**Pichia sorbitophila**, an Interspecies Yeast Hybrid, Reveals Early Steps of Genome Resolution After Polyploidization

Véronique Leh Louis,† Laurence Desponts, Anne Friedrich, Tiphaine Martin, Pascal Durrens,* Serge Casarégola, Cécile Neuvéglise, Cécile Fairhead, Christian Marck, José A. Cruz,† Marie-Laure Straub, Valérie Kugler, Christine Sacerdot, Zlatyo Uzunov,‡‡‡‡ Agnes Thierry, Stéphanie Weiss,‡ Claudine Bleykasten, Jacky De Montigny, Noemie Jacques, Paul Jung, Marc Lemaire, Sandrine Mallet, Guillaume Morel, Guy-Franck Richard, Anasua Sarkar, Guilhem Savel,†† Joseph Schacherer, Marie-Line Seret, Emmanuel Talla, Gaelle Samson, Claire Jubin, Julie Poulain, Benoît Vacherie, Valérie Barbe, Eric Pelletier, David J. Sherman, Eric Westhof, Jean Weissenbach, Philippe V. Baret, Patrick Wincker, Claude Gaillardin, Bernard Dujon, and Jean-Luc Souciet.

*Université de Strasbourg, CNRS UMR7156, F-67000 Strasbourg, France; †Université de Bordeaux 1, LaBRI INRIA Bordeaux Sud-Ouest (MAGNOME), F-33405 Talence, France; ‡INRA UMR 1319 Micalis, AgroParisTech, Bat. CBAI, F-78850 Thiverval-Grignon, France; §Institut de Génétique et Microbiologie, Université Paris-Sud, UMR CNRS 8621, F-91405 Orsay CEDEX, France; ‡‡‡‡ Université de Strasbourg, Architecture et Réactivité de l’ARN, Institut de Biologie Moléculaire et Cellulaire du CNRS, F-67084 Strasbourg, France; †††† Institut Pasteur, CNRS URA2171, Université Pierre et Marie Curie, Paris 6 UFR927, F-75724, Paris-CEDEX 15, France; ††††† Universidad de La República, Departamento de Biología, F-33405 Talence, France; ††‡‡‡‡ Universidad de la República, Departamento de Biología, F-33405 Talence, France; ††‡‡‡‡ facultad de biología, Universidad de la República, F-33405 Talence, France.

ABSTRACT

Polyploidization is an important process in the evolution of eukaryotic genomes, but ensuing molecular mechanisms remain to be clarified. Autopolyploidization or whole-genome duplication events frequently are resolved in resulting lineages by the loss of single genes from most duplicated pairs, causing transient gene dosage imbalance and accelerating speciation through meiotic infertility. Allopolyploidization or formation of interspecies hybrids raises the problem of genetic incompatibility (Bateson-Dobzhansky-Muller effect) and may be resolved by the accumulation of mutational changes in resulting lineages. In this article, we show that an osmotolerant yeast species, *Pichia sorbitophila*, recently isolated in a concentrated sorbitol solution in industry, illustrates this last situation. Its genome is a mosaic of homologous and homeologous chromosomes, or parts thereof, that corresponds to a recently formed hybrid in the process of evolution. The respective parental contributions to this genome were characterized using existing variations in GC content. The genomic changes that occurred during the short period since hybrid formation were identified (e.g., loss of heterozygosity, unilateral loss of rDNA, reciprocal exchange) and distinguished from those undergone by the two parental genomes after separation from their common ancestor (e.g., NUMT (Nucleolar organizer region) insertions, gene acquisitions, gene location movements, reciprocal translocation). We found that the physiological characteristics of this new yeast species are determined by specific but unequal contributions of its two parents, one of which could be identified as very closely related to an extant *Pichia farinosa* strain.

KEYWORDS

osmotolerant yeast *P. sorbitophila* allopolyploidy hybridization genome evolution loss of heterozygosity

Copyright © 2012 Louis et al.
doi: 10.1534/g3.111.000745

Manuscript received July 21, 2011; accepted for publication December 16, 2011

This is an open-access article distributed under the terms of the Creative Commons Attribution Unported License (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supporting information is available online at http://www.g3journal.org/lookup/suppl/doi:10.1534/g3.111.000745/-/DC1

P. sorbitophila genome sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. FO082046-FO082059.

†Corresponding author: Université de Strasbourg, Génétique Moléculaire, Génomique, Microbiologie CNRS UMR7156, Institut de Botanique, 28 rue Goethe, F-67000 Strasbourg, France. E-mail: jsouciet@unistra.fr; vleh@unistra.fr

A new species could arise via a process of interspecific hybridization, that is, the union of different organisms across a species barrier. At the early stage, hybridization produces allopolyploid or allopolyploid hybrids, depending on the ploidy of parents. This situation gives rise to a period of genome resolution where diverse genetic events act simultaneously and successively to form a stable chimerical genome. Recombination, loss of chromosomes and loss of heterozygosity (LOH) are frequent mechanisms involved in this genome shuffling that stabilizes hybrids (Belloch et al. 2009; Forche et al. 2005; Gerstein et al. 2006; Scannell et al. 2006; Sipiczki 2008). Because of the combination and modification of at least two distinct gene pools, hybrids...
could display better adaptive properties than their parental species, called heterosis (Arnold and Martin 2010; Belloch et al. 2008). The limitation on the amount of such interspecies hybrids is often related to genetic incompatibilities, leading to fitness and survival decrease and/or fertility reduction (Johnson 2008). Hybridization occurs approximately in 25% of plant species and 10% of animal species (Mallet 2005).

Recent genomic data and experimental analyses suggest that interspecies hybrids are frequent in yeast populations. Several interspecific hybrids have been described among species that belong to the genus Saccharomyces and play key roles in industrial fermentations (Bond et al. 2009; Dunn and Sherlock 2008; Nakao et al. 2009; Querol and Bond 2009; Rainieri et al. 2006; Sipiczki 2008). In contrast, only few cases of interspecific hybrids have been reported among other yeasts or fungi: the fungal pathogens Candida albicans and Candida dubliniensis can mate to produce tetraploid hybrids (Pujol et al. 2004), a random genomic sequencing of a Zygosaccharomyces rouxii wild isolate showed that it contains two different sets of genes (Gordon and Wolfe 2008), and finally, among Basidiomycota, anomalous Cryptococcus neoformans strains isolated from patients appeared to be hybrids between C. neoformans and C. gattii (Bowers et al. 2006). If stress conditions in cultures have been invoked to stimulate the formation of yeast hybrids (Belloch et al. 2008; Querol and Bond 2009), experiments show that yeast species in general tend to have no or limited prezygotic barriers (Chou et al. 2010; Greig et al. 2002; Liti et al. 2006; Marinoni and Lachance 2004; Murphy et al. 2006). In addition, only rare cases of heterosis or Bateson-Dobzhansky-Muller incompatibility are reported in yeasts (Belloch et al. 2008; Chou et al. 2010; Kao et al. 2010; Lee et al. 2008). Therefore, the role of interspecies crosses in yeast populations need to be better clarified, as well as the events involved in the hybrid genome formation and stabilization.

The yeast P. sorbitophila (de Miranda et al. 1980), a member of the “CTG” group of Saccharomycotina (De Montigny et al. 2000; Kuczman and Suzuki 2010), has been largely studied for its resistance to osmotic and salt stress (Banuelos et al. 2002; Benito et al. 2004; Lages et al. 1999; Lages and Lucas 1995; Maresova and Sychrova 2003; Neves et al. 2004; Oliveira et al. 1996; Prista et al. 2005). This osmotolerant yeast was isolated as a contaminant of a 70% sorbitol solution. By the complete sequencing and the detailed analysis of its genome, we demonstrate here that P. sorbitophila is, in fact, a hybrid yeast. From its genome, we extracted the two subgenomes, inherited from both parents. We identified genes for specific metabolism pathways, acquired either from both parents or from a sole parent, giving now the opportunity to study possible cases of heterosis. We finally demonstrate that P. sorbitophila is a very recent hybrid in the process of its resolution. The genomic rearrangement characterizing the early steps of genome stabilization were distinguished from those appeared in the parental genomes before the hybridization event. The hybrid state of the P. sorbitophila genome raises now the question about the fertility of this yeast.

**METHODS**

Full methods are available in the online version with the supporting information. Here are described the methods used for the genome sequencing and assembly, and the determination of the global GC content variations between P4 and P5 subgenomes, depicted in Figure 1.

**Sequencing and assembly**

The sequencing of the nuclear genome of Pichia sorbitophila CBS 7064 (de Miranda et al. 1980) was performed using a whole-genome shotgun strategy, with two plasmid libraries (4-kb inserts and 40-kb inserts) and ABI/Sanger technology to 10X-15X depth. In a first read assembly performed using ARACHNE assembler_version 03 (Batzoglou et al. 2002), we obtained 17 supercontigs, from 0.144 Mb to 2.114 Mb in size, with a read coverage of 7.5X on average for 11 contigs and 14X for 6 contigs (supporting information, Figure S1). The weakly covered contigs were joined into five distinct pairs showing approximately 85% sequence identity between contigs forming a pair, reflecting the partial heterozygosity of the genome (Figure S3). The homozygous genomic regions were represented by the six contigs 14X covered. Some of the weakly and highly covered contigs could also be joined on the basis of the sequence identity shared at the end of the contigs, corresponding to partly heterozygous and partly homozygous chromosomes (Figure S1). We confirmed this assembly by reiterating all the process using ARACHNE assembler_version 04 and with a manual finishing step to join contigs and resolve low-quality sequences. Homozygous sequences were doubled, and some of them concatenated with heterozygous sequences if necessary, to be representative of the genome state. The presence of small heterozygous regions inside homozygous areas, potentially masked during the assembly process, was checked by analyzing single-nucleotide polymorphism (SNP) distribution (see Table S2 for method). No area with a complex pattern of heterozygosity/homozygosity was highlighted inside the homozygous regions. At the end of that process, 14 contigs were obtained and named in alphabetical order according to their size. Only two sequence gaps remained in chromosomes G and H telomeric repeats. This assembly was validated by comparing with the number and the size of chromosomes estimated by pulsed-field gel electrophoresis (Figure S2), one contig corresponding to one chromosome.

**GC content calculation**

The GC content along chromosomes was computed using all protein coding genes present in two syntenic copies between chromosomes forming a pair and having identical sizes to minimize the effect of insertions/deletions on the GC values (2499 gene pairs considered). The two GC% obtained for a gene pair were compared with each other by calculating their ratio (dGC) from the mean GC% value:

\[
dGC1 = \frac{GC1}{(GC1 + GC2)/2}, \quad dGC2 = \frac{GC2}{(GC1 + GC2)/2},
\]

where GC1 and GC2 are the GC% for two nonidentical copies, the gene copy that contains the highest GC% corresponding to the start codon of one gene located on an arbitrarily chosen chromosome. The GC content raises now the question about the fertility of this yeast.

\[
dGC1 = \frac{GC1}{(GC1 + GC2)/2} \quad \text{and} \quad dGC2 = \frac{GC2}{(GC1 + GC2)/2},
\]

with GC1 and GC2 corresponding to the GC% values obtained for copies 1 and 2, respectively. This normalization does not allow us to take into account the GC content variation between genes. Therefore, when both copies of a gene are identical at the nucleotide level (a situation observed in homozygous regions), the following identity relationship is attributed to these two alleles: dGC1 = dGC2 = 1. Conversely, for two nonidentical copies, the gene copy that contains the highest GC content has its dGC > 1. The two dGC values obtained for each gene pair were plotted along chromosome pairs using a sliding window of 11 genes with a step of 1 for representing a moving average of 11 adjacent genes along each chromosome. Curve superposition was performed using a single chromosomal coordinate. This coordinate corresponds to the start codon of one gene located on an arbitrarily chosen chromosome of each pair. The GC content variations in all protein coding genes were also calculated using rRNA that pair with two codons (Crick 1966). The codon usage was determined for each heterozygous region, independently from its parental origin, using genes that are present in two syntenic copies (5516 gene pairs...
RESULTS AND DISCUSSION

Structure of the Pichia sorbitophila hybrid genome

Size of the genome, number of chromosome pairs, and global heterozygosity: The nuclear genome of Pichia sorbitophila CBS 7064 (also referenced as Pichia or Millerzyma farinosa) (Kurtzman and Suzuki 2010) was completely sequenced using a whole-genome shotgun strategy and initially assembled into 17 supercontigs (see Materials and Methods and Figure S1). Sequence alignments revealed nearly perfect synteny conservation and approximately 85% identity at nucleotide level between some pairs of supercontigs (11 supercontigs), suggesting a partial heterozygosity of the genome. The other supercontigs (6 supercontigs) remained unique. Interestingly, sequencing coverage was double for the latter compared with the former (Figure S1), suggesting that they represent homozygous regions with two identical or almost identical copies of sequence and that the read assembly led to the production of a single consensus sequence for each homozygous part. A manually finishing process confirmed that the P. sorbitophila genome is a mosaic of homologous and homoeolo-gous chromosome pairs (see Materials and Methods). We obtained 14 contigs (named chr. A to N), from 1.05 Mb to 2.12 Mb, representing the 14 chromosomes of the P. sorbitophila nuclear genome and giving a global genome size of 21.5 Mb (Table S1). Contig sizes largely match chromosome sizes estimated by pulsed-field gel electrophoresis (Figure S2). Telomeric repeats were found at the ends of almost all contigs (Table S1). As suspected, the 14 chromosomes could be gathered into seven pairs of homologous or homoeologous chromosomes (Figure 1) on the basis of the sequence identity along the chromosomes (Figure S3). The chromosome pairs A/B and C/D were found to be partly heterozygous and partly homozygous, G/H and K/L homozygous, and M/N and E/F/I/J heterozygous. These four chromosomes could not be separated into two clearly distinct pairs since a switch of identity was observed from I/E to E/F and from F/J to I/J (Figure 1 and Figure S3), suggesting that a translocation event has occurred between two of these chromosomes.

The global nucleotide sequence identity between heterozygous regions ranges from 70.6 to 88.6% (see Table S1). Some local loss of identity (hemizygous regions) are detected at the end of some chromosomes and in dispersed intrachromosomal areas (see Figure S3). If we do not consider these hemizygous regions, the remaining syntenic parts of the genome (11.85 Mb) show no more than 89.16% nucleotide identity, revealing that P. sorbitophila is an hybrid genome derived from two distinct progenitors that have a high level of nucleotide polymorphism (10.84% of divergence) but a very well conserved synteny.

Homozygous regions and junctions between homozygous and heterozygous regions: Forty percent of the genome is homozygous (two identical or almost identical copies of sequence) and result from LOH events (Figure 1 and Table S1). LOHs concern chromosome pairs either in their totality (chr. G/H and K/L) or in part (chr. A/B and C/D). We quantified the number of SNPs in homozygous regions by realigning initial reads against the consensus sequences given by the supercontigs (see Table S2 for data and method). As a whole, the polymorphism is very low (0.46 SNP per 10 kb) and reveals a LOH process recently started. According to the chromosome pairs considered, different polymorphism levels are observed, from 1 SNP per 51,474 bp within C/D to 1 SNP per 18,000 bp within G/H. The first LOH event involved probably the G/H pair, the most polymorphic region, and may have been followed by the additional events in successive order: K/L, A/B, and C/D.

We also examined the junctions between the heterozygous and homozygous regions in the A/B and C/D chromosome pairs. In both cases, they are located inside protein coding genes and the two allelic copies of these genes are highly conserved at the nucleotide level (95.7% for A/B and 99.3% for C/D). It contrasts with the average identity observed for the others heterozygous genes (90.5%) and for the surrounding heterozygous regions (91%). No repetitive element or low complexity sequence is present at these junctions.

Position of centromeres: In Debaryomyces hansenii (Dujon et al. 2004), the most closely related yeast to P. sorbitophila (Figure 3), each chromosome has a unique island of highly repeated and degenerated sequences of retrotransposons Tdh5 with a poor GC content, which probably corresponds to the centromere (Butler et al. 2009; Dujon 2010; Lynch et al. 2010). In P. sorbitophila, we did not find any repetitive elements such as retrotransposons, but we identified for each chromosome a unique island with a poor GC content (10.8% less than the global GC content, see Figure S4), corresponding likely to the centromere position. These GC-poor regions range from 2.2 to 3.1 kb in size, are devoid of protein-coding genes or other features and are at equivalent positions in chromosomes forming a pair (Figure 1 and Figure S4). For homoeologous pairs of chromosomes, all the proposed centromeres are located in heterozygous regions and share a very weak sequence identity or none, although surrounding allelic regions are well conserved (see Figure S3).

Extraction of Pγ and Pt parental genomes

In the case of P. sorbitophila and contrary to the analyses conducted so far on other yeast hybrids (Bovers et al. 2006; Dunn and Sherlock 2008; Gordon and Wolfe 2008; Nakao et al. 2009; Rainieri et al. 2006; Spiczki 2008), the extraction method of the two parental subgenomes cannot be based on knowledge of the genomic sequence of at least one species closely related to one parent. Therefore, we used two complementary strategies to extract the parental subgenomes. A first strategy, determined by the GC content, led us to propose a subdivision of the genome. The proposed assignation of genomic regions to each subgenome was then tested and completed by a comparative analysis performed with the species Pichia farinosa CBS 2001.

GC variation and bias in codon usage: The global GC content calculated along chromosomes systematically revealed two different GC tendencies in heterozygous regions: each homoeologous pair shows one chromosome with a greater GC content (average dGC = 1.006; σ = 0.012) and one chromosome with a lower GC content (average dGC = 0.994; σ = 0.012), as shown in Figure 1 using protein coding genes present in two syntenic copies between chromosomes forming a pair (Method described in the Materials and Methods). We hypothesized that it may reflect the subtle differences in nucleotide composition between the two parental genomes at the origin of the P. sorbitophila hybrid genome. We arbitrarily named “Pγ” the parental subgenome with the highest GC content and “Pτ” the parental subgenome with the lowest one. We tested that hypothesis by analyzing the GC content variations in all protein coding genes (independently of their size) using tRNAs that pair with two different codons (Crick 1966). As shown in Table 1, genes in heterozygous regions defined as Pγ contain on average 1.57%
more codons with a C at the third position than genes of Pe regions (mean = 1.56 [1.08-1.95]), allowing to propose an assignment of heterozygous regions to each subgenome (Figure 1, see also Figure S5). Applied to homozygous regions, the GC-based analyses suggested that both G/H and K/L pairs belong to Pg subgenome (Figure S5). As for A/B and C/D homozygous parts, the discrimination was not obvious.

We finally observed a large exchange between GC% trend curves in the C/D chromosomal pair (Figure 1), with a bias of 2.92% (mean = 2.44 [1.39-4.36]) in the codon usage between upstream and downstream regions (Figure 2). We checked that it was not relevant to a sequence assembly error by confirming both chromosomal sequences using polymerase chain reaction amplifications and resequencing. Thus, the GC trend curve profile suggests that
a reciprocal exchange of sequences took place between both chromosomes. The left border of this chromosomal exchange (Figure 2) is located into a putative isoleucine tRNA synthetase coding sequence, with 100% sequence identity between both alleles. Other GC% trend curve exchanges were not taken into account in the following analyses, because they were of limited size and some of them likely corresponded to noise.

**Py subgenome allelic sequences in *P. farinosa CBS 2001***: A taxonomical study of the *Pichia* (*Millerozyma*) *farinosa* group of species (S. Mallet *et al.*, unpublished results) revealed that several sequences from the *P. sorbitophila* Py subgenome were nearly 100% identical to the haploid yeast *Pichia farinosa var. farinosa* CBS 2001. To confirm the origin of each part of the *P. sorbitophila* genome, we completed this study by sequencing 17 other chromosomal sites in *P. farinosa* CBS 2001 (Table S4). Special attention was paid to cover regions of the E/F/I/J translocation breakpoint as well as the C/D chromosomal exchange and NUMTs (NUclear sequences of MiTochondrial origin) positions (see section Unequal acquisition of mitochondrial DNA sequences, below). We observed that all markers in regions defined as belonging to Py subgenome shared 99% to 100% identity with *P. farinosa* CBS 2001, whereas all markers in regions attributed to Pe subgenome shared only 90% to 96% identity (Figure 1 and Table S4). Sequencing data for chr. A/B and C/D showed also that their homologous regions belonged to Pe and Py, respectively. As a whole, these results confirm that: (1) the *P. sorbitophila* genome is an admixture of two parental subgenomes (Py and Pe); (2) the A/B, C/D, G/H, and K/L pairs were subjected to LOH processes; and (3) a C/D chromosomal exchange took place in the *P. sorbitophila* hybrid genome after its formation. Indeed, SNPs identified between both *P. sorbitophila* subgenomes and the genome of *P. farinosa* CBS 2001 (Figure 2C) are strongly correlated with the GC% trend curves exchange observed for chr. C/D (Figure 2A).

**Phylogenetic position of Py and Pe**: The sequence identity shared by *P. farinosa* CBS 2001 and Py for several positions on the genome suggests that one of the *P. sorbitophila* progenitors belongs to the *Millerozyma* group of species and is closely related to *P. farinosa* CBS 2001. We determined the phylogenetic position of Py and Pe in the CTG group of yeasts (Butler *et al.* 2009; Dujon *et al.* 2004; Jackson *et al.* 2009; Jeffries *et al.* 2007; Iones *et al.* 2004) (Table S5), using protein coding genes simultaneously present in both subgenomes (Figure S6). As shown in Figure 3, most of the branches of the phylogenetic tree obtained agree with previously published trees (Butler *et al.* 2009; Kurtzman and Suzuki 2010). Py and Pe are positioned in distinct strongly supported branches (bootstrap value = 100), with a phylogenetic distance corresponding to the separation of the species and with *D. hansenii* as the nearest known yeast of the CTG group with a fully sequenced genome. According to recent phylogenetic studies (Kurtzman and Suzuki 2010; S. Mallet *et al.*, unpublished results) Py and Pe correspond to two distinct species of the *Millerozyma* (*Pichia*) *farinosa* group of species.

**Comparison between Py and Pe subgenomes**

**Divergence of protein-coding gene alleles**: The genome of *P. sorbitophila* proves to be one of the smallest identified in the CTG group according to its size (10.75 Mb) and the number of annotated genes

### Table 1 Bias in codon usage between Py and Pe subgenomes

| Amino Acid | Codon | No. Codons | Usage, % | Pγ-Pe |
|------------|-------|------------|---------|--------|
|            |       | Py         | Pe      |        |
|            |       |            |         |        |
| Phe        | TTT   | 35,134     | 37,411  | 50.56  | 53.55  | 2.99  |
|            | TTC   | 34,354     | 32,448  | 49.44  | 46.45  |       |
|            | GCC   | 34,498     | 36,153  | 64.45  | 66.26  | 1.80  |
|            | TCC   | 19,030     | 18,412  | 35.55  | 33.74  |       |
|            | CTT   | 24,501     | 25,360  | 66.37  | 67.88  | 1.51  |
|            | GCT   | 12,414     | 11,999  | 33.63  | 32.12  |       |
|            | GCC   | 25,044     | 26,245  | 58.24  | 60.78  | 2.55  |
|            | CAT   | 30,768     | 31,479  | 61.14  | 63.54  | 2.4   |
|            | GAC   | 19,557     | 18,060  | 38.86  | 36.46  |       |
|            | Asn   | 20,343     | 20,787  | 58.51  | 60.07  | 1.56  |
|            | Asp   | 48,995     | 51,063  | 53.57  | 55.28  | 1.71  |
|            | Cys   | 14,426     | 13,818  | 41.49  | 39.93  |       |
|            | Val   | 55,556     | 56,476  | 57.36  | 58.56  | 1.21  |
|            | Gly   | 10,196     | 10,161  | 55.71  | 55.19  | 0.52  |
|            | Ser   | 8105       | 8250    | 44.29  | 44.81  |       |
|            | Glu   | 21,219     | 21,365  | 50.48  | 50.81  | 0.33  |
|            | Pro   | 20,817     | 20,682  | 49.52  | 49.19  |       |
|            | Thr   | 30,801     | 29,869  | 64.52  | 63.82  | 0.70  |
|            | Ser   | 16,937     | 16,931  | 35.48  | 36.18  |       |
| Avg.       |       |            |         | 1.57   |        |       |
| Mean       |       |            |         | 1.56   |        |       |
| Q1-Q3      |       |            |         | [1.08-1.95] |   |   |

*For tRNA species that pair with two codons, the usage % of each codon was determined as follows: (number of one codon/number of both codon x100). The values were calculated for all chromosomal regions defined as belonging to Py or Pe.*
(5626 protein-coding genes), if we consider the equivalent haploid genome (Table 2 and Table S7). The gene redundancy is limited for both subgenomes compared to D. hansenii with a total of 33.2% protein-coding genes belonging to multigene families against 51.5% and three times less tandemly duplicated gene arrays (Table S8). In each chromosomal pair, a protein coding gene is present in most cases in two allelic copies coming either from both parents (in heterozygous regions) or from a sole parent (in homozygous regions), giving a total of 11,252 loci (Table 2 and supporting information). When genes with introns are located in heterozygous regions, both alleles of the genes contain the same number of introns, almost identical in size (Table S6).

Among the 3425 genes in heterozygous regions, 3205 (93.6%) are represented by two nonidentical coding-alleles (Table 2) demonstrating on average 92.1% of identity at the protein level (Idprot) and a mean ratio of nonsynonymous substitutions per synonymous substitutions (dN/dS) of 0.121 (Figure S9). Three subsets of heterozygous genes are particularly interesting. First, genes showing highly conserved alleles (44 genes, dN/dS, 0.0046, Idprot. 99.8%) likely encode for essential functions (Table S9). Second, some genes with highly divergent alleles (15 among 90 genes) correspond to "Millerzyma-specific genes" found only in Py and Pe subgenomes but not in the other yeasts of the CTG group (Table S10). They probably appeared in the Py and Pe common ancestor after its separation with the ancestor of D. hansenii and are under a relaxed selection pressure. The third subset is composed of 38 CDS-pseudogene allele pairs (CDS or CoDing Sequence is defined here as the region of nucleotides that corresponds to the sequence of amino acids in the predicted protein and that is not interrupted by internal frameshift or stop codon). The allele distribution for these genes is not homogeneous since almost all pseudogene alleles (81.6%) are located in the Pe subgenome whereas the corresponding coding alleles are in the Pg subgenome. Finally, the two groups of genes with highly divergent alleles (90 CDS-CDS pairs and 38 CDS-pseudogene pairs) may constitute an interesting pool of genes required for specific adaptations to environmental conditions, since they show bias in Gene Ontology frequencies in favor of transporters, phosphatases, oxidoreductases, and cell wall proteins (Table S11 and Table S12).

Only a limited number of genes (220 genes) in one subgenome has no allelic counterpart in the other subgenome. Among them, 102 correspond probably to "dubious open reading frames" (Figure S8), reducing the number of single allele genes to 118. They are located in three different genomic areas: 83 genes (70.3%) at the end of contigs, 12 genes (10.2%) in the E/F/I/J translocation breakpoint and 23 genes (19.5%) in the heterozygous regions. The distribution of single allele genes between Pg and Pe is slightly in favor of Pe with 78 genes (66.7%) against 40 for Pg (33.3%). Despite their small number (2% of the total number of genes), they are at the origin of some metabolic pathways in the P. sorbitophila hybrid as described in the section Unilateral acquisition of genes for sugar degradation.

Uniparental conservation of ribosomal RNA genes: Despite the fact that the P. sorbitophila nuclear genome has three different ribosomal DNA (rDNA) loci, the rDNA is at a hemizygous state with D and P. farinosa CBS 2001 are indicated in red. As shown, P. farinosa chromosomal sequence is first identical to P. sorbitophila chr. D sequence and then to chr. C (after the 1934-nt long sequence), confirming that the GC exchange ensues from a reciprocal translocation event.
Phylogenetic positions of the two subgenome sequences identified in *P. sorbitophila* hybrid. *Yarrowia lipolytica* is used as outgroup to root the CTG tree. The tree was built from the alignment of 233 protein families (87,181 amino acids per species) having a single member in each analyzed species. Amino acid sequences for each family were aligned with MAFFT (Katoh et al. 2009) and cleaned with Gblocks (Talavera and Castresana 2007). The tree was built from the resulting alignment with the maximum likelihood method using PHYML with a JTT substitution model corrected for heterogeneity among sites by a G-parameter of the G-law distribution using four different categories of evolution rates (Guindon and Gascuel 2003). The proportion of invariable sites and the α-parameter of the G-law distribution were optimized according to the data. Bootstraps were calculated from 100 replicates. They are indicated before each node and the scale for branch length at the bottom of the figure.

Unequal acquisition of mitochondrial DNA sequences: Most eukaryotic nuclear genomes contain pieces of mitochondrial sequences (Lenglez et al. 2010; Sacerdot et al. 2008), designated as NUMTs for NUClear sequences of MiTochondrial origin. They result from the transfer of fragments of mitochondrial DNA into chromosomes. The number and size of the NUMTs vary significantly between yeasts within the monophyletic group of hemiascomycetes (Sacerdot et al. 2008). For the *P. sorbitophila* nuclear genome, we observed a highly unequal distribution of NUMTs (Figure 1 and Table S3). First, of the 24 NUMTs identified from its extant mitochondrial genome (Jung et al. 2009), 20 are exclusively located in the subgenome deriving from the *Pc* parent. Therefore, most of the NUMTs (14 NUMTs) are at a hemizygous state with the equivalent allelic sequence being devoid of NUMT. Three arguments lead us to propose that the mitochondrial sequences have been inserted into the nuclear genomes of both parents before the hybridization event: (1) all NUMTs in heterozygous regions are at a hemizygous state; (2) the present *P. sorbitophila* hybrid is separated from its ancestral hybrid by only few thousand successive generations (estimated from the polymorphism level in homoyzogous regions and based on the method described by Rolland and Dujon 2011), whereas the number of NUMTs insertions is equivalent to what observed for other yeast genomes; and (3) the NUMT located on the G/H homozygous pair in two identical alleles is also present in *P. farinosa* CBS 2001 (Table S4), closely related to the *Py* progenitor. The nonuniform distribution of NUMTs reveals also that *Pc* has undergone four times more mitochondrial DNA insertions than *Py*. The sequence divergence between the *P. sorbitophila* and *P. farinosa* CBS 2001 mitochondrial COX2 genes suggests that the hybrid formation, the *Py*-like parental chromosome (chr. E) underwent an entire loss of the rDNA repeats. The two rDNA relics on chr. E and issued from *Pc* share 90.3% sequence identity (Figure S10) suggesting that they have already been present in the *Pc* parental genome before the hybrid formation. These two loci are located in highly polymorphic subtelomeric regions (Figure S11), that could have contributed to the loss of the rDNA repeats. The hemizygous state of the ribosomal DNA contrasts with the distribution of other noncoding RNA genes (ncRNA and tRNA), which are all represented by two well-conserved alleles (see supporting information).

---

| Table 2  | *P. sorbitophila* genomic features in *Py* and *Pc* subgenomes |
|----------|---------------------------------------------------------|
| Parental Contribution | Chromosomal Region | CDS Without Intron | CDS With Introns | Pseudo-Gene | Total Gene | tRNA | snoRNA | snRNA | Pol III ncRNA | Ribosomal DNA* | Total No. Other Elements NUMTs Loci |
| **Py** | | 95 | 2 | 7 | 104 | 0 | 0 | 0 | 73 | 1 |
| Solo | | 2973 | 219 | 13 | 3205 | 88 | 21 | 3 | 5 | 0 | 0 |
| Heterozygous | | 3834 | 246 | 10 | 4090 | 92 | 26 | 2 | 0 | 0 | 2 |
| Homozygous | | 6902 | 467 | 30 | 7399 | 180 | 47 | 5 | 5 | 3 | 3 |
| Total | | 106 | 4 | 6 | 116 | 0 | 0 | 0 | 8 | 0 | 8 |
| **Pc** | | 2950 | 218 | 37 | 3205 | 88 | 21 | 3 | 5 | 0 | 0 |
| Solo | | 2973 | 219 | 13 | 3205 | 92 | 26 | 2 | 0 | 0 | 2 |
| Heterozygous | | 482 | 46 | 4 | 532 | 20 | 6 | 0 | 0 | 0 | 4 |
| Homozygous | | 3538 | 268 | 47 | 3853 | 108 | 27 | 3 | 5 | 0 | 12 |
| Total | | 5,220 | 367.5 | 38.5 | 5626 | 144 | 37 | 4 | 5 | 36.5 | 7.5 |

CDS, CoDing Sequence; Pol III, polymerase III; NUMTs, NUClear sequences of MiTochondrial origin.

a Tandemly repeated units.

b Loci (320S pairs) containing both parental genes (a total of 6410 genes). They correspond to CDS/CDS pairs for 3161 pairs, to CDS/pseudogene pairs for 38 pairs, and to pseudogene/pseudogene pairs for 6 pairs (supporting information).
mitochondrial genome (mtDNA) of *P. sorbitophila* was inherited from the *Pc* progenitor (see supporting information).

Chromosomal rearrangements in parental genomes: Three distinct chromosomal areas of the *P. sorbitophila* genome were analyzed to identify molecular events at the origin of synteny breaks: (1) E/F/I/J chromosomes, (2) subtelomeric regions, and (3) internal chromosomal positions where single allele genes are located.

Gene orders at the E/F/I/J translocation breakpoint were compared with the orthologous regions identified in *D. hansenii* and *Candida guilliermondii* (Figure S12). The fact that the gene orders on chr. I and F (Py) are identical or almost identical to those of *D. hansenii* and *C. guilliermondii* suggests that the chromosomal arm exchange likely took place between the two other chromosomes, chr. E and J (Pc). The short evolutionary period of *P. sorbitophila* since its formation suggests that this event appeared in *Pc* before the hybridization event, although we cannot exclude that it was part of the early chromosomal rearrangements occurred in the hybrid genome. The translocation breakpoint is particularly enriched in tandemly duplicated genes. Among the 12 single-allele genes located in chr. E and J, eight are tandemly duplicated and form four distinct tandem gene arrays (Figure S12). Insertion and amplification of those duplicated genes probably either induced the chromosomal exchange or were initiated by the exchange.

The plasticity of subtelomeric regions and their capacity to harbor large gene families make their comparison between subgenomes extremely difficult. However, we found that some of the 83 single allele genes located in subtelomeric regions were the result of gene location movements in parental genomes. An example of such cases is given in Figure S13.

The remaining 23 single allele genes in heterozygous regions are mainly concentrated in 10 internal chromosomal areas (Figure S3). Compared with the corresponding gene order with *D. hansenii*, *C. guilliermondii*, *Pichia stipitis*, and *C. albicans*, the single allele genes are missing in all other species at these specific locations whereas the surrounding regions are highly syntenic (Figure S14). Consequently, most of these genes were probably inserted into the corresponding genomic areas in one parental genome after the separation of both parents from their common ancestor. These insertions may result from diverse molecular events: gene location movement as observed in subtelomeric regions, gene acquisition (four single-allele genes are species-specific), gene duplication in cases of multigene families (eight single-allele genes), and tandem duplication as observed for chr. F (Figure S14) or more complex events.

**Acquisition of metabolic pathways**

Unilateral acquisition of genes for sugar degradation: Analysis of single-allele genes at the level of their putative encoded functions shows that some key genes for *P. sorbitophila* metabolism have been acquired from the *Pc* parent. They are still conserved in the *Pc* progenitor of *P. sorbitophila*, although the *Pc* subgenome constitutes no more than 32% of the total. This is noteworthy for maltose degradation (Table 3): *P. sorbitophila* is able to hydrolyze maltose (de Miranda et al. 1980) in contrast to *P. farinosa* CBS 2001 (strain closely related to Py). The MAL genes (Alves et al. 2008; Chow et al. 1989) are exclusively single allele genes inherited from *Pc* and exist in several copies at four different areas of synteny break in heterozygous regions: for example, the E/F/I/J translocation breakpoint contains two tandemly duplicated MALX3 genes on chr. E and two tandemly duplicated MALX2 genes on chr. J (Figure S12). At a synteny breakpoint between chr. M and N, four of the five additional genes on chr. M are MAL genes, with the corresponding gene order MALX2-MALX1-MALX3 (pseudogene)–MALX3 (Figure S14). As a whole, the *Pc* subgenome encodes three...
MALX1 permeases, four MALX2 maltases and three MALX3 activators whereas Pg has none (Table 3).

We also observed a bias in favor of the Pe subgenome for the invertase SUC2 gene (Carlson et al. 1983; Taussig and Carlson 1983) with one tandem array of two SUC2 genes at the translocation breakpoint (Figure S12) and no copies in Pg subgenome. The Pe subgenome also contributes for seven of the ten allantoate permease genes and the Pg for three of five sorbitol dehydrogenase genes (Table 3). In conclusion, a limited number of genes have been deleted or acquired in both parents since their separation from their common ancestor. These minor differences may however have significant consequences on P. sorbitophila metabolism.

**Biparental acquisition of genes for stress resistance, mating type, and meiosis:** P. sorbitophila was isolated as a contaminant of a 70% sorbitol solution (de Miranda et al. 1980). Compared with S. cerevisiae and D. hansenii, P. sorbitophila is more tolerant to NaCl (4M), LiCl (0.8 M), and KCl (2.5 M) (Maresova and Sychrova 2003). Previous studies on Na⁺ intracellular limitation pathways in P. sorbitophila allowed the identification of three key transporters: the NHA1 and NHA2 cation/H⁺ antiporters (Banuelos et al. 2002), which correspond to two alleles of the same gene; the P-type ATPase (Benito et al. 2004); and a H⁺/glycerol symport activity (Lages and Lucas 1995). To complete those data, we searched for the presence of more than 50 known genes involved in ion, glycerol, and water transport or in osmotolerance-related pathways (Table S18), identified in S. cerevisiae (obtained from Saccharomyces Genome Database, http://www.yeastgenome.org/) and also well documented for D. hansenii (Prista et al. 2005).

We found that, in contrast to the metabolism of sugar degradation, osmotolerance genes were not specific to one subgenome but rather identified with their two related coding alleles in heterozygous regions. To achieve Na⁺ efflux, several ENA genes ("Exitus NAtru" genes) encoding Na⁺-ATPases are found in D. hansenii and S. cerevisiae (Benito et al. 2004), whereas only one ENA gene has been identified in P. sorbitophila. The latter also lacks the HAL1 gene (Rios et al. 1997), which decreases intracellular Na⁺ via ENA1. Potassium is the most abundant intracellular cation in living cells and plays important roles in biological processes. The NHA1-Na⁺/H⁺ antiporter is not specific for Na⁺ but also mediates K⁺ efflux in addition to Na⁺, allowing a possible intracellular K⁺ depletion (Benito et al. 2004). In P. sorbitophila, this depletion may be intensified by the presence of TOK1 (Table S18) a permeable channel for potassium efflux (Ketchum et al. 1995; Rios et al. 1997) missing in D. hansenii. To counterbalance the K⁺ efflux at high NaCl concentrations, it has been

---

**Table 3 Distribution of single-allele genes between Py and Pe for sugar degradation and other transports**

| Putative Function     | Gene Name | Locus in Py | Locus in Pe |
|-----------------------|-----------|-------------|-------------|
| Sugar metabolism      |           |             |             |
| Maltose permease      | MALX1     | PISO0M16930g| PISO0J21547g|
|                       |           | PISO0M00166g| PISO0M16888g|
| Maltase               | MALX2     | PISO0M00188g| PISO0J35515g|
|                       |           | PISO0M00166g| PISO0J34419g|
| MAL activator         | MALX3     | PISO0E02028g| PISO0E02050g|
| Invertase             | SUC2      | PISO0N22123g| PISO0M16974g|
| Sorbitol dehydrogenase| SOR1      | PISO0K006604g/| PISO0J03639g|
|                       |           | PISO0K006043g/| PISO0J03573g|
| Glutathione metabolism| OXP1      | PISO0I02648g | PISO0J04431g|
| 5-oxoprolinase        |           | PISO0C10078g/PISO0D10145g | PISO0E04690g/PISO0J03595g|
| Allantoate transport  | DAL5      | PISO0K23022g/PISO0L23023g | PISO0J21503g/PISO0J10019g|
|                       |           | PISO0I08192g | PISO0E04646g|
| Nicotinic acid transport| NTA1    | PISO0I02626g | PISO0J04409g/PISO0J03617g|
|                       |           | PISO0K00318g/PISO0L00319g/PISO0K23023g/PISO0L23023g | PISO0J03485g/PISO0J10019g/PISO0E04646g|
|                       |           | PISO0K00406g/PISO0L11407g/PISO0K23023g/PISO0L23023g | PISO0J03485g/PISO0J10019g/PISO0E04646g|
|                       |           | PISO0N15105g | PISO0M14708g|

**a** Pseudogene.

**b** Identical alleles located in homozygous regions.
proposed that an efficient K\textsuperscript{+} uptake system must exist in \textit{P. sorbitophila} (Banuelos \textit{et al.} 2002). We identified the two potassium transporters \textit{HAK1} (high-affinity K transporter) and \textit{TRK1} (Transport of K) and the P-type ATPase \textit{ACU1} (Benito \textit{et al.} 2004). \textit{P. sorbitophila} lacks the \textit{PHO89} Na\textsuperscript{+}/Pi cotransporter (Martinez and Persson 1998) in contrast to \textit{D. hansenii}, which has \textit{PHO89} but not \textit{ACU1}. Osmoregulation is also achieved by the production of glycerol or other osmolytes (e.g., arabitol, erythritol) and the capacity to maintain them into the cells (Kayingo \textit{et al.} 2001; Lages and Lucas 1995; Lages \textit{et al.} 1999). Glycerol leaks through the plasma membrane and its retention therefore needs an active transport system. In \textit{P. sorbitophila}, H\textsuperscript{+}/glycerol symport allows the intracellular accumulation of glycerol (Lages and Lucas 1995) and we also show that the aquaglyceroporin \textit{FPsi}, the glycerol permease responsible for glycerol leakage (Tamas \textit{et al.} 1999), is missing.

As for stress resistance, \textit{P. sorbitophila} genes involved for mating, meiosis, and spore formation are in two allelic versions, suggesting a conservation of these genes from the common ancestor of \textit{Py} and \textit{Pc} progenitors until the present hybrid. In \textit{P. sorbitophila}, we found two similar mating-type loci at the heterozygous allelic positions in chr. M and N (Figure S15). Both contain \textit{MATa2} and \textit{MATa1} genes but lack \textit{MATa1} and \textit{MATa2}. Thus, as \textit{P. stiptis} and \textit{D. hansenii} (Butler 2010; Fabre \textit{et al.} 2005; Jaffres \textit{et al.} 2007; Lee \textit{et al.} 2010), \textit{P. sorbitophila} has its mating-type loci fused with no \textit{MATa1}. \textit{MATa1} is missing only in \textit{P. sorbitophila}, a gene loss that likely took place in the common ancestral species of both parents. We also searched for the presence of 227 orthologs of \textit{S. cerevisiae} genes involved for mating, meiosis and spore formation since key genes for the sexual development and meiosis in \textit{S. cerevisiae} are missing in \textit{Candida} species (Butler \textit{et al.} 2009), and we also revisited the gene annotations of the 2nd version of the \textit{D. hansenii} genome (Souciet \textit{et al.} 2009). All genes identified in \textit{D. hansenii} (187 genes) were detected in two copies in \textit{P. sorbitophila} (Table S19). Three additional genes are present in \textit{P. sorbitophila}: \textit{DIT1} and \textit{DIT2} involved in dityrosine synthesis (Briza \textit{et al.} 1994) and \textit{YEL023C} which may participate in this process (Butler \textit{et al.} 2009). Dityrosine is the major component of the outer layer of spore wall. It forms an insoluble scaffold on the surface of the spore (Briza \textit{et al.} 1988, 1990) and contributes to its resistance. There are no data available for spore resistance concerning \textit{P. sorbitophila}. However, we can speculate that the presence of these three genes may contribute to the resistance of spores in specific environmental conditions.

**CONCLUSION**

Allopolyploid hybrids, which are the result of cell fusions between different species, undergo generally rapid and extensive genomic modifications after their formation (Rainieri \textit{et al.} 2006; Sipiczki 2008). With the aim of deciphering hybrid genome evolution, these posthybridization rearrangements have to be distinguished from those which took place in the parental genomes during the long period that preceded the hybrid formation. The complete sequencing of the genome of the osmotolerant yeast \textit{P. sorbitophila} strain CBS 7064 (de Miranda \textit{et al.} 1980) revealed that it is a hybrid genome. Despite the fact that no complete genome sequence closely related to that of one or both parents is available so far, a detailed analysis of its genome led us to retrace its evolution. The \textit{P. sorbitophila} nuclear genome (21.5 Mb) is actually composed of seven pairs of chromosomes issued from two progenitors named \textit{Py} and \textit{Pc}, with \textit{Py} being closely related to \textit{P. farinosa} CBS 2001. In contrast, the origin of the \textit{Pc} parent remains unknown because of the lack of close sequence similarity among existing yeasts. The sequence divergence ob-
rDNA genes derived from one progenitor are silent (Lewis et al. 2007; Pikaard 2000). These observations raise the problem of the origin of the rDNA instability and the viability of meiotic products. Our preliminary results of sporulation experiments show that *P. sorbitophila* is able to produce ascii containing one to four ascopores, as observed previously by de Miranda et al. (1980) and Oliveira et al. (1996). Results also suggest that at most two spores per ascus are viable (data not shown). These preliminary results could be consistent with the hemizygous state of the rDNA if we assume that *P. sorbitophila* is able to undergo meiosis, a hypothesis that requires additional data to be confirmed.

The reconstruction of the complete parental subgenomes allowed us to decipher the recent *P. sorbitophila* genome history and, therefore, to depict, gene by gene, how two divergent genomes put together into a viable hybrid are rearranged during the process of genome stabilization. Our results also show that interspecies hybrids, because of poor prezygotic barrier, are widespread in the *Saccharomycytina* group of species. Human activities in industrial contexts provide unusual substrates (70% sorbitol, for example) that may act as bottlenecks for the selection of particularly resistant species of yeasts and fungi or hybrids. This was probably the case for *P. sorbitophila*. Its genome analysis gave us therefore the opportunity to highlight, in the present times, the early steps of genome evolution after the formation of a hybrid.

ACKNOWLEDGMENTS
We thank Henri Grosjean and Fredj Tekaia for useful discussions. An interactive website can be found at [http://www.genolevures.org/](http://www.genolevures.org/). This work was supported in part by funding from the Consortium National de Recherche en Génomique (CNRG) to Génoscope, from CNRS (GDR 2354, Génolevures), ANR (ANR-05-BLAN-0331, GENARISE). The computer framework was supported by the University of Bordeaux 1, the Aquitaine Région in the program “Génotypage et Génomique Comparée,” and the ACI IMPBIO “Génolevures En Ligne.” We thank the System and Network Administration team in LaBRI for excellent help and advice. J.A.C. is supported by the PhD Program in Computational Biology of the Instituto Gulbenkian de Ciência, Portugal (sponsored by Fundação Calouste Gulbenkian, Siemens SA, and Fundação para a Ciência e Tecnologia; SFRH/BD/35328/2008). B.D. is a member of Institut Universitaire de France.

LITERATURE CITED
Alves, S. L. J., R. A. Herberts, C. Hollatz, D. Trichez, L. C. Miletii et al., 2008 Molecular analysis of malthotriose active transport and fermentation by *Saccharomyces cerevisiae* reveals a determinant role for the AGT1 permease. Appl. Environ. Microbiol. 74: 1494–1501.

Arnold, M. L., and N. H. Martin, 2010 Hybrid fitness across time and habitats. Trends Ecol. Evol. 25: 530–536.

Banuelos, M. A., J. Ramos, F. Calero, V. Braun, and S. Potier, 2002 Cation/H+ antiporters mediate potassium and sodium fluxes in *Pichia sorbitophila*. Cloning of the PanH1A and PanH2A genes and expression in *Saccharomyces cerevisiae*. Yeast 19: 1365–1372.

Batzoglou, S., D. B. Jaffe, K. Stanley, J. Butler, S. Gnerre et al., 2002 ARACHNE: a whole-genome shotgun assembler. Genome Res. 12: 177–189.

Belloch, C., S. Orlic, E. Barrio, and A. Querol, 2008 Fermentative stress adaptation of hybrids within the *Saccharomyces* sensu stricto complex. Int. J. Food Microbiol. 122: 188–195.

Belloch, C., R. Perez-Torrado, S. S. Gonzalez, J. E. Perez-Ortin, J. Garcia-Martinez et al., 2009 Chimeric genomes of natural hybrids of *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii*. Appl. Environ. Microbiol. 75: 2534–2544.

Benito, B., B. Garcia-deblas, P. Schreier, and A. Rodriguez-Navarro, 2004 Novel p-type ATPases mediate high-affinity potassium or sodium uptake in fungi. Eukaryot. Cell 3: 359–368.

Bond, U., I. L. Allen, S. Sima, and M. G. Geoffrey, 2009 The genomes of lager yeasts, pp. 159–182 in *Advances in Applied Microbiology*, Vol. 69, edited by A. I. Laskin, S. Sariaslani, and G. M. Gadd. Academic Press, San Diego.

Bovers, M., F. Hagen, E. E. Kuramae, M. R. Diaz, L. Spanjaard and B. de la Torre, 2008 Molecular analysis of maltotriose active transport and fermentative stress in Pichia sorbitophila. Yeast 25: 330–338.

Butler, G., A. Ellinger, G. Winkler, and M. Breitenbach, 1988 Chemical composition of the yeast ascospore wall. The second outer layer consists of chitosan. J. Biol. Chem. 263: 11569–11574.

Butler, G., A. Ellinger, G. Winkler, and M. Breitenbach, 1990 Characterization of a DL-dityrosine-containing macromolecule from yeast ascospore walls. J. Biol. Chem. 265: 15118–15123.

Butiza, P., A. Briza, P. Romulo, and J. Garcia-Martinez, 2009 Unique hybrids between the fungal pathogens *Cryptococcus neoformans* and *Cryptococcus gattii*. FEM. Yeast Res. 6: 599–607.

Brisa, P., A. Ellinger, T. G. Winkler, and M. Breitenbach, 1988 Chemical composition of the yeast ascospore wall. The second outer layer consists of chitosan. J. Biol. Chem. 263: 11569–11574.

Brisa, P., A. Ellinger, G. Winkler, and M. Breitenbach, 1990 Characterization of a DL-dityrosine-containing macromolecule from yeast ascospore walls. J. Biol. Chem. 265: 15118–15123.

Brisa, P., M. Eckerstorfer, and M. Breitenbach, 1994 The sporulation-specific enzymes encoded by the DIT1 and DIT2 genes catalyze a two-step reaction leading to a soluble LL-dityrosine-containing precursor of the yeast spore wall. Proc. Natl. Acad. Sci. USA91: 4524–4528.

Butler, G., 2010 Fungal sex and pathogenesis. Clin. Microbiol. Rev. 23: 140–159.

Butler, G., M. D. Rasmussen, M. F. Lin, M. A. Santos, S. Saktihumar et al., 2009 Evolution of pathogenicity and sexual reproduction in eight Candida genomes. Nature 459: 657–662.

Carlson, M., R. Taussig, S. Kustu, and D. Botstein, 1983 The secreted form of invertase in *Saccharomyces cerevisiae* is synthesized from mRNA encoding a signal sequence. Mol. Cell. Biol. 3: 439–447.
Kao, K. C., K. Schwartz, and G. Sherlock, 2010 A genome-wide analysis of yeast osmo-sensing. Arch. Microbiol. 177: 29–35.
Rainieri, S., Y. Kodama, Y. Kaneko, K. Mikata, Y. Nakao et al., 2006 Pure and mixed genetic lines of Saccharomyces bayanus and Saccharomyces pastorianus and their contribution to the lager brewing strain genome. Appl. Environ. Microbiol. 72: 3968–3974.

Rios, G., A. Ferrando, and R. Serrano, 1997 Mechanisms of salt tolerance conferred by overexpression of the HAL1 gene in Saccharomyces cerevisiae. Yeast 13: 515–528.

Rolland, T., and B. Dujon, 2011 Yeasty clocks: dating genomic changes in yeasts. C. R. Biol. 334: 620–628.

Sacerdot, C., S. Casaregola, I. Lafontaine, F. Tekaiia, B. Dujon et al., 2008 Promiscuous DNA in the nuclear genomes of hemiascomycetous yeasts. FEM. Yeast Res. 8: 846–857.

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.

Sipiczki, M., 2008 Interspecies hybridization and recombination in Saccharomyces wine yeasts. FEM. Yeast Res. 8: 996–1007.

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.

Talavera, G., and J. Castresana, 2007 Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56: 564–577.

Tamás, M. J., K. Luyten, F. C. Sutherland, A. Hernandez, J. Albertyn et al., 1999 Fps1p controls the accumulation and release of the compatible solute glycerol in yeast osmoregulation. Mol. Microbiol. 31: 1087–1104.

Taussig, R., and M. Carlson, 1983 Nucleotide sequence of the yeast SUC2 gene for invertase. Nucleic Acids Res. 11: 1943–1954.

Communicating editor: T. R. Hughes