The evolutionary history of *Saccharomyces* species inferred from completed mitochondrial genomes and revision in the ‘yeast mitochondrial genetic code’

Pavol Sulo1,*, Dana Szabóová1, Peter Bielik1, Silvia Poláková1,†, Katarína Soltys2, Katarína Jatzová1, and Tomáš Szemes2,3,4

1Department of Biochemistry, Faculty of Natural Sciences, Comenius University, Bratislava 842 15, Slovakia, 2Comenius University Science Park, Bratislava 841 04, Slovakia, 3Department of Molecular Biology, Faculty of Natural Sciences, Comenius University, Bratislava 842 15, Slovakia, and 4Geneton s.r.o., Galvaniho 7, Bratislava 821 04, Slovakia

*To whom correspondence should be addressed. Tel. +421 2 6029 6611. Fax. +421 2 6029 6452. Email: sulo@fns.uniba.sk

†Present address: Department of Membrane Biochemistry, Centre of Biosciences, Slovak Academy of Sciences, Bratislava 84005, Slovakia

Edited by Prof. Takashi Ito

Received 16 December 2016; Editorial decision 19 May 2017; Accepted 23 May 2017

Abstract

The yeast *Saccharomyces* are widely used to test ecological and evolutionary hypotheses. A large number of nuclear genomic DNA sequences are available, but mitochondrial genomic data are insufficient. We completed mitochondrial DNA (mtDNA) sequencing from Illumina MiSeq reads for all *Saccharomyces* species. All are circularly mapped molecules decreasing in size with phylogenetic distance from *Saccharomyces cerevisiae* but with similar gene content including regulatory and selfish elements like origins of replication, introns, free-standing open reading frames or GC clusters. Their most profound feature is species-specific alteration in gene order. The genetic code slightly differs from well-established yeast mitochondrial code as GUG is used rarely as the translation start and CGA and CGC code for arginine. The multilocus phylogeny, inferred from mtDNA, does not correlate with the trees derived from nuclear genes. mtDNA data demonstrate that *Saccharomyces cariocanus* should be assigned as a separate species and *Saccharomyces bayanus* CBS 380T should not be considered as a distinct species due to mtDNA nearly identical to *Saccharomyces uvarum* mtDNA. Apparently, comparison of mtDNAs should not be neglected in genomic studies as it is an important tool to understand the origin and evolutionary history of some yeast species.

Key words: mitochondrial DNA, yeasts, genetic code, introns, phylogeny

1. Introduction

Because of advanced genetics and simple and easy production of mutants, the yeast *Saccharomyces cerevisiae* is the most widely used unicellular eukaryotic model organism. It is the best known member of the genus *Saccharomyces* with accepted species: *S. cerevisiae*, *S. cariocanus*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevi*, *S. arboricolus*,...
S. bayanus var. bayanus, S. bayanus var. uvarum and S. pastorianus. These yeast species are involved in fermenting sugars and produce enormous amounts of ethanol; however, most of the strains used to manufacture beer, wine or fuel ethanol are interspecific hybrids. Therefore, the taxonomic classifications of allotopic species S. pastorianus and S. bayanus do not agree with the outcome of whole genome sequencing.

These discrepancies arose exclusively from the comparison of nuclear chromosomes. The data for much smaller mitochondrial genomes encoding only eight proteins required for oxidative phosporylation, one ribosomal protein, two ribosomal RNAs (rRNA), the RNA component of RNase P, and a set of 24 tRNA genes were largely overlooked or unavailable. Despite large numbers of nuclear genome sequences now accessible for different Saccharomyces species and isolates, completed mitochondrial genome sequences are known only for S. cerevisiae, S. paradoxus, S. pastorianus and Saccharomyces eubayanus. Recently, mitochondrial DNA (mtDNA) from S. bayanus and S. pastorianus were sequenced but not annotated. The problem is associated with the mitochondrial genes were largely overlooked or unavailable. Nevertheless, large numbers of nuclear genome sequences now accessible for different Saccharomyces species and isolates, completed mitochondrial genome sequences are known only for S. cerevisiae, S. paradoxus, S. pastorianus and Saccharomyces eubayanus. Recently, mitochondrial DNA (mtDNA) from S. bayanus and S. pastorianus were sequenced but not annotated. The problem is associated with the short DNA reads (100–200 nt) generated by the next generation sequencing platforms in genome studies, especially by Illumina MiSeq. This approach can be used routinely for sequencing mtDNA from different yeast species but not from many species of Saccharomyces. Their mitochondrial genomes can be assembled only by combination of different gap-filling programmes or by extensive manual editing.

The purpose of this work was to obtain complete mitochondrial genomes for all species assigned to Saccharomyces with the aim of comparing their genetic content to that of S. cerevisiae, as many different Saccharomyces mtDNAs can substitute, to some extent, for the original molecule in S. cerevisiae.

2. Materials and methods

2.1. Yeast strains and mtDNA sequencing

Sequenced strains are being used as taxonomic standards (type strains) in whole genome studies, or their mtDNA is being transferred to S. cerevisiae in order to prepare xenomitochondrial hybrids. Their origin is described in detail in reference. Saccharomyces arboriculus NRRL Y-63701 and S. pastorianus NRRL Y-27171NT were kindly provided by the Agricultural Research Service Culture Collection, US Department of Agriculture, Peoria, IL, USA. Yeast cells were grown at 28°C in YPD medium (2% glucose, 1% yeast extract and 1% peptone). The mtDNA was purified mainly by differential centrifugation or alternatively by bishenizimide/CsCl buoyant density centrifugation. mtDNA sequencing libraries were prepared using the Nextera Library Preparation Kit (Illumina, San Diego, USA) and sequenced with average genome coverage between 176 and 3,698x on an Illumina MiSeq sequencing system using a MiSeq Reagent kit v3 (Illumina) with reads 2 x 100 bp in length. The sequencing summary is listed in Table 1.

2.2. mtDNA assembly and annotation

Sequencing provided several hundred megabytes of data in more than one million reads, 90–190 nt long. Paired reads were trimmed, short reads (often with erroneous sequences) removed and assembled into individual contigs using CLC Genomics Workbench 9.5 (Qiagen, Hilden, DE). From those containing mtDNA by BLASTN comparison to already known mtDNA from related yeasts (S. cerevisiae NC001224; S. paradoxus JQ862335, S. pastorianus EU852811) were selected. Individual contigs with mtDNA segments were assembled into a single molecule using the Vector NTI v.9.0 (v.10) software package from InforMax, Inc. Overall, 1–10 contigs were obtained covering 98–101% of the entire sequence, depending on the purification method (Table 1). All mtDNAs were interrupted in the redundant intergenic regions, mostly in the origins of replication (ori ~270 nt) and GC clusters. Contigs were manually edited and linked together from individual reads as previously reported. Some gaps were sealed by direct sequencing using fluorescent dye terminator sequencing chemistry. Assembly of S. mikatae CBS 8839 (NRRL Y-27341) mtDNA was confirmed by direct sequencing as previously described. The assembled mitochondrial genomes were evaluated by mapping back the raw reads using CLC Genomics Workbench 9.5. However, a few substitutions, indels, and the occasional GC cluster can be found if our mtDNA is aligned to the mtDNA obtained from the same strains by different groups.

Rough gene annotation was carried out with MFannot (http://meg asun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl). Precise exon and intron positions were corrected when compared with annotated sequences from several strains of S. cerevisiae and S. paradoxus. Intron and endonuclease nomenclature was accounted for and the intergenic open reading frames (ORFs) are numbered according to references. The sizes of rRNA were inferred from the model elaborated for their equivalents from S. cerevisiae mitochondria. The 5′ and 3′ ends of rRNA were deduced from the secondary structure consensus. Gene nomenclature followed the rules described in GOBASE.

Saccharomyces arboriculus CBS 10644 mtDNA was assembled from the sequence CM001579 by manual editing from individual reads (EMBL: ERP001702, ERP001703, ERP001704). Saccharomyces uvarum CBS 395 was assembled from the Sequence Read Archive, NCBI.

2.3. Phylogeny

Phylogenetic relationships were analysed by the Maximum likelihood phylogeny PhyML program included in the CLC Genomics Workbench 9.5 package. Phylogenetic trees were constructed from unambiguously aligned portions of the concatenated DNA sequences coding for proteins using the neighbor-joining algorithm and the best model (GTR + G + I) selected by model testing tool (hLRT, BIC, AIC, AICc). The stability of individual branches was assessed by the bootstrap method. DNA concatemers for mitochondrial protein coding sequences belonging to other Saccharomyces strains were extracted from GenBank deposits. The nuclear DNA-derived phylogenetic tree was constructed from protein-coding genes used in population studies. Because of the hybrid allotopic nature of S. bayanus, S. uvarum and S. pastorianus we used gene sequences from the species that are the least related to S. cerevisiae variants. Corresponding DNA was obtained from assembled contigs occasionally sealed manually from reads (S. cerevisiae NRRL Y-12632, S. cariocanus CBS 7994, S. arboriculus NRRL Y-63701, S. paradoxus CBS 2908, S. bayanus CBS 380). Other concatemers were obtained from sequences, scaffolds (contigs) deposited in the Saccharomyces genome database (http://www.yeastgenome.org/) for S. cerevisiae S288c, S. paradoxus CBS 432, S. pastorianus Weihenstephan 34/70, S. kudriavzevii CBS 8840 and S. mikatae CBS 8839 and for S. arboriculus CBS 10644. Data for S. bayanus CBS 380 were obtained from contigs assembled from SRR147291 and for...
Table 1. Sequencing summary

| Strain                  | Paired reads | Average length | Coverage X4 | DNA | Contigs/nts | Coverage % | mtDNA size |
|-------------------------|--------------|----------------|-------------|-----|-------------|------------|------------|
| S. cerevisiae NRRL Y-12632T | 2,578,141    | 147            | 4,234       | Df  | 5/87,765    | 99         | 89,507     |
| S. paradoxus CBS 2908    | 1,580,034    | 96             | 2,283       | Gr  | 5/66,147    | 99         | 66,436     |
| S. paradoxus CBS 7400    | 2,242,369    | 96             | 3,014       | Gr  | 2/70,538    | 98         | 71,419     |
| S. cariocanus CBS 7994T  | 1,897,384    | 96             | 2,354       | Gr  | 1/70,729    | 98         | 77,380     |
| S. mikatae CBS 8839T     | 413,405      | 189            | 917         | Df  | 2/85,137    | 99         | 85,211     |
| S. kudriavzevi CBS 8840T | 2,259,934    | 96             | 2,692       | Gr  | 3/80,240    | 99         | 80,588     |
| S. arboricolus CBS 8839T | 535,328      | 139            | 1,073       | Df  | 9/70,040    | 101        | 69,363     |
| S. paradoxus NRRL Y-27171T | 555,165     | 158            | 1,271       | Df  | 7/68,800    | 99         | 69,019     |
| S. bayanus CBS 380T      | 1,221,753    | 85             | 1,604       | Gr  | 1/64,742    | 100        | 64,736     |
| S. uvarum CBS 395T       | 25,298,820   | 36             | 14,059      | –   | 34/51,450   | 79         | 64,779     |

4Coverage calculated as LN/G, where L is the read length, N is the number of reads and G is the mitochondrial genome length.

3. Results and discussion

3.1. General features of genes of the genetic code and GUG as a translation start

We have successfully completed mitochondrial genomes for nine Saccharomyces species, and their general features are listed in Table 2 (GenBank accession numbers KX657740–KX657750). All mtDNAs are circularly mapped and their size decreases with phylogenetic distance from S. cerevisiae (89.5 kb) to ~64.7 kb for S. bayanus. GC content varies from 14% for S. paradoxus strains to 19.3% for S. pastorianus and correlates positively with the presence of introns and negatively with the size of AT-rich intergenic spacers. The mtDNAs contain the basic set of genes known in Saccharomyces, coding for the RNA subunit of RNase P and the tRNA packaged in the rntB gene for the RNA subunit of RNase P and the tRNA package (Fig. 1).

The genetic code in Saccharomyces mtDNA differs from the universal code by TGA being read as tryptophan, CTN as threonine, ATA as methionine and CGN for arginine, with CGA and CGC absent, that is recognized as the ‘yeast mitochondrial code’ (transl_table = 3).41–43 The trnK2 gene (anticodon ACG) was found in all Saccharomyces mtDNAs. Thus, all four CGN codons possess coding potential. However, the CGU codon is present in exons only once (besides S. arboricolus) coding for a non-conserved arginine residue in rps3 protein (Supplementary Material S1). In addition, the CGU codon occurs quite frequently in intron-coded or free-standing ORFs, whereas the CGG codon is present only in GC clusters located inside them. The arginine CGA and CGC codons considered as unassigned are also used sporadically in maturase/homing endonucleases (HE) reading frames coded by introns (Supplementary Material S1). The most convincing evidence is a CGC codon present in the cox1II reading frame in a number of S. cerevisiae strains including type strain NRRL Y-12632 and the well-studied model strain S288c. This reverse transcriptase/maturase protein is required for group II intron splicing from cox1 pre-mRNA as well as for intron mobility.45,46 Apparently, tRNA coded by the trnR2 gene is required in Saccharomyces predominantly for the expression of mobile elements and is therefore gained and lost in different clades from the Saccharomyces Kluyveromyces complex.47,48 Protein-coding genes start with ATG,
Figure 1. The genetic organization of *Saccharomyces* mtDNA. For simpler comparison, the circular genomes, exported from Vector NTI, were aligned at the beginning of the large rRNA subunit (rnl). Protein-coding genes, ribosomal RNA, *mpB* are marked as arrows and bar, tRNA genes as black lines, introns with white rectangles, intronic and free-standing ORFs by gray arrows and replication origins with black circles. Gene nomenclature follows the rules described in GOBASE (atp for ATP synthetase subunits, *cox* for cytochrome oxidase subunits, *cob* for cytochrome b, *ms* for small rRNA ribosomal subunit, *ml* for large rRNA ribosomal subunit, T2, C, H, etc. for particular tRNA coding genes, *rps3* for ribosomal protein and *mpB* for the RNA subunit of RNase P). Sizes are given on the bottom line in kbp.
with some exceptions. The translation start in the *S. cariocanus* cox3 gene is GTG (GUG), reported to be nearly as effective as the original ATG in the *S. cerevisiae* cox2 gene.\textsuperscript{49} Translation initiation at the nearby downstream AUG is not an option because mutagenesis of the regular initiation codon does not allow translation from this site.\textsuperscript{49} Instead, a stem loop structure in the mRNA sequence of the first six codons plays a role in the recognition of the translation start.\textsuperscript{50} The first 50 nucleotides in the cox3 genes of all sequenced Saccharomyces are identical, which emphasizes the reliability of GUG as the initiation start codon (Fig. 2A, B). Atypical start codons are present in many mitochondrial systems, but GUG as a start codon is known only in some birds, invertebrates, plants and protists, but not in yeast and molds.\textsuperscript{51,64} Also, a maternally inherited and practically homoplasmic mtDNA mutation, 8527 A > G, changing the initiation codon AUG into GUG, was observed in the human *atp6* gene. The patient harboring this mutation exhibited clinical symptoms of Leber hereditary optic neuropathy, but his mother was healthy.\textsuperscript{51} The second exception is free-standing ORFs in *S. kudriavzevi*, *S. mikatae*, *S. bayanus*, *S. uvarum*, *S. arboricola* and *S. pastorianus*, where GUG is used as a translation start site. This gene with unknown function codes for a hypothetical maturase-like protein, the start codon overlaps with the 3’ end of the cox2 gene and the reading frame extends 16 bp downstream of the cox2 stop codon. Consequently, the ‘yeast mitochondrial code’ should be updated to accept CGA and CGG as arginine codons and GTG as the alternative initiation codon (Fig. 2C). The preferable termination codon in exons is UAA, and UAG is used only occasionally in cox2 of *S. kudriavzevi* and *S. mikatae* and more often in intron-coded ORFs. Codon usage in the reading frames shows a strong bias towards codons ending with T or A (in 10:1), as previously observed in the mitochondrial genomes of other yeasts.\textsuperscript{61} (Supplementary Material S1).

All sequenced mitochondrial genomes exhibit sequence conservation at the nucleotide level within exons (excluding extremely variably rtp3), with at least 94% sequence identity between the most unrelated species and a substitution rate of 11%. The substitutions are mostly neutral and have a low impact on the protein sequence, as they generate nearly identical amino acid residues. Therefore, the overall protein identity, even among less-related species, is 97%. Interestingly, the *atp8* protein sequence is identical in all sequenced species. The same DNA substitution rate (6%) in the other small gene, *atp6*, is responsible for the three changes at the protein level (Supplementary Material S2).

3.2 Introns and free-standing ORFs

Introns are present in Saccharomyces only within *cox1*, *cob* and *rnl* genes.\textsuperscript{22,27} Although the majority are I introns, four of them belong to group II, and only two of these (coxI1, coxI2) code for a protein with the maturase/reverse transcriptase motif. The intervening sequences are inserted into the same sites as has been described in other Saccharomyces, especially more than 100 strains of *S. cerevisiae*.\textsuperscript{11,19,20,21,22} (Fig. 3). However, a detailed analysis indicates that the insertion site for *coxI3α* should be position shifted to position 242. Position 243 would generate lysine instead of an asparagine residue in the GGFGN section in both *S. arboricola* molecules. This motif is conserved across the tree of life from yeast to humans. G-G-F-G-N-[WY]-[FL]-[MV] is a part of the well-recognized water membrane interface motif in cytochrome c and quinol oxidases (MeMotif—http://projects.bioteo.tu-dresden.de/memotif).\textsuperscript{53}

A number of introns also contain ORFs in phase with the upstream exon, sometimes interrupted by a frame-shift stop codon, often due to the insertion of GC clusters. Only those possessing at least one motif characteristic for the HE (LAGLIDADG, GIY-YIG, etc.) were considered as potentially active\textsuperscript{54,55} (Fig. 3).

However, there are two exceptions: *coxI5β* and *cobI1α*, where an ORF coding for a nuclease/maturase motif is not fused to the upstream exon. *coxI5β* is not known as mobile and the splicing requires at least five different nuclei-coded factors and an ORF from *cobI3*.\textsuperscript{56} The reason is that the 3’ splice site lies unusually far from the catalytic core of the intron; therefore, the ORF 357 amino acids (AA) long containing two LAGLIDADG motifs in the 28844 strain was not considered as potentially active.\textsuperscript{56} But a recent *S. cerevisiae* mitochondrial transcriptome study revealed an alternative splice site that puts the *coxI5β* ORF in frame with the upstream exon.\textsuperscript{57} We found the same ACTATTATATATT consensus of the alternative splice site in several other Saccharomyces species (S. mikatae, S. cariocanus, S. bayanus, S. wuvarum, S. pastorianus strains and S. eubayanus and S. paradoxus CBS 432; Fig. 4A).

Perhaps, a similar mechanism can be used in *cobI1α*. We found a hypothetical alternative splice site sharing sequence similarity AAATTTAG(AGT(AGT)AAATTTA allowing the translation of an intron reading frame with an upstream exon (Fig. 4B). In *S. cariocanus*, *S. pastorianus*, *S. mikatae* and *S. eubayanus* it contains a 169–297 AA-long reading frame with one GIY-YIG endonuclease maturase motif.\textsuperscript{58}

Introns, especially those containing HEs, represent a class of mobile elements spread horizontally into an intron-less allele by ‘homing’. They can be found in different species in the same position of the same gene in a chain-involved cyclical gain and loss.\textsuperscript{19,59,60} Consequently, they are randomly distributed among the different species.\textsuperscript{19,20} However, there are a few remarkable exceptions. Intron *coxI3β* has not been found in *S. cerevisiae* although more than a hundred strains have been sequenced. Besides both *S. arboricola* strains, *coxI3β* is present in all other Saccharomyces species.\textsuperscript{19–21} Other evidence of uneven *coxI3β* distribution comes from *S. cerevisiae* xenomitochondrial hybrids containing *S. paradoxus* mtDNA. They exhibit a slower growth rate on a non-fermentable carbon source, decreased respiration capacity and reduced cytochrome aa3 content associated with the inefficient splicing of *coxI3β*.\textsuperscript{22} Apparently, the occurrence of *coxI3β* is related to Dobzhansky–Muller nuclear incompatibilities in nucleomitochondrial communication that may result in the divergence of yeast species.\textsuperscript{22}

Very rare is the orphan *coxI4γ* intron that occurs only in *S. bayanus* CBS 380 and *S. wuvarum* CBS 393. We were unable to find homologs in spite of extensive search in the literature and databases.\textsuperscript{19} Also, *coxI4γ*, well known as coding for HE I-Scel with a LAGLIDADG motif,\textsuperscript{61} was not identified in any other Saccharomyces but *S. cerevisiae*. Unlike *coxI4γ* it can be found in less related species from the Saccharomyces/Keyleveromyces complex *Namptoua castellii*\textsuperscript{62} and *Kleyeveromyces marxianus* species.\textsuperscript{19,63}

Intron homing is catalyzed by HEs\textsuperscript{54,65} that have different evolutionary history and can exist alone.\textsuperscript{56,68} Such free-standing ‘endonucleases’, namely ORFs,\textsuperscript{29} have already been described in *S. cerevisiae* and *S. pastorianus* mtDNA, but their occurrence is strain dependent.\textsuperscript{6,29} ORF1 orthologs are most frequent in different Saccharomyces species (Fig. 3). They code for one or two LAGLIDADG motifs and a DNA-binding helix-turn-helix domain.\textsuperscript{68} Because of the number of stop codons or a GC cluster insertion ORF2 is present mostly in truncated form.\textsuperscript{19,22} A reading frame coding for one LAGLIDADG motif was found only in *S. cariocanus*. We were unable to detect ORF3 in any ‘non-cerevisiae’ Saccharomyces.
Apparently this dubious ORF does not encode a functional protein. Full size ORF4 can be found in *S. cerevisiae* and *S. arboricola*. Among all previously identified small ORFs, regions with significant homology to ORF5 were found only in *S. paradoxus* and *S. carioca*. Therefore, ORF5 is considered as a probable genomic artifact, although very small proteins coded by mtDNA are expected, since at least two small proteins were reported to be coded by *rns* in mammalian mitochondria. Figure 2. GTG as the initiation codon and revised yeast mitochondrial code. (A) Initiation codons in *cox2* gene. (B) Initiation codons in *ORF1* gene. Sequences overlapping with *cox2* 3’ end are underlined. (C) Revised yeast mitochondrial code. Exceptional codons are grey-shadowed.
Figure 3. Occurrence of introns and free-standing ORFs. Introns and free-standing ORFs with at least one motif characteristic for the homing endonucleases (LAGLIDADG, GIY-YIG, etc.) are in black; without reading frame are white; with truncated ORF interrupted by stop codon are spotted. Numbers indicate the base preceding the intron insertion site in the CDS. Positions of introns, known as mobile, in *S. cerevisiae* are marked in bold; I—group I introns, II—group II introns. ORF1, ORF2 and ORF4 correspond to the nomenclature used in reference29 (their ORF5 is an ORF coded by the *ml* intron). Only ORFs containing at least one endonuclease/maturase motif were considered.

Figure 4. Alternative splicing of *cox1* (A) and *cob1* (B). Arrows mark exons–introns boundaries. Consensus of alternative splice site as described in reference57 and hypothetical alternative splice site in *cob1* are underlined.
3.3. Intergenic DNA, GC clusters and origins of replication

The putative origins of replication in *S. cerevisiae* and *S. paradoxus* are ~270 bp in length and contain GC blocks separated by AT-rich stretches. According to novel bioinformatics analysis (strain S288c) it is a five-part consensus sequence [GGGGGAAGGGGT GGGTGAT −200 A/T-rich GGGTCCC 29 A/T-rich GGACC] downstream of the 20-base promoter [DDWDWTAWAA GT]ARTADDDD. Saccharomyces cerevisiae mtDNA contains seven or eight ori elements, depending on the strain, but only ori 2, 3 and 5 are active. They contain a transcription initiation site adjacent to the five-part consensus sequence providing a contiguous transcript of >350 nucleotides. The other ori sites are disabled by the insertion of a G-class GC cluster CCCGGTTTCTTACGAAACCGGGACC TCGGAGA(C/A)GT into the promoter site. Recently, the M2 class consensus was reduced to 14 bases (TCCGGCC GAAGGAG). Therefore, we took into consideration as GC clusters only regions longer than 20 nt with GC content ≥ 40%. The average number of GC clusters in *S. cerevisiae* is above 100 and their occurrence decreases in other *Saccharomyces* species from 90 to 30 (Table 3). A G-class ori-specific GC cluster has been found in any species from the *S. bayanus*, *S. wuvarum*, *S. pastorius* clade, but remote from ori. The most frequent are M1 and M2 classes, although members of other classes can be found. In addition, a species-specific GC-rich sequence (GGGGTCCC N(28–30) GGGGTCCC) was recognized in both *S. arboricola* strains and a GC cluster with the consensus TCGTNWCGYACCCTCAATWGACGTTACG in species from the lager beer species/*S. wuvarum* clade.

| Table 3. Replication origins and GC clusters |
|----------------------------------------------|
| *Saccharomyces* strain | ori  | Number of GC clusters |
|------------------------|------|----------------------|
|                        | G    | V       | M1  | M1' | M2  | M2' | M2''| M3  | M4  | U   | SpS* | Classified/total |
| *S. cerevisiae*        | 7/2  | 5      | 48  | 7   | 46  | 1   | 6   | 11  | 4   | 50  | –    | 127/177         |
| NRRL Y-12632T          |      |        |     |     |     |     |     |     |     |     |      |                  |
| *S. paradoxus*         | 8/8  | –      | 20  | 1   | 12  | –   | 1   | 1   | 1   | 30  | –    | 46/76           |
| CBS 2908               |      |        |     |     |     |     |     |     |     |     |      |                  |
| *S. paradoxus*         | 8/8  | –      | 19  | 1   | 12  | 1   | 1   | 1   | 1   | 32  | –    | 45/77           |
| CBS 7400               |      |        |     |     |     |     |     |     |     |     |      |                  |
| *S. cariocanus*        | 7/7  | –      | 10  | 2   | 23  | 5   | –   | –   | –   | 72  | –    | 77/149          |
| CBS 7994T              |      |        |     |     |     |     |     |     |     |     |      |                  |
| *S. mikatae*           | 7/7  | –      | 32  | 7   | 30  | 1   | –   | –   | 10  | 50  | –    | 90/143          |
| CBS 8839T              |      |        |     |     |     |     |     |     |     |     |      |                  |
| *S. kudriavzevi*       | 8/8  | –      | 26  | 9   | 37  | 5   | –   | –   | –   | 51  | –    | 85/136          |
| CBS 8840T              |      |        |     |     |     |     |     |     |     |     |      |                  |
| *S. arboricola*        | 6/6  | –      | 6   | 6   | 2   | 5   | 2   | –   | –   | 76  | 7    | 36/112          |
| NRRL Y-63701           |      |        |     |     |     |     |     |     |     |     |      |                  |
| *S. arboricola*        | 6/6  | –      | 7   | 2   | 9   | 5   | 6   | –   | 2   | 63  | 6    | 38/105          |
| CBS 10644T             |      |        |     |     |     |     |     |     |     |     |      |                  |
| *S. pastorinus*        | 4/4  | 3     | 13  | –   | 44  | 1   | 6   | 3   | 1   | 101 | 6    | 82/183          |
| NRRL Y-27171T          |      |        |     |     |     |     |     |     |     |     |      |                  |
| *S. bayanus*           | 4/4  | 5     | 13  | 16  | –   | 17  | –   | –   | –   | 38  | 7    | 57/95           |
| CBS 380T               |      |        |     |     |     |     |     |     |     |     |      |                  |
| *S. wuvarum*           | 4/4  | 5     | 13  | 17  | –   | 17  | –   | –   | –   | 38  | 7    | 58/96           |
| CBS 395T               |      |        |     |     |     |     |     |     |     |     |      |                  |
| *S. cerevisiae*        | 8/3  | 5     | 55  | 7   | 37  | 9   | 13  | 3   | 9   | 48  | –    | 127/177         |
| S288c                  |      |        |     |     |     |     |     |     |     |     |      |                  |
| *S. embayanus*         | 4/4  | 1     | 26  | –   | 33  | –   | –   | –   | –   | 56  | 5    | 64/120          |
| CBS 12357T             |      |        |     |     |     |     |     |     |     |     |      |                  |
| *S. pastorinus*        | 4/4  | 6     | 13  | 1   | 50  | 3   | 10  | –   | –   | 85  | 9    | 86/171          |
| WS 34/70               |      |        |     |     |     |     |     |     |     |     |      |                  |
| *S. paradoxus*         | 8/8  | –      | 14  | 8   | –   | 1   | 1   | 1   | 50  | –   | 32/82          |
| CBS 432T               |      |        |     |     |     |     |     |     |     |     |      |                  |

*ori2 considered as active is absent; G, V, M1, M1', M2, M2', M2'', M3 and M4 are GC cluster classes identified according to consensus described in detail in reference 20 with various degrees of degeneration (Supplementary Material S1C); U—unclassified GC clusters defined as longer than ≥ 20 nt with GC content ≥ 40%; SpS—species-specific GC clusters found in the *S. bayanus*, *S. wuvarum*, *S. pastorinus* clade consensus sequence TCGTNWCGYACCCTCAATWGACGTTACG and in *S. arboricola* GGGGTCCC N(28–30) GGGGTCCC. Strains assembled in this work are marked in bold.*

* GC clusters were originally defined as regions 35 bp long on average and 45–62% GC and are in general the major source of noticeable intergenic region polymorphism. According to the characteristic primary structures, they have been assigned to seven families which exhibit varying degrees of homology. Recently, the M2 class consensus was reduced to 14 bases (TCCGGCC GAAGGAG). Therefore, we took into consideration as GC clusters only regions longer than ≥ 20 nt with GC content ≥ 40%. The average number of GC clusters in *S. cerevisiae* is above 100 and their occurrence decreases in other *Saccharomyces* species from 90 to 30 (Table 3). A G-class ori-specific GC cluster has been found in any species from the *S. bayanus*, *S. wuvarum*, *S. pastorinus* clade, but remote from ori. The most frequent are M1 and M2 classes, although members of other classes can be found. In addition, a species-specific GC-rich sequence (GGGGTCCC N(28–30) GGGGTCCC) was recognized in both *S. arboricola* strains and a GC cluster with the consensus TCGTNWCGYACCCTCAATWGACGTTACG in species from the lager beer species/*S. wuvarum* clade.
3.4. Transcription units and adjacent motifs
The highly-conserved sequence motif WTATAAGTA is known as the yeast mitochondrial transcription initiation site. Recently, transcriptome analysis confirmed experimentally the sequence DDWDWTAWAA GT | ARTADD DD as the consensus promoter site in S. cerevisiae and 19 potential transcription initiation sites were reduced to 11. We found these consensus promoter sites in all sequenced Saccharomyces at approximately the same position as they were reported for S. cerevisiae. Because of the rearrangement of the trnS1- rps3 block in the mtDNA of the S. bayanus, S. uvarum, S. pastorianus clade, an additional transcription site was found upstream of the trnS1 gene (Supplementary Material S4A). When all hypothetical promoter sites are compared, the S. cerevisiae consensus is slightly degenerated mainly at the 5′ site to DNDNDNTA WAA GT | ARTADD DD that is more related to the original nonanucleotide WTATAAGTA in spite of the phage T7 origin of yeast mitochondrial RNA polymerase. Apparently, the 23 base-long phage T7 consensus promoter is reduced in Saccharomyces mtDNA.

The 3′ termini of mitochondrial mRNA cleavage sites in budding yeast are recognized according to the postulated dodecamer motif 5′-AAUAA (UC) AUU CUU-3′ in the 3′ UTR region. A processing site was confirmed by transcriptome analysis where an alternative heptakaidecamer 5′-AATAATATTCTTAT|AGTCGCGCC|CG CCC with part of the M2 GC cluster is recognized instead. Variations to the dodecamer have been found in all protein-coding transcription units at the 3′ UTR region of all Saccharomyces species, in most cases up to 200 nucleotides downstream from the terminon codon (Supplementary Material S4B). A putative atp8 transcription termination site exhibits the most degenerated consensus sequence. Interesting features are the multiple cleavage sites for ORF1 present in S. mikatae, S. kudriavzevii and S. arboricola. An alternative heptakaidecamer transcription processing site was found only in S. cerevisiae.

3.5. Gene order and synteny
The most profound feature is the alteration in gene order that involves mainly trnF-M-rpB-trnP, rns-rps3, trnL-cob and trnS1-rps3 gene clusters. S. cerevisiae NRRL Y-12632 mtDNA with a length of almost 90 kb belongs among the largest genomes and has the same gene order as is known in 105 other strains. S. paradoxus CBS 2908 and CBS 7400 mtDNA differs in size (66 kb versus 71 kb), associated with the intron occurrence in the cob gene. The gene order is the same as in 14 other strains. They differ from S. cerevisiae by the excision of the trnFT1V-cox3-trnF-M-rpB-trnP-rns-trnW segment that is translocated behind the trnS1-rps3 gene cluster where rns-trnW is inverted (Figs 1 and 5). Surprisingly, the 77 kb mtDNA of S. cariocanus is syntenic to S. cerevisiae, although these species can be distinguished according to the coxI3β intron present in S. cariocanus (Figs 1 and 3). S. mikatae mtDNA is >85 kb long and the gene architecture is altered from S. cerevisiae by the movement and inversion of the trnF-M-rpB-trnP gene cluster behind the rns-trnW locus (Fig. 5). S. kudriavzevii mtDNA is ~80 kb long and the gene order differs from S. cerevisiae by the excision of rns-trnW and its transplacement to the complementary strand at the 3′ end of the trnS1-rps3 gene cluster (Fig. 5). Both S. arboricola mtDNAs are around 70 kb in size and the genome architecture exhibits very extensive reshuffling in relation to S. cerevisiae. The rns-trnW cluster is excised and transferred at the 3′ end of the atp9-trnS1-rps3 gene cluster. The remaining trnF-M-rpB-trnP-coc1-atp8-atp6 genome segment is translocated at the end of the cob gene (Figs 1 and 5). The characteristic transcription unit atp9-trnS1-rps3 is broken and the trnS1-rps3 part is inverted at the same location (Figs 1 and 5). Strains from the lager beer species S. uvarum clade despite their hybrid nature are classified as separate species, although they exhibit strikingly similar mtDNA sequences. Nearly identical are mtDNAs from S. uvarum CBS 395T and S. bayanus CBS 380T.

Figure 5. Comparison of mitochondrial gene order in Saccharomyces. Individual gene clusters are highlighted by different shades of grey. Pentagon symbols indicate the transcription direction.
The alteration in gene order within yeast genera is not so frequent and is associated with the genome size. The majority of yeast mtDNA is relatively small (<40 kbp) with the genes located on the same strand. The size of <15% of mtDNA exceeds 50 kb and the genes are scattered in conserved blocks (e.g. cox1-atp8-atp6, mt-trn) coded in both strands with altered transcription orientation.7 Obviously, larger genomes like in Saccharomyces are more prone to rearrangements. An extremely high degree of gene conservation and synteny is the characteristic feature among and within the species possessing smaller mtDNA of the Lachanceae,80 Torulaspora81 and Yarrowia clades.82 However, significant mitochondrial genome rearrangement was observed in yeasts with larger mtDNA from the Nakaseomyces83 and Dekkera/Brettanomyces genera.84 Apparently, alteration in mtDNA gene order depends most of all on the size of intergenic regions and is obvious if their ratio to genes exceeds 60%.7 The gene order alteration mechanism was deduced exclusively from the DNA sequence, proposing an intermediate with duplicate genes, associated with conserved gene blocks mimicking transcription units.11,84,85 This is not the case in S. bayanus, S. eubayanus, S. uvarum and S. pastorianus where the atp9-trnS1-rps3 transcription unit is broken apart and the transposition of trnS1-rps3 to the opposite strand requires formation of a de novo promoter.

3.6. Phylogeny and taxonomy
As mentioned above, currently S. cerevisiae, S. paradoxus, S. mikatae, S. kudriavzevi, S. arboricola, S. bayanus, S. uvarum and S. pastorianus are the only species accepted by taxonomists as Saccharomyces.1 However, whole (nuclear) genome studies demonstrated that S. pastorianus is an allopolyploid of S. cerevisiae and the newly described species S. eubayanus as well as S. bayanus is a hybrid between S. uvarum and S. uvarum with a small contribution of S. cerevisiae.4,5,13,26,86 Saccharomyces cariocanus is reproducibly isolated by four chromosomal translocations from S. paradoxus but not by sequence and therefore is considered to be a S. paradoxus subspecies.4,5,86 All these data came exclusively from nuclear genetic information, although mitochondrial genomes are more susceptible to mutations than their nuclear counterparts.

The accuracy of taxonomic classification as well as the evolution history can be inferred from phylogenetic trees derived from concatenated DNA or protein sequences of mitochondrial or nuclear origin. However, support for many lineages in present phylogenetic trees is often weak and more robust analyses of relationships will require whole genome comparisons.87 To shed light on these taxonomy–phylogeny pitfalls we constructed phylogenetic trees from unambiguously aligned concatenated DNA sequences using the Maximum likelihood phylogeny PhyML program. (A) mtDNA-derived phylogeny from DNA sequences coding for proteins in the order cox1, atp8, atp6, cox2, cox3, atp9, atp8, and atp9 (CBS 395) was isolated from turbid beer. Only S. uvarum (CBS 395) was isolated from natural substrate (blackcurrant).26 The origin of S. pastorianus was studied in detail and has been elucidated by the discovery of a new species S. eubayanus.5 All modern S. pastorianus are

Figure 6. Saccharomyces phylogeny. Both trees were constructed from unambiguously aligned concatenated DNA sequences using the Maximum likelihood phylogeny PhyML program. (A) mtDNA-derived phylogeny from DNA sequences coding for proteins in the order cox1, atp8, atp6, cox2, cox3, atp9, atp8, and atp9 (CBS 395) was isolated from turbid beer. Only S. uvarum (CBS 395) was isolated from natural substrate (blackcurrant).26 The branch length is proportional to the nucleotide differences indicated by the bar. The numbers given at the nodes are the frequencies of a given branch appearing in 1,000 bootstrap replications. All are above 90%, indicating good statistical support. Naumovia castelli NRRL Y-12630 was used as an outgroup.
aneuploid descendants of a tetraploid S. eubayanus/S. cerevisiae hybrid (2n, 2e) with limited loss of contribution from either parent.4,5,13,89 Apparently, the S. pastorianus strain sequenced in this work as well as a number of other strains inherited mtDNA exclusively from S. eubayanus mtDNA.13,89 The systematics of S. bayanus and S. warum has been confusing and controversial for decades. S. warum (S. bayanus var. warum) strains represent a nearly pure lineage that contains very little genetic input from other Saccharomyces species although type strain CBS 395 does not produce viable spores and sequencing revealed that the strain is aneuploid.26 However, a number of different strains (including CBS 7001 ‘genomic type strain’) are good sporulants and introgression from other Saccharomyces species is low.4,6 S. bayanus is not a ‘species’ in an evolutionary sense and should be rather understood as a product of the artificial brewing environment with no occurrence in nature. Therefore, all known strains of S. bayanus (including the type strain CBS 380) are hybrids of S. eubayanus and S. warum that contain contributions from S. cerevisiae in at least some cases.26 Comparison of mtDNA clearly demonstrates its S. warum origin in S. bayanus CBS 380. A similar conclusion was reported by 91 when they compared part of the cox2 gene in the collection of S. eubayanus and its interspecies hybrids. A model of S. bayanus formation involves multiple hybridization events of S. pastorianus with wild strains of S. warum. In the case of S. bayanus CBS 380 it happened only recently as mtDNA from S. warum CBS 395 and S. bayanus CBS 380 is nearly identical. They only differ in four substitutions and five indels of which one is the insertion of a GC cluster and one a TTATTATAC repeat. Therefore, comparison of mtDNA should not be neglected in genomic studies as it is an important tool to understand the origin and evolutionary history of some yeast species.

4. Conclusions

We sequenced mtDNA from a variety of Saccharomyces species by Illumina MiSeq. All are circularly mapped molecules decreasing in size with phylogenetic distance from S. cerevisiae but with similar gene content including regulatory and selfish elements like origins of replication, introns, free-standing ORFs or GC clusters. Their most profound feature is species-specific alteration in gene order, apparently accompanying the speciation process in yeasts. Their most profound feature is species-specific alteration in gene origins of replication, introns, free-standing ORFs or GC clusters.

Supplementary data are available at DNARES online.

Acknowledgements

The authors thank two anonymous reviewers for their valuable comments. The study is a result of implementation of project REVGENE – Research Center for Molecular Genetics (ITMS 26240220067) supported by the Research & Development Operational Program funded by the ERDF. Part of this work was funded by grants from VEGA 1/0360/12 and 1/0048/16.

Conflict of interest

None declared.

Accession numbers

KX657740, KX657741, KX657742, KX657743, KX657744, KX657745, KX657746, KX657747, KX657748, KX657749, KX657750

References

1. Vaughan-Martini, A., Martini, A. 2011, Saccharomyces meyen ex rees (1870). In: Kurtzman C.P., ed. The Yeasts, A Taxonomic Study, Elsevier: Amsterdam, NL. pp. 733–46.
2. Replansky, T., Koufopanou, V., Greig, D., Bell, G. 2008, Saccharomyces sensu stricto as a model system for evolution and ecology, Trends Ecol. Evol., 23, 494–501.
3. Spicizci, M. 2008, Interspecies hybridization and recombination in Saccharomyces wine yeasts, FEMS Yeast Res., 8, 996–1007.
4. Borneman, A.R., Pretorius, I.S. 2015, Genomic insights into the Saccharomyces sensu stricto complex, Genetics, 199, 281–91.
5. Baker, E., Wang, B., Bellora, N., et al. 2015, The genome sequence of Saccharomyces eubayanus and the domestication of lager-brewing yeasts, Mol. Biol. Evol., 32, 2818–31.
6. Fourry, F., Roganti, T., Lecrenier, N., Purnelle, B. 1998, The complete sequence of the mitochondrial genome of Saccharomyces cerevisiae, FEBS Lett., 440, 325–31.
7. Freel, K.C., Friedrich, A., Schacherer, J. 2015, Mitochondrial genome evolution in yeasts: an all-encompassing view, FEMS Yeast Res., 15, 1–9.
8. Kellis, M., Patterson, N., Endrizzi, M., Birren, B., Lander, E.S. 2003, Sequencing and comparison of yeast species to identify genes and regulatory elements, Nature, 423, 241–54.
9. Liti, G., Ba, A.N.N., Blythe, M., et al. 2013, High quality de novo sequencing and assembly of the Saccharomyces arboriculus genome, BMC Genomics, 14, 1–14.
10. Hittinger, C.T., Rokas, A., Bai, F.-Y., et al. 2015, Genomics and the making of yeast biodiversity, Curr. Opin. Genet. Dev., 35, 100–9.
11. Procházká, E., Franko, F., Poláková, S., Sulo, P. 2012, A complete sequence of Saccharomyces paradoxus mitochondrial genome that restores the respiration in S. cerevisiae, FEMS Yeast Res., 12, 819–30.
12. Nakao, Y., Kanamori, T., Itoh, T., et al. 2009, Genome sequence of the lager brewing yeast, an interspecies hybrid, DNA Res., 16, 115–29.
13. Okuno, M., Kajitani, R., Ryusui, R., Morimoto, H., Kodama, Y., Itoh, T. 2016, Next-generation sequencing analysis of lager brewing yeast strains reveals the evolutionary history of interspecies hybridization, DNA Res., 23, 67–80.
14. Liu, L., Li, Y., Li, S., et al. 2012, Comparison of next generation sequencing systems, J. Biomed. Biotechnol., 2012, 1–11.
15. Quinl, M.A., Smith, M., Coupaland, P., et al. 2012, A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers, BMC Genomics, 13, 1–13.
16. Skelly, D.A., Merrihew, G.E., Ruffle, M., et al. 2013. Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast, *Genome Res.*, 23, 1496–504.

17. Bergstrom, A., Simpson, J.T., Salinas, F., et al. 2014. A high-definition view of functional genetic variation from natural yeast genomes. *Mol. Biol. Evol.*, 31, 872–88.

18. Lang, B.F., Jakubkova, M., Hegedusova, E., et al. 2014. Massive programmed translational jumping in mitochondria, *Proc. Natl. Acad. Sci. U. S. A.*, 111, 5926–31.

19. Wu, B., Hao, W. 2015. A dynamic mobile DNA family in the yeast mitochondrial genome, *FEMS Yeast Res.*, 15, 1–21.

20. Libkind, D., Hittinger, C.T., Valeiro, E., et al. 2011, Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast, *Proc. Natl. Acad. Sci. U. S. A.*, 108, 14539–44.

21. Strope, P.K., Skelly, D.A., Kozmin, S.G., et al. 2015, The 100-genomes project: the power of reagents for targeted DNA modification, *PLoS ONE*, 1–21.

22. Libkind, D., Hittinger, C.T., Valeiro, E., et al. 2011. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast, *Proc. Natl. Acad. Sci. U. S. A.*, 108, 14539–44.

23. de Zamaroczy, M., Bernardi, G. 1986b, The primary structure of the mitochondrial RNA landscape of *Saccharomyces cerevisiae*, *Annu. Rev. Biochem.*, 55, 559–74.

24. Libkind, D., Hittinger, C.T., Valeiro, E., et al. 2011. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast, *Proc. Natl. Acad. Sci. U. S. A.*, 108, 14539–44.

25. Libkind, D., Hittinger, C.T., Valeiro, E., et al. 2011. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast, *Proc. Natl. Acad. Sci. U. S. A.*, 108, 14539–44.

26. Libkind, D., Hittinger, C.T., Valeiro, E., et al. 2011. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast, *Proc. Natl. Acad. Sci. U. S. A.*, 108, 14539–44.

27. Libkind, D., Hittinger, C.T., Valeiro, E., et al. 2011. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast, *Proc. Natl. Acad. Sci. U. S. A.*, 108, 14539–44.

28. Libkind, D., Hittinger, C.T., Valeiro, E., et al. 2011. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast, *Proc. Natl. Acad. Sci. U. S. A.*, 108, 14539–44.
two homologous intron encoded proteins from yeast mitochondria, Cell, 56, 431–41.

62. Petersen, R.F., Langkjær, R.B., Hvidfeldt, J., et al. 2002, Inheritance and organisation of the mitochondrial genome differ between two Saccharomyces yeasts, J. Mol. Biol., 318, 627–36.

63. Talla, E., Anthouard, V., Bouchier, C., Frangeul, L., Dujon, B. 2005, The complete mitochondrial genome of the yeast Kluyveromyces thermotolerans, FEBS Lett., 579, 30–40.

64. Haugen, P., Simon, D.M., Bhattacharya, D. 2005, The natural history of group I introns, Trends Genet., 21, 111–9.

65. Stoddard, B.L. 2006, Homing endonuclease structure & function, Q. Rev. Biophys., 38, 49–95.

66. Edgell, D.R. 2009, Selfish DNA: homing endonucleases find a home, Curr. Biol., 19, R115–7.

67. Belfort, M. 2003, Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase, Genes Dev., 17, 2860–3.

68. Marchler-Bauer, A., Lu, S., Anderson, J.B., et al. 2011, CDD: a Conserved Domain Database for the functional annotation of proteins, Nucleic Acids Res., 39, D225–9.

69. Fisk, D.G., Ball, C.A., Dolinski, K., et al. 2006, The Conserved Domain Database for the functional annotation of proteins, Nucleic Acids Res., 34, 71–9.

70. Lee, C., Kim, K.H., Cohen, P. 2016, MOTS-c: a novel mitochondrial-derived peptide regulating muscle and fat metabolism, Free Radical Biol. Med., 100, 1–6.

71. Lee, C., Zeng, J., Drew, B.G., et al. 2013, The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance, Cell Metab., 21, 443–54.

72. Lecrenier, N., Foury, F. 2000, New features of mitochondrial DNA replication system in yeast and man, Gene, 246, 37–48.

73. Weiller, G., Schueller, C.M.E., Schweyen, R.J. 1989, Putative target sites for mobile G+C rich clusters in yeast mitochondrial DNA: single elements and tandem arrays, Mol. Gen. Genet., 218, 272–83.

74. de Zamaroczy, M., Bernardi, G. 1986a, The GC clusters of the mitochondrial genome of yeast and their evolutionary origin, Gene, 41, 1–22.

75. Christianson, T., Edwards, J., Levens, D., Locker, J., Rabinowitz, M. 1982, Transcriptional initiation and processing of the small ribosomal RNA of yeast mitochondria, J. Biol. Chem., 257, 6494–500.

76. Osinga, K.A., De Vries, E., Van der Horst, G.T., Tabak, H.F. 1984a, Initiation of transcription in yeast mitochondria: analysis of origins of replication and of genes coding for a messenger RNA and a transfer RNA, Nucleic Acids Res., 12, 1889–900.

77. Schäfer, B. 2005, RNA maturation in mitochondria of S. cerevisiae and S. pombe, Gene, 354, 80–5.

78. Osinga, K.A., De Vries, E., Van der Horst, G., Tabak, H.F. 1984b, Processing of yeast mitochondrial messenger RNAs at a conserved dodecamer sequence, EMBO J., 3, 829–34.

79. Hofmann, T.J., Min, J., Zassenhaus, H.P. 1993, Formation of the 3′ end of yeast mitochondrial mRNAs occurs by site-specific cleavage two bases downstream of a conserved dodecamer sequence, Yeast, 9, 1319–30.

80. Friel, K.C., Friedrich, A., Hou, J., Schacherer, J. 2014, Population genomic analysis reveals highly conserved mitochondrial genomes in the yeast species Lachancea thermotolerans, Genome Biol. Evol., 6, 2586–94.

81. Wu, B., Buljc, A., Hao, W. 2015, Extensive horizontal transfer and homologous recombination generate highly chimeric mitochondrial genomes in yeast, Mol. Biol. Evol., 32, 2539–70.

82. Gaillardin, C., Neurgeise, C., Kerscher, S., Nicaud, J.M. 2012, Mitochondrial genomes of yeasts of the Yarrowia clade, FEMS Yeast Res., 12, 317–31.

83. Bouchier, C., Ma, L., Creno, S., Dujon, B., Fairhead, C. 2009, Complete mitochondrial genome sequences of three Nakaseomyces species reveal invasion by palindromic GC clusters and considerable size expansion, FEMS Yeast Res., 9, 1283–92.

84. Procházková, E., Poláková, S., Piškur, J., Sulo, P. 2010, Mitochondrial genome of the facultative anaerobe and petite-positive yeast Dekkera bruxellensis contains the NADH dehydrogenase subunit genes, FEMS Yeast Res., 10, 545–57.

85. Clark-Walker, G.D. 1992, Evolution of mitochondrial genomes in fungi, Int. Rev. Cytol., 141, 89–127.

86. Hittinger, C.T. 2013, Saccharomyces diversity and evolution: a budding model genus, Trends Genet., 29, 309–17.

87. Kurtzman, C.P., Mateo, R.Q., Kolecka, A., Theelen, B., Robert, V., Boukhout, T. 2015, Advances in yeast systematics and phylogeny and their use as predictors of biotechnologically important metabolic pathways, FEMS Yeast Res., 15, 1–17.

88. Peris, D., Langdon, Q.K., Moriarty, R.V., et al. 2016, Complex ancestries of lager-brewing hybrids were shaped by standing variation in the wild yeast Saccharomyces uvarum, PLoS Genet., 12, 1–20.

89. Rainieri, S., Kodama, Y., Nakao, Y., Pulvirenti, A., Giudici, P. 2008, The inheritance of mtDNA in lager brewing strains, FEMS Yeast Res., 8, 86–96.

90. Almeida, P., Gonçalves, C., Teixeira, S., et al. 2014, A Gondwanan imprint on global diversity and domestication of wine and cider yeast Saccharomyces uvarum, Nat. Commun., 5, 1–12.

91. Peris, D., Sylvester, K., Libkind, D., Gonçalves, P., Sampaio, J.P., Alexander, W.G., Hittinger, C.T. 2014, Population structure and reticulate evolution of Saccharomyces eubayanus and its lager-brewing hybrids, Mol. Ecol., 23, 2031–45.