 2011), are also likely to be due to bursts of Ty activity.

### **Gene losses and differentiation of function**

The individual genome sequences from one strain each of some of the species supported these earlier physical studies but also led to some interesting findings. For example, the entire galactose-utilization (GAL) pathway in *S. kudriavzevii* (Hittinger *et al.* 2004) was inactivated due to various mutations in the genes, which could have been due to the independent accumulation in each gene, leading to pseudogenes. This was found in an isolate of the species from the Far East. The more recent finding of a European population supports

the nature of this loss as this population has a functioning GAL pathway (Hittinger *et al.* 2010). Finding this population also solved one of the mysteries of where yeasts originated, given that many European wine strains are *S. cerevisiae* × *S. kudriavzevii* hybrids but no European *S. kudriavzevii* population was known before. Another example where mutation accumulation between species occurs came from the use of genome annotations as a tool for comparative analysis. When the *S. uvarum* gene annotations were compared to those of *S. cerevisiae*, as many as 35 apparent breakpoints were discovered (Fischer *et al.* 2001), yet the physical analysis of the chromosomes described only 8 due to four reciprocal translocations (Fischer *et al.* 2000). This was rectified by direct sequence comparisons where relics of genes were found between open reading frames (ORFs) in one species compared to the other, resulting in loss of nearest neighbor synteny contacts when only annotated ORFs are used (Lafontaine *et al.* 2004). There can clearly be rapid accumulation in genes and sets of genes in one lineage compared to another.

As these species are derived from a WGD ancestor, there are many examples of duplicated genes, and these can lead to differentiation of function when one or the other is not lost. Two examples are the *GAL1-GAL3* (Hittinger and Carroll 2007) and the *SIR2-HST1* (Rusche and Rine 2001) pairs of duplicates. Other examples have been found in comparison of the reference genomes for each species in the clade (Scannell *et al.* 2011). Most of the differentiation occurred prior to the expansion of the clade but there is still further differentiation among the species (Kuang *et al.* 2016).

### Population genomics

The sequencing of representatives of some of these species was followed after a while by the sequencing of multiple individuals within species, and this has recently been followed by whole-genome sequencing of hundreds to thousands of individuals, mostly of *S. cerevisiae*. In the days prior to large-scale, second-generation sequencing, sample sequencing of several genes was used to look at diversity and phylogenetic relationships. A few individuals were completely sequenced for specific purposes. EC1118, a wine strain (Novo *et al.* 2009); YJM789, a clinical isolate (Wei *et al.* 2007); W303, another laboratory strain (Ralser *et al.* 2012); Sigma1278b, another laboratory strain (Dowell *et al.* 2010); and CEN.PK, an industrial strain (Nijkamp *et al.* 2012); were sequenced. This was in addition to several studies of partial sequences in *S. cerevisiae* (Fay and Benavides 2005) and *S. paradoxus* (Tsai *et al.* 2008), as well as across the clade (Liti *et al.* 2006). Whole-genome diversity was also assessed using microarrays based on the S288C sequence (Gresham *et al.* 2006; Schacherer *et al.* 2007a, 2009).

The early population genomics survey of *S. cerevisiae* led to the conclusion that there have been two domestication events, one in wine strains and one in sake strains (Fay and Benavides 2005; Schacherer *et al.* 2009). A comparison of *S. cerevisiae* variation to that seen in other species revealed much less genetic variation among *S. cerevisiae* isolates from

diverse sources, which was a paradox given the phenotypic diversity that was observed. Populations of *S. paradoxus* were much more diverse from each other and correlated with geographic boundaries, with no such correlation seen in *S. cerevisiae* (Liti *et al.* 2006). The two *S. cariocanus* isolates were very close in sequence to the North American *S. paradoxus* isolates, bringing into question their designation as a species as there must have been very recent gene flow between them (Liti *et al.* 2006). Within *S. paradoxus* there was determination of the population genetics parameters of the species with a measure of how much sex and outbreeding occurred within a population (Tsai *et al.* 2008). It is clear that most evolution in the clade is by clonal expansion from vegetative reproduction with a minor contribution of sexual reproduction, however there is some sexual reproduction and recombination and they can have a great deal of influence.

The first large-scale sampling of genomes from *S. cerevisiae* as well as *S. paradoxus* was done using Sanger sequencing and resulted in the start of increased efforts into population genomics surveys (Liti *et al.* 2009a). At the same time, a microarray survey was done on a larger number of *S. cerevisiae* strains (Schacherer *et al.* 2009). The two studies were consistent with each other and there was some overlap in the strains analyzed. The main findings included: there are several distinct populations of *S. cerevisiae* that are equidistant from each other across their genomes, so called clean lineages; there are several mosaic strains that appear to be the result of recent outbreeding between these clean lineages; and that genome wide there is nothing special about the wine or sake populations, even if they have specific variants associated with adaptation to the fermentation process in which they are usually found (Perez-Ortin *et al.* 2002). One hypothesis is that the fermentation properties desired already existed in extant populations of yeast and that our activity sampled and selected the strains with the best properties. The mosaic strains clearly had some sequence variation that came from outside the five clean populations described, indicating that there were likely more populations to be discovered. The survey was consistent with the previous sample sequencing and found that the *S. paradoxus* populations were clean with little or no gene flow between them. The originally characterized *S. cariocanus* was clearly embedded within the North American population based on shared sequence variation, despite the reproductive isolation.

### Clean lineages vs. mosaics

The concept of clean lineages merits some additional discussion, especially in *S. cerevisiae*. Clean lineages were defined as those populations where the phylogenetic relationship with other populations was the same across the genomes, where the topology of the phylogenetic tree remains the same for any segment of the genome being compared (Liti *et al.* 2009a). This would be expected for populations that are evolving independently with little or no gene flow between them, as is seen for the *S. paradoxus* populations in which the genetic diversity is completely correlated with geographic

location. The survey of 36 *S. cerevisiae* genomes revealed that about half fell into five clean lineages by this definition, while the other half were comprised of segments from two or more of the clean lineages with a few segments from unknown populations. As many of these mosaic strains were associated with human activity—such as bread making, food spoilage, fermentation, and even clinical cases—the origins of these recent outbred strains could be due to human activity providing the opportunity for interbreeding. The contrast with *S. paradoxus* could be a partial explanation for the higher levels of phenotypic diversity seen in *S. cerevisiae* despite lower overall genetic diversity. The genetic diversity that exists in *S. cerevisiae* is more mixed up due to outbreeding, which may result in an expansion of the phenotypic space.

Such a global genome analysis has not yet been done on other populations, so it is not known whether the wild populations from China (Wang *et al.* 2012) or those associated with other human activities (Cromie *et al.* 2013; Arana-Sanchez *et al.* 2015; Ludlow *et al.* 2016) are clean lineages or mosaics, though admixture was detected using genome-wide analysis of variants using restriction site associated DNA sequencing (Cromie *et al.* 2013; Ludlow *et al.* 2016). It would be very interesting if some populations associated with human activity are mosaic with genetic combinations that make them useful for particular activities, such as bread making. It is clear that some fermentation-associated populations are clean lineages, such as the sake and the wine populations. A large number of *S. cerevisiae* strains are now sequenced or about to be sequenced (Strope *et al.* 2015b; <http://1002genomes.u-strasbg.fr/index.html>; <http://www.ncyc.co.uk/yeast-treasure-trove-goes-live/>) and this collection will provide an unparalleled view of diversity within the species, addressing issues of selection and movement by human activity as well as origins of particular adaptations.

### **Genetic vs. phenotypic diversity**

Another result of the survey was confirmation that there was greater genetic diversity in *S. paradoxus* than in *S. cerevisiae*. This result was surprising as the strains were all tested for growth under a large number of conditions and it was found that the phenotypic diversity of *S. cerevisiae* was much greater than that of *S. paradoxus*. Despite this paradox, there was a significant correlation between genetic diversity and phenotypic diversity in the two species. One hypothesis explaining this paradox is the level of outbreeding seen in *S. cerevisiae*. Perhaps the outcrossing, likely facilitated by human activity, has led to the increased phenotypic diversity by bringing together novel combinations of variants at different loci.

More diversity has been found in *S. cerevisiae* with the discovery of at least 10 new populations from China, many from primeval forests (Wang *et al.* 2012). Whole-genome analysis has yet to be completed but it is clear there is greater diversity in these isolates as seen in Figure 2. Some of the populations from China are close to those previously described (Liti *et al.* 2009a), but others, particularly the ones from primeval forests (populations CHN I, II, and III in Figure 2)

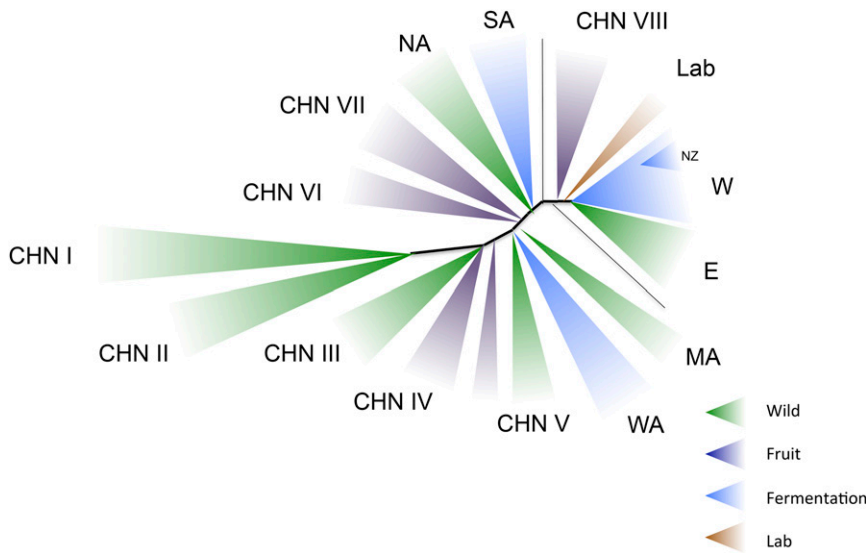
nearly doubles the known diversity. In addition, distinct populations of *S. cerevisiae* are being found, which are associated with other fermentation activities such as cacao (Cromie *et al.* 2013; Arana-Sanchez *et al.* 2015; Ludlow *et al.* 2016) and coffee (Ludlow *et al.* 2016) processing.

There are now thousands of *S. cerevisiae* isolates from a variety of sources. Many cluster with their source even if not with location. Figure 2 is a composite of *S. cerevisiae* populations sampled over the years and analyzed by different means. The most sampled population is the wine/European, which has subpopulations found around the world wherever wine is produced. One subpopulation analyzed extensively is that from New Zealand which must have arrived with human migration (Goddard *et al.* 2010; Gayevskiy and Goddard 2012). A sister population was recently found; it is associated with oaks in Europe but not used in fermentation (Almeida *et al.* 2015). This strain differs by the lack of alien DNA from other species that confers useful properties for wine making (Novo *et al.* 2009). There are numerous wild populations from around the world as well as distinct populations associated with various human fermentation activities including beer, chocolate, coffee, *etc.* (Legras *et al.* 2007; Sicard and Legras 2011; Tapsoba *et al.* 2015; Ludlow *et al.* 2016). There are also intermediate populations associated with fruit trees and orchards. As can be seen in Figure 2, almost every population associated with fermentation or fruit trees has a close wild sister population.

As more populations of different species are being found, whole-genome surveys are being expanded, and now there is good data on populations of *S. uvarum* (Almeida *et al.* 2014), *S. eubayanus* (Almeida *et al.* 2014; Peris *et al.* 2014, 2016a), and *S. kudriavzevii* (Hittinger *et al.* 2010; Peris *et al.* 2016b), in addition to *S. cerevisiae* and *S. paradoxus*. These will make for interesting comparisons with *S. cerevisiae*, particularly for those species with little human influence.

### **Domestication or harnessing existing potential**

Given the large number of surveys of *S. cerevisiae* from a wide variety of locations and environments, we can address the issue of domestication more thoroughly. The wine/European population is the most widespread due to human wine production, and the introduction and spread of these yeasts can be observed in some cases (Gayevskiy and Goddard 2012, 2016). The question of domestication, at least with the wine/European population, can be informed by the very recent discovery of a wild population from oak trees that is sister to the wine/European population (Almeida *et al.* 2015). This finding supports the hypothesis that rather than domestication, our use in fermentation took advantage of the properties of an existing population of yeast with perhaps some adaptations to the wine-making process (see below for discussion of introgressions). However, there are phenotypes and specific genetic variants associated with use in wine fermentation (Gallone *et al.* 2016). In ale and lager strains, there is stronger evidence of selection on specific genes (Gallone *et al.* 2016; Goncalves *et al.* 2016) as well as on



**Figure 2** The populations of *S. cerevisiae* as an unrooted cladogram (not to scale) composite from several studies (Liti *et al.* 2009a; Wang *et al.* 2012; Almeida *et al.* 2015). Populations shown include North American (NA), sake (SA), West African (WA), Malaysian (MA), and wine/European (W) as previously defined (Liti *et al.* 2009a), with E as the wild European sister clade to the wine group, recently described (Almeida *et al.* 2015). Eight additional populations from China (CHN I–CHN VIII) were recently described (Wang *et al.* 2012). In addition a New Zealand (NZ) subpopulation of the wine population has been described (Gayevskiy and Goddard 2012). The original distinction of two domestication events leading to wine and sake fermentation (Fay and Benavides, 2005; Cromie *et al.* 2013; Arana-Sanchez *et al.* 2015; Ludlow *et al.* 2016) has been extended as more populations have been sampled. These include wild populations from oak and related trees, populations associated with fruit trees, and populations associated with fermentation activities. It is clear that many of the

populations associated with fermentation are very close to wild populations exemplified by the wine vs. European sister groups, which differ (Almeida *et al.* 2015) in the presence/absence of the genes acquired by HGT from outside the clade (Novo *et al.* 2009). The wine population has been spread around the world from its European origins and in some cases, such as New Zealand (Gayevskiy and Goddard 2012), the subpopulation can be placed within the larger population. There are now descriptions of populations associated with ale fermentation not distinguished here (Gallone *et al.* 2016; Goncalves *et al.* 2016). Lab, laboratory.

the whole genome (Baker *et al.* 2015) from sequence comparisons. Of the other species used in human fermentation activities, there are probably not enough populations sampled to address the domestication issue. Many strains used in fermentation are hybrids of other *Saccharomyces* species with *S. cerevisiae* (Figure 1C), and it may be that the existence of hybrids preceded their use in fermentation.

### Subtelomeres and the pan-genome

One thing that became apparent from whole-genome sequence surveys is that you cannot study what you do not know exists. With microarrays and other hybridization techniques, you can only assess variation in known sequences. Several genes and gene families were known from yeast studies that are not found in the S288C genome. Six of these gene families involved in carbon-source utilization are located in the subtelomeres (Liti *et al.* 2009a). Subtelomere variation and dynamics was already recognized (Louis and Haber 1990, 1992; Louis *et al.* 1994; Louis 1995) and it was known that the region was a hotbed of gene-family evolution (Brown *et al.* 2010). The analysis of the *Saccharomyces* Genome Resequencing Project collection revealed 38 new gene families not previously known to be in *S. cerevisiae*, in addition to the six known families (Liti *et al.* 2009a). These varied in presence/absence as well as in copy number across the *S. cerevisiae* and *S. paradoxus* isolates. It is clear that as a species, *S. cerevisiae* is much more than merely represented by S288C.

This pan-genome has become important in the study of genetic variation underlying phenotypic variation among strains. A recent set of tools has been developed to help assess genetic material not found in the reference genome (Song

*et al.* 2015). Quantitative genetic studies have found that much of the causal genetic variation for a given trait maps beyond the ends of assembled contigs into the subtelomeres. In any given strain, there may be ~8% of the genome that is considered subtelomeric. In most quantitative traits analyzed, up to 25% of the causal segregating variation maps to the subtelomeres (Cubillos *et al.* 2011; Liti and Louis 2012). One explanation of the higher phenotypic diversity on *S. cerevisiae* compared to *S. paradoxus* is subtelomeric variation. A total of 50% of the phenotypic variation difference can be explained by copy number and presence/absence variation in subtelomeric genes, which is greater in *S. cerevisiae* than in *S. paradoxus* (Bergstrom *et al.* 2014). This leaves 50% to be explained perhaps by outbreeding, supported by the expansion of the phenotypic variation in every cross between clean lineages as described above.

### Horizontal gene transfers

The sequence of a wine strain, EC1118, revealed several genome segments whose origin was from outside *S. cerevisiae*, but within the *Saccharomyces*, and some outside the clade (Novo *et al.* 2009; Marsit *et al.* 2015). These segments conferred useful properties to the wine niche and therefore may be considered adaptive. One of these appears to have originated from a wine spoilage yeast and so it is possible that somehow DNA from this was incorporated into a wine strain either by some level of cellular interaction or by uptake from the environment. It is certainly possible that the two were occupying the same vessel at the same time during fermentation. The structure of the insertions in various strains indicated a circular intermediate that integrated independently in different configurations (Borneman *et al.* 2011; Galeote *et al.* 2011). All wine strains

have some permutation of this configuration and the other horizontal gene transfers (HGTs), and so it has spread through the yeast used by the wine community. Another example is in Ty2 elements, which are not thought to be infective though they do generate virus-like particles within the yeast cell. *S. cerevisiae* and *S. mikatae* have many copies, yet *S. paradoxus* in between has none. Either *S. paradoxus* has lost Ty2 or never had it. The comparison consensus sequences of Ty2 in *S. cerevisiae* and *S. mikatae* indicate a time of divergence much more recently than the rest of the genome, and more recently than *S. paradoxus* and *S. cerevisiae* diverged (Liti *et al.* 2006). The most consistent explanation is that Ty2s moved from one to the other in the recent past, postspeciation.

### **Hybrids, introgressions, and reticulate evolution**

One of the surprising findings of population genomics surveys is the introgression of sequence (homologous replacement) seen between species. Such introgression requires the ability to hybridize, which all members of the clade have, and some breakdown of the reproductive isolation allowing gene flow from one species into the other. The first examples were between *S. paradoxus* and *S. cerevisiae*. When one genome sequence of *S. paradoxus* was available it became clear that there was a segment with little (0.1%) divergence compared to the S288C sequence, whereas the rest of the genome was 10–15% divergent (Liti *et al.* 2006). Upon analysis of the region in several isolates, it was shown that the entire European population of *S. paradoxus* had an introgression from *S. cerevisiae* not seen in the other populations. Interestingly, this was in a subtelomere. Following this observation, introgressions in both directions have been found in surveys of *S. cerevisiae* and *S. paradoxus* (Muller and McCusker 2009). As more populations of different species are surveyed, more examples of such introgressions are found (Almeida *et al.* 2014). For example, there are numerous introgressions of *S. uvarum* into *S. eubayanus* and vice versa (Peris *et al.* 2014). There are introgressions from *S. kudriavzevii* into these two species as well. Given that these species are used in human fermentation activities, the potential opportunity for genetic exchange exists.

There are numerous examples of hybrids between yeast species (see *What Did We Learn from Comparative Genomics of Other Saccharomycotina?*). They are generally sterile and therefore dead ends in terms of long-term evolution, as no sexual recombination is possible anymore. The question then arises as to how segments from one species can move into the genome of another, clearly via homologous recombination, when the hybrids are sterile? The usual thought of how this occurs is through a rare viable spore backcrossing with one of the parents, with subsequent backcrosses becoming easier and easier. The alternative is a single step homologous introgression in a hybrid that does not have to undergo numerous backcrosses, as has been observed in artificial *S. cerevisiae* × *S. uvarum* hybrids (Dunn *et al.* 2013).

The evolution of the *Saccharomyces* clade appears to be reticulate rather than bifurcating as is usually portrayed in

phylogenetic trees. Within populations there is clearly mostly vegetative reproduction. However, there is some outbreeding and sexual recombination so that within a population there is breakdown of linkage disequilibrium and some level of outcrossing (Tsai *et al.* 2008). Between populations there is less gene flow, however there is admixture in some cases and certainly the movement of yeast through human activity has provided the opportunity for gene flow between populations. This is reticulate evolution and expected for sexually reproducing species. We generally think that once speciation has occurred then gene flow stops and we move from reticulate to bifurcating phylogenies, but the population genomics surveys of *Saccharomyces* species indicates otherwise. Perhaps gene flow is possible, even if rare, across the whole of the clade, leading to connections between even distant branches. This requires the exploration of reproductive isolation and its role in speciation.

### **Reproductive isolation and speciation**

What is a species and how will we move into the future with species concepts? We have seen above that under one definition, the biological species definition, we define *S. cariocanus* as a new species distinct from the others in the clade. However under another definition, the phylogenetic species definition, we place these isolates into the American population of *S. paradoxus*. There are numerous examples of individuals and populations having GCRs relative to closely related other strains, which would make them a species under one definition but not the other. An example is the Malaysian population of *S. cerevisiae*. It has not been labeled as a new species but as *S. cerevisiae* based on its sequence and physiological characteristics, yet it has at least 10 breakpoints from GCRs that preclude fertility when crossed to any other *S. cerevisiae* (Liti *et al.* 2009a; Cubillos *et al.* 2011). If *S. cariocanus* is a new species then this population should also be one. If it remains *S. cerevisiae*, then *S. cariocanus* should be relabeled *S. paradoxus*.

Speciation is generally thought to arise from reproductively isolated populations that continuously diverge from each other. There are three major hypotheses for the underlying mechanism of speciation, two of which are found in other systems and are well established. As discussed above there are numerous examples of introgressions between species, indicating that reproductive isolation is not complete which may bring us to question the concept of species in these yeasts (Louis 2011).

**GCRs:** It is clear that there are many GCRs segregating within populations of a species and these can result in partial to near complete reproductive isolation, which could then lead to eventual speciation. Is this a major driver of speciation within the *Saccharomyces* clade? In other Ascomycetes clades, such as the *Lachancea* clade (Vakirlis *et al.* 2016), there are large numbers of GCRs between each pair of species and it cannot be ruled out that GCRs could have been an important driver of the speciation process. In the *Saccharomyces* clade, there are several species that are collinear yet are species by all



definitions, and so GCRs could not have been involved in their speciation. There are segregating translocations and large inversions that do cause spore inviability within species and populations and it may be that they are leading to eventual speciation (Hou *et al.* 2014), but as a clade these are not the major drivers of speciation. Even in cases with many GCRs such as the South American *S. paradoxus* vs. the North American population, or the Malaysian population of *S. cerevisiae* vs. other populations, there is clear evidence of gene flow as they share SNPs with one population but not others (Liti *et al.* 2009a).

**Bateson–Dobzhansky–Müller incompatibilities:** The search for speciation genes has not been very successful in many taxa, and in the *Saccharomyces* clade there are only a few examples of Bateson–Dobzhansky–Müller (B-D-M) incompatibilities leading to inviability. Dominant B-D-M incompatibilities between species have been ruled out by testing the meiotic fertility of allotetraploids (Greig *et al.* 2002a,b) where hybrid diploids exhibit little or no spore viability, but tetraploidy restores high levels of spore viability. Individual chromosome replacements of *S. paradoxus* into *S. cerevisiae* failed to find any recessive B-D-M incompatibilities (Greig 2007, 2009). In some pairs of species, a nuclear–mitochondrial incompatibility was found (Lee *et al.* 2008; Chou and Leu 2010) but it is not clear whether this arose postspeciation or not (Louis 2009, 2011). Recently a classical B-D-M incompatibility pair was found segregating in *S. cerevisiae* (Hou *et al.* 2015). This was between a nonsense mutation in a nuclear-encoded mitochondrial gene and a transfer RNA (tRNA) suppressor of the mutation. This example would never lead to speciation as the tRNA suppressor would have many detrimental pleiotropic effects. Strong divergent selection in laboratory evolution experiments can lead to B-D-M incompatibilities, as has been demonstrated (Dettman *et al.* 2010), and so there is potential for such in nature.

Despite this paucity of inviable B-D-M incompatibilities, there are numerous examples of incompatibilities resulting in lack of function of important processes. The first example was in the mismatch-repair system where a particular combination of segregating variants of *PMS1* and *MLH1* products in *S. cerevisiae*, which act together as a heterodimer, exhibited a null phenotype (Heck *et al.* 2006). Similarly segregating variants of *YKU70* and *YKU80* in *S. paradoxus* exhibited a null phenotype in one specific combination (Liti *et al.* 2009b). In crosses between the clean lineages of *S. cerevisiae*, two of six pairwise combinations resulted in a global reduction of homologous recombination during meiosis, with one of these exhibiting reduced spore viability due to loss of genetic interference (Cubillos *et al.* 2011). In any species, such as in the *Saccharomyces* clade, where populations diverge from each other yet maintain coadaptation within, it might be expected that incompatibilities arise for combinations of variants that break up coadapted gene complexes (Liti and Louis 2012). Whether these will be drivers of reproductive isolation leading to speciation remains to be demonstrated.

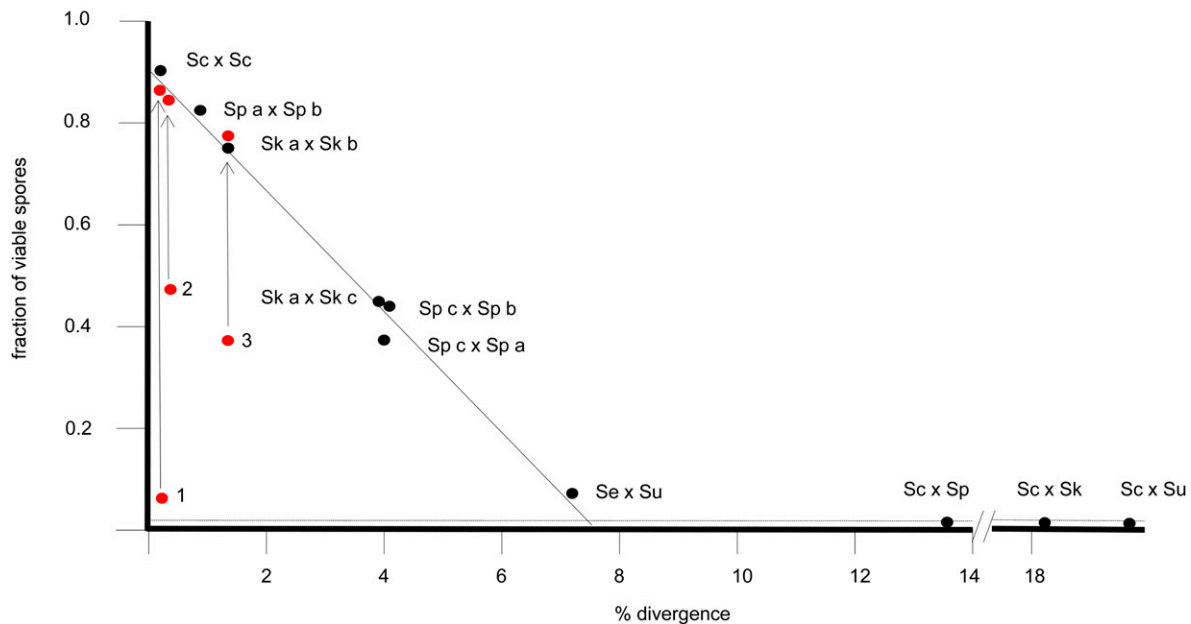
**Sequence divergence and mismatch repair:** A third mechanism that applies across the entire clade is simple sequence divergence and the action of the mismatch-repair system (Liti *et al.* 2006; Louis 2011). Here the intermediates of homologous recombination; required for crossing over, chiasmata formation, and proper chromosome segregation; contain numerous sequence differences between diverged parents. These are recognized by the mismatch-repair system and rather than being “repaired” to one or the other parental sequence, as would happen for a single mismatch, the large number of differences lead to abortion of the intermediate. The end result is no recombination and aneuploidy due to random segregation of the chromosomes (Chambers *et al.* 1996; Hunter *et al.* 1996). The test of this hypothesis is the restoration of some spore viability associated with increased crossing over when the mismatch repair system is deleted (Chambers *et al.* 1996; Hunter *et al.* 1996). This notion holds for the closely related species *S. paradoxus* and *S. cerevisiae* (Chambers *et al.* 1996; Hunter *et al.* 1996) but also for diverged populations within each of the species (Greig *et al.* 2003), indicating that it is acting at the right stage for incipient speciation.

In Figure 3 there is a clear linear relationship between sequence divergence and fertility/spore viability. The off-line examples in specific crosses are accounted for by one or more reciprocal translocations and, when corrected for these, the expected fertility fits the linear relationship. In a recent survey of many isolates (Hou *et al.* 2014), where 60 were crossed to the S288C background to measure spore viability, the conclusion was that sequence divergence was not correlated with the level of spore viability and that GCRs were the major mechanism underlying the observations. There are two problems with this conclusion. One is that they surveyed a narrow range of divergence (0.1–0.5%) where not much change in viability is expected and the divergence in each measure of viability would be greater than any signal of reduced viability with increased divergence. The second and more important issue is that the S288C strain is a mosaic sharing large segments of the genome with low sequence divergence with most other isolates rather than evenly distributed divergence. These low diversity regions will allow normal recombination and chromosome segregation, alleviating any problems elsewhere in the genome.

### **Remaining issues raised by *Saccharomyces***

The yeast community needs to deal with the definition of species. On the one hand, the biological species definition provides a functional definition, which coincides with phylogeny in the majority of cases. On the other hand, there are reproductively isolated populations whose isolation is due to GCRs yet there has not been sufficient time for accumulated sequence divergence. Phylogenetically, *S. cariocanus* is in the American population of *S. paradoxus* yet is reproductively isolated by infertility. Similarly, the Malaysian population of *S. cerevisiae* is reproductively isolated yet has not been designated as a new species. Perhaps the definition has to be a practical one fit for the specific purpose at hand.

# Fertility versus Divergence



**Figure 3** Fertility (spore viability) vs. sequence divergence. The biological species definition is based on within-species fertility but between-species sterility after mating. The potential causes of postzygotic reproductive isolation include GCRs, B-D-M incompatibilities, and sequence divergence itself acted upon by the mismatch-repair system during meiotic recombination. Most of the interspecies hybrids exhibit levels of spore viability that are in the range of random segregation of chromosomes leading to an accidental viable combination of <1%. In populations that have diverged yet are not separate species [three populations of *S. paradoxus* (Sp a, b, and c) and three populations of *S. kudriavzevii* (Sk a, b, and c) as well as in the hybrid of the very closely related species *S. uvarum* and *S. eubayanus*, there is a linear relationship between sequence divergence and spore viability (Liti *et al.* 2006; Hittinger 2013). Some close strains by sequence do exhibit spore viability less than expected by the linear relationship; however, when corrected for known translocations, the viability returns to the expected relationship. This is particularly relevant for the *S. cariocanus* by North American *S. paradoxus* (1) where the sequence divergence is small and there are four previously described reciprocal translocations (Fischer *et al.* 2000) and other large GCRs discovered by complete assembly of the genome (D. Delneri, personal communication). Other examples where the correction of GCRs reestablishes the sequence divergence-spore viability relationship include strains of *S. paradoxus* from the Far East population (2) where one has a translocation, and strains from the European population and Far East populations differing by one translocation (3).

Are species truly reproductively isolated? There is increasing evidence of rampant reticulate evolution in the *Saccharomyces* yeasts, with introgressions jumping across long-established separated lineages. These examples are clear evidence of the opportunities for interbreeding and indeed hybrids are found for many pairs of the species in various sets of samples. At what stage of reproductive isolation do we decide we have a species? Perhaps the level of divergence where the spore viability is as low as that of random segregation of chromosomes would be one molecular measure of phylogenetic species. Although the frequencies of such reticulate evolution in different taxa have not been analyzed sufficiently for comparison, there is evidence that such gene flow is more prevalent in many taxa than previously thought (Mallet *et al.* 2016).

What is domestication? Are the *Saccharomyces* yeasts used in fermentation activities like dogs while the wild cousins are wolves, or is the clade more like cats moving in and out of feral and tamed niches? It is clear in the wine population of *S. cerevisiae* that the utility in wine is partly due to the introgression and HGT of genes from within and outside the clade, conferring useful properties such as those beneficial to

fermentation (Marsit *et al.* 2015). A wild sister population exists in Europe which is not tamed.

What is the genome of a species? It is clear that as a whole, *S. cerevisiae* has many more genes/sequences than present in the first genome sequenced. How can we describe this pan-genome and how relevant is it to understanding yeast ecology, evolution, and adaptation to specific niches, in particular those associated with human activity? How prevalent is presence/absence and copy number variation in subtelomeres of the other *Saccharomyces* species and of Ascomycetes yeasts as a whole?

## What Did We Learn from Comparative Genomics of Other Saccharomycotina?

### *The rise of comparative yeast genomics: a brief history*

Considerable progress in our understanding of yeast genomes and their evolution has also been made over the last decade by comparison between distantly related species. This strategy, however, faced several difficulties at the start, not the least of

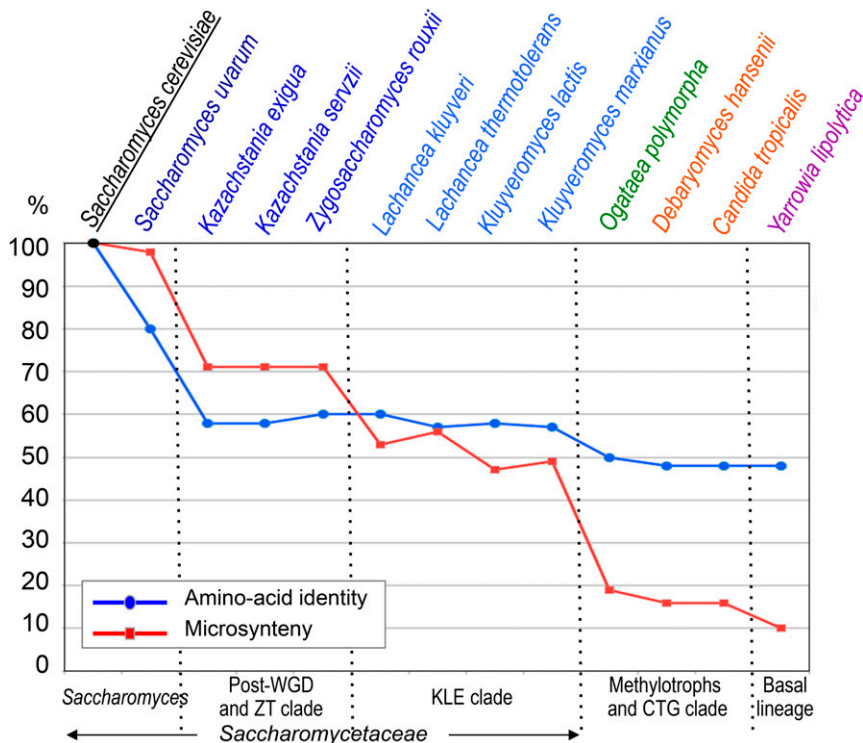
which being the application of relevant criteria for appropriate taxon sampling in a field traditionally dominated by practical considerations in which ecology, metabolic properties, or pathogenic characters of each yeast attracted greater attention than the evolution of their genomes. Even the definition of species in yeasts has remained an open question until now, and their taxonomy has considerably evolved over time. Today, >1500 yeast species have been described and classified in a variety of lineages (Kurtzman *et al.* 2011), of which only a subset has been studied at the genomic level so far. In addition, when the complete sequence of the *S. cerevisiae* genome was determined, back in 1996 (Goffeau *et al.* 1996), most scientists involved in this project were still deeply influenced by the concept of the universality of life that formed the basis of early molecular biology and, consequently, they were paying little attention to natural diversity and its evolutionary origin (see Dujon 2015b). What was true for *S. cerevisiae* had to also be true for other yeasts and, beyond, for other eukaryotes as well. Logically, therefore, *S. cerevisiae* immediately became the system of choice to develop functional genomics, a very successful endeavor indeed (Giaever and Nislow 2014), but it provided no insight on the evolutionary origin of its genome. For many people, it was not obvious what could be learned from the comparison of other yeasts in absence of experimental tools for many of them. Following this same logic, the second yeast genome ever completely sequenced was the other favored experimental model of yeast molecular geneticists, the fission yeast *S. pombe* (Wood *et al.* 2002), a representative of the Taphrinomycotina subphylum of Ascomycota. Beside its great intrinsic interest, this genome was, however, much too different from that of *S. cerevisiae* to identify many common traces representative of evolutionary conservation from their very distant common ancestry and to distinguish them from convergence; a phenomenon more frequent than generally anticipated (see below).

In the end, it took no less than 8 years after the completion of the *S. cerevisiae* sequencing project before the first complete sequences of other Saccharomycotina genomes, seven in total, became available for analysis and comparisons. Actually, some of these species had been selected for reasons other than to examine genome evolution (see below). Yet, altogether, these genomes happened to represent interesting samples across the evolutionary range of Saccharomycotina and, consequently, often served as references in subsequent publications. This favorable situation was due in part to the results of an earlier exploration of this subphylum in which 13 different species had been sequenced at low coverage, as permitted by the technology of this time, and compared to *S. cerevisiae* (Souciet *et al.* 2000). This work gave us the first quantitative estimates of the evolutionary spectrum covered by the Saccharomycotina based on sequence divergence between orthologous genes (Malpertuy *et al.* 2000) and loss of microsynteny (Llorente *et al.* 2000). The presence of thresholds in the distributions observed (Figure 4) were the first clues about the existence of several subgroups of species worthy of further examination. If *S. uvarum* showed high

sequence conservation with *S. cerevisiae* (~80% mean amino-acid identity between orthologous gene products), the figure rapidly dropped to <60% for seven other yeasts and to <50% for the last four. Similarly, the stepwise loss of microsynteny (average gene adjacency) separated the *Kazachstania* and *Zygosaccharomyces* group (with figures close to 70% conservation) from the *Lachancea* and *Kluyveromyces* group (~50%), and from the other yeasts (with figures dropping below 20%).

The first eight complete genome sequences of Saccharomycotina, plus *S. pombe* used as outgroup, are summarized on Figure 5 with indication of the major criterion used for their selection: *Candida albicans* (Jones *et al.* 2004) and *C. glabrata* (Dujon *et al.* 2004) are human pathogens; *Kluyveromyces lactis*, *Debaryomyces hansenii*, *Yarrowia lipolytica* (Dujon *et al.* 2004), and *Eremothecium gossypii* (Dietrich *et al.* 2004) have biotechnological or agricultural importance. Only *Lachancea waltii* (Kellis *et al.* 2004) was primarily selected as representative of a novel lineage. Initially, *L. waltii*, *E. gossypii*, and *C. albicans* were each compared alone against *S. cerevisiae*, and the first two revealed a characteristic dual synteny supporting the idea that *S. cerevisiae* inherited from WGD followed by extensive gene loss, as initially proposed by Wolfe and Shields (1997). The last four, *C. glabrata*, *K. lactis*, *D. hansenii*, and *Y. lipolytica*, allowed the first multidimensional comparisons of yeast genomes at a global scale and revealed major signatures of the distinct lineages. Combining these data, it became clear that the human pathogen *C. glabrata*, the plant pathogen *E. gossypii*, the lactose-utilizing yeast *K. lactis*, as well as *L. waltii*, shared many genomic features with *S. cerevisiae*; features now known as characteristics of the Saccharomycetaceae family (see below). Alternatively, the human pathogen *C. albicans*, the halotolerant and occasional pathogen *D. hansenii*, and the methylotrophic, alkane-utilizing yeast *Y. lipolytica*, representative of other families, showed very distinct characteristics including, for the first two, the usage of the alternative genetic code (see Santos *et al.* 2011) and for the last one a larger and less compact genome (see Kelkar and Ochman 2011).

As the importance of genomics rapidly grew in subsequent years, additional yeast genomes were sequenced including, for the earliest ones: *Scheffersomyces stipitis*, a xylose-fermenting yeast (Jeffries *et al.* 2007); *V. polyspora*, representing an early branching lineage after the duplication (Scannell *et al.* 2007b); *Komagataella phaffii*, the commonly used host for heterologous protein production previously designated *Pichia pastoris* (De Schutter *et al.* 2009; Mattanovich *et al.* 2009); *Lachancea thermotolerans*, *Lachancea kluyveri*, and *Zygosaccharomyces rouxii*, three other Saccharomycetaceae that did not inherit the WGD (Souciet *et al.* 2009); and several pathogenic *Candida* (*C. albicans*, *C. tropicalis*, *C. dubliniensis*, and *C. parapsilosis*) plus the related species *Lodderomyces elongisporus*, *Meyerozyma guilliermondi*, and *Clavispora lusitaniae* (Butler *et al.* 2009; Jackson *et al.* 2009). Other, incomplete genome sequences were also released during the same period, some revealing interspecies hybridizations. By 2010, >20 Saccharomycotina species had been fully sequenced (Dujon 2010). This number has now risen to >100 (ignoring the many hybrids) if one only



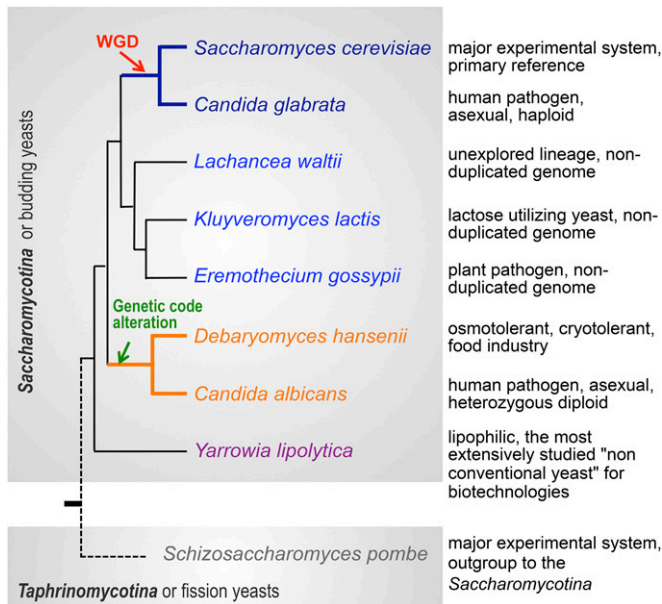
**Figure 4** The first exploration of the evolutionary diversity of Saccharomycotina genomes, from back in 2000. Data taken from Llorente *et al.* (2000), Malpertuy *et al.* (2000), and Souciet *et al.* (2000). The figure summarizes the mean amino-acid identity (blue line) and conservation of gene adjacency (red line) when random sequence tags from each of the yeast species indicated at top were compared to *S. cerevisiae*. Ordinate in percentage. Vertical lines point to major thresholds, separating groups of species whose corresponding affiliation to present subgroups of Saccharomycotina is indicated at bottom.

counts the completely assembled sequences and permanent drafts with sufficiently limited numbers of scaffolds (Table 1, Table 2, and Table 3). Many more are lying in the form of published draft sequences with too large numbers of scaffolds to be useful for genome analyses, or remain publicly nonavailable. A project to sequence the genomes of all described yeast species is planned (Hittinger *et al.* 2015).

#### Four major genomic architectures within the Saccharomycotina

From presently available genome sequences, four major subgroups can be recognized within the Saccharomycotina subphylum of yeasts: (i) the Saccharomycetaceae, by far the most extensively studied family; (ii) the “CTG clade,” a diversified subgroup made of yeast species using an alternative genetic code; (iii) the “methylotroph clade,” exemplified by a few recently sequenced yeasts bearing novel signatures; and (iv) several species belonging to distinct and probably distant lineages altogether regarded as “basal” to the Saccharomycotina subphylum but actually very heterogeneous. The four subgroups differ from one another by the presence/absence of specific genomic signatures (Figure 6) and are clearly separable from one another by overall proteome comparisons. Additional species may eventually illustrate novel genomic architectures or interesting evolutionary intermediates between the four subgroups, but data are presently too dispersed for definitive conclusions. The most interesting ones so far are represented by members of the Ascoidaceae and Lipomycetaceae families (see Table 3), which represent early branching lineages at the basis of the Saccharomycetaceae and all Saccharomycotina, respectively (Shen *et al.* 2016).

Genomes of the Saccharomycetaceae family (Table 1) harbor several signatures that distinguish them from all other subgroups: the presence of point centromeres, the triplication of mating loci, and the absence of genes for complex I subunits of the respiratory chain in their mitochondrial DNA (mtDNA). In addition, they generally contain a single rDNA locus either internal to a chromosome arm or, less frequently, located in subtelomeric regions as a result of independent translocation events during the evolution of the family (Proux-Wera *et al.* 2013). The rDNA locus is made of large tandem arrays of the 35S precursor transcript of ribosomal RNAs (rRNAs), like in all eukaryotes, but also includes the 5S RNA gene, most often in opposite orientation and sometimes duplicated (Bergeron and Drouin 2008). Remarkably, almost all yeasts of this family use the bacterial mode of tRNA decoding instead of the eukaryotic mode for the arginine CGN and the leucine CUN codon families (Marck *et al.* 2006; Grosjean *et al.* 2010; see below). Genomes vary in size between ~9 and 14 Mb (haploid equivalent, ignoring the rDNA locus, mitochondria, and plasmids) but are always highly compact (~65–75% coding) and contain ~4500–5900 protein-coding genes with very few spliceosomal introns [3–5% of coding DNA sequences (CDS) are split]. This family is the most extensively studied, all known genera having now at least one representative fully sequenced with the sole exception of *Zygorhynchus* and *Cyniclomyces* (the latter is basal to the family). The 2- $\mu$ m family of self-replicating plasmids, initially discovered in some isolates of *S. cerevisiae* (Guerineau *et al.* 1971), seems to be circumscribed to this family where it was found in several genera (Blaisonneau *et al.* 1997; Strobe *et al.* 2015a). However, plasmids in general have unfortunately not



**Figure 5** The historical set of fully sequenced yeast genomes, from back in 2004. The figure lists the nine fully sequenced yeast genomes published by year 2004 (original references in text) with indication of their major selection criterion. Tree topology from Kurtzman *et al.* (2011), branch lengths ignored.

attracted sufficient attention in recent genomic works to have a precise image of their evolutionary distribution.

By contrast, sequenced members of the CTG clade (Table 2) have regional centromeres (~3–4 kb long), often made of remnants of mobile elements, generally poorer in GC than average genomes and lacking conserved sequence motif. They have typically only one mating-type locus. Their moderately larger genomes (~10–16 Mb, same definition as above) are also highly compact (~55–70% coding) with ~5600–6400 protein-coding genes and few spliceosomal introns (6–7% of CDS are split). Their major distinctive signature, relative to other subgroups, is the usage of the alternative genetic code in which the CUG codon is used to specify serine instead of leucine (or in addition to it), due to an additional tRNA species (Santos *et al.* 2011). Their mitochondrial genomes contain the seven genes encoding subunits of complex I of the respiratory chain. The different lineages in this large and heterogeneous clade have been unequally studied so far, and numerous genera have not yet been explored at the genomic level.

A smaller number of species have been sequenced so far in the methylotroph subgroup (Table 3). They share common signatures with the CTG clade (regional centromeres, high coding density, and the complete set of genes in mtDNA) except for their usage of the universal genetic code and moderately smaller genomes (~9–13 Mb). They also exhibit a slightly higher number of spliceosomal introns (10–15% of CDS are split). Some species of this subgroup (*K. phaffii*, *Ogataea polymorpha*, and *Ogataea parapolymorpha*) have duplicated mating loci, and haploid cells undergo mating-type switching (Hanson *et al.* 2014; Maekawa and Kaneko 2014; see below). Methylotroph yeasts appear as evolutionary

intermediates between Saccharomycetaceae and basal lineages by their eukaryotic mode of tRNA decoding for the leucine CUN codon family, and bacterial mode for the arginine CGN codon family (Morales *et al.* 2013). Although associated to this subgroup (Kurtzman *et al.* 2011), *K. phaffii* shares common genomic signatures with basal lineages such as the eukaryotic mode of tRNA decoding for both codon families and the dispersion of 5S RNA gene copies in the genome independently of the rDNA locus.

Basal lineages of the Saccharomycotina subphylum (Table 3) have long been represented solely by *Y. lipolytica*, whose genome exhibits a number of important differences with the three previous subgroups, such as its significantly larger size (20.5 Mb), lower compactness (46% coding), multiple rDNA loci located in subtelomeric regions, and dispersed 5S RNA gene (116 copies of this gene are distributed across the genome) like in most Pezizomycotina and Taphrinomycotina (Bergeron and Drouin 2008). With nearly 6600 protein-coding genes, 14% of which are split, it also possesses the largest protein repertoire. Additional species of *Yarrowia* that are now sequenced (see Table 3) confirm these trends. Historically, the second genome sequenced within the basal lineages was *Blastobotrys adenivorans* (Kunze *et al.* 2014). It shares 49% of conserved microsynteny with *Y. lipolytica*, indicating a distant but still recognizable common ancestry. But its small size (12 Mb), high compactness (74% coding), and the presence of a single rDNA locus internal to a chromosome arm are more reminiscent of the evolved subgroups of Saccharomycotina. It has, however, a high number (~6150) of protein-coding genes, 11% of which are split by introns. The genome of *Geotrichum candidum* (Morel *et al.* 2015) provided another example of the diversity of the basal lineages and the long evolutionary distances separating them. With a size (24.2 Mb), compactness (45% coding), and gene number (~6800) comparable to *Yarrowia*, it is remarkably intron rich (35% of CDS are split) for a Saccharomycotina. Similar size (16 Mb), compactness (50% coding), and gene number (6820) were also found for *Sugiyamaella lignohabitans* (Bellasio *et al.* 2016), but only 5% of its protein-coding genes are split by introns. The recently published genomes of *Nadsonia fulvescens* and *Tortispora caseinolytica* (Riley *et al.* 2016) are comparable to *G. candidum* for their high numbers of introns, but are much smaller in size (13.7 and 9.2 Mb) and gene number. The genome of *Lipomyces starkeyi* (Riley *et al.* 2016), representative of the most basal lineage to all Saccharomycotina, is extremely intron rich with an average of three introns per protein-coding gene. The other sequences of the basal lineages reported in Table 3 (mostly unpublished) have not yet been analyzed in detail. *S. lignohabitans*, *N. fulvescens*, *T. caseinolytica*, and *L. starkeyi* have dispersed 5S RNA genes, like *Y. lipolytica*.

#### Origin of the distinctive features of Saccharomycotina genomes

**Introns:** The limited number of spliceosomal introns is one of the most universal signatures of Saccharomycotina genomes

**Table 1 Major reference genomes of the Saccharomycetaceae and related families**

Species	Strain	Genome size (Mb) <sup>a</sup>	No. of chromosomes <sup>b</sup>	Total no. of CDS <sup>c</sup>	Reference or source
<b>Saccharomycetaceae (post-WGD)</b>					
<i>Saccharomyces cerevisiae</i>	S288C	<b>12.2</b>	<b>16</b>	5771	Goffeau <i>et al.</i> (1996)
<i>Saccharomyces paradoxus</i>	CBS432 <sup>T</sup>	<b>12.0</b>	<b>16</b>	5527	unpublished data <sup>d</sup>
<i>Saccharomyces mikatae</i>	IFO1815 <sup>T</sup>	(11.4)	<b>16</b>	6384	Scannell <i>et al.</i> (2011)
<i>Saccharomyces kudriavzevii</i>	IFO1812 <sup>T</sup>	(11.3)	<b>16</b>	5968	Scannell <i>et al.</i> (2011)
<i>Saccharomyces arboricola</i>	CBS10644 <sup>T</sup>	(11.6)	<b>16</b>	5413	Liti <i>et al.</i> (2013)
<i>Saccharomyces uvarum</i>	CBS7001	(11.5)	<b>16</b>	5915	Scannell <i>et al.</i> (2011)
<i>Saccharomyces eubayanus</i>	CBS12357 <sup>T</sup>	(11.7)	<b>16</b>	5515	Baker <i>et al.</i> (2015)
<i>Candida glabrata</i>	CBS138 <sup>T</sup>	<b>12.3</b>	<b>13</b>	5203	Dujon <i>et al.</i> (2004)
<i>Candida braccarensis</i>	CBS10154 <sup>T</sup>	(12.2)	<b>12</b>	5315	Gabaldón <i>et al.</i> (2013)
<i>Nakaseomyces delphensis</i>	CBS2170 <sup>T</sup>	(10.7)	<b>10</b>	5168	Gabaldón <i>et al.</i> (2013)
<i>Candida nivariensis</i>	CBS9983 <sup>T</sup>	(11.5)	<b>13</b>	5238	Gabaldón <i>et al.</i> (2013)
<i>Candida castelii</i>	CBS4332 <sup>T</sup>	(10.1)	<b>8</b>	4875	Gabaldón <i>et al.</i> (2013)
<i>Nakaseomyces bacillisporus</i>	CBS7720 <sup>T</sup>	(10.9)	<b>15</b>	5086	Gabaldón <i>et al.</i> (2013)
<i>Kazachstania africana</i> <sup>e</sup>	CBS 2517 <sup>T</sup>	<b>11.1</b>	<b>12</b>	5378	Gordon <i>et al.</i> (2011)
<i>Kazachstania naganishii</i> <sup>e</sup>	CBS8797 <sup>T</sup>	<b>10.8</b>	<b>13</b>	5321	Gordon <i>et al.</i> (2011)
<i>Naumovozyma castelii</i> <sup>e</sup>	CBS 4309 <sup>T</sup>	<b>11.2</b>	<b>10</b>	5592	Gordon <i>et al.</i> (2011)
<i>Naumovozyma dairenensis</i> <sup>e</sup>	CBS 421 <sup>T</sup>	<b>13.5</b>	<b>11</b>	5548	Gordon <i>et al.</i> (2011)
<i>Tetrapisispora phaffii</i> <sup>e</sup>	CBS 4417 <sup>T</sup>	<b>12.1</b>	<b>16</b>	5250	Gordon <i>et al.</i> (2011)
<i>Tetrapisispora blattae</i> <sup>e</sup>	CBS 6284 <sup>T</sup>	<b>14.0</b>	<b>10</b>	5389	Gordon <i>et al.</i> (2011)
<i>Vanderwaltozyma polyspora</i> <sup>e</sup>	DSMZ70294 <sup>T</sup>	(14.7)	(41)	5652	Scannell <i>et al.</i> (2007a,b)
<b>Saccharomycetaceae (ZT clade)</b>					
<i>Zygosaccharomyces rouxii</i>	CBS732 <sup>T</sup>	<b>9.8</b>	<b>7</b>	4997	Souciet <i>et al.</i> (2009)
<i>Zygosaccharomyces bailii</i>	CLIB213 <sup>T</sup>	(10.3)	(27)	5084	Galeote <i>et al.</i> (2013)
<i>Torulasporea delbrueckii</i> <sup>e</sup>	CBS1146 <sup>T</sup>	<b>9.2</b>	<b>8</b>	4972	Gordon <i>et al.</i> (2011)
<b>Saccharomycetaceae (KLE clade)</b>					
<i>Lachancea fantastica</i> nom. nud.	CBS6924	<b>11.3</b>	<b>7</b>	5060	Vakirlis <i>et al.</i> (2016)
<i>Lachancea meyersii</i>	CBS8951 <sup>T</sup>	<b>11.3</b>	<b>8</b>	4997	Vakirlis <i>et al.</i> (2016)
<i>Lachancea dasiensis</i>	CBS10888	<b>10.7</b>	<b>8</b>	5099	Vakirlis <i>et al.</i> (2016)
<i>Lachancea nothofagi</i>	CBS11611 <sup>T</sup>	<b>11.3</b>	<b>8</b>	5153	Vakirlis <i>et al.</i> (2016)
<i>Lachancea thermotolerans</i>	CBS6340 <sup>T</sup>	<b>10.4</b>	<b>8</b>	5177	Souciet <i>et al.</i> (2009)
<i>Lachancea quebecensis</i>	CBS14088	(10.2)	(51)	5075	Freel <i>et al.</i> (2016)
<i>Lachancea waltii</i>	NCYC2644	(10.2)	(8)	(4768)	Kellis <i>et al.</i> (2004)
<i>Lachancea lanzarotensis</i>	CBS12615 <sup>T</sup>	(11.5)	(24)	5058	Sarilar <i>et al.</i> (2015)
<i>Lachancea mirantina</i>	CBS11717	<b>10.1</b>	<b>8</b>	5057	Vakirlis <i>et al.</i> (2016)
<i>Lachancea fermentati</i>	CBS6772	<b>10.3</b>	<b>8</b>	5233	Vakirlis <i>et al.</i> (2016)
<i>Lachancea cidri</i>	CBS2950	<b>10.1</b>	<b>8</b>	5188	Vakirlis <i>et al.</i> (2016)
<i>Lachancea kluyveri</i>	CBS3082 <sup>T</sup>	<b>11.3</b>	<b>8</b>	5378	Souciet <i>et al.</i> (2009)
<i>Kluyveromyces lactis</i>	CLIB210	<b>10.6</b>	<b>6</b>	5108	Dujon <i>et al.</i> (2004)
<i>Kluyveromyces dobzhanskii</i>	CBS 2104	(10.7)	(86)	4957	B. Nystedt and S. Astrom; unpublished data <sup>f</sup>
<i>Kluyveromyces marxianus</i>	NBRC 1777	<b>10.9</b>	<b>8</b>	4912	Inokuma <i>et al.</i> (2015)
—	DMKU3-1042	<b>11.0</b>	<b>8</b>	4952	Lertwattanasakul <i>et al.</i> (2015)
<i>Eremothecium gossypii</i>	ATCC 10895	<b>8.7</b>	<b>7</b>	4718	Dietrich <i>et al.</i> (2004), (2013)
<i>Eremothecium cymbalariae</i>	DBVPG 7215	<b>9.7</b>	<b>8</b>	4712	Wendland and Walther, (2011)
<i>Eremothecium coryli</i>	CBS5749	(9.1)	<b>6</b>	4682	Wendland and Walther, (2014)
<i>Eremothecium aceri</i>	FM-2008	<b>8.9</b>	<b>7</b>	4479	Dietrich (2013)
<i>Eremothecium sinecaudum</i>	ATCC 58844	<b>8.9</b>	<b>7</b>	4528	F. S. Dietrich; unpublished data <sup>g</sup>
<b>Saccharomycodaceae</b>					
<i>Hanseniaspora opuntiae</i>	AWRI3578	(8.8)	(18)	4176	Sternes <i>et al.</i> (2016)
<i>Hanseniaspora osmophila</i>	AWRI3579	(11.4)	(17)	4660	Sternes <i>et al.</i> (2016)
<i>Hanseniaspora uvarum</i>	AWRI3580	(8.8)	(18)	4061	Sternes <i>et al.</i> (2016)
<b>Wickerhamomycetaceae</b>					
<i>Wickerhamomyces anomalus</i>	NRRL Y-366-8	(14.1)	(46)	6423	Riley <i>et al.</i> (2016)
<i>Cyberlindnera jadinii</i>	NRRL Y-1542	(13.0)	(76)	6038	Riley <i>et al.</i> (2016)
<i>Cyberlindnera fabianii</i>	JCM3601	(12.3)	(12)	5874	R. Manabe - unpublished data <sup>h</sup>

(continued)

**Table 1, continued**

Species	Strain	Genome size (Mb) <sup>a</sup>	No. of chromosomes <sup>b</sup>	Total no. of CDS <sup>c</sup>	Reference or source
<b>Saccharomycopsidaceae</b>					
<i>Saccharomycopsis malanga</i>	JCM7620	(16.7)	(44)	6280	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma; unpublished data <sup>h</sup>
<i>Saccharomycopsis fibuligera</i>	KPH12		<b>7</b>	6155	Choo <i>et al.</i> (2016)

The table lists the yeast species of indicated families whose genomes were published in the form of complete sequences or permanent drafts with reasonable numbers of scaffolds (<100). Exploratory sequences (low coverage), sequence read archives, and preliminary assemblies with excessive numbers of contigs or scaffolds were ignored. Readers can find some of these additional sequences with their references in Hittinger *et al.* (2015). Strain numbers correspond to the reference indicated. Species in which several isolates have been sequenced in complete form or permanent drafts are indicated by bold type, and only one isolate is listed (the first one or the best one). Note that an exception was made for *K. marxianus* because two independent sequences of equivalent quality were published simultaneously. Species designation follows Kurtzman *et al.* (2011) and may, therefore, differ from original publications. Hybrid genomes and unspecified isolates are ignored.

<sup>a</sup> Genome sizes are indicated in bold type when determined from complete sequences and in brackets when deduced from scaffolds in assemblies (figures are haploid equivalent in case of known diploid strains). Sizes ignore rDNA, mtDNA, and plasmids.

<sup>b</sup> Numbers of chromosomes are indicated in bold type when known. Figures in brackets correspond to numbers of scaffolds in assemblies.

<sup>c</sup> Total numbers of predicted protein-coding genes (CDS) are taken from original publications or subsequent annotations, as most appropriate. Figures in brackets represent estimate from incomplete sequence.

<sup>d</sup> Unpublished data taken from [https://yix1217.github.io/Yeast\\_PacBio\\_2016/data/](https://yix1217.github.io/Yeast_PacBio_2016/data/).

<sup>e</sup> Sequence data taken from <http://ygob.ucd.ie>.

<sup>f</sup> Unpublished data taken from GenBank: CCBQ000000000.1.

<sup>g</sup> Unpublished data taken from GenBank: CP014248.1.

<sup>h</sup> Unpublished data taken from the National BioResource Project ([www.jcm.riken.jp/cgi-bin/nbrp/nbrp\\_list.cgi](http://www.jcm.riken.jp/cgi-bin/nbrp/nbrp_list.cgi)).

(Bon *et al.* 2003; Neuvéglise *et al.* 2011). It contrasts with their intron-rich fungal ancestors (Stajich *et al.* 2007). If some members of the basal lineages of Saccharomycotina have more than a third of their CDS split by introns (an already low figure), regular figures for other yeasts are 3–5 times smaller (methylotrophs and some members of the basal lineages) or even 10 times smaller (Saccharomycetaceae). This scarcity is specific for introns of Pol-II transcripts (spliceosomal introns). Yeast nuclear tRNA genes have introns in comparable numbers to other eukaryotes (with only few exceptions, see Morales *et al.* 2013), and group-I and group-II introns are frequent in yeast mitochondrial genes (see below).

The reason for the severe reduction of the number of spliceosomal introns in budding yeasts is not completely understood. An important loss may have occurred during the early stages of their evolution, perhaps associated to the loss of RNA interference (RNAi) (see below), but a tendency to additional losses seems to have continued in the different lineages. In the Saccharomycetaceae family, most introns appear to have existed in the last common ancestor and then to have been differentially lost during speciation. The overall rate of intron loss exceeds that of intron gain by two orders of magnitude, even if important quantitative differences exist in the various branches (Hooks *et al.* 2014). The mechanism by which intron-less genes are formed is not directly demonstrated, but can be imagined. The possible reinsertion of reversed-transcribed mature messenger RNAs (mRNAs) has been evoked many years ago for *S. cerevisiae* (Fink 1987), consistent with its efficient homologous recombination machinery, with the presence of numerous type-I transposable elements (encoding reverse transcriptases), and with the fact that the majority of its rare introns are located in the 5' part of CDS. But active type-I transposable elements are not uniformly present in all yeast lineages, and the role of microhomology or non-

homologous end joining mechanisms has also been recently considered (Hooks *et al.* 2014).

Whatever the causes, the reduced prevalence of introns during yeast genome evolution has been accompanied by an increased standardization of the 5' splice sites and branch points compared to other eukaryotes, with strong consensus sequences in both cases (6- and 7-nt long, respectively) that have only moderately evolved across the entire Saccharomycotina spectrum (Neuvéglise *et al.* 2011). Note that only U2-type introns exist because the U12 minor spliceosome has been lost in all Ascomycota (Russell *et al.* 2006). Similarly, the distance between the 3' splice sites and branch points is also often reduced, probably by action of the splicing factor *U2AF1*, because longer distances are observed in species that have lost this factor.

The reduction of intron number in yeast genome evolution may also be correlated with the fact that, in *S. cerevisiae*, a majority of introns can be experimentally removed with no or only minor consequences for cellular fitness (Parenteau *et al.* 2008). This is not true, however, for all introns, probably explaining why a complex splicing machinery has been conserved. For example, expression of genes encoding ribosomal proteins, where introns have been frequently preserved, is affected by the presence of introns (Parenteau *et al.* 2011). Similarly, the role of introns is important for genes involved in mating (Müller *et al.* 2007). Finally, a few introns encode essential small nucleolar RNA (snoRNA) molecules and, as expected, they have been preserved during evolution (Souciet *et al.* 2009; Morales *et al.* 2013).

**Centromeres:** The presence of point centromeres, originally discovered in *S. cerevisiae* from their ability to properly partition circular plasmids during cellular divisions (Clark and Carbon 1980), is a specific landmark of the Saccharomycetaceae family (Meraldi *et al.* 2006) that raises the question of their origin (Figure 7). These short structures, made of two sequence motifs (CDE I and CDE III) separated from each

**Table 2 Major reference genomes of the CTG clade (Debaryomycetaceae, Metchnikowiaceae, and related species)**

Species	Strain	Genome size (Mb) <sup>a</sup>	No. of chromosomes <sup>b</sup>	Total no. of CDS <sup>c</sup>	Reference or source
<b>Debaryomycetaceae</b>					
<i>Debaryomyces hansenii</i>	CBS767 <sup>T</sup>	<b>12.2</b>	<b>7</b>	6411	Dujon <i>et al.</i> (2004)
<i>Priceomyces haplophilus</i>	JCM1635	(10.5)	(9)	5253	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma unpublished data <sup>d</sup>
<i>Meyerozyma guilliermondii</i>	ATCC6260	(10.6)	(8)	6135	Butler <i>et al.</i> (2009)
<i>Meyerozyma caribbica</i>	MG20W	(10.6)	(9)	7472 <sup>e</sup>	Kim <i>et al.</i> (2015)
<i>Candida carpophila</i>	JCM9396	(10.2)	(10)	5418	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma unpublished data <sup>d</sup>
<i>Millerozyma acaciae</i>	JCM10732	(11.1)	(10)	5217	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma unpublished data <sup>d</sup>
<i>Lodderomyces elongisporus</i>	CBS2605 <sup>T</sup>	(15.4)	(11)	5931	Butler <i>et al.</i> (2009)
<b><i>Candida albicans</i></b>	SC5314	<b>14.3</b>	<b>2 × 8<sup>f</sup></b>	6207	Jones <i>et al.</i> (2004)
<i>Candida dubliniensis</i>	CD36	<b>14.6</b>	<b>2 × 8<sup>f</sup></b>	6070	Jackson <i>et al.</i> (2009)
<i>Candida tropicalis</i>	MYA3404	(14.6)	(24)	6445	Butler <i>et al.</i> (2009)
<i>Candida orthopsilosis</i>	90-125	<b>12.6</b>	<b>2 × 8<sup>f</sup></b>	5707	Riccombeni <i>et al.</i> (2012)
<b><i>Candida parapsilosis</i></b>	CDC317	(13.1)	(24)	5843	Butler <i>et al.</i> (2009)
<b><i>Scheffersomyces stipitis</i></b>	CBS6054	<b>15.4</b>	<b>8</b>	6026	Jeffries <i>et al.</i> (2007)
<i>Scheffersomyces lignosum</i>	JCM9837	(16.6)	(19)	6330	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma unpublished data <sup>d</sup>
<i>Spathaspora passalidarum</i>	NRRL Y-27907	(13.2)	(8)	6071	Wohlbach <i>et al.</i> (2011)
<i>Spathaspora arborariae</i>	UFMG-HM19.1A <sup>T</sup>	(12.7)	(41)	5625	Lobo <i>et al.</i> (2014)
<i>Candida tanzawaensis</i>	NRRL Y-17324 <sup>T</sup>	(13.1)	(16)	5895	Riley <i>et al.</i> (2016)
<b><i>Candida (Yamadazyma) tenuis</i></b>	NRRL Y-1498	(10.7)	(61)	5533	Wohlbach <i>et al.</i> (2011)
<b>Metchnikowiaceae</b>					
<i>Clavispora lusitanae</i>	ATCC42720	<b>12.1</b>	<b>8</b>	6116	Butler <i>et al.</i> (2009)
<i>Metchnikovia bicuspidata</i>	NRRL YB-4993	(15.1)	(48)	5851	Riley <i>et al.</i> (2016)
<i>Metchnikovia fructicula</i>	277	(26.1)	(93)	9631 <sup>g</sup>	Hershkovitz <i>et al.</i> (2013)
<i>Candida auris</i>	Ci6684	(12.5)	(99)	8358 <sup>h</sup>	Chatterjee <i>et al.</i> (2015)
<i>Candida intermedia</i>	JCM1607	(13.0)	(12)	5823	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma unpublished data <sup>d</sup>
<b>Related species</b>					
<i>Babjeviella inositovora</i>	NRRL Y-12698 <sup>T</sup>	(15.2)	(49)	6403	Riley <i>et al.</i> (2016)
<i>Cephaloascus albidus</i>	ATCC66658	(15.3)	(25)	6196	J. K. Magnuson unpublished data <sup>i</sup>
<i>Cephaloascus fragrans</i>	12-1022	(13.7)	(16)	5717	J. K. Magnuson unpublished data <sup>i</sup>
<i>Hyphopichia burtonii</i>	NRRL Y-1933 <sup>T</sup>	(12.4)	(27)	6002	Riley <i>et al.</i> (2016)
<i>Candida homilientoma</i>	JCM1507	(12.2)	(8)	5661	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma unpublished data <sup>d</sup>
<i>Wickerhamia fluorescens</i>	JCM1821	(13.2)	(20)	5676	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma unpublished data <sup>d</sup>

The table lists the yeast species of indicated families under the same conditions as specified in Table 1. Strain numbers correspond to the reference indicated. Species in which several isolates have been sequenced in complete form or permanent drafts are indicated by bold type, and only one isolate is listed (the first one or the best one).

<sup>a</sup> Genome sizes are indicated in bold type when determined from complete sequences and in brackets when deduced from scaffolds in assemblies (figures are haploid equivalent in case of known diploid strains). Sizes ignore rDNA, mtDNA, and plasmids.

<sup>b</sup> Numbers of chromosomes are indicated in bold type when known. Figures in brackets correspond to numbers of scaffolds in assemblies.

<sup>c</sup> Total numbers of predicted protein-coding genes (CDS) are taken from original publications or subsequent annotations, as most appropriate.

<sup>d</sup> Unpublished data taken from the National BioResource Project ([www.jcm.riken.jp/cgi-bin/nbrp/nbrp\\_list.cgi](http://www.jcm.riken.jp/cgi-bin/nbrp/nbrp_list.cgi)).

<sup>e</sup> Automated prediction.

<sup>f</sup> Diploids.

<sup>g</sup> Number of *unigenes* determined from RNA sequencing, genome sequence unpublished, data taken from GenBank: ANFW000000000.2. Assembly size suggests hybrid genome.

<sup>h</sup> Dubious number, average CDS size (1024 bp) is significantly shorter than that of other yeast genomes (~1500 bp), suggesting sequencing errors (frameshifts).

<sup>i</sup> Unpublished data taken from <http://genome.jgi.doe.gov> (nonpublic).



**Table 3 Major reference genomes of the methylotroph and basal lineages**

Species	Strain	Genome size (Mb) <sup>a</sup>	No. of chromosomes <sup>b</sup>	Total no. of CDS <sup>c</sup>	Reference or source
<b>Methylotrophs</b>					
<b>Phaffomycetaceae</b>					
<i>Komagataella phaffii</i>	CBS 7435	<b>9.4</b>	<b>4</b>	5325	Küberl <i>et al.</i> (2011)
<i>Candida sorboxylosa</i>	JCM1536	(10.7)	(37)	4724	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma unpublished data <sup>d</sup>
<b>Pichiaceae and related Saccharomycetales incertae sedis</b>					
<b><i>Dekkera bruxellensis</i></b>	CBS 2499	<b>13.4</b>	—	5208	Piškur <i>et al.</i> , (2012)
<b><i>Dekkera anomala</i></b>	YV396	(12.9)	(30)	5241	Y. Vervoort, B. Herrera-Malaver, S. Mertens, V. Guadalupe Medina, J. Duitama, L. Michiels, G. Derdelinckx, K. Voordeckers and K. J. Verstrepen unpublished data <sup>e</sup>
<i>Dekkera naardensis</i>	CBS7540	(11.3)	(76)	—	H. Jiang, unpublished data <sup>f</sup>
<i>Nakazawaea peltata</i>	JCM9829	(11.7)	(11)	5620	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma unpublished data <sup>d</sup>
<i>Pichia membranifaciens</i>	NRRL Y-2248 <sup>T</sup>	(11.6)	(11)	5546	Riley <i>et al.</i> (2016)
<i>Kuraishia capsulata</i>	CBS 1993 <sup>T</sup>	<b>11.4</b>	<b>7</b>	6029	Morales <i>et al.</i> (2013)
<b><i>Ogataea polymorpha</i></b>	NCYC495 leu1.1	(9.0)	<b>7</b>	5177	Riley <i>et al.</i> (2016)
<i>Ogataea parapolyomorpha</i>	DL-1	<b>9.1</b>	<b>7</b>	5325	Ravin <i>et al.</i> (2013)
<i>Ogataea methanolica</i>	JCM10240	(15.1)	(32)	6063	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma unpublished data <sup>d</sup>
<i>Candida boidinii</i>	JCM9604	(19.4)	(32)	6053	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma, unpublished data <sup>d</sup>
<i>Ambrosiozyma kashinagacola</i>	JCM15019	(12.3)	(23)	5787	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma, unpublished data <sup>d</sup>
<i>Candida succiphila</i>	JCM9445	(12.1)	(22)	5455	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma, unpublished data <sup>d</sup>
<i>Candida arabinofementans</i>	NRRL YB-2248 <sup>T</sup>	<b>(13.2)</b>	(62)	5861	Riley <i>et al.</i> (2016)
<b><i>Pachysolen tannophilus</i></b>	CBS 4044	(12.2)	(34)	5546	Liu <i>et al.</i> (2012)
<b>Basal lineages</b>					
<b>Dipodascaceae, Trichomonascaceae, and related Saccharomycetales incertae sedis</b>					
<i>Geotrichum candidum</i>	CLIB918	<b>24.2</b>	(134) <sup>g</sup>	6804	Morel <i>et al.</i> (2015)
<b><i>Yarrowia lipolytica</i></b>	CLIB122 (E150)	<b>20.6</b>	<b>6</b>	6582	Dujon <i>et al.</i> (2004)
<i>Yarrowia keelungensis</i>	JCM14894	(21.8)	(41)	6618	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma unpublished data <sup>d</sup>
<i>Yarrowia deformans</i>	JCM1694	(20.9)	(44)	6704	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma unpublished data <sup>d</sup>
<i>Starmerella bombicola</i>	JCM9596	(9.6)	(16)	4887	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma unpublished data <sup>d</sup>
<i>Starmerella (Candida) apicola</i>	NRRL Y-50540	(9.8)	(40)	3818 <sup>h</sup>	Vega-Alvarado <i>et al.</i> (2015)
<i>Sporopachydermia quercuum</i>	JCM9486	(16.4)	(15)	5992	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma unpublished data <sup>d</sup>
<i>Blastobotrys adenivorans</i>	LS3	<b>11.8</b>	<b>4</b>	6150	Kunze <i>et al.</i> (2014)
<i>Blastobotrys attinorum</i>	NRRL Y27639	(14.0)	(14)	6184	J. K. Magnuson unpublished data <sup>i</sup>
<i>Trichomonascus petasosporus</i>	NRRL YB2093	(14.5)	(79)	6567	J. K. Magnuson unpublished data <sup>i</sup>
<i>Sugiyamaella lignohabitans</i>	CBS10342	<b>16.0</b>	<b>4</b>	6820	Bellasio <i>et al.</i> (2016)
<i>Sugiyamaella americana</i>	NRRL YB2067	(16.5)	(48)	6288	J. K. Magnuson unpublished data <sup>i</sup>
<i>Wickerhamiella domercqiae</i>	JCM9478	<b>8.5</b>	<b>4</b>	4928	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma unpublished data <sup>d</sup>

(continued)

Table 3, continued

Species	Strain	Genome size (Mb) <sup>a</sup>	No. of chromosomes <sup>b</sup>	Total no. of CDS <sup>c</sup>	Reference or source
<i>Zygoascus hellenicus</i>	Y7136	(12.2)	(11)	5430	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma unpublished data <sup>d</sup>
<i>Candida infanticola</i>	DS02	(8.1)	(22)	—	H. Lee, C. Han, G. Park, W. Jeon, H. Lee and J. Ahn. unpublished data <sup>i</sup>
<i>Nadsonia fulvescens</i> var. <i>elongata</i>	DSM6958	(13.7)	(20)	5657	Riley <i>et al.</i> (2016)
<i>Tortispora caseinolytica</i>	NRRL Y-17796 <sup>T</sup>	<b>9.2</b>	<b>6</b>	4657	Riley <i>et al.</i> (2016)
<b>Early branching<sup>k</sup></b>					
<b>Ascoideaceae</b>					
<i>Ascoidea rubescens</i>	NRRL Y-17699 <sup>T</sup>	(17.5)	(63)	6802	Riley <i>et al.</i> (2016)
<i>Ascoidea asiatica</i>	JCM7603	(20.3)	(71)	7694	R. Manabe, R. Endoh, S. Uzuhashi, G. Okada, M. Takashima and M. Ohkuma unpublished data <sup>d</sup>
<b>Lipomycetaceae</b>					
<i>Lipomyces starkeyi</i>	NRRL Y-11557	(21.3)	(117) <sup>g</sup>	8192	Riley <i>et al.</i> (2016)

The table lists the yeast species of indicated families under the same conditions as specified in Table 1. Strain numbers correspond to the reference indicated. Species in which several isolates have been sequenced in complete form or permanent drafts are indicated by bold type, and only one isolate is listed (the first one or the best one).

<sup>a</sup> Genome sizes are indicated in bold type when determined from complete sequences and in brackets when deduced from scaffolds in assemblies (figures are haploid equivalent in case of known diploid strains). Sizes ignore rDNA, mtDNA, and plasmids.

<sup>b</sup> Numbers of chromosomes are indicated in bold type when known. Figures in brackets correspond to numbers of scaffolds in assemblies.

<sup>c</sup> Total numbers of predicted protein-coding genes (CDS) are taken from original publications or subsequent annotations, as most appropriate. Data not found indicated by —.

<sup>d</sup> Unpublished data taken from the National BioResource Project ([www.jcm.riken.jp/cgi-bin/nbrp/nbrp\\_list.cgi](http://www.jcm.riken.jp/cgi-bin/nbrp/nbrp_list.cgi)).

<sup>e</sup> Unpublished data taken from GenBank: LCTY000000000.1.

<sup>f</sup> Unpublished data taken from GenBank: MDSA000000000.1.

<sup>g</sup> *G. candidum* and *L. starkeyi* are listed here despite their large number of scaffolds because each genome represents the only reference available, so far, in its respective lineage.

<sup>h</sup> Automated prediction based on comparisons to the CTG clade.

<sup>i</sup> Unpublished data taken from <http://genome.jgi.doe.gov> (nonpublic).

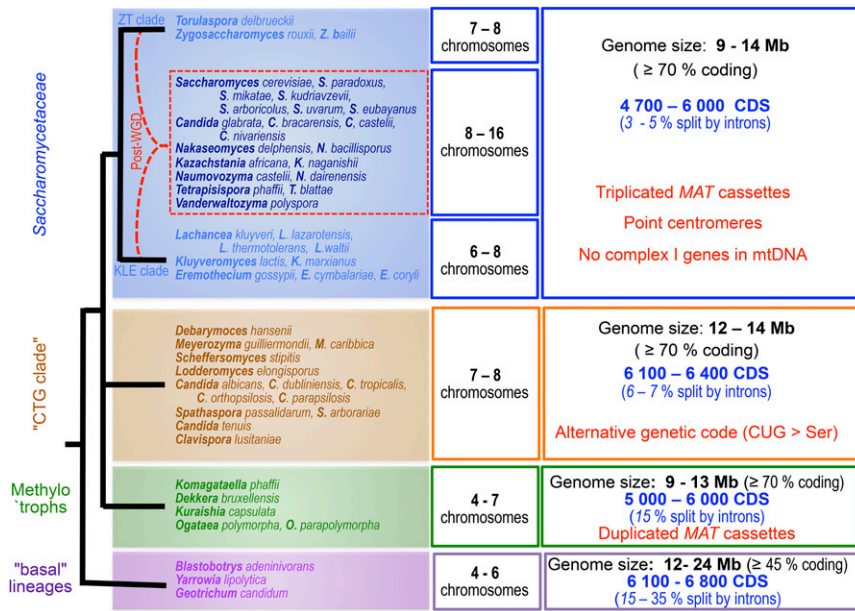
<sup>j</sup> Unpublished data taken from GenBank: LWLF000000000.1.

<sup>k</sup> These are listed here for convenience and are the few sequenced yeast species not included in the four major subgroups of Saccharomycotina defined in this review. See Shen *et al.* (2016) for a recent genome-based phylogenetic reconstruction. The Lipomycetaceae family is basal to all Saccharomycotina.

other by a short AT-rich interval (CDEII), contrast with the 3- to 5-kb-long regional centromeres observed in the CTG clade (Sanyal *et al.* 2004; Padmanabhan *et al.* 2008; Kapoor *et al.* 2015; Chatterjee *et al.* 2016), in *Y. lipolytica* (Vernis *et al.* 1997, 2001; Lynch *et al.* 2010), and in the methylotrophs (Morales *et al.* 2013; Coughlan *et al.* 2016), and which are made of variable sequences often associated to clustered traces of mobile elements or flanked by inverted repeats. Regional centromeres of Saccharomycotina yeasts are reminiscent of the epigenetic centromeres of multicellular eukaryotes made of large arrays of DNA lacking sequence specificity (Brown and O'Neill 2014), but are not identical to them. Compared to *S. pombe* centromeres, whose central region made of CenH3 nucleosomes (the variant form of histone H3 essential for centromere function) is flanked by heterochromatin like in most eukaryotes, the CenH3 nucleosomes of the regional centromeres of *C. albicans* or *C. lusitaniae* are not embedded in heterochromatin (Sanyal *et al.* 2004; Kapoor *et al.* 2015). This property may facilitate their neofunctionalization (Ketel *et al.* 2009), although the phenomenon has also been observed for large epigenetic centromeres.

It has been proposed that the loss of RNAi and of classical heterochromatin in Saccharomycotina (see below) was at the origin of the evolution of their centromeres (Baum *et al.* 2006;

Padmanabhan *et al.* 2008; Ketel *et al.* 2009; Malik and Henikoff 2009). The emergence of point centromeres in the Saccharomycetaceae, in which the CenH3 protein is conserved, must have taken place by recruiting novel components such as *Ndc10p* and *Ctf13p* to bind the evolutionarily conserved kinetochore machinery to the CDEIII motif, forming the CBF3 complex (Meraldi *et al.* 2006). Malik and Henikoff (2009) have argued that these two genes, which have no homologs in other organisms, might have originated from the capture of the *REP1* and *REP2* genes of the 2- $\mu$ m plasmids specifically found in the Saccharomycetaceae yeasts (Blaisonneau *et al.* 1997). According to this view, the first point centromeres may have derived from the chromosomal integration of the *cis*-acting STB locus of these plasmids, hence alleviating the need of the ancestral regional centromeres. How this initial structure spread to all chromosomes remains an open question but, once established, point centromeres have been highly conserved in Saccharomycetaceae chromosomes as judged from the conservation of their flanking genes. This evolutionary process, however, has not been unique as revealed by the new type of point centromeres recently discovered at different chromosomal locations in *Naumovozya castelli* and *Naumovozya dairensis* (Kobayashi *et al.* 2015). The fact that the overexpression of *CSE4* (encoding the CenH3 protein) reveals new centromere-like



**Figure 6** The four major subgroups of Saccharomycotina as defined from genome architectures. The major species whose genomes served to define the subgroups are indicated (see Table 1, Table 2, and Table 3 for details). Tree topology based on Kurtzman *et al.* (2011). For more recent global phylogenies, refer to Shen *et al.* (2016). Most significant properties of each subgroup are summarized in boxes on the right (see text). Red dotted lines symbolize hybridization between members of the ZT and KLE clades of Saccharomycetaceae prior to WGD (Marcet-Houben and Gabaldón 2015; Wolfe 2015).

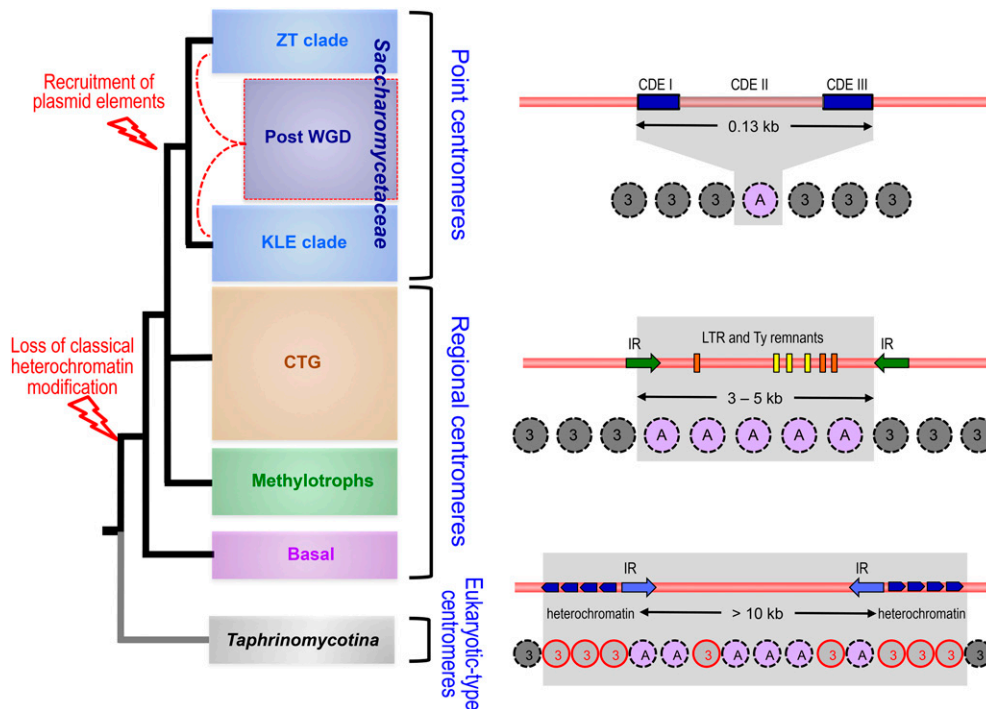
regions (CLR) in the genome of *S. cerevisiae* suggests a possible mechanism at the origin of novel point centromeres (Lefrançois *et al.* 2013). CLR are similar in size to point centromeres but lack sequence specificity.

**Mating cassettes and mating-type switching:** Mating-type switching by the conversion of one idiomorph to its opposite at the *MAT* locus during mitotic cycles of yeasts has retained considerable attention during the last decades (Heitman *et al.* 2013). It is a characteristic feature of yeasts that differentiates them from their multicellular fungal ancestors. In *S. cerevisiae* and *S. pombe*, the molecular mechanisms at the basis of the phenomenon have been precisely elucidated and they extensively differ (Haber 2012; Klar *et al.* 2014). Despite the fact that these mechanisms are inherently destructive for genomes (Gordon *et al.* 2011), mating-type switching has emerged independently several times during yeast evolution, suggesting some selective advantage (Figure 8). One of them might be that switching facilitates the finding of a mating partner of the same species in harsh environmental conditions (Hanson *et al.* 2014). Another one might be that, by forming homozygous diploids, switching alleviates the risk of dissociation of favorable allelic combinations at the next meiosis, such combinations resulting from insidious selection during the long phases of clonal propagation that are so characteristic of yeasts.

In *S. cerevisiae*, mating-type switching involves two extra copies of the *MAT* locus (designated *HML* and *HMR* cassettes, respectively), located close to telomeres on the same chromosome and silenced by a combination of the *SIR1–4* proteins. Note that the common histone modification (H3K9me), characteristic of silent chromatin in most eukaryotes, is absent in Saccharomycotina due to their original loss of the *CLR4* and *SWI6/HP1* genes (Hickman *et al.* 2011). Such a triplication is another intrinsic landmark of genomes of the Saccharomycetaceae family not found in other Saccharomy-

cotina (Figure 8). Members of the CTG clade as well as *Y. lipolytica* have only one *MAT* locus per haploid genome, designated *MTL*, and do not switch mating type. *MTL* bears either the *MATa* or *MATα* idiomorphs with some rare cases of mixing of both (Dujon *et al.* 2004; Butler *et al.* 2009). In the heterothallic members of the methylotrophic subgroup capable of switching, the *MAT* locus is duplicated (see below). Triplication of the *MAT* locus also occurred in *S. pombe* and other *Schizosaccharomyces* species, but this represents an evolutionary convergence (Klar *et al.* 2014), not an ancestral feature. During evolution of the Saccharomycetaceae family, the triplication has received secondary modifications such as the loss of one or two cassettes in some species with the concomitant loss of switching or even sex (Fabre *et al.* 2005; Müller *et al.* 2007; Wolfe *et al.* 2015), or the expansion to a fourth locus in members of the *Eremothecium* clade (Dietrich *et al.* 2013).

In *S. cerevisiae*, the *HML* cassette bears copies of the *MATα1* and *MATα2* genes, and the *HMR* cassette has a copy of the *MATa1* gene (encoding a homeodomain protein that drives meiosis and sporulation). The *MATa2* gene, encoding a transcriptional activator belonging to the HMG family conserved in most yeast species, has been lost in the post-WGD species of the Saccharomycetaceae family but is present in members of the ZT and KLE clades (see below and Table 1). In *S. cerevisiae*, the switching mechanism requires (i) a specific double-strand endonuclease (*HO*), cleaving DNA within the active *MAT* locus while leaving intact the silenced copies; (ii) a double-strand break repair machinery using the sequence homology of the flanking regions (designated X and Z); and (iii) a donor preference enhancer (RE) to ensure repair using the appropriate silent cassette (Haber 2012). The *HO* gene is also found in *Z. rouxii*, *Torulaspore delbrueckii* (ZT clade), and almost all post-WGD species of Saccharomycetaceae (lost in *Kazachstania africana*), but it is not found in species of the KLE clade of Saccharomycetaceae. It encodes



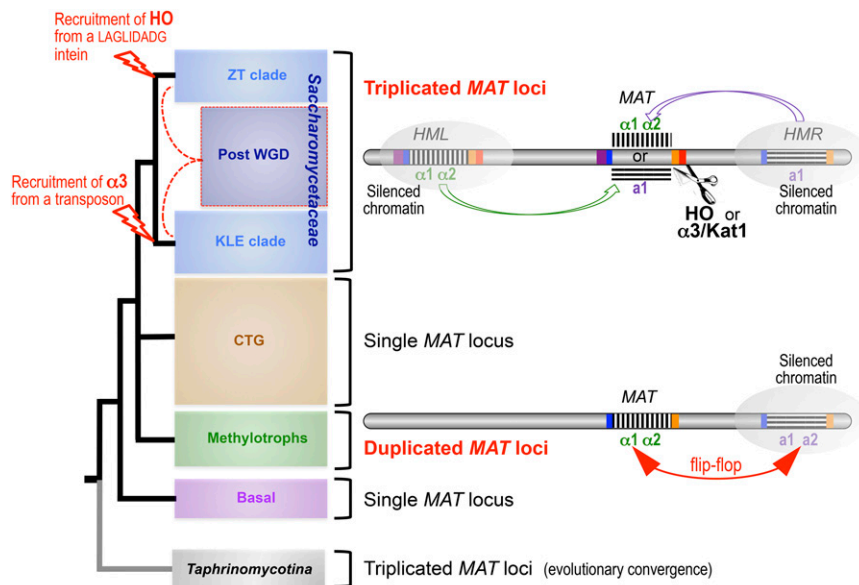
**Figure 7** Evolution of centromeres. Left: tree topology of major subgroups of Saccharomycotina derived from Figure 6 with major evolutionary events (red lightning bolts) relative to centromeres. Taphrinomycotina used as outgroup. The loss of classical heterochromatin modification occurred very early in Saccharomycotina after the branching (not shown here) of *L. starkeyi* (Riley *et al.* 2016). Right: Centromeres (gray shade) are symbolized on chromosomes (red lines) by their characteristic elements (boxes and arrows). Symbolized below are nucleosomes (circles) of corresponding regions. Dark gray background, normal histone H3; void red circles, classical heterochromatin modifications of histone H3; purple background, CENP-A centromeric histone H3 variant (A). Adapted from Ishii (2009). Structures vary between species for regional centromeres (Chatterjee *et al.* 2016; Coughlan *et al.* 2016). IR, inverted repeats.

an endonuclease belonging to the LAGLIDADG family of intein-related proteins that has probably been recruited for mating-type switching in the ancestor of the ZT clade and transmitted to the post-WGD species, or, perhaps, recruited in the ancestor of the entire family but then lost in members of the KLE clade (no species has a complete *HO* gene). The presence of sequences similar to *HO* in some of these last species (Fabre *et al.* 2005) corresponds to relics of an ancestral gene of the same LAGLIDADG family that may or may not have been involved in mating-type switching.

The entirely different switching mechanism discovered in *K. lactis* (Barsoum *et al.* 2010; Rajaei *et al.* 2014) indicates the independent recruitment of a completely different machinery in this lineage. In this case, the *MAT $\alpha$*  to *MATa* switching events and the reverse *MATa* to *MAT $\alpha$*  events involve two domesticated transposases,  $\alpha 3$  and Kat1, respectively.  $\alpha 3$  is encoded by the *MAT $\alpha 3$*  gene located next to the *MAT $\alpha$*  locus and shares homology to the MULE family of transposable elements. Kat1 is the product of the *KAT1* gene containing a programmed  $-1$  frameshift and shares homology to the hAT family of transposable elements. Both are under the transcriptional control of the *MTS1/RME1* gene product, induced by nutrient limitation. In *MAT $\alpha$*  strains, binding of Mts1 to sites close to the *MAT $\alpha 3$*  gene stimulates the excision of the MULE, inducing *MAT $\alpha$*  to *MATa* switching to repair the chromosome break. In *MATa* strains, Kat1 generates two hairpin-capped, double-strand breaks at the fossil imprints of an ancient transposon located within the *MATa* locus; inducing *MATa* to *MAT $\alpha$*  switching to repair the chromosome break. Within the KLE clade, the *KAT1* gene is specific to the *Kluyveromyces* species (absent from *L. thermotolerans* and *E. gossypii*), suggesting a domestication event of the transposon at the origin of this genus (Rajaei *et al.* 2014).

The mechanisms of mating-type switching in some methylotrophic yeasts may help us understand the origin of the complex three-locus system that evolved in the Saccharomycetaceae (Hanson *et al.* 2014; Maekawa and Kaneko 2014). In *O. polymorpha* and *K. phaffii* (see Table 3), two *MAT* loci bearing opposite idiomorphs lie in opposite orientation on the same chromosome. They are flanked by inverted repeats. One is the active locus that determines cell type, the other is silenced by its proximity to the centromere (*O. polymorpha*) or the telomere (*K. phaffii*). In both cases, switching occurs by a flip-flop mechanism that inverts the chromosomal segment between the two loci at the same time as it exchanges idiomorphs between active and silent locations. The phenomenon is induced by nutrient starvation. The same situation was recently found in *Pachysolen tannophilus* and *Ascoidea rubescens* (Riley *et al.* 2016). Whether or not this two-locus system has a common origin with the three-locus system of the Saccharomycetaceae remains an open question given the important differences between the molecular mechanisms involved. But its presence in *A. rubescens*, a member of the Ascoideaceae family, is consistent with an evolutionary intermediate between the methylotrophs and the Saccharomycetaceae.

**tRNA decoding and alternative genetic codes:** The alteration of the genetic code in *C. albicans* and related species, with the CUN codon read as serine instead of leucine, has attracted attention long (Santos and Tuite, 1995; Sugita and Nakase 1999) before being recognized as a common characteristic of >75 species of the CTG clade (Santos *et al.* 2011). Contrary to several other alterations of the code that often concern nonsense codons (the reading of UGA as tryptophan in yeast mitochondria is one such example), this sense-to-sense alteration is mediated by a Ser tRNA containing a 5'-CAG-3' anticodon (Ser-tRNA<sup>[CAG]</sup>), which



**Figure 8** Evolution of mating-type switching. Left: same as Figure 7. Red lightning bolts refer to capture of site-specific nucleases. Right: *MAT* cassettes (colored boxes) are symbolized on chromosomes (gray bars) with *MAT $\alpha$*  (horizontal black stripes) or *MAT $\alpha$*  (vertical black stripes) idiomorphs. Silenced cassettes are shaded. Shown are the topologies observed in *S. cerevisiae* (for the Saccharomycetaceae) and *K. phaffii* (for the methylotrophs), but significant variations are found in both subgroups (see text). The  $\alpha 3$  and *Kat1* transposases of *K. lactis* are shown acting at the same place as the HO endonuclease of *S. cerevisiae*, despite important mechanistic differences.

is efficiently charged by the seryl-tRNA synthetase and occasionally charged by the leucyl-tRNA synthetase. Ratios of 97 vs. 3% are reported under normal conditions. This unusual tRNA appeared prior to the split between the CTG clade and other subgroups of Saccharomycotina, and must have competed for a long time with the normal Leu-tRNA<sup>[CAG]</sup> for mRNA decoding before the latter was eventually lost (Yokogawa *et al.* 1992; Massey *et al.* 2003). Its formation involved three mutational steps around the anticodon, plus one in the variable loop (Santos *et al.* 2011) that must have occurred in a stepwise manner because basal representatives of this clade, such as *Metchnikowia bicuspidata* and *Babjeviella inositovora*, have only some of these changes (Riley *et al.* 2016). The CUG reassignment was associated to massive mutational change of CUG codons to UUG or UUA codons in protein-coding genes to preserve leucine in proteins, whereas new CUG codons were created at positions corresponding to serine or amino acids of similar chemical properties in other yeast proteins. It remains that the ambiguous identity of the CUG codons in these yeasts generates “statistical” proteins whose biological implications may have been important for the evolution of the CTG clade.

Alteration of the CUG codon significance in the nucleus is not unique to yeasts of the CTG clade. It has occurred several times during yeast evolution as a result of distinct mechanisms (Mühlhausen *et al.* 2016). In *P. tannophilus*, representing a basal lineage to the methylotroph clade, as well as in *A. rubescens*, a member of the Ascoidaceae family, the CUG codon specifies alanine instead of leucine or serine (Riley *et al.* 2016). In *P. tannophilus*, reassignment followed tRNA loss (Mühlhausen *et al.* 2016). Several different sense-to-sense codon reassignments are also observed in the mitochondrial genetic code of the Saccharomycetaceae yeasts (Ling *et al.* 2014, for a recent overview).

Alteration of tRNA molecules was not rare in the evolution of Saccharomycotina yeasts, and the total number of genes encoding them varied to a large extent between ~80 for several members of the methylotroph subgroup (Morales

*et al.* 2013) to >500 for *Y. lipolytica* (Dujon *et al.* 2004; Marck *et al.* 2006). Whereas the global degree of redundancy is generally conserved (some tRNA species are always encoded by multiple genes, others by few), the repertoire of tRNA species has evolved between the different yeast lineages. This is particularly remarkable for the leucine and arginine four-codon families. When *Y. lipolytica* and *K. phaffii* use the eukaryotic mode of tRNA decoding for both CUN and CGN codon families, as expected of their fungal ancestry, most members of the Saccharomycetaceae family use a bacterial mode for the same codons (the difference corresponds to the nucleotide at position 34 which is used to interpret the wobble position of the codons; Marck *et al.* 2006). Interestingly, members of the methylotroph subgroup use the ancestral eukaryotic mode for the Leu-CUN family, but the bacterial mode for the Arg-CGN family. As members of the CTG clade also use the bacterial mode of decoding for the Arg-CGN family, the transition from the eukaryotic mode to the bacterial mode for this family must have occurred very early during the evolution of the Saccharomycotina, immediately after the separation of the basal lineages; whereas the transition for the Leu-CUN family seems to be specific to the Saccharomycetaceae.

**RNAi machinery:** It is sometimes argued that, owing to the power of its genetic tools, RNAi would probably have been discovered earlier if it had existed in *S. cerevisiae*. But it does not. And the same is true for most Saccharomycotina, but not all. The evolution of this important eukaryotic function in budding yeasts appears complex if one judges from the patchy distributions of the *Dicer* and *Argonaute* genes in the Saccharomycetaceae family (Wolfe *et al.* 2015), a phenomenon also observed in other fungi where the variable numbers of these two genes suggest recent duplications and losses (Hu *et al.* 2013).

The existence of RNAi in Saccharomycotina was discovered by analysis of *C. albicans*, the most intensively studied pathogen of the CTG clade; and *N. castellii*, a poorly studied

member of the Saccharomycetaceae family (Drinnenberg *et al.* 2009). Contrary to *S. cerevisiae*, these two yeasts are able to generate small interfering RNAs, mostly corresponding to transposable elements and subtelomeric Y' elements, using noncanonical Dicer proteins. They also have an Argonaute gene. The same is true for *V. polyspora* and several other post-WGD genera of Saccharomycetaceae, but not for *Saccharomyces* (Wolfe *et al.* 2015). Detailed structural analysis, however, indicates that the canonical Dicer (ancestral) was probably lost at the origin of the entire Saccharomycotina lineage; the noncanonical dicer activity being reestablished after duplication of genes of the RNase-III family and separation of the rRNA-, snoRNA-, and small nuclear RNA-processing activity conserved in all yeasts (*RNT1* gene in *S. cerevisiae*) from a neo-dicer activity (*DCR1* gene in *N. castellii*) conserved in only some lineages (Bernstein *et al.* 2012a). In *C. albicans*, the *DCR1* gene also ensures the maturation of the ribosomal and spliceosomal RNAs, allowing the secondary inactivation of the *RNT1* gene (Bernstein *et al.* 2012b).

The loss of RNAi activity in some lineages of Saccharomycotina has probably greatly influenced their mode of evolution. Besides the loss of the multiple controls and regulations, it has probably played a role in their ability to host double-stranded, viral-like RNA elements such as the killer particles in *S. cerevisiae* (Wickner *et al.* 2013), as was experimentally demonstrated in *S. cerevisiae* (Drinnenberg *et al.* 2011). Similarly, the remarkable lengthening in yeasts of some peripheral branches in otherwise conserved noncoding RNA molecules (Kachouri *et al.* 2005; Mitrovich and Guthrie 2007) may be another direct consequence of this phenomenon. The telomerase RNA, for example, is >2-kb long in *C. glabrata* (Kachouri-Lafond *et al.* 2009).

**mtDNA:** Like in most eukaryotes, yeast mtDNA encodes the RNA molecules necessary to ensure an independent protein synthesis from the cytoplasm: genes for the large and small rRNAs, a set of 23–25 distinct tRNAs sufficient to read the entire genetic code (with some differences from the “universal” one) and, often, the RNA subunit of the mitochondrial RNase P. This machinery translates RNA from only a few mitochondrial genes encoding subunits of the respiratory chain (4–11 genes, depending on the species) and of the ATP synthase complex (2–3 genes), as well as one protein (*Var1*) involved in the late-step assembly of the small subunit of the mitochondrial ribosome. Mitochondrial ribosomes and tRNAs also translate the reading frames of group-I and group-II introns which, although optional, may equal or outnumber the above-mentioned protein-coding genes. These notions have been known for a very long time, thanks in particular to the pioneering work performed >3 decades ago on *S. cerevisiae* (reviewed in Dujon 1981). In this species, the mitochondrial genome consists of a few dozens of copies of a ~80-kb circular mtDNA molecule containing, in addition to the above, long AT-rich intergenic regions including short GC-rich palindromes. They recombine between themselves (Dujon *et al.* 1974; Fritsch *et al.* 2014). Today, the evolution

of the mitochondrial genomes of Saccharomycotina is illustrated by the complete mtDNA sequences of >80 different species (reviewed in Freel *et al.* 2015) and it shows a remarkable consistency with the four subdivisions deduced from nuclear genomes (above).

The *VAR1* gene is present in all subgroups except the CTG clade. Its presence in the basal groups indicates its ancestral origin and long-range conservation despite the fact that its sequence evolves very rapidly. The reason for the loss of *VAR1* in the CTG clade remains unclear. It may be that assembly of the mitochondrial ribosome in which all other proteins are encoded by the nucleus was modified by the alteration of the genetic code. *VAR1* has also been lost from the mtDNA of *Hanseniopsis uvarum*, but this species, a member of the Saccharomycodaceae family, has a linear mitochondrial genome extremely reduced in size (~11 kb) (Pramateftaki *et al.* 2006). The loss of genes encoding complex-I subunits of the respiratory chain in mtDNA is a landmark of the Saccharomycetaceae genomes that differentiates them from other subgroups of Saccharomycotina in which, with only few exceptions, mtDNA always encodes seven subunits of the NADH:ubiquinone oxidoreductase complex like in most eukaryotes (reviewed in Freel *et al.* 2015). Remarkably, the same loss occurred in the *Schizosaccharomyces* genomes; another example of evolutionary convergence between the two most extensively studied yeast species. The loss of complex I was found coupled to the duplication of nuclear genes for alternative dehydrogenases, a phenomenon possibly related with the adaptation to fermentative metabolism (Marcet-Houben *et al.* 2009). Beside these important differences, several other variations occurred during the evolution of yeast mitochondrial genomes. In the *Saccharomyces*, the presence of an abnormal tRNA molecule reads the CUN codons as threonine (Dujon 1981). In *C. glabrata*, the CGN-codon family has been entirely eliminated due to the loss of the cognate tRNA molecule (Koszul *et al.* 2003). Considerable size variation of mtDNA molecules is also observed between related species, ranging from ~20 kb in *C. glabrata* to almost 110 kb in *Nakaseomyces bacillisporus* (Bouchier *et al.* 2009). Size variation is due in part to the presence/absence of group-I and group-II introns. But it also results from the invasion of short GC-rich palindromic clusters of unclear origin associated with the expansion of large AT-rich intergenic regions (Bouchier *et al.* 2009). These elements were discovered long ago in *S. cerevisiae* (Dujon 1981), but their patchy distribution across the entire phylogeny of Saccharomycotina species (Freel *et al.* 2015) raises an interesting problem to examine.

### **Tentative reconstruction of the evolution of Saccharomycotina genomes**

Considering the above elements, the important evolutionary changes separating the four major subgroups of Saccharomycotina genomes can be summarized as illustrated by Figure 9 using parsimony as a guiding principle, but keeping in mind that more complex scenarios, including multiple and reversible changes, are not excluded. As discussed above, some

changes concern RNA molecules and others concern chromosome structure. Among the latter, some events correspond to the intrinsic dynamics of chromosomes in vertical lineages and others are related to horizontal exchanges such as the capture of extrachromosomal elements or interspecific hybridizations (see below). The loss of ancestral elements and functions also played a critical role in subsequent evolutionary events of the Saccharomycotina subphylum. Among those, the early losses of HP1-mediated formation of heterochromatin (Hickman *et al.* 2011; Riley *et al.* 2016) and of canonical eukaryotic RNAi (Bernstein *et al.* 2012a) were probably the two most important events. The decreasing number of spliceosomal introns and the increasing compactness of genomes are general trends that seem to have persisted throughout the evolution of the Saccharomycotina subphylum, but were already significant at its origin.

Among the RNA-related events, a fixation of the 5S RNA gene into the rDNA locus must have occurred very early, as only some members of the basal lineages have dispersed 5S RNA genes (a normal situation for eukaryotes) and no 5S gene in the rDNA locus. Most other yeasts have one (or rarely two) 5S gene(s) in the rDNA locus and no (or few) dispersed copies. But this fixation was probably reversible since some members of the CTG and methylotroph clades, such as *M. bicuspicata*, *Hyphophichia burtonii*, and *K. phaffii*, only have dispersed copies in multiple numbers (Riley *et al.* 2016). Note, however, that these species represent lateral branches to these clades. The switch to the bacterial decoding mode of the Arg-CGN codon family also occurred very early, whereas the situation appears more complex for the Leu-CUN codon family (with some reversions and multiple reassignments of the CUG codon).

Most chromosome-related events concern the Saccharomycetaceae family, but this is also where they have been most extensively studied, particularly in *S. cerevisiae*. The WGD was recognized very early (Wolfe and Shield 1997) and its origin and consequences have been extensively studied to the point that post-WGD species are overrepresented among sequenced yeasts (see Table 1). The event, suspected to have occurred from an ancestor with eight chromosomes (Gordon *et al.* 2009), was followed by extensive gene loss (Scannell *et al.* 2007a,b), while neofunctionalization or subfunctionalization of the remaining paralogous copies (ohnologs) have played an important functional role in the evolution of these yeasts. The existence in post-WGD genomes of numerous paralogs predating the WGD event was recently interpreted as an indication that the duplication event followed an interspecies hybridization event that probably occurred between a member of the ZT clade (which appears phylogenetically closer to post-WGD) and a member of the KLE clade of Saccharomycetaceae (Marcet-Houben and Gabaldón 2015; Wolfe 2015).

The triplication of the *MAT* cassettes exists in both clades (with variations in some species) but the recruitment of the HO endonuclease for mating-type switching (probably from a LAGLIDADG intein) is specific to the ZT clade and, consequently, post-WGD species. Less data are unfortunately available about

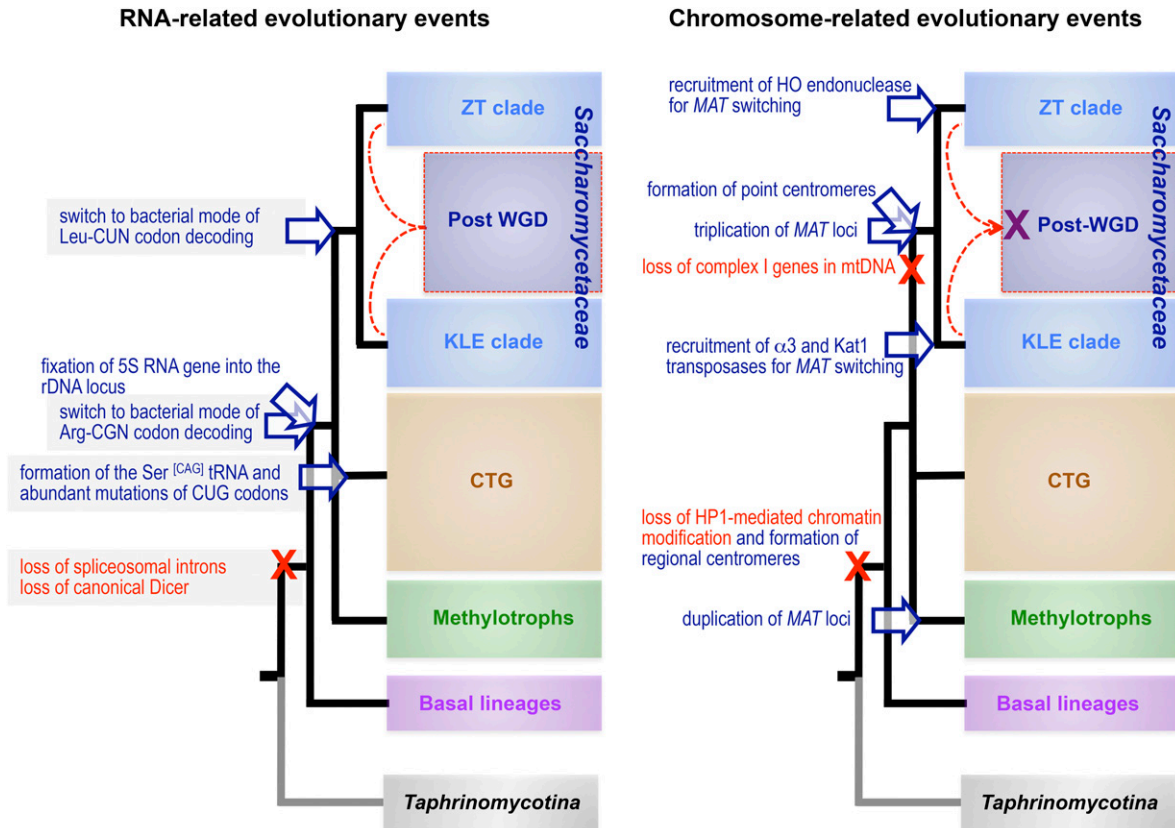
the recruitment of the  $\alpha 3$  and Kat1 transposases discovered in *K. lactis* (Barsoum *et al.* 2010; Rajaei *et al.* 2014) and about the mechanism of mating-type switching in other members of the KLE clade. How the duplication of the *MAT* cassettes in *A. rubescens* and some methylotroph yeasts (Hanson *et al.* 2014; Maekawa and Kaneko 2014; Riley *et al.* 2016) relate to the triplication in Saccharomycetaceae is not yet entirely solved, but it was inferred that mating-type switching evolved soon after the loss of the classical H3K9 chromatin modification, a very early event at the origin of all Saccharomycotina after the separation of *L. starkeyi* (Riley *et al.* 2016).

Another highly spectacular evolutionary transition at the chromosomal level is the replacement of the regional centromeres found in all Saccharomycotina, except Saccharomycetaceae, by the point centromeres specific to the latter family; a phenomenon proposed to result from the capture of genes and *cis*-acting elements from the circular 2- $\mu$ m plasmids into chromosomes (Malik and Henikoff 2009). Other captures have been reported which, together with other mechanisms (below), have contributed at various levels during the evolution of Saccharomycotina.

### **Molecular basis of the evolutionary dynamics of yeast genomes**

Yeasts remarkably illustrate the fact that the evolution of genomes is driven by a combination of intrinsic mechanisms, operating during the successive generations in vertical lineages, and horizontal exchanges with other contemporary lineages. The latter represent either regular events, as in sexual reproduction, or accidental events as in interspecies hybridization or foreign gene capture. If the two types of phenomena are necessarily entangled in eukaryotic organisms with obligate sexual reproduction, they are clearly separable in organisms, such as yeasts, capable of unlimited clonal propagation under favorable conditions and of surviving the stresses of complex ecosystems.

**Intrinsic genome dynamics: sequence divergence and chromosomal rearrangements:** Sequence divergence between yeast genomes is always much higher than anticipated by the similarities between the species. Within single genera, two species such as *S. cerevisiae* and *S. uvarum*, for example, can differ from each other by  $\sim 33\%$  at the nucleotidic level. Compare this to the 1.8% separating genomes of human and chimpanzee, classified in two distinct genera! Additionally, their orthologous proteins share, on average,  $\sim 80\%$  aa identity, a figure similar to the average identity between human and chicken proteins (Dujon 2006). Other genera are even more diverged. In *Lachancea*, the average amino-acid identity between orthologous proteins of distinct species range from 83 to 69% (Vakirlis *et al.* 2016). In *Eremothecium*, the same measure drops to  $\sim 60\%$  (Wendland and Walther 2011, 2014), and even lower figures (51–53%) were found within the *Nakaseomyces* genus (Gabaldón *et al.* 2013). These high sequence divergences indicate that, contrary to what is frequently written in the literature, there is no such thing as



**Figure 9** Overview of Saccharomycotina genome evolution. Evolutionary innovations (blue arrows) or losses (red crosses) attributed from parsimony to major separations between subgroups of Saccharomycotina (secondary losses of preceding innovations, not shown for simplicity). Left: RNA-related events. Note that, in some lineages, noncanonical enzymes replaced the ancestral Dicer and that the phylogenetic distribution of Dicer and Argonaute proteins in Saccharomycotina is discontinuous (see text). Right: chromosome-related events. Note that the recruitment of  $\alpha 3$  and Kat1 transposases concerns the *Kluyveromyces*, not the entire KLE clade, and that the loss of H3K9me chromatin modification occurred after the separation of *L. starkeyi* (data not shown) from other basal lineages. Purple cross: massive post-WGD gene loss. Note that the loss of complex-I genes in mtDNA is not a chromosomal event, but its occurrence is indicated because of its association with the duplication of nuclear genes encoding alternative dehydrogenases (see text).

“closely related yeasts,” except in very specific instances. In agreement with their primarily clonal mode of propagation, the distinct species of yeasts only represent the remnants of separated lineages after very high numbers of successive mitotic generations and numerous population bottlenecks.

The same conclusion can be drawn from examination of synteny. If extensive conservation of gene orders can be observed between members of some genera (see *What did We Learn from Comparative and Population Genomics of the Saccharomyces Species Complex?*), chromosomal breakpoints rapidly accumulate at greater evolutionary distances, rapidly reducing the size of conserved blocks to smaller and smaller numbers of genes. In pairwise comparisons between members of the KLE clade of Saccharomycetaceae, for example, average block sizes of only 10–20 genes are generally observed (Souciet *et al.* 2009), with frequent insertion/deletion or inversion of genes within them (Byrne and Wolfe 2006). A similar situation applies for the methylotrophic yeasts (Morales *et al.* 2013; Ravin *et al.* 2013). If yeast genome maps are more stable than those of vertebrates or insects for equivalent degrees of sequence divergence (Drillon and Fischer

2011; Rolland and Dujon 2011), nearly all traces of conserved synteny disappear when members of the distinct subgroups are compared, *i.e.*, the total number of chromosomal breakpoints accumulated in genomes during their evolution approaches the number of genes. The nature and extent of chromosomal rearrangements have recently been documented in detail within the single yeast genus of *Lachancea* (Vakirlis *et al.* 2016). It was found that unbalanced rearrangements, primarily made of gene duplications and losses, outnumber the balanced ones, such as inversions and translocations, and frequently disrupt genes at breakpoints.

**Intrinsic genome dynamics: loss, duplication, and neoformation of genes:** Yeasts also differ extensively from one another by their protein repertoires, due to gene losses and gains (Souciet *et al.* 2009; Libkind *et al.* 2011; Scannell *et al.* 2011; Liti *et al.* 2013; Wendland and Walther 2014; Riley *et al.* 2016; Vakirlis *et al.* 2016). Even within each of the four subgroups defined above, core genomes are significantly limited relative to pan-genomes. This is obviously not unique to yeasts, but their genomes remarkably illustrate the



importance of gene loss in evolution, as well as its compensation by gene duplication and *de novo* formation.

Yeast genomes offer numerous examples of gene loss, in agreement with the fact that only a minority of genes (~18%) are essential for cell viability, at least under laboratory conditions as demonstrated by the systematic gene deletions in *S. cerevisiae* (Giaever and Nislow 2014). If part of this phenomenon may be explained by the presence of paralogs, numerous losses of unique genes can also be observed when comparing genomes, leading to functional losses that differentiate the lineages. Species adapted to special niches such as, for example, the human pathogen *C. glabrata* (Dujon *et al.* 2004) tend to lose more genes. Remarkably, gene loss may result in the gain of novel adaptive functions by alteration of regulatory networks (Gabaldón *et al.* 2013). Most gene losses correspond to entire deletions of the corresponding chromosome segments. Pseudogenes, exhibiting various degrees of sequence alteration, are also found but remain rare in yeasts (Lafontaine and Dujon 2010). Interestingly, functionally coordinated genes, even scattered in genomes, tend to be lost in parallel in given lineages (Hittinger *et al.* 2010; Gabaldón *et al.* 2013). The importance of gene loss in yeast genome evolution is consistent with their clonal mode of propagation.

In recent work on *G. candidum*, a member of the basal lineages of Saccharomycotina, it was observed that a significant number of its genes (263 or 3.9% of total) were “phylogenetically discordant,” *i.e.*, were absent from other Saccharomycotina but present in Pezizomycotina or other fungal lineages (Morel *et al.* 2015). Their similar patristic distance with conserved genes from the common ancestor of both subphyla indicated, however, common vertical lineage rather than recent horizontal acquisitions; leading to the conclusion that these genes had been lost in all Saccharomycotina other than *G. candidum*. They were designated as specifically retained ancestral genes (SRAG). SRAGs were also observed in the genomes of *Y. lipolytica* and *D. hansenii*, where they represent 3.6 and 2.0% of the total numbers of protein-coding genes, respectively.

Because total numbers of protein-coding genes vary within only moderate limits in yeast genomes ( $<1.4\times$  in general and  $\sim 1.1\times$  within each subgroup; Figure 6), it follows that gene gains must roughly equilibrate losses in number over long evolutionary times. Gene gains have at least four entirely distinct origins: duplications of preexisting genes; horizontal acquisitions from other genomes; capture from nonchromosomal sources such as plasmids, viruses, or mitochondria; and, finally, *de novo* gene formation. These mechanisms are of unequal importance numerically but not biologically. Duplications, *de novo* formation, and intracellular captures from mtDNA are intrinsic mechanisms of vertical inheritance; the other mechanisms involve a horizontal dimension of inheritance. In contrast to the other mechanisms, duplications do not initially expand gene family repertoires but generate collections of paralogs, altering gene number equilibria, upon which subsequent evolutionary changes may eventually operate. The duplications themselves result from different

mechanisms, as reflected from the locations of paralogous genes in yeast genomes and by the plurimodal distributions of sequence identity between their products (Dujon *et al.* 2004; Souciet *et al.* 2009; Marcet-Houben and Gabaldón 2015). These mechanisms range from the formation of tandem gene arrays to the accidental duplication of an entire genome (above) or, more frequently, to the duplication or higher-order amplification of large chromosomal segments embedding series of adjacent genes. Tandem gene arrays are observed in all yeast genomes, albeit in different proportions (Despons *et al.* 2011), and are generally not conserved between species. Segmental amplifications are frequently observed in experiments using *S. cerevisiae* (Kozul *et al.* 2004; Schacherer *et al.* 2007b; Araya *et al.* 2010; Payen *et al.* 2014; Thierry *et al.* 2015, 2016) but leave few traces in natural yeast genomes, except in subtelomeric regions (Fairhead and Dujon 2006).

The overall result of all duplication mechanisms is such that no yeast genome is minimal. Families of two or more members with varying degrees of sequence divergence represent, in general, 30–45% of total gene numbers (Butler *et al.* 2009; Souciet *et al.* 2009). Some families correspond to very ancient duplications, conserved across enormous numbers of successive generations. Others represent the result of recent duplications that may or may not be conserved in subsequent generations. The comparison between five protoploid species of the Saccharomycetaceae family (members of the ZT and KLE clades) performed some years ago (Souciet *et al.* 2009) gave us the first indication of the kinetics involved. A little more than 100 pairs of dispersed paralogs (114) were completely conserved between *Z. zouxii*, *L. kluyveri*, *L. thermotolerans*, *K. lactis*, and *E. gossypii*; compared to 9–22 multigene families specific to each species (total 84). Interestingly, if ~60% (51) of the latter correspond to genes present in single copies in the four other yeast species, suggesting lineage-specific duplications of ancestral genes, the rest (33) are absent from the four other yeast species, suggesting lineage-specific acquisitions. Tandem gene arrays show faster kinetics of evolutionary change consistent with their adaptive role (Brown *et al.* 1998; Müller *et al.* 2009). In the same five species above, only eight arrays were found entirely conserved, compared to 9–16 arrays specific to each species (total 63) and 32 others showing partial conservation.

The *de novo* formation of genes has traditionally been regarded as a nearly impossible mechanism because of the extremely low probability of obtaining a genetically meaningful sequence by single nucleotide changes from a random sequence. But ideas along this line are rapidly progressing for two reasons. First, there is no such thing in nature as a random DNA sequence, in yeasts as well as any other genome. Noncoding parts of genomes are not random nucleotide sequences but sequences inherited from previous generations that, in eukaryotes at least, are almost entirely transcribed. Using the ribosome profiling method, it was recently shown that the genome of *S. cerevisiae* contains a large number of *protogenes*, *i.e.*, short sequences susceptible to create new

active genes after a limited number of point mutations, the logical symmetry to pseudogenes (Carvunis *et al.* 2012). Second, actual examples of *de novo* gene formation have now been experimentally demonstrated in *S. cerevisiae*: the *BSC4* gene encoding a DNA-repair enzyme active in stationary phase, and the *MDF1* gene encoding a pleiotropic protein regulating glucose assimilation and the budding pathway. The first gene has emerged in *S. cerevisiae* from the conserved transcribed intergene between *LYP1* and *ALP1*, two genes encoding amino-acid transporters themselves originated from an inverted tandem duplication of an ancestral gene followed by subfunctionalization (Cai *et al.* 2008). *BSC4* has 132 codons. The second gene was created in *S. cerevisiae* on the opposite strand of *ADF1*, a gene considered as its antisense regulator before it was discovered to be, in reality, the evolutionarily conserved ancestral gene present in all other species (Li *et al.* 2010, 2014). *MDF1* has 153 codons. The presence of a few *de novo* created genes in each yeast lineage is also expected from careful comparison between the *Lachancea* genomes (Vakirlis *et al.* 2016).

Finally, the capture of nonchromosomal elements can be regarded as an intrinsic mechanism of genome evolution when permanent elements are involved, or as involving horizontal exchanges (below) in other cases. The most conspicuous examples of the first case are represented by the copies of mtDNA fragments (NUMT) found in chromosomes of almost all yeast species studied so far (Sacerdot *et al.* 2008; Leh-Louis *et al.* 2012). Depending upon their nature and precise location, such elements may alter genes in a variety of manners, including the possibility to extend ancestral reading frames. Their insertion occurs during repair of chromosomal double-strand breaks, as experimentally demonstrated in *S. cerevisiae* several years ago (Ricchetti *et al.* 1999).

**Horizontal exchanges and interspecies hybridization:** Similarly, traces of DNA plasmids or RNA viruses (designated NUPAV) were recognized in 40% of the 20 Saccharomycotina genomes examined (Frank and Wolfe 2009), probably representing ancient captures. Genomes of several species of the CTG clade contain a capsid protein gene originating from double-stranded RNA viruses (Taylor and Bruenn 2009). The evolutionary role of such phenomena in yeasts is not yet fully understood but, as mentioned above, such mechanisms were probably at the origin of point centromeres and mating-type switching mechanisms in the Saccharomycetaceae.

In addition to the above cases, yeast genomes show numerous traces of other horizontal inheritance without which their evolution would not have been the same. First, a number of genes, or sometimes gene clusters, have been acquired from other organisms. Second, a rapidly growing number of interspecies hybrids have now been recognized in various yeast lineages, raising the question of the degree of reticulation in yeast phylogenies.

Genes of bacterial origin, as judged from molecular phylogeny, have been recognized in almost every yeast genome sequenced so far, most of which correspond to basic metabolic

functions. Their function in yeasts is not always demonstrated, but spectacular examples can be mentioned such as the acquisition of the *URA1* gene in the Saccharomycetaceae family (Gojković *et al.* 2004) or of  $\beta$ -lactamases in *Kuraishia capsulata* (Morales *et al.* 2013) and *B. adeninovorans* (Kunze *et al.* 2014). The *URA1* gene, encoding a dihydroorotate dehydrogenase, has been critical in the evolution of the Saccharomycetaceae by allowing their propagation under strictly anaerobic conditions; the enzyme encoded by the ancestral gene *URA9* requiring oxygen for the synthesis of uracil. Similarly, the acquisition of genes encoding amino-acid racemases was also probably significant in the evolution of the lineages where it occurred (Fitzpatrick *et al.* 2008). The fact that horizontally acquired genes have a tendency to duplicate in several copies in the host genomes (Rolland *et al.* 2009) also supports the idea of functional selection.

Next to single genes, whose small size distribution suggests that they were acquired by natural DNA transformation (Rolland *et al.* 2009), yeast genomes also bear long chromosomal segments of alien origin. As already mentioned in *What Did We Learn from Comparative and Population Genomics of the Saccharomyces Species Complex?*, some wine strains of *S. cerevisiae* were found to contain a 17-kb-long segment containing 5 genes from *Zygosaccharomyces bailii* (Galeote *et al.* 2011) and a 65-kb-long segment containing 18 genes from *Torulaspora microellipsoides* (Marsit *et al.* 2015), two species from the ZT clade of Saccharomycetaceae present as contaminants in wine fermentations. The near complete nucleotide sequence identity of these segments (>99.7%) with respective donors demonstrate recent transfers into *S. cerevisiae* by a mechanism that remains to be identified. The circular permutations observed for the 17-kb segment of *Z. bailii* suggests that, in this case, an extrachromosomal circular element served as a vector between the two yeast species. Another remarkable example of introgression of long chromosomal segments of alien origin is represented by the presence of nitrate-assimilating gene clusters in several members of the methylotroph subgroup of Saccharomycotina (Morales *et al.* 2013). As a result, these yeasts assimilate nitrate, an important trait acquired by an ancestral introgression from plants or other nitrate-assimilating fungi. Again, the mechanism at the origin of such a large introgression remains to be elucidated.

Finally, it is now clear that interspecies hybrids, for a long time considered as evolutionary dead ends because of their low meiotic fertility, play a more important role than anticipated in yeast genome evolution (Morales and Dujon 2012). The existence of hybrids in *Saccharomyces* yeasts was suspected long ago for major brewing strains (see *What Did We Learn from Comparative and Population Genomics of the Saccharomyces Species Complex?*). It has now been fully demonstrated by genome sequencing (Nakao *et al.* 2009; Wahlter *et al.* 2014). Spontaneous interspecies hybridization appears so frequent in this genus, at least under industrial conditions of fermentation, that a complex hybrid was for long time mistaken for the type strain of *S. bayanus* before actual parental species were discovered (Libkind *et al.* 2011; Nguyen

*et al.* 2011). This raises the important question of how to recognize hybrid yeast genomes after they have been resolved in mosaics of parental pieces (see below) in the absence of information on the parents. It is quite possible, therefore, that the evolution of yeast genomes includes a higher degree of reticulation than generally expected. This is particularly true because meiotically fertile lines can rapidly emerge from initially unfertile hybrids, using different mechanisms (Greig *et al.* 2002a,b; Sebastiani *et al.* 2002; Antunovics *et al.* 2005).

The formation of interspecies hybrids appears as a general phenomenon in all subgroups of Saccharomycotina. Natural hybrid genomes were found in *Z. rouxii* (Gordon and Wolfe 2008) and *Z. bailii* (Mira *et al.* 2014), two members of the ZT clade of Saccharomycetaceae; in *Saccharomycopsis fibuligera* (Choo *et al.* 2016); in members of the CTG clade such as *Millerozyma sorbitophila* (Leh-Louis *et al.* 2012) and *Candida orthopsilosis* (Pryszcz *et al.* 2014; Schröder *et al.* 2016); as well as in *Dekkera bruxellensis* (Curtin *et al.* 2012; Borneman *et al.* 2014), a representative of the methylotroph subgroup. The phenomenon also extends to Basidiomycota yeasts where virulent hybrids were found in *Cryptococcus* (Boekhout *et al.* 2001; Cogliati *et al.* 2006; Bovers *et al.* 2008; Li *et al.* 2012).

As for the *Saccharomyces*, hybridizations involved parents of distinct ploidy, creating diploid or triploid hybrid genomes that may subsequently evolve into a variety of aneuploidies by partial chromosome loss. In general, the evolution of hybrid strains during their clonal development remains to be better studied. In this regard, the osmotolerant yeast *M. sorbitophila* proved informative because this genome, whose parents differed from each other by 12–15% nucleotide sequence divergence, is in the process of resolution by a mechanism of loss of heterozygosity that transformed some segments of chromosomes or even entire chromosomes in homozygous pairs (Leh-Louis *et al.* 2012). The loss-of-heterozygosity mechanism must also operate in other hybrids and even in wild diploid *S. cerevisiae* strains if one judges from the frequent presence of long homozygous segments in pairs of chromosomes (Belloch *et al.* 2009; Nakao *et al.* 2009; Magwene *et al.* 2011). The fact that these segments almost always extend to chromosome ends suggests that they result from a break-induced mechanism during mitotic divisions (Morales and Dujon 2012) or from reversed meiotic prophase (Laureau *et al.* 2016). Interestingly, in *M. sorbitophila* as well as in other hybrids carefully studied, only one of the two parental rDNA loci persists, suggesting a rapid resolution in favor of one parent immediately after hybridization. This is interesting to keep in mind because the sequences of the rDNA are being used to define yeast species (Kurtzman *et al.* 2011) and therefore a number of hybrid genomes may have remained undetected so far. The 1-Mb-long, GC-rich segment in the genome of all *L. kluyveri* isolates sequenced so far (Friedrich *et al.* 2015) suggests such a “species” with several ancestries. Subsequent studies will be necessary to precisely determine how reticulated yeast genome evolution was.

## Perspectives

Much more has been learned in a short time from the remarkable progress of yeast genomics than what could be anticipated at the time of the original sequencing of the *S. cerevisiae* genome, with lessons largely applicable to all eukaryotic organisms. Reasons for this success lie of course in the facility to sequence many yeast genomes using modern technologies but also because yeasts offer a rare combination of exquisite experimental systems and widespread natural populations spanning a broad range of evolutionary distances, ecological diversity, and interactions with human activities. It can easily be anticipated that this short review will be quickly outdated by a rapidly expanding set of novel data. But the basic questions elucidated or unraveled by yeast genomics offer a solid background for the future of genome biology. What is the relative importance of sequence variation, chromosomal rearrangements, copy number variation, and gene gain and loss in speciation and in the phenotypic variation within populations? What is the importance of horizontal exchanges, therefore of extant biodiversity and ecological proximity, in these phenomena and in the origin of successful lineages? And what is the impact of human activities on these questions? At the same time, yeast genomes continue to spring surprises. The genetic code is more variable than expected. Mobile elements find their way across distinct lineages. Novel genes are generated *de novo*. And all this in a group of organisms, the Saccharomycotina, whose long-term success seems to have been based on the ancestral loss of essential functions relative to RNA processing or chromatin modification that appear so critical to other eukaryotes. It can only be hoped that more research in this field will continue for a long time.

## Acknowledgments

While writing this review article, the authors' laboratories were supported in part by the Centre National de la Recherche Scientifique (UMR3525), Université Pierre et Marie Curie (UFR927) and the Agence Nationale de la Recherche (contract DYGEVO from program blanc 2011 SVSE6) for B.D.; and by the Biotechnology and Biological Sciences Research Council and the University of Leicester for E.J.L. The authors declare no conflict of interest.

## Literature Cited

- Almeida, P., C. Goncalves, S. Teixeira, D. Libkind, M. Bontrager *et al.*, 2014 A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. *Nat. Commun.* 5: 4044.
- Almeida, P., R. Barbosa, P. Zalar, Y. Imanishi, K. Shimizu *et al.*, 2015 A population genomics insight into the Mediterranean origins of wine yeast domestication. *Mol. Ecol.* 24: 5412–5427.
- Antunovics, Z., H.-V. Nguyen, C. Gaillardin, and M. Sipiczki, 2005 Gradual genome stabilisation by progressive reduction of the *Saccharomyces uvarum* genome in an interspecific hybrid with *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 5: 1141–1150.

- Arana-Sanchez, A., L. E. Segura-Garcia, M. Kirchmayr, I. Orozco-Avila, E. Lugo-Cervantes *et al.*, 2015 Identification of predominant yeasts associated with artisan Mexican cocoa fermentations using culture-dependent and culture-independent approaches. *World J. Microbiol. Biotechnol.* 31: 359–369.
- Araya, C. L., C. Payen, M. J. Dunham, and S. Fields, 2010 Whole-genome sequencing of a laboratory-evolved yeast strain. *BMC Genomics* 11: 88.
- Baker, E., B. Wang, N. Bellora, D. Peris, A. B. Hulfachor *et al.*, 2015 The genome sequence of *Saccharomyces eubayanus* and the domestication of lager-brewing yeasts. *Mol. Biol. Evol.* 32: 2818–2831.
- Barsoum, E., P. Martinez, and S. U. Aström, 2010 Alpha3, a transposable element that promotes host sexual reproduction. *Genes Dev.* 24: 33–44.
- Baum, M., K. Sanyal, P. K. Mishra, N. Thaler, and J. Carbon, 2006 Formation of functional centromeric chromatin is specified epigenetically in *Candida albicans*. *Proc. Natl. Acad. Sci. USA* 103: 14877–14882.
- Bellasio, M., A. Peymann, M. G. Steiger, M. Valli, M. Sipiczki *et al.*, 2016 Complete genome sequence and transcriptome regulation of the pentose utilizing yeast *Sugiyamaella lignihabitans*. *FEMS Yeast Res.* 16: fow037.
- Belloch, C., R. Pérez-Torrado, S. S. González, J. E. Pérez-Ortín, J. García-Martínez *et al.*, 2009 Chimeric genomes of natural hybrids of *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii*. *Appl. Environ. Microbiol.* 75: 2534–2544.
- Bergeron, J., and G. Drouin, 2008 The evolution of 5S ribosomal RNA genes linked to the rDNA units of fungal species. *Curr. Genet.* 54: 123–131.
- Bergstrom, A., J. T. Simpson, F. Salinas, B. Barre, L. Parts *et al.*, 2014 A high-definition view of functional genetic variation from natural yeast genomes. *Mol. Biol. Evol.* 31: 872–888.
- Bernstein, D. A., V. K. Vyas, and G. R. Fink, 2012a Genes come and go. *RNA Biol.* 9: 1123–1128.
- Bernstein, D. A., V. K. Vyas, D. E. Weinberg, I. A. Drinnenberg, D. P. Bartel *et al.*, 2012b *Candida albicans* dicer (*CaDcr1*) is required for efficient ribosomal and spliceosomal RNA maturation. *Proc. Natl. Acad. Sci. USA* 109: 523–528.
- Bing, J., P. J. Han, W. Q. Liu, Q. M. Wang, and F. Y. Bai, 2014 Evidence for a Far East Asian origin of lager beer yeast. *Curr. Biol.* 24: R380–R381.
- Blaisonneau, J., F. Sor, G. Cheret, D. Yarrow, and H. Fukuhara, 1997 A circular plasmid from the yeast *Torulaspora delbrueckii*. *Plasmid* 38: 202–209.
- Boekhout, T., B. Theelen, M. Diaz, J. W. Fell, W. C. Hop *et al.*, 2001 Hybrid genotypes in the pathogenic yeast *Cryptococcus neoformans*. *Microbiology* 147: 891–907.
- Bon, E., S. Casaregola, G. Blandin, B. Llorente, C. Neuvéglise *et al.*, 2003 Molecular evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns. *Nucleic Acids Res.* 31: 1121–1135.
- Borneman, A. R., B. A. Desany, D. Riches, J. P. Affourtit, A. H. Forgan *et al.*, 2011 Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. *PLoS Genet.* 7: e1001287.
- Borneman, A. R., R. Zeppel, P. J. Chambers, and C. D. Curtin, 2014 Insights into the *Dekkera bruxellensis* genomic landscape: comparative genomics reveals variations in ploidy and nutrient utilisation potential amongst wine isolates. *PLoS Genet.* 10: e1004161.
- Bouchier, C., L. Ma, S. Créno, B. Dujon, and C. Fairhead, 2009 Complete mitochondrial genome sequences of three *Nakaseomyces* species reveal invasion by palindromic GC clusters and considerable size expansion. *FEMS Yeast Res.* 9: 1283–1292.
- Bovers, M., F. Hagen, E. E. Kuramae, H. L. Hoogveld, F. Dromer *et al.*, 2008 AIDS patient death caused by novel *Cryptococcus neoformans* x *C. gattii* hybrid. *Emerg. Infect. Dis.* 14: 1105–1108.
- Brown, C. A., A. W. Murray, and K. J. Verstrepen, 2010 Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Curr. Biol.* 20: 895–903.
- Brown, C. J., K. M. Todd, and R. F. Rosenzweig, 1998 Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. *Mol. Biol. Evol.* 15: 931–942.
- Brown, J. D., and R. O'Neill, 2014 The evolution of centromeric DNA sequences in *eLS*. John Wiley and Sons, Chichester, United Kingdom.
- Butler, G., M. D. Rasmussen, M. F. Lin, M. A. Santos, S. Sakthikumar *et al.*, 2009 Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature* 459: 657–662.
- Byrne, K. P., and K. H. Wolfe, 2006 Visualizing syntenic relationships among the hemiascomycetes with the yeast gene order browser. *Nucleic Acids Res.* 34: D452–D455.
- Cai, J., R. Zhao, H. Jiang, and W. Wang, 2008 *De novo* origination of a new protein-coding gene in *Saccharomyces cerevisiae*. *Genetics* 179: 487–496.
- Carvunis, A. R., T. Rolland, I. Wapinski, M. A. Calderwood, M. A. Yildirim *et al.*, 2012 Proto-genes and *de novo* gene birth. *Nature* 487: 370–374.
- Casey, G. P., and M. B. Pedersen, 1988 DNA sequence polymorphisms in the genus *Saccharomyces*. V. Cloning and characterization of a LEU2 gene from *S. carlsbergensis*. *Carlsberg Res. Commun.* 53: 209–219.
- Chambers, S. R., N. Hunter, E. J. Louis, and R. H. Borts, 1996 The mismatch repair system reduces meiotic homeologous recombination and stimulates recombination-dependent chromosome loss. *Mol. Cell. Biol.* 16: 6110–6120.
- Chatterjee, G., S. R. Sankaranarayanan, K. Guin, Y. Thattikota, S. Padmanabhan *et al.*, 2016 Repeat-associated fission yeast-like centromeres in the ascomycetous budding yeast *Candida tropicalis*. *PLoS Genet.* 12: e1005839.
- Chatterjee, S., S. V. Alampalli, R. K. Nageshan, S. T. Chettiar, S. Joshi *et al.*, 2015 Draft genome of a commonly misdiagnosed multi-drug resistant pathogen *Candida auris*. *BMC Genomics* 16: 686.
- Choo, J. H., C. P. Hong, J. Y. Lim, J.-A. Seo, Y.-S. Kim *et al.*, 2016 Whole-genome *de novo* sequencing, combined with RNA-Seq analysis, reveals unique genome and physiological features of the amylolytic yeast *Saccharomycopsis fibuligera* and its interspecies hybrid. *Biotechnol. Biofuels* 9: 246.
- Chou, J. Y., and J. Y. Leu, 2010 Speciation through cytonuclear incompatibility: Insights from yeast and implications for higher eukaryotes. *Bioessays* 32: 401–411.
- Clark, L., and J. Carbon, 1980 Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature* 287: 504–509.
- Cliften, P., P. Sudarsanam, A. Desikan, L. Fulton, B. Fulton *et al.*, 2003 Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* 301: 71–76.
- Cogliati, M., M. C. Esposto, A. M. Tortorano, and M. A. Viviani, 2006 *Cryptococcus neoformans* population includes hybrid strains homozygous at mating-type locus. *FEMS Yeast Res.* 6: 608–613.
- Coughlan, A. Y., S. J. Hanson, K. P. Byrne, and K. H. Wolfe, 2016 Centromeres of the yeast *Komagataella phaffii* (*Pichia pastoris*) have a simple inverted-repeat structure. *Genome Biol. Evol.* 8: 2482–2492.
- Cromie, G. A., K. E. Hyma, C. L. Ludlow, C. Garmendia-Torres, T. L. Gilbert *et al.*, 2013 Genomic sequence diversity and population structure of *Saccharomyces cerevisiae* assessed by RAD-seq. *G3 (Bethesda)* 3: 2163–2171.
- Cubillos, F. A., E. Billi, E. Zorgo, L. Parts, P. Fargier *et al.*, 2011 Assessing the complex architecture of polygenic traits in diverged yeast populations. *Mol. Ecol.* 20: 1401–1413.

- Curtin, C. D., A. R. Borneman, P. J. Chambers, and I. S. Pretorius, 2012 *De-novo* assembly and analysis of the heterozygous triploid genome of the wine spoilage yeast *Dekkera bruxellensis* AWRI1499. *PLoS One* 7: e33840.
- Despons, L., Z. Uzunov, and V. L. Louis, 2011 Tandem gene arrays, plastic chromosomal organizations. *C. R. Biol.* 334: 639–646.
- De Schutter, K., Y. C. Lin, P. Tiels, A. Van Hecke, S. Glinka *et al.*, 2009 Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nat. Biotechnol.* 27: 561–566.
- Dettman, J. R., J. B. Anderson, and L. M. Kohn, 2010 Genome-wide investigation of reproductive isolation in experimental lineages and natural species of *Neurospora*: identifying candidate regions by microarray-based genotyping and mapping. *Evolution* 64: 694–709.
- Dietrich, F. S., S. Voegeli, S. Brachat, A. Lerch, K. Gates *et al.*, 2004 The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* 304: 304–307.
- Dietrich, F. S., S. Voegeli, S. Kuo, and P. Philippsen, 2013 Genomes of *Ashbya* fungi isolated from insects reveal four mating-type loci, numerous translocations, lack of transposons, and distinct gene duplications. *G3 (Bethesda)* 3: 1225–1239.
- Dowell, R. D., O. Ryan, A. Jansen, D. Cheung, S. Agarwala *et al.*, 2010 Genotype to phenotype: a complex problem. *Science* 328: 469.
- Drillon, G., and G. Fischer, 2011 Comparative study on synteny between yeasts and vertebrates. *C. R. Biol.* 334: 629–638.
- Drinnenberg, I. A., D. E. Weinberg, X. T. Xie, J. P. Mower, K. H. Wolfe *et al.*, 2009 RNAi in budding yeast. *Science* 326: 544–550.
- Drinnenberg, I. A., G. R. Fink, and D. P. Bartel, 2011 Compatibility with killer explains the rise of RNAi-deficient fungi. *Science* 333: 1592.
- Dujon, B., 1981 Mitochondrial genetics and functions, pp. 505–635 in *Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*, edited by J. N. Strathern, E. W. Jones, and J. Broach. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Dujon, B., 2006 Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends Genet.* 22: 375–387.
- Dujon, B., 2010 Yeast evolutionary genomics. *Nat. Rev. Genet.* 11: 512–524.
- Dujon, B., 2015a Genome evolution in yeasts in *eLS*. John Wiley & Sons, Ltd., Chichester, United Kingdom.
- Dujon, B., 2015b Basic principles of yeast genomics, a personal recollection. *FEMS Yeast Res.* 15: fov047.
- Dujon, B., P. P. Slonimski, and L. Weill, 1974 Mitochondrial genetics IX: a model for recombination and segregation of mitochondrial genomes in *Saccharomyces cerevisiae*. *Genetics* 78: 415–437.
- Dujon, B., F. Sherman, G. Fischer, P. Durrrens, S. Casaregola *et al.*, 2004 Genome evolution in yeasts. *Nature* 430: 35–44.
- Dunn, B., T. Paulish, A. Stanbery, J. Piotrowski, G. Koniges *et al.*, 2013 Recurrent rearrangement during adaptive evolution in an interspecific yeast hybrid suggests a model for rapid introgressions. *PLoS Genet.* 9: e1003366.
- Fabre, E., H. Muller, P. Therizols, I. Lafontaine, B. Dujon *et al.*, 2005 Comparative genomics in hemiascomycete yeasts: evolution of sex, silencing, and subtelomeres. *Mol. Biol. Evol.* 22: 856–873.
- Fairhead, C., and B. Dujon, 2006 Structure of *Kluyveromyces lactis* subtelomeres: duplications and gene content. *FEMS Yeast Res.* 6: 428–441.
- Farrer, R. A., C. A. Desjardins, S. Sakthikumar, S. Gujja, S. Saif *et al.*, 2015 Genome evolution and innovation across the four major lineages of *Cryptococcus gattii*. *MBio* 6: e00868-e15.
- Fay, J. C., and J. A. Benavides, 2005 Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genet.* 1: 66–71.
- Fink, G. R., 1987 Pseudogenes in yeast? *Cell* 49: 5–6.
- Fischer, G., S. A. James, I. N. Roberts, S. G. Oliver, and E. J. Louis, 2000 Chromosomal evolution in *Saccharomyces*. *Nature* 405: 451–454.
- Fischer, G., C. Neuvéglise, P. Durrrens, C. Gaillardin, and B. Dujon, 2001 Evolution of gene order in the genomes of two related yeast species. *Genome Res.* 11: 2009–2019.
- Fitzpatrick, D. A., M. E. Logue, and G. Butler, 2008 Evidence of recent interkingdom horizontal gene transfer between bacteria and *Candida parapsilosis*. *BMC Evol. Biol.* 8: 181.
- Frank, A. C., and K. H. Wolfe, 2009 Evolutionary capture of viral and plasmid DNA by yeast nuclear chromosomes. *Eukaryot. Cell* 8: 1521–1531.
- Freel, K. C., A. Friedrich, and J. Schacherer, 2015 Mitochondrial genome evolution in yeasts: an all-encompassing view. *FEMS Yeast Res.* 15: fov023.
- Freel, K. C., A. Friedrich, V. Sarilar, H. Devillers, C. Neuvéglise *et al.*, 2016 Whole-genome sequencing and intraspecific analysis of the yeast species *Lachancea quebecensis*. *Genome Biol. Evol.* 8: 733–741.
- Friedrich, A., P. Jung, C. Reisser, G. Fischer, and J. Schacherer, 2015 Population genomics reveals chromosome-scale heterogeneous evolution in a protoploid yeast. *Mol. Biol. Evol.* 32: 184–192.
- Fritsch, E. S., C. D. Chabbert, B. Klaus, and L. M. Steinmetz, 2014 A genome-wide map of mitochondrial DNA recombination in yeast. *Genetics* 198: 755–771.
- Gabaldón, T., T. Martin, M. Marcet-Houben, P. Durrrens, M. Bolotin-Fukuhara *et al.*, 2013 Comparative genomics of emerging pathogens in the *Candida glabrata* clade. *BMC Genomics* 14: 623.
- Galeote, V., F. Bigey, E. Beyne, M. Novo, J.-L. Legras *et al.*, 2011 Amplification of a *Zygosaccharomyces bailii* DNA segment in wine yeast genomes by extrachromosomal circular DNA formation. *PLoS One* 6: e17872.
- Galeote, V., F. Bigey, H. Devillers, C. Neuvéglise, and S. Dequin, 2013 Genome sequence of the food spoilage yeast *Zygosaccharomyces bailii* CLIB213<sup>T</sup>. *Genome Announc.* 1: e00606–e613.
- Gallone, B., J. Steensels, T. Prah, L. Soriaga, V. Saels *et al.*, 2016 Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* 166: 1397–1410.
- Gayevskiy, V., and M. R. Goddard, 2012 Geographic delineations of yeast communities and populations associated with vines and wines in New Zealand. *ISME J.* 6: 1281–1290.
- Gayevskiy, V., and M. R. Goddard, 2016 *Saccharomyces eubayanus* and *Saccharomyces arboricola* reside in North Island native New Zealand forests. *Environ. Microbiol.* 18: 1137–1147.
- Giaever, G., and C. Nislow, 2014 The yeast deletion collection: a decade of functional genomics. *Genetics* 197: 451–465.
- Goddard, M. R., N. Anfang, R. Tang, R. C. Gardner, and C. Jun, 2010 A distinct population of *Saccharomyces cerevisiae* in New Zealand: evidence for local dispersal by insects and human-aided global dispersal in oak barrels. *Environ. Microbiol.* 12: 63–73.
- Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon *et al.*, 1996 Life with 6000 genes. *Science* 274: 546, 563–547.
- Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon *et al.*, 1997 The yeast genome directory. *Nature* 387: 5–105.
- Gojković, Z., W. Knecht, E. Zameitat, J. Warneboldt, J. B. Coutelis *et al.*, 2004 Horizontal gene transfer promoted evolution of

- the ability to propagate under anaerobic conditions in yeasts. *Mol. Genet. Genomics* 271: 387–393.
- Goncalves, M., A. Pontes, P. Almeida, R. Barbosa, M. Serra *et al.*, 2016 Distinct domestication trajectories in top-fermenting beer yeasts and wine yeasts. *Curr. Biol.* 26: 2750–2761.
- Gonzalez, S. S., E. Barrio, J. Gafner, and A. Querol, 2006 Natural hybrids from *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations. *FEMS Yeast Res.* 6: 1221–1234.
- Gordon, J. L., and K. H. Wolfe, 2008 Recent allopolyploid origin of *Zygosaccharomyces rouxii* strain ATCC 42981. *Yeast* 25: 449–456.
- Gordon, J. L., K. P. Byrne, and K. H. Wolfe, 2009 Additions, losses and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. *PLoS Genet.* 5: e1000485.
- Gordon, J. L., D. Armisén, E. Proux-Wéra, S. S. ÓhÉigeartaigh, K.P. Byrne *et al.*, 2011 Evolutionary erosion of yeast sex chromosomes by mating-type switching accidents. *Proc. Natl. Acad. Sci. USA* 108: 20024–20029.
- Greig, D., 2007 A screen for recessive speciation genes expressed in the gametes of F1 hybrid yeast. *PLoS Genet.* 3: e21.
- Greig, D., 2009 Reproductive isolation in *Saccharomyces*. *Heredity* 102: 39–44.
- Greig, D., R. H. Borts, E. J. Louis, and M. Travisano, 2002a Epistasis and hybrid sterility in *Saccharomyces*. *Proc. Biol. Sci.* 269: 1167–1171.
- Greig, D., E. Louis, R. Borts, and M. Travisano, 2002b Hybrid speciation in experimental populations of yeast. *Science* 298: 1773–1775.
- Greig, D., M. Travisano, E. J. Louis, and R. H. Borts, 2003 A role for the mismatch repair system during incipient speciation in *Saccharomyces*. *J. Evol. Biol.* 16: 429–437.
- Gresham, D., D. M. Ruderfer, S. C. Pratt, J. Schacherer, M. J. Dunham *et al.*, 2006 Genome-wide detection of polymorphisms at nucleotide resolution with a single DNA microarray. *Science* 311: 1932–1936.
- Grosjean, H., V. de Crecy-Lagard, and C. Marck, 2010 Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. *FEBS Lett.* 584: 252–264.
- Guerineau, M., C. Grandchamp, and P. P. Slonimski, 1971 Characterization of a new class of circular DNA molecules in yeast. *Biochem. Biophys. Res. Commun.* 42: 550–557.
- Haber, J., 2012 Mating-type genes and MAT switching in *Saccharomyces cerevisiae*. *Genetics* 191: 33–64.
- Hansen, J., and M. C. Kielland-Brandt, 1994 *Saccharomyces carlsbergensis* contains two functional MET2 alleles similar to homologues from *S. cerevisiae* and *S. monacensis*. *Gene* 140: 33–40.
- Hansen, J., H. Cherest, and M. C. Kielland-Brandt, 1994 Two divergent MET10 genes, one from *Saccharomyces cerevisiae* and one from *Saccharomyces carlsbergensis*, encode the alpha subunit of sulfite reductase and specify potential binding sites for FAD and NADPH. *J. Bacteriol.* 176: 6050–6058.
- Hanson, S. J., K. P. Byrne, and K. H. Wolfe, 2014 Mating-type switching by chromosomal inversion in methylotrophic yeasts suggest an origin for the three-locus *Saccharomyces cerevisiae* system. *Proc. Natl. Acad. Sci. USA* 111: E4851–E4858.
- Heck, J. A., J. L. Argueso, Z. Gemici, R. G. Reeves, A. Bernard *et al.*, 2006 Negative epistasis between natural variants of the *Saccharomyces cerevisiae* MLH1 and PMS1 genes results in a defect in mismatch repair. *Proc. Natl. Acad. Sci. USA* 103: 3256–3261.
- Heitman, J., S. Sun, and T. Y. James, 2013 Evolution of fungal sexual reproduction. *Mycologia* 105: 1–27.
- Hershkovitz, V., N. Sela, L. Taha-Salaime, J. Liu, G. Rafael *et al.*, 2013 *De-novo* assembly and characterization of the transcriptome of *Metchnikowia fructicola* reveals differences in gene expression following interaction with *Penicillium digitatum* and grapefruit peel. *BMC Genomics* 14: 168.
- Hickman, M., C. A. Froyd, and L. N. Rusche, 2011 Reinventing heterochromatin in budding yeasts: Sir2 and the origin recognition complex take center stage. *Eukaryot. Cell* 10: 1183–1192.
- Hittinger, C. T., 2013 *Saccharomyces* diversity and evolution: a budding model genus. *Trends Genet.* 29: 309–317.
- Hittinger, C. T., and S. B. Carroll, 2007 Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449: 677–681.
- Hittinger, C. T., A. Rokas, and S. B. Carroll, 2004 Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proc. Natl. Acad. Sci. USA* 101: 14144–14149.
- Hittinger, C. T., P. Goncalves, J. P. Sampaio, J. Dover, M. Johnston *et al.*, 2010 Remarkably ancient balanced polymorphisms in a multi-locus gene network. *Nature* 464: 54–58.
- Hittinger, C. T., A. Rokas, F. Y. Bai, T. Boekhout, P. Gonçalves *et al.*, 2015 Genomics and the making of yeast biodiversity. *Curr. Opin. Genet. Dev.* 35: 100–109.
- Hooks, K. B., D. Delneri, and S. Griffiths-Jones, 2014 Intron evolution in *Saccharomycetaceae*. *Genome Biol. Evol.* 6: 2543–2556.
- Hou, J., A. Friedrich, J. de Montigny, and J. Schacherer, 2014 Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in *Saccharomyces cerevisiae*. *Curr. Biol.* 24: 1153–1159.
- Hou, J., A. Friedrich, J. S. Gounot, and J. Schacherer, 2015 Comprehensive survey of condition-specific reproductive isolation reveals genetic incompatibility in yeast. *Nat. Commun.* 6: 7214.
- Hu, Y., J. Stenlid, M. Elfstrand, and A. Olson, 2013 Evolution of RNA interference proteins dicer and argonaute in *Basidiomycota*. *Mycologia* 105: 1489–1498.
- Hunter, N., S. R. Chambers, E. J. Louis, and R. H. Borts, 1996 The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. *EMBO J.* 15: 1726–1733.
- Inokuma, K., J. Ishii, K. Y. Hara, M. Mochizuki, T. Hasunuma *et al.*, 2015 Complete genome sequence of *Kluyveromyces marxianus* NBRC1777, a nonconventional thermotolerant yeast. *Genome Announc.* 3: e00389-15 [erratum: *Genome Announc.* 5: e00058-17 (2017)].
- Ishii, K., 2009 Conservation and divergence of centromere specification in yeast. *Curr. Op. Microb.* 12: 616–622.
- Jackson, A. P., J. A. Gamble, T. Yeomans, G. P. Moran, D. Saunders *et al.*, 2009 Comparative genomics of the fungal pathogens *Candida dubliniensis* and *Candida albicans*. *Genome Res.* 19: 2231–2244.
- Jeffries, T. W., I. V. Grigoriev, J. Grimwood, J. M. Laplaza, A. Aerts *et al.*, 2007 Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*. *Nat. Biotechnol.* 25: 319–326.
- Jones, T., N. A. Federspiel, H. Chibana, J. Dungan, S. Kalman *et al.*, 2004 The diploid genome sequence of *Candida albicans*. *Proc. Natl. Acad. Sci. USA* 101: 7329–7334.
- Kachouri, R., V. Stribinskis, Y. Zhu, K. S. Ramos, E. Westhof *et al.*, 2005 A surprisingly large RNase P RNA in *Candida glabrata*. *RNA* 11: 1064–1072.
- Kachouri-Lafond, R., B. Dujon, E. Gilson, E. Westhof, C. Fairhead *et al.*, 2009 Large telomerase RNA, telomere length heterogeneity and escape from senescence in *Candida glabrata*. *FEBS Lett.* 583: 3605–3610.
- Kapoor, S., L. Zhu, C. Froyd, T. Liu, and L. N. Rusche, 2015 Regional centromeres in the yeast *Candida lusitanae* lack pericentromeric heterochromatin. *Proc. Natl. Acad. Sci. USA* 112: 12139–12144.
- Kellis, M., N. Patterson, M. Endrizzi, B. Birren, and E. S. Lander, 2003 Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423: 241–254.

- Kellis, M., B. W. Birren, and E. S. Lander, 2004 Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428: 617–624.
- Kelkar, Y. D., and H. Ochman, 2011 Causes and consequences of genome expansion in fungi. *Genome Biol. Evol.* 4: 13–23.
- Ketel, C., H. S. Wang, M. McClellan, K. Bouchonville, A. Selmecki *et al.*, 2009 Neocentromeres form efficiently at multiple possible loci in *Candida albicans*. *PLoS Genet.* 5: e1000400.
- Kim, J. S., J. H. Baek, N. H. Park, and C. Kim, 2015 Complete genome sequence of rhizophilic yeast *Meyerozyma caribbica* MG20W isolated from rhizosphere soil. *Genome Announc.* 3: e00127-15.
- Klar, A. J., K. Ishikawa, and S. Moore 2014 A unique DNA recombination mechanism of the mating/cell-type switching of fission yeasts: a review. *Microbiol. Spectr.* 2. DOI: 10.1128/microbiolspec.MDNA3-0003-2014.
- Kobayashi, N., Y. Suzuki, L. W. Schoenfeld, C. A. Müller, C. Nieduszynski *et al.*, 2015 Discovery of an unconventional centromere in budding yeasts redefines evolution of point centromeres. *Curr. Biol.* 25: 2026–2033.
- Kozul, R., A. Malpertuy, L. Frangeul, C. Bouchier, P. Wincker *et al.*, 2003 The complete mitochondrial genome sequence of the pathogenic yeast *Candida (Torulopsis) glabrata*. *FEBS Lett.* 534: 39–48.
- Kozul, R., S. Caburet, B. Dujon, and G. Fischer, 2004 Eukaryotic genome evolution through the spontaneous duplication of large chromosomal segments. *EMBO J.* 23: 234–243.
- Kuang, M. C., P. D. Hutchins, J. D. Russell, J. J. Coon, and C. T. Hittinger, 2016 Ongoing resolution of duplicate gene functions shapes the diversification of a metabolic network. *Elife* 5: 10.7554/eLife.19027.
- Kuberl, A., J. Schneider, G. G. Thallinger, I. Anderl, D. Wibberg *et al.*, 2011 High-quality genome sequence of *Pichia pastoris* CBS7435. *J. Biotechnol.* 154: 312–320.
- Kunze, G., C. Gaillardin, M. Czernicka, P. Durrens, T. Martin *et al.*, 2014 The complete genome of *Blastobotrys (Arxula) adenivorans* LS3 – a yeast of biotechnological interest. *Biotechnol. Biofuels* 7: 66.
- Kurtzman, C. P., J. W. Fell, and T. Boekhout, 2011 *The Yeasts: A Taxonomic Study*, Ed. 5. Elsevier, Amsterdam.
- Lafontaine, I., and B. Dujon, 2010 Origin and fate of pseudogenes in hemiascomycetes: a comparative analysis. *BMC Genomics* 11: 260.
- Lafontaine, I., G. Fischer, E. Talla, and B. Dujon, 2004 Gene relics in the genome of the yeast *Saccharomyces cerevisiae*. *Gene* 335: 1–17.
- Laureau, R., S. Loeillet, F. Salinas, A. Bergström, P. Legoix-Né *et al.*, 2016 Extensive recombination of a yeast diploid hybrid through meiotic reversion. *PLoS Genet.* 12: e1005781.
- Lenassi, M., C. Gostincar, S. Jackman, M. Turk, I. Sadowski *et al.*, 2013 Whole genome duplication and enrichment of metal cation transporters revealed by *de novo* genome sequencing of extremely halotolerant black yeast *Hortaea werneckii*. *PLoS One* 8: e71328.
- Lee, H. Y., J. Y. Chou, L. Cheong, N. H. Chang, S. Y. Yang *et al.*, 2008 Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* 135: 1065–1073.
- Lefrancois, P., R. K. Auerbach, C. M. Yellman, G. S. Roeder, and M. Snyder, 2013 Centromere-like regions in the budding yeast genome. *PLoS Genet.* 9: e1003209.
- Legras, J. L., D. Merdinoglu, J. M. Cornuet, and F. Karst, 2007 Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Mol. Ecol.* 16: 2091–2102.
- Leh-Louis, V., L. Despons, A. Friedrich, T. Martin, P. Durrens *et al.*, 2012 *Pichia sorbitophila*, an interspecies yeast hybrid, reveals early steps of genome resolution after polyploidization. *G3 (Bethesda)* 2: 299–311.
- Lertwattanasakul, N., T. Kosaka, A. Hosoyama, Y. Suzuki, N. Rodrussamee *et al.*, 2015 Genetic basis of the highly efficient yeast *Kluyveromyces marxianus*: complete genome sequence and transcriptome analyses. *Biotechnol. Biofuels* 8: 47.
- Li, D., Y. Dong, Y. Jiang, H. Jiang, J. Cai *et al.*, 2010 A *de novo* originated gene depresses budding yeast mating pathway and is repressed by the protein encoded by its antisense strand. *Cell Res.* 20: 408–420.
- Li, D., Z. Yan, L. Lu, H. Jiang, and W. Wang, 2014 Pleiotropy of the *de novo* originated gene *MDF1*. *Sci. Rep.* 4: 7280.
- Li, W., A. F. Averette, M. Desnos-Ollivier, M. Ni, F. Dromer *et al.*, 2012 Genetic diversity and genomic plasticity of *Cryptococcus neoformans* AD hybrid strains. *G3 (Bethesda)* 2: 83–97.
- Libkind, D., C. T. Hittinger, E. Valerio, C. Goncalves, J. Dover *et al.*, 2011 Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc. Natl. Acad. Sci. USA* 108: 14539–14544.
- Ling, J., R. Daoud, M. J. Lajoie, G. M. Church, D. Söll *et al.*, 2014 Natural reassignment of CUU and CUA sense codons to alanine in *Ashbya* mitochondria. *Nucleic Acids Res.* 42: 499–508.
- Liti, G., and E. J. Louis, 2012 Advances in quantitative trait analysis in yeast. *PLoS Genet.* 8: e1002912.
- Liti, G., D. B. Barton, and E. J. Louis, 2006 Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics* 174: 839–850.
- Liti, G., D. M. Carter, A. M. Moses, J. Warringer, L. Parts *et al.*, 2009a Population genomics of domestic and wild yeasts. *Nature* 458: 337–341.
- Liti, G., S. Haricharan, F. A. Cubillos, A. L. Tierney, S. Sharp *et al.*, 2009b Segregating YKU80 and TLC1 alleles underlying natural variation in telomere properties in wild yeast. *PLoS Genet.* 5: e1000659.
- Liti, G., B. A. Nguyen, M. Blythe, C. A. Müller, A. Bergström *et al.*, 2013 High quality *de novo* sequencing and assembly of the *Saccharomyces arboricolus* genome. *BMC Genomics* 14: 69.
- Liu, X., R. S. Kaas, R. Jensen, and M. Workman, 2012 Draft genome sequence of the yeast *Pachysolen tannophilus* CBS4044/NRR1 Y-2460. *Eukaryot. Cell* 11: 827.
- Llorente, B., A. Malpertuy, C. Neuvéglise, J. de Montigny, M. Aigle *et al.*, 2000 Genomic exploration of the hemiascomycetous yeasts: 18. Comparative analysis of chromosome maps and synteny with *Saccharomyces cerevisiae*. *FEBS Lett.* 487: 101–112.
- Lobo, F. P., D. L. Gonçalves, S. L. Alves Jr., A. L. Gerber, A. T. de Vasconcelos *et al.*, 2014 Draft genome sequence of the D-xylose-fermenting yeast *Spathaspora arborariae* UFMG-HM19.1AT. *Genome Announc.* 2: e01163-13.
- Lopandic, K., H. Gangl, E. Wallner, G. Tschek, G. Leitner *et al.*, 2007 Genetically different wine yeasts isolated from Austrian vine-growing regions influence wine aroma differently and contain putative hybrids between *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii*. *FEMS Yeast Res.* 7: 953–965.
- Louis, E. J., 1995 The chromosome ends of *Saccharomyces cerevisiae*. *Yeast* 11: 1553–1573.
- Louis, E. J., 2009 Evolutionary genetics: origins of reproductive isolation. *Nature* 457: 549–550.
- Louis, E. J., 2011 Population genomics and speciation in yeasts. *Fungal Biol. Rev.* 25: 136–142.
- Louis, E. J., and J. E. Haber, 1990 Mitotic recombination among subtelomeric Y' repeats in *Saccharomyces cerevisiae*. *Genetics* 124: 547–559.
- Louis, E. J., and J. E. Haber, 1992 The structure and evolution of subtelomeric Y' repeats in *Saccharomyces cerevisiae*. *Genetics* 131: 559–574.
- Louis, E. J., E. S. Naumova, A. Lee, G. Naumov, and J. E. Haber, 1994 The chromosome end in yeast: its mosaic nature and influence on recombinational dynamics. *Genetics* 136: 789–802.

- Ludlow, C. L., G. A. Cromie, C. Garmendia-Torres, A. Sirr, M. Hays *et al.*, 2016 Independent origins of yeast associated with coffee and cacao fermentation. *Curr. Biol.* 26: 965–971.
- Lynch, D. B., M. E. Logue, G. Butler, and K. H. Wolfe, 2010 Chromosomal G + C content evolution in yeasts: systematic interspecies differences, and GC-poor troughs at centromeres. *Genome Biol. Evol.* 2: 572–583.
- Maekawa, H., and Y. Kaneko, 2014 Inversion of the chromosomal region between two mating type loci switches the mating type in *Hansenula polymorpha*. *PLoS Genet.* 10: e1004796.
- Magwene, P. M., Ö. Kayıkçı, J. A. Granek, J. M. Reininga, Z. Scholl *et al.*, 2011 Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 108: 1987–1992.
- Malik, H. S., and S. Henikoff, 2009 Major evolutionary transitions in centromere complexity. *Cell* 138: 1067–1082.
- Mallet, J., N. Besansky, and M. W. Hahn, 2016 How reticulated are species? *Bioessays* 38: 140–149.
- Malpertuy, A., F. Tekaiia, S. Casarégola, M. Aigle, F. Artiguenave *et al.*, 2000 Genomic exploration of the hemiascomycetous yeasts: 19. Ascomycete-specific genes. *FEBS Lett.* 487: 113–121.
- Marcet-Houben, M., and T. Gabaldón 2015 Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLoS Biol.* 13: e1002220.
- Marcet-Houben, M., G. Marceddu, and T. Gabaldón, 2009 Phylogenomics of the oxidative phosphorylation in fungi reveals extensive gene duplication followed by functional divergence. *BMC Evol. Biol.* 9: 295.
- Marck, C., R. Kachouri-Lafond, I. Lafontaine, E. Westhof, B. Dujon *et al.*, 2006 The RNA polymerase III-dependent family of genes in hemiascomycetes: comparative RNomics, decoding strategies, transcription and evolutionary implications. *Nucleic Acids Res.* 34: 1816–1835.
- Marsit, S., A. Mena, F. Bigey, F. X. Sauvage, A. Couloux *et al.*, 2015 Evolutionary advantage conferred by an eukaryote-to-eukaryote gene transfer event in wine yeasts. *Mol. Biol. Evol.* 32: 1695–1707.
- Martin, K. J., and P. T. Rygielwicz, 2005 Fungal-specific PCR primers developed for analysis of the ITS region of environmental DNA extracts. *BMC Microbiol.* 5: 28.
- Massey, S. E., G. Moura, P. Beltrao, R. Almeida, J. R. Garey *et al.*, 2003 Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in *Candida* spp. *Genome Res.* 13: 544–557.
- Mattanovich, D., A. Graf, J. Stadlmann, M. Dragosits, A. Redl *et al.*, 2009 Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris*. *Microb. Cell Fact.* 8: 29.
- Meraldi, P., A. D. McAinsch, E. Rheinaby, and P. K. Sorger, 2006 Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. *Genome Biol.* 7: R23.
- Mira, N. P., M. Münsterkötter, F. Dias-Valada, J. Santos, M. Palma *et al.*, 2014 The genome sequence of the highly acetic-tolerant *Zygosaccharomyces bailii*-derived interspecies hybrid strain ISA1307, isolated from a sparkling wine plant. *DNA Res.* 21: 299–313.
- Mitrovich, Q. M., and C. Guthrie, 2007 Evolution of small nuclear RNAs in *S. cerevisiae*, *C. albicans*, and other hemiascomycetous yeasts. *RNA* 13: 2066–2080.
- Morales, L., and B. Dujon, 2012 Evolutionary role of interspecies hybridization and genetic exchanges in yeasts. *Microbiol. Mol. Biol. Rev.* 76: 721–739.
- Morales, L., B. Noel, B. Porcel, M. Marcet-Houben, M.-F. Hullo *et al.*, 2013 Complete DNA sequence of *Kuraishia capsulata* illustrates novel genomic features among budding yeasts (*Saccharomycotina*). *Genome Biol. Evol.* 5: 2524–2539.
- Morel, G., L. Sterck, D. Swennen, M. Marcet-Houben, D. Onesime *et al.*, 2015 Differential gene retention as an evolutionary mechanism to generate biodiversity and adaptation in yeasts. *Sci. Rep.* 5: 11571.
- Mühlhausen, S., P. Findensen, U. Plessmann, H. Urlaub, and M. Kolmar, 2016 A novel nuclear genetic code alteration in yeasts and the evolution of codon reassignment in eukaryotes. *Genome Res.* 25: 945–955.
- Müller, H., C. Hennequin, B. Dujon, and C. Fairhead, 2007 Ascomycetes: the *Candida MAT* locus. Comparing *MAT* in the genomes of Hemiascomycetous yeasts, pp. 247–263 in *Sex in Fungi: Molecular Determination and Evolutionary Implications*, edited by J. Heitman, J. W. Kronstad, J. W. Taylor, and L. A. Casselton. American Society for Microbiology, Washington, DC.
- Müller, H., A. Thierry, J. Y. Coppée, C. Gouyette, C. Hennequin *et al.*, 2009 Genomic polymorphism in the population of *Candida glabrata*: gene copy-number variation and chromosomal translocations. *Fungal Genet. Biol.* 46: 264–276.
- Muller, L. A., and J. H. McCusker, 2009 A multispecies-based taxonomic microarray reveals interspecies hybridization and introgression in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 9: 143–152.
- Nagy, L. G., R. A. Ohm, G. M. Kovács, D. Floudas, R. Riley *et al.*, 2014 Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. *Nat. Commun.* 5: 4471.
- Nakao, Y., T. Kanamori, T. Itoh, Y. Kodama, S. Rainieri *et al.*, 2009 Genome sequence of the lager brewing yeast, an interspecies hybrid. *DNA Res.* 16: 115–129.
- Naseeb, S., S. A. James, H. Alsammar, C. Michaels, B. Gini *et al.*, 2017 *Saccharomyces jurei* sp. nov., Isolation and genetic identification of a novel yeast species from *Quercus robur*. *Int. J. Syst. Evol. Microbiol.* (in press).
- Naumov, G. I., 1987 Genetic basis for classification and identification of the ascomycetes yeasts. *Stud. Mycol.* 30: 469–475.
- Naumov, G. I., E. S. Naumova, R. A. Lantto, E. J. Louis, and M. Korhola, 1992 Genetic homology between *Saccharomyces cerevisiae* and its sibling species *S. paradoxus* and *S. bayanus*: electrophoretic karyotypes. *Yeast* 8: 599–612.
- Naumov, G. I., E. S. Naumova, A. N. Hagler, L. C. Mendonca-Hagler, and E. J. Louis, 1995a A new genetically isolated population of the *Saccharomyces sensu stricto* complex from Brazil. *Antonie van Leeuwenhoek* 67: 351–355.
- Naumov, G. I., E. S. Naumova, and E. J. Louis, 1995b Two new genetically isolated populations of the *Saccharomyces sensu stricto* complex from Japan. *J. Gen. Appl. Microbiol.* 41: 499–505.
- Naumov, G. I., S. A. James, E. S. Naumova, E. J. Louis, and I. N. Roberts, 2000 Three new species in the *Saccharomyces sensu stricto* complex: *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*. *Int. J. Syst. Evol. Microbiol.* 50: 1931–1942.
- Neuvéglise, C., C. Marck, and C. Gaillardin, 2011 The intronome of budding yeasts. *C. R. Biol.* 334: 662–670.
- Nguyen, H. V., J. L. Legras, C. Neuvéglise, and C. Gaillardin, 2011 Deciphering the hybridisation history leading to the Lager lineage based on the mosaic genomes of *Saccharomyces bayanus* strains NBRC1948 and CBS380. *PLoS One* 6: e25821.
- Nijkamp, J. F., M. van den Broek, E. Datema, S. de Kok, L. Bosman *et al.*, 2012 De novo sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN. PK113-7D, a model for modern industrial biotechnology. *Microb. Cell Fact.* 11: 36.
- Novo, M., F. Bigey, E. Beyne, V. Galeote, F. Gavory *et al.*, 2009 Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proc. Natl. Acad. Sci. USA* 106: 16333–16338.



- Padmanabhan, S., J. Thakur, R. Siddharthan, and K. Sanyal, 2008 Rapid evolution of Cse4p-rich centromeric DNA sequences in closely related pathogenic yeasts, *Candida albicans* and *Candida dubliniensis*. *Proc. Natl. Acad. Sci. USA* 105: 19797–19802.
- Parenteau, J., M. Durand, S. Véronneau, A. A. Lacombe, G. Morin *et al.*, 2008 Deletion of many yeast introns reveals a minority of genes that require splicing for function. *Mol. Biol. Cell* 19: 1932–1941.
- Parenteau, J., M. Durand, G. Morin, J. Gagnon, J. F. Lucier *et al.*, 2011 Introns within ribosomal protein genes regulate the production and function of yeast ribosomes. *Cell* 147: 320–331.
- Payen, C., S. C. Di Rienzi, G. T. Ong, J. L. Pogachar, J. C. Sanchez *et al.*, 2014 The dynamics of diverse segmental amplifications in populations of *Saccharomyces cerevisiae* adapting to strong selection. *G3 (Bethesda)* 4: 399–409.
- Perez-Ortin, J. E., A. Querol, S. Puig, and E. Barrio, 2002 Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. *Genome Res.* 12: 1533–1539.
- Peris, D., K. Sylvester, D. Libkind, P. Goncalves, J. P. Sampaio *et al.*, 2014 Population structure and reticulate evolution of *Saccharomyces eubayanus* and its lager-brewing hybrids. *Mol. Ecol.* 23: 2031–2045.
- Peris, D., Q. K. Langdon, R. V. Moriarty, K. Sylvester, M. Bontrager *et al.*, 2016a Complex ancestries of lager-brewing hybrids were shaped by standing variation in the wild yeast *Saccharomyces eubayanus*. *PLoS Genet.* 12: e1006155.
- Peris, D., L. Perez-Traves, C. Belloch, and A. Querol 2016b Enological characterization of Spanish *Saccharomyces kudriavzevii* strains, one of the closest relatives to parental strains of wine-making and brewing *Saccharomyces cerevisiae* x *S. kudriavzevii* hybrids. *Food Microbiol.* 53: 31–40.
- Piškur, J., Z. Ling, M. Marcet-Houben, O. P. Ishchuk, A. Aerts *et al.*, 2012 The genome of wine yeast *Dekkera bruxellensis* provides a tool to explore its food-related properties. *Int. J. Food Microbiol.* 157: 202–209.
- Pramateftaki, P. V., V. N. Kourelis, P. Lanaridis, and M. A. Typas, 2006 The mitochondrial genome of the wine yeast *Hanseniaspora uvarum*: a unique gene organization among yeast/fungal counterparts. *FEMS Yeast Res.* 6: 77–90.
- Proux-Wera, E., K. P. Byrne, and K. H. Wolfe, 2013 Evolutionary mobility of the ribosomal DNA array in yeasts. *Genome Biol. Evol.* 5: 525–531.
- Pryszcz, L. P., T. Nemeth, A. Gacser, and T. Gabaldón, 2014 Genome comparison of *Candida orthopsilosis* clinical strains reveal the existence of hybrids between two distinct subspecies. *Genome Biol. Evol.* 6: 1069–1078.
- Rajaei, N., K. Chiruvella, F. Lin, and S. U. Astrom, 2014 Domesticated transposase Kat1 and its fossil imprints induce sexual differentiation in yeast. *Proc. Natl. Acad. Sci. USA* 111: 15491–15496.
- Ralsler, M., H. Kuhl, M. Werber, H. Lehrach, M. Breitenbach *et al.*, 2012 The *Saccharomyces cerevisiae* W303–K6001 cross-platform genome sequence: insights into ancestry and physiology of a laboratory mutt. *Open Biol.* 2: 120093.
- Ravin, N. V., M. A. Eldarov, V. V. Kadnikov, A. V. Beletsky, J. Schneider *et al.*, 2013 Genome sequence and analysis of methylophilic yeast *Hansenula polymorpha* DL1. *BMC Genomics* 14: 837.
- Rhind, N., Z. Chen, M. Yassour, D. A. Thompson, B. J. Haas *et al.*, 2011 Comparative functional genomics of the fission yeasts. *Science* 332: 930–966.
- Ricchetti, M., C. Fairhead, and B. Dujon, 1999 Mitochondrial DNA repairs double strand breaks in yeast chromosomes. *Nature* 402: 96–100.
- Riccombeni, A., G. Vidanes, E. Proux-Wéra, K. H. Wolfe, and G. Butler, 2012 Sequence and analysis of the genome of the pathogenic yeast *Candida orthopsilosis*. *PLoS One* 7: e35750.
- Riley, R., S. Haridas, K. H. Wolfe, M. R. Lopes, C. T. Hittinger *et al.*, 2016 Comparative genomics of biotechnologically important yeasts. *Proc. Natl. Acad. Sci. USA* 113: 9882–9887.
- Rolland, T., and B. Dujon, 2011 Yeasty clocks: dating genomic changes in yeasts. *C. R. Biol.* 334: 620–628.
- Rolland, T., C. Neuvéglise, C. Sacerdot, and B. Dujon, 2009 Insertion of horizontally transferred genes within conserved syntenic regions of yeast genomes. *PLoS One* 4: e6515.
- Rusche, L. N., and J. Rine, 2001 Conversion of a gene-specific repressor to a regional silencer. *Genes Dev.* 15: 955–967.
- Russell, A. G., J. M. Charette, D. F. Spencer, and M. W. Gray, 2006 An early evolutionary origin for the minor spliceosome. *Nature* 443: 863–866.
- Sacerdot, C., S. Casaregola, I. Lafontaine, F. Tekaia, B. Dujon *et al.*, 2008 Promiscuous DNA in the nuclear genomes of hemiascomycetous yeasts. *FEMS Yeast Res.* 8: 846–857.
- Sampaio, J. P., and P. Goncalves, 2008 Natural populations of *Saccharomyces kudriavzevii* in Portugal are associated with oak bark and are sympatric with *S. cerevisiae* and *S. paradoxus*. *Appl. Environ. Microbiol.* 74: 2144–2152.
- Santos, M. A., and M. F. Tuite, 1995 The CUG codon is decoded in vivo as serine and not leucine in *Candida albicans*. *Nucleic Acids Res.* 23: 1481–1486.
- Santos, M. A., A. C. Gomes, M. C. Santos, L. C. Carreto, and G. R. Moura, 2011 The genetic code of the fungal CTG clade. *C. R. Biol.* 334: 607–611.
- Sanyal, K., M. Baum, and J. Carbon, 2004 Centromeric DNA sequences in the pathogenic yeast *Candida albicans* are all different and unique. *Proc. Natl. Acad. Sci. USA* 101: 11374–11379.
- Sarilar, V., H. Devillers, K. C. Freel, J. Schacherer, and C. Neuvéglise, 2015 Draft genome sequence of *Lachancea lanzarotensis* CBS 12615<sup>T</sup>, an ascomycetous yeast isolated from grapes. *Genome Announc.* 3: e00292-15.
- Scannell, D. R., K. P. Byrne, J. L. Gordon, S. Wong, and K. H. Wolfe, 2006 Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* 440: 341–345.
- Scannell, D. R., G. Butler, and K. H. Wolfe, 2007a Yeast genome evolution—the origin of the species. *Yeast* 24: 929–942.
- Scannell, D. R., A. C. Frank, G. C. Conant, K. P. Byrne, M. Woolfit *et al.*, 2007b Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proc. Natl. Acad. Sci. USA* 104: 8397–8402.
- Scannell, D. R., O. A. Zill, A. Rokas, C. Payen, M. J. Dunham *et al.*, 2011 The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. *G3 (Bethesda)* 1: 11–25.
- Schacherer, J., D. M. Ruderfer, D. Gresham, K. Dolinski, D. Botstein *et al.*, 2007a Genome-wide analysis of nucleotide-level variation in commonly used *Saccharomyces cerevisiae* strains. *PLoS One* 2: e322.
- Schacherer, J., Y. Tourrette, S. Potier, J.-L. Souciet, and J. de Montigny, 2007b Spontaneous duplications in diploid *Saccharomyces cerevisiae* cells. *DNA Repair (Amst.)* 6: 1441–1452.
- Schacherer, J., J. A. Shapiro, D. M. Ruderfer, and L. Kruglyak, 2009 Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* 458: 342–345.
- Schröder, M., K. Martinez de San Vicente, T. H. R. Prandini, S. Hammel, D. G. Higgins *et al.*, 2016 Multiple origins of the pathogenic yeast *Candida orthopsilosis* by separate hybridizations between two parental species. *PLoS Genet.* 12: e1006404.
- Sebastiani, F., C. Barberio, E. Casalone, D. Cavalieri, and M. Polinelli, 2002 Crosses between *Saccharomyces cerevisiae* and

- Saccharomyces bayanus* generate fertile hybrids. *Res. Microbiol.* 153: 53–58.
- Shen, X.-X., X. Zhou, J. Komiek, C. P. Kurtzman, C. T. Hittinger *et al.*, 2016 Reconstructing the backbone of the *Saccharomyces* yeast phylogeny using genomes-scale data. *G3 (Bethesda)* 3: 3927–3939.
- Sicard, D., and J. L. Legras, 2011 Bread, beer and wine: yeast domestication in the *Saccharomyces sensu stricto* complex. *C. R. Biol.* 334: 229–236.
- Sniegowski, P. D., P. G. Dombrowski, and E. Fingerman, 2002 *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. *FEMS Yeast Res.* 1: 299–306.
- Song, G., B. J. Dickins, J. Demeter, S. Engel, J. Gallagher *et al.*, 2015 AGAPE (Automated Genome Analysis PipelinE) for pan-genome analysis of *Saccharomyces cerevisiae*. *PLoS One* 10: e0120671.
- Souciet, J.-L., M. Aigle, F. Artiguenave, G. Blandin, M. Bolotin-Fukuhara *et al.*, 2000 Genomic exploration of the hemiascomycetous yeasts. *FEBS Lett.* 487: 3–12.
- Souciet, J.-L., B. Dujon, C. Gaillardin, M. Johnston, P. V. Baret *et al.*, 2009 Comparative genomics of protoploid *Saccharomycetaceae*. *Genome Res.* 19: 1696–1709.
- Stajich, J. E., F. S. Dietrich, and S. W. Roy, 2007 Comparative genomic analysis of fungal genomes reveals intron-rich ancestors. *Genome Biol.* 8: R223.
- Sternes, P. R., D. Lee, D. R. Kutyna, and A. R. Bornemann, 2016 Genome sequences of 3 species of *Hanseniaspora* isolated from spontaneous wine fermentation. *Genome Announc.* 4: e01287-16.
- Strope, P. K., S. G. Kozmin, D. A. Skelly, P. M. Magwene, F. S. Dietrich *et al.*, 2015a 2 $\mu$  plasmid in *Saccharomyces* species and in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 15: fov090.
- Strope, P. K., D. A. Skelly, S. G. Kozmin, G. Mahadevan, E. A. Stone *et al.*, 2015b The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res.* 25: 762–774.
- Sugita, T., and T. Nakase, 1999 Non-universal usage of the leucine CUG codon and the molecular phylogeny of the genus *Candida*. *Syst. Appl. Microbiol.* 22: 79–86.
- Tapsoba, F., J. L. Legras, A. Savadogo, S. Dequin, and A. S. Traore, 2015 Diversity of *Saccharomyces cerevisiae* strains isolated from *Borassus akeassii* palm wines from Burkina Faso in comparison to other African beverages. *Int. J. Food Microbiol.* 211: 128–133.
- Taylor, D. J., and J. Bruenn, 2009 The evolution of novel fungal genes from non-retroviral RNA viruses. *BMC Biol.* 7: 88.
- Thierry, A., V. Khanna, S. Créno, I. Lafontaine, L. Ma *et al.*, 2015 Macrotene chromosomes provide insights to a new mechanism of high-order gene amplification in eukaryotes. *Nat. Commun.* 6: 6154.
- Thierry, A., V. Khanna, and B. Dujon, 2016 Massive amplification at an unselected locus accompanies complex chromosomal rearrangements in yeast. *G3 (Bethesda)* 6: 1201–1215.
- Thomsen, P. F., and E. Willerslev, 2015 Environmental DNA – an emerging tool in conservation for monitoring past and present biodiversity. *Biol. Conserv.* 183: 4–18.
- Tsai, I. J., D. Bensasson, A. Burt, and V. Koufopanou, 2008 Population genomics of the wild yeast *Saccharomyces paradoxus*: quantifying the life cycle. *Proc. Natl. Acad. Sci. USA* 105: 4957–4962.
- Vakirlis, N., V. Sarilar, G. Drillon, A. Fleiss, N. Agier *et al.*, 2016 Reconstruction of ancestral chromosome architecture and gene repertoire reveals principles of genome evolution in a model yeast genus. *Genome Res.* 26: 918–932.
- Vaughan Martini, A., 1989 *Saccharomyces paradoxus* comb. nov., a newly separated species of the *Saccharomyces sensu stricto* complex based upon nDNA/nDNA homologies. *Syst. Appl. Microbiol.* 12: 179–182.
- Vaughan Martini, A., and A. Martini, 1987 Three newly delimited species of *Saccharomyces sensu stricto*. *Antonie van Leeuwenhoek* 53: 77–84.
- Vega-Alvarado, L., J. Gómez-Angulo, Z. Escalante-García, R. Grande, A. Gschaedler-Mathis *et al.*, 2015 High-quality draft genome sequence of *Candida apicola* NRRL Y-50540. *Genome Announc.* 3: e00437-15 (erratum: *Genome Announc.* 3: e01043-15).
- Vernis, L., A. Abbas, M. Chasles, C. M. Gaillardin, C. Brun *et al.*, 1997 An origin of replication and a centromere are both needed to establish replicative plasmid in the yeast *Yarrowia lipolytica*. *Mol. Cell. Biol.* 17: 1995–2004.
- Vernis, L., L. Poljak, M. Chasles, K. Uchida, S. Casarégola *et al.*, 2001 Only centromeres can supply the partition system required for ARS function in the yeast *Yarrowia lipolytica*. *J. Mol. Biol.* 305: 203–217.
- Wahlter, A., A. Hesselbart, and J. Wendland, 2014 Genome sequence of *Saccharomyces carlsbergensis*, the world's first pure culture lager yeasts. *G3 (Bethesda)* 4: 783–793.
- Wang, Q. M., W. Q. Liu, G. Liti, S. A. Wang, and F. Y. Bai, 2012 Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Mol. Ecol.* 21: 5404–5417.
- Wang, S. A., and F. Y. Bai, 2008 *Saccharomyces arboricolus* sp. nov., a yeast species from tree bark. *Int. J. Syst. Evol. Microbiol.* 58: 510–514.
- Wei, W., J. H. McCusker, R. W. Hyman, T. Jones, Y. Ning *et al.*, 2007 Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. *Proc. Natl. Acad. Sci. USA* 104: 12825–12830.
- Wendland, J., and A. Walther, 2011 Genome evolution in the *Eremothecium* clade of the *Saccharomyces* complex revealed by comparative genomics. *G3 (Bethesda)* 1: 539–548.
- Wendland, J., and A. Walther, 2014 Chromosome number reduction in *Eremothecium coryli* by two telomere-to-telomere fusions. *Genome Biol. Evol.* 6: 1186–1198.
- Wickner, R. B., T. Fujimura, and R. Esteban, 2013 Viruses and prions of *Saccharomyces cerevisiae*. *Adv. Virus Res.* 86: 1–36.
- Wohlbach, D. J., A. Kuo, T. K. Sato, K. M. Potts, A. A. Salamov *et al.*, 2011 Comparative genomics of xylose-fermenting fungi for enhanced biofuel production. *Proc. Natl. Acad. Sci. USA* 108: 13212–13217.
- Wolfe, K. H., 2015 Origin of the yeast whole-genome duplication. *PLoS Biol.* 13: e1002221.
- Wolfe, K. H., and D. C. Shields, 1997 Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387: 708–713.
- Wolfe, K. H., D. Armisén, E. Proux-Wera, S. S. ÓhÉigeartaigh, H. Azam *et al.*, 2015 Clade- and species-specific features of genome evolution in the *Saccharomycetaceae*. *FEMS Yeast Res.* 15: fov035.
- Woo, P. C. Y., S. K. P. Lau, B. Liu, J. J. Cai, K. T. K. Chong *et al.*, 2011 Draft genome sequence of *Penicillium marneffeii* strain PM1. *Eukaryot. Cell* 10: 1740–1741.
- Wood, V., R. Gwilliam, M. A. Rajandream, M. Lyne, R. Lyne *et al.*, 2002 The genome sequence of *Schizosaccharomyces pombe*. *Nature* 415: 871–880.
- Yokogawa, T., T. Suzuki, T. Ueda, M. Mori, T. Ohama *et al.*, 1992 Serine tRNA complementary to the nonuniversal serine codon CUG in *Candida cylindracea*: evolutionary implications. *Proc. Natl. Acad. Sci. USA* 89: 7408–7411.

Communicating editor: R. Rothstein