

Phylogenetic relationships among yeasts of the ‘*Saccharomyces* complex’ determined from multigene sequence analyses

Cletus P. Kurtzman *, Christie J. Robnett

Microbial Genomics and Bioprocessing Research Unit, National Center for Agricultural Utilization Research, Agricultural Research Service,
U.S. Department of Agriculture, 1815 N. University Street, Peoria, IL 61604, USA

Received 2 October 2002; received in revised form 16 December 2002; accepted 17 December 2002

First published online 12 February 2003

Abstract

Species of *Saccharomyces*, *Arxiozyma*, *Eremothecium*, *Hanseniaspora* (anamorph *Kloeckera*), *Kazachstania*, *Kluyveromyces*, *Pachytichospora*, *Saccharomycodes*, *Tetrapispora*, *Torulaspora*, and *Zygosaccharomyces*, as well as three related anamorphic species assigned to *Candida* (*C. castellii*, *C. glabrata*, *C. humilis*), were phylogenetically analyzed from divergence in genes of the rDNA repeat (18S, 26S, ITS), single copy nuclear genes (translation elongation factor 1 α , actin-1, RNA polymerase II) and mitochondrially encoded genes (small-subunit rDNA, cytochrome oxidase II). Single-gene phylogenies were congruent for well-supported terminal lineages but deeper branches were not well resolved. Analysis of combined gene sequences resolved the 75 species compared into 14 clades, many of which differ from currently circumscribed genera.

© 2003 Federation of European Microbiological Societies. Published by Elsevier Science B.V. All rights reserved.

Keywords: *Saccharomyces*; *Kluyveromyces*; Molecular systematics; Multigene phylogeny; Ribosomal DNA; Elongation factor 1 α ; Actin-1; RNA polymerase II; Cytochrome oxidase II

1. Introduction

The assignment of yeast species to genera and families has been based primarily on morphology of vegetative cells and sexual states, and secondarily on physiological responses on the ca. 60–80 fermentation and growth tests commonly used in yeast systematics. Application of gene sequence analyses to yeast systematics has shown conflict between the placement of species on gene trees and their classification from phenotype. For example, the genus *Wingea* had been described because of the uniqueness of its lenticular ascospores, but phylogenetic analysis of rRNA sequences has placed it in the genus *Debaryomyces*, where species are generally characterized by roughened, spheroidal ascospores [1]. There is now a widespread pattern of disparity between phenotype and genotype as means for classifying yeasts, and these differences have been demonstrated from analyses of 18S ribosomal DNA (rDNA) [2–4], internal transcribed spacer (ITS)

rDNA [5], 26S rDNA [6,7], and cytochrome oxidase II (COX II) [8], leaving little doubt that phenotype is a poor predictor of genetic relationships among species. At present, most molecular comparisons have been made from single genes and it is unclear whether the relationships portrayed are representative of the evolutionary history of the organisms themselves. Additionally, basal branches in phylogenetic trees derived from single genes are often only weakly supported, resulting in uncertain relationships among more divergent species. This is particularly evident in studies of the ‘*Saccharomyces* complex’ where 18S rDNA [2,4] and 26S rDNA [7] comparisons show that many genera are incorrectly circumscribed, but these individual datasets strongly resolve only the most closely related species, leaving less closely related species with uncertain placement.

In the present study, we have compared species of *Saccharomyces*, *Kluyveromyces* and other genera associated with these two taxa, using the combined sequences from groups of genes that are believed to be unlinked. These groups include genes of the rDNA repeat (18S, 26S, ITS), single-copy nuclear genes (translation elongation factor 1 α (EF-1 α), actin-1, RNA polymerase II) and mitochondrially encoded genes (small-subunit rDNA, COX II). The spe-

* Corresponding author. Tel.: +1 (309) 681-6561;

Fax: +1 (309) 681-6672.

E-mail address: kurtzman@ncaur.usda.gov (C.P. Kurtzman).

cies included in our study appear to be a natural group broadly defined earlier from analyses of 18S rDNA [2,4] and partial 26S rDNA [7]. The latter study included all ascomycetous yeasts (anamorphs and teleomorphs) then known and showed the ‘*Saccharomyces* complex’ to include, in addition to *Saccharomyces*, the genera *Arxiozyma*, *Eremothecium*, *Hanseniaspora* and its anamorph *Kloeckera*, *Kazachstania*, *Kluyveromyces*, *Pachytichospora*, *Saccharomycodes*, *Tetrapisispora*, *Torulaspora*, *Zygosaccharomyces* and three anamorphic species assigned to the genus *Candida*. Most of the genera reproduce vegetatively by multilateral budding, as is typical of *Saccharomyces*. However, *Hanseniaspora* and *Saccharomycodes* reproduce by bipolar budding, and three of the species assigned to *Eremothecium* do not bud at all. In this study, we present multigene phylogenetic analyses that place species of the ‘*Saccharomyces* complex’ into well-supported clades (i.e., monophyletic groups).

2. Materials and methods

2.1. Organisms

The strains studied are listed in Table 1, and all are maintained in the ARS Culture Collection (NRRL), National Center for Agricultural Utilization Research, Peoria, IL, USA. Three recently described *Saccharomyces* species, i.e., *S. naganishii*, *S. humaticus*, and *S. yakushimaensis* [9], were not included in this work.

2.2. Growth of cultures and DNA isolation

Cells used for DNA extraction were grown for approximately 24 h at 25°C in 50 ml of Wickerham’s [10] YM broth (3 g yeast extract, 3 g malt extract, 5 g peptone, and 10 g glucose per liter of distilled water) on a rotary shaker at 200 rpm and harvested by centrifugation. The cells were washed once with distilled water, resuspended in 2 ml of distilled water and aliquoted to two 1.5-ml microcentrifuge tubes. After centrifugation, excess water was decanted from the microcentrifuge tubes, and the packed cells were freeze-dried for 1–2 days and stored at –20°C until used. DNA isolation for polymerase chain reaction (PCR) was performed using either the sodium dodecyl sulfate method of Raeder and Broda [11] or the CTAB (hexadecyltrimethyl-ammonium bromide) procedure, both of which were described in detail by Kurtzman and Robnett [7].

2.3. PCR and DNA sequencing reactions

Oligonucleotide primers for symmetrical amplification of gene sequences and sequencing of the genes compared are given in the following sections. Unless otherwise stated, symmetrical amplifications were performed for 36 PCR cycles with denaturation at 94°C for 1 min, anneal-

ing at 52°C, and extension at 72°C for 2 min, with the final extension for 10 min. Both strands of the DNAs compared were sequenced with the ABI TaqDyeDeoxy Terminator Cycle sequencing kit (Applied Biosystems, Foster City, CA, USA) using either ABI 377 (gel) or ABI 3100 (capillary) automated DNA sequencers, following the manufacturer’s instructions. Procedures for purification of DNA from reaction mixtures were given previously [7].

18S rDNA. Amplicons comprising the nearly entire 18S rRNA gene (nucleotides 22–1771; nucleotide locations here and elsewhere refer to *Saccharomyces cerevisiae*) were produced using primers NS-1 (5′-GTAGCATATGCTTGTCTC) (F, forward) and either NS-8 (5′-TCGCGAGGTTACCTACGGA) (R, reverse) or NS-8A (5′-CCTTCCGCGAGGTTACCTACGGAAACC) (R). The following sequencing primers were used: NS-1, NS-2 (5′-GGCTGCTGGCACCAGACTTGC) (R), NS-3 (5′-GCAAGTCTGGTGCCAGCAGCC) (F), NS-4 (5′-CTTCCGTCAATTCCTTTAAG) (R), NS-5B (5′-GGGTTCTGGGGGGAGTATGGTCGCAAGGC) (F), NS-6A (5′-GTCAGTGTAGCGCGCGTGC GGCCAG) (R), NS-6B (5′-GCAGACAAATCACTCCACCAAC) (R), NS-7A (5′-CTGGGCGCGACGCGCGCTACACTGAC) (F), NS-8 and NS-8A. GenBank accession numbers for the 18S rDNA sequences are the following: X77930, X83822–X83824, X84638, X84639, X89520, X89523, X89526–X89528, X90756, X91083–X91086, X97805–X97807, X98120, X98868, Z75577–Z75581, AF339889–AF339891, and AY046224–AY046272. Sequences with the prefixes X and Z are from the work of Cai et al. [2] and James et al. [4], AF is from Kurtzman et al. [12] and AY is from the present study.

26S rDNA. Three sections of the 26S rRNA gene were sequenced for all species and nearly the entire gene was sequenced for 18 species. Following are the regions sequenced and the primers used.

Domains 1 and 2 (nucleotides 63–642): primers NL-1 (5′-GCATATCAATAAGCGGAGGAAAAG) (F) and NL-4 (5′-GGTCCGTGTTTCAAGACGG) (R) were used for synthesis of amplicons and for sequencing. The following internal primers were used if needed: NL-2A (5′-CTTGTTTCGCTATCGGTCTC) (R) and NL-3A (5′-GAGACCGATAGCGAACAAG) (F). Many of the D1/D2 sequences used in this study were deposited earlier in GenBank [7]. Those new to this work have the following numbers: AF339888, AF339904, AF398478–AF398490, AY048172, and AY130336–AY130346.

Region 1498–2152. Various combinations of the following primers were used for symmetrical amplifications: NL-1, NL-E27R (5′-TGACGAGGCATTGGCTACC) (R), NL-13R (5′-GCGTTATCGTTTAAACAGATGTGCCG) (R), NL-8AF (5′-CACGTCAACAGCAGTTGGAC) (F), and NL-11BR (5′-GCTATGTTTAAATTAGACAGTCAG) (R). Sequencing primers were: NL-3A, NL-E27R, NL-5A (5′-GACTAATCGAACCATCTAGTAGCTGG)

Table 1

Species and strains compared and intraspecific nucleotide differences for genes of selected taxa

| Species ^a | Accession numbers ^{b,c} | | Genes ^{d,e} | | | | | |
|---|----------------------------------|------------|----------------------|------|------|---------------|---------------------|--------|
| | NRRL | Other | 26S D1/D2 | ITS1 | ITS2 | EF-1 α | mtSm rDNA | COX II |
| <i>Arxiozyma telluris</i> | YB-4302 ^T | CBS 2685 | | | | | | |
| <i>Candida castellii</i> | Y-17070 ^T | CBS 4332 | | | | | | |
| <i>C. glabrata</i> | Y-65 ^T | CBS 138 | | | | | | |
| | Y-17815 | | 1 | 3 | 3 | | | |
| <i>C. humilis</i> | Y-17074 ^T | CBS 5658 | | | | | | |
| (<i>C. milleri</i> ^T) | Y-7245 | CBS 6897 | 1 | 4 | 6 | 11 | 0 | 0 |
| <i>Eremothecium ashbyi</i> | Y-1363 ^A | | | | | | | |
| | Y-7249 | | 1 | 2 | 1 | | | |
| <i>E. (Nematospora) coryli</i> | Y-12970 ^T | CBS 2608 | | | | | | |
| | Y-1618 | CBS 2599 | 0 | 0 | 3 | | | |
| <i>E. cymbalariae</i> | Y-17582 ^A | CBS 270.75 | | | | | | |
| <i>E. (Ashbya) gossypii</i> | Y-1056 ^A | CBS 109.51 | | | | | | |
| | Y-1810 | | 0 | 0 | 0 | 0 | 0 | |
| <i>E. (Holleya) sinecaudum</i> | Y-17231 ^T | CBS 8199 | | | | | | |
| <i>Hanseniaspora guilliermondii</i> | Y-1625 ^T | CBS 465 | | | | | | |
| <i>H. (Kloeckeraspora) occidentalis</i> | Y-7946 ^T | CBS 2592 | | | | | | |
| | YB-4040 | | 0 | 0 | 2 | | | |
| <i>H. (Kloeckeraspora) osmophila</i> | Y-1613 ^T | CBS 313 | | | | | | |
| <i>H. uvarum</i> | Y-1614 ^T | CBS 314 | | | | | | |
| | Y-1612 | CBS 287 | 1 | 0 | 0 | | | |
| <i>H. valbyensis</i> | Y-1626 ^T | CBS 479 | | | | | | |
| | Y-7575 | CBS 6558 | 0 | | | | | |
| <i>H. (Kloeckeraspora) vineae</i> | Y-17529 ^T | CBS 2171 | | | | | | |
| <i>Kazachstania viticola</i> | Y-27206 ^T | CBS 6463 | | | | | | |
| <i>Kloeckera lindneri</i> | Y-17531 ^T | CBS 285 | | | | | | |
| <i>Kluyveromyces aestuarii</i> | YB-4510 ^T | CBS 4438 | | | | | | |
| <i>K. africanus</i> | Y-8276 ^T | CBS 2517 | | | | | | |
| <i>K. bacillisporus</i> | Y-17846 ^T | CBS 7720 | | | | | | |
| <i>K. blattae</i> | Y-10934 ^T | CBS 6284 | | | | | | |
| <i>K. delphensis</i> | Y-2379 ^T | CBS 2170 | | | | | | |
| <i>K. dobzhanskii</i> | Y-1974 ^T | CBS 2104 | | | | | | |
| <i>K. lactis</i> var. <i>lactis</i> | Y-8279 ^T | CBS 683 | | | | | | |
| | Y-7618 | | 0 | 1 | 1 | | | |
| var. <i>drosophilae</i> | Y-8278 ^T | CBS 2105 | 0 | 1 | 1 | 5 | 12(13) ^f | |
| (<i>S. vanudenii</i> ^T) | Y-6682 | CBS 4372 | 0 | | | | | |
| <i>K. lodderae</i> | Y-8280 ^T | CBS 2757 | | | | | | |
| <i>K. marxianus</i> | Y-8281 ^T | CBS 712 | | | | | | |
| | Y-2415 | CBS 397 | 0 | 0 | 0 | | | |
| <i>K. nonfermentans</i> | Y-27343 ^T | JCM 10232 | | | | | | |
| <i>K. piceae</i> | Y-17977 ^T | CBS 7738 | | | | | | |
| <i>K. polysporus</i> | Y-8283 ^T | CBS 2163 | | | | | | |
| <i>K. sinensis</i> | Y-27222 ^T | CBS 7660 | | | | | | |
| <i>K. thermotolerans</i> | Y-8284 ^T | CBS 6340 | | | | | | |
| | Y-2233 | CBS 137 | 0 | 0 | 0 | | | |
| <i>K. waltii</i> | Y-8285 ^T | CBS 6430 | | | | | | |
| <i>K. wickerhamii</i> | Y-8286 ^T | CBS 2745 | | | | | | |
| | YB-1490 | | 0 | 0 | 0 | | 0 | |
| <i>K. yarrowii</i> | Y-17763 ^T | CBS 8242 | | | | | | |
| <i>Saccharomyces barnettii</i> | Y-27223 ^T | CBS 6946 | | | | | | |
| | Y-17918 | CBS 5648 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Y-27227 | CBS 8515 | 0 | 0 | 0 | | | |
| <i>S. bayanus</i> | Y-12624 ^T | CBS 380 | | | | | | |
| (<i>S. uvarum</i> ^T) | Y-17034 | CBS 395 | 1 | 2 | 0 | | 0 | 0 |
| <i>S. bulderi</i> | Y-27203 ^T | CBS 8638 | | | | | | |
| | Y-27204 | CBS 8639 | 0 | 0 | 0 | | | |
| | Y-27205 | | 0 | 0 | 0 | | | |
| <i>S. cariocanus</i> | Y-27337 ^T | NCYC 2890 | | | | | | |
| | Y-27338 | N50791-2D | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>S. castellii</i> | Y-12630 ^T | CBS 4309 | | | | | | |
| | Y-12631 | CBS 4310 | 0 | 0 | 0 | | 0 | |
| | ETC NY-51 | | 0 | 0 | 1 | | | |

Table 1 (Continued).

| Species ^a | Accession numbers ^{b,c} | | Genes ^{d,e} | | | | | |
|--|----------------------------------|-----------|----------------------|--------------------|-------------------|---------------|-------------------|--------|
| | NRRL | Other | 26S D1/D2 | ITS1 | ITS2 | EF-1 α | mtSm rDNA | COX II |
| <i>S. cerevisiae</i> | Y-12632 ^{NT} | CBS 1171 | | | | | | |
| | Y-1375 | | 0 | 2(4) ^f | 1(2) ^f | 0 | | |
| | Y-2034 | | 0 | 0 | 0 | 0 | 6(7) ^f | 5 |
| (<i>S. diastaticus</i>) | Y-2416 | CBS 1636 | 0 | 0 | 1 | 0 | 3 | 18 |
| (<i>S. beticus</i> ^T) | Y-12625 | CBS 6203 | 2 | 2(26) ^f | 2(8) ^f | 1 | 4(5) ^f | 13 |
| (<i>S. italicus</i> ^T) | Y-12649 | CBS 459 | 0 | 1 | 0 | 0 | 5(6) ^f | 18 |
| | YB-210 = Y-1350 | CBS 6333 | 0 | 0 | 1 | 0 | 4 | 0 |
| <i>S. dairenensis</i> | Y-12639 ^T | CBS 421 | | | | | | |
| <i>S. exiguus</i> | Y-12640 ^{NT} | CBS 379 | | | | | | |
| <i>S. kluyveri</i> | Y-12651 ^T | CBS 3082 | | | | | | |
| | YB-4288 | CBS 6545 | 0 | 0 | 0 | | | |
| <i>S. kudriavzevii</i> | Y-27339 ^T | IFO 1802 | | | | | | |
| | Y-27340 | IFO 1803 | 5 | 2 | 0 | 4 | 2 | 19 |
| <i>S. kunashirensis</i> | Y-27209 ^T | CBS 7662 | | | | | | |
| <i>S. martiniae</i> | Y-409 ^T | CBS 6334 | | | | | | |
| <i>S. mikatae</i> | Y-27341 ^T | IFO 1815 | | | | | | |
| | Y-27342 | IFO 1816 | 0 | 0 | 0 | 0 | 1 | 11 |
| <i>S. paradoxus</i> | Y-17217 ^{NT} | CBS 432 | | | | | | |
| | Y-1548 | CBS 406 | 0 | 0 | 0 | | 0 | |
| <i>S. pastorianus</i> | Y-27171 ^{NT} | CBS 1538 | | | | | | |
| | Y-1551 | | 0 | 0 | 0 | 3 | 0 | |
| (<i>S. monocensis</i> ^T) | Y-1525 | CBS 1503 | 0 | 0 | 0 | 0 | 1 | 0 |
| (<i>S. carlsbergensis</i> ^T) | Y-12693 | CBS 1513 | 0 | 0 | 0 | 2 | 1 | 0 |
| | Y-27172 | CBS 1542 | 0 | 0 | 0 | 3 | 0 | 0 |
| <i>S. rosinii</i> | Y-17919 ^T | CBS 7127 | | | | | | |
| <i>S. servazzii</i> | Y-12661 ^T | CBS 4311 | | | | | | |
| <i>S. spencerorum</i> | Y-17920 ^T | CBS 3019 | | | | | | |
| <i>S. (Pachytichospora) transvaalensis</i> | Y-17245 ^T | CBS 2186 | | | | | | |
| <i>S. turicensis</i> | Y-27345 ^T | CBS 8665 | | | | | | |
| <i>S. unisporus</i> | Y-1556 ^T | CBS 398 | | | | | | |
| | Y-1565 | CBS 399 | 0 | 0 | 0 | | | |
| <i>Saccharomyces ludwigii</i> | Y-12793 ^T | CBS 821 | | | | | | |
| | Y-12860 | CBS 820 | 0 | 3 | 0 | | 0 | |
| <i>Tetrapisispora arboricola</i> | Y-27308 ^T | IFO 10925 | | | | | | |
| <i>T. iriomotensis</i> | Y-27309 ^T | IFO 10929 | | | | | | |
| <i>T. nanseiensis</i> | Y-27310 ^T | IFO 10899 | | | | | | |
| <i>T. phaffii</i> | Y-8282 ^T | CBS 4417 | | | | | | |
| <i>Torulaspora delbrueckii</i> | Y-866 ^T | CBS 1146 | | | | | | |
| (<i>T. rosei</i> ^{NT}) | Y-1567 | CBS 817 | 0 | 0 | 0 | | | |
| <i>T. franciscae</i> | Y-17532 ^T | CBS 2926 | | | | | | |
| <i>T. globosa</i> | Y-12650 ^T | CBS 764 | | | | | | |
| <i>T. pretoriensis</i> | Y-17251 ^T | CBS 2187 | | | | | | |
| <i>Zygosaccharomyces bailii</i> | Y-2227 ^T | CBS 680 | | | | | | |
| | Y-787 | | 0 | 6 | 1 | | | |
| <i>Z. bisporus</i> | Y-12626 ^T | CBS 702 | | | | | | |
| | Y-7253 | | 0 | 1 | 0 | | | |
| <i>Z. cidri</i> | Y-12634 ^T | CBS 4575 | | | | | | |
| | Y-12635 | CBS 2950 | 0 | 0 | 0 | | | |
| <i>Z. fermentati</i> | Y-1559 ^T | CBS 707 | | | | | | |
| | Y-7434 | CBS 4506 | 0 | 0 | 0 | | | |
| | Y-11844 | CBS 7004 | 0 | 0 | 0 | | | |
| | Y-11847 | | 0 | 0 | 0 | | | |
| | Y-12620 | CBS 4686 | 0 | 0 | 0 | | | |
| | Y-17054 | CBS 6544 | 0 | 0 | 0 | | | |
| | Y-17055 | CBS 6711 | 0 | 0 | 0 | | | |
| <i>Z. florentinus</i> | Y-1560 ^T | CBS 746 | | | | | | |
| | Y-12642 | CBS 6081 | 1 | 0 | 0 | | | |
| <i>Z. kombuchaensis</i> | YB-4811 ^T | CBS 8849 | | | | | | |
| | YB-4810 | | 0 | 0 | 0 | 1 | 2 | 0 |
| | Y-27162 | | 0 | 0 | 0 | | | |
| | Y-27163 | CBS 8850 | 0 | 0 | 0 | | | |

Table 1 (Continued).

| Species ^a | Accession numbers ^{b,c} | | Genes ^{d,e} | | | | | | |
|----------------------------------|----------------------------------|----------|----------------------|------|-------------------|-------|-----------|--------|--|
| | NRRL | Other | 26S D1/D2 | ITS1 | ITS2 | EF-1α | mtSm rDNA | COX II | |
| <i>Z. lentus</i> | Y-27276 ^T | CBS 8574 | | | | | | | |
| | Y-27275 | CBS 8517 | 0 | 1 | 0 | | | | |
| <i>Z. mellis</i> | Y-12628 ^T | CBS 736 | | | | | | | |
| | Y-1024 | | 0 | 0 | 0 | | | | |
| <i>Z. microellipsoides</i> | Y-1549 ^T | CBS 427 | | | | | | | |
| | Y-17058 | CBS 6142 | 0 | | | | | | |
| <i>Z. mrakii</i> | Y-12654 ^T | CBS 4218 | | | | | | | |
| | Y-12655 | CBS 4219 | 0 | 0 | 0 | | | | |
| <i>Z. rouxii</i> | Y-229 ^T | CBS 732 | | | | | | | |
| | Y-55 | | 0 | 0 | 2(5) ^f | | | | |
| | Y-998 | | 0 | 1 | 2(5) ^f | 3 | 0 | 0 | |
| | Y-1481 | | 0 | 0 | 2(5) ^f | | | | |
| | Y-12691 | | 0 | | | | 0 | 1 | |
| | Y-27326 | | 0 | 1 | 2(5) ^f | | | | |
| | YB-3050 | | 0 | 1 | 0 | | | | |
| | ETC RY-208 | | 0 | 1 | 0 | | | | |
| Reference species | | | | | | | | | |
| <i>Pichia anomala</i> | Y-366 ^{NT} | CBS 5759 | | | | | | | |
| <i>Schizosaccharomyces pombe</i> | Y-12796 ^T | CBS 356 | | | | | | | |
| <i>Neurospora crassa</i> | 13141 ^A | | | | | | | | |

^a Commonly recognized synonym names are given in parentheses.

^b T = type strain, NT = neotype strain, A = authentic strain, the reference strain used when there is no living type or ex-type strain.

^c NRRL = ARS Culture Collection, National Center for Agricultural Utilization Research, Peoria, IL, USA; CBS = Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands; JCM = Japan Collection of Microorganisms, Saitama, Japan; IFO = Institute for Fermentation, Osaka, Japan; NCYC = National Collection of Yeast Cultures, Norwich, UK; N = G.H. Naumov State Institute for Genetics and Selection of Industrial Microorganisms, Moscow, Russia.

^d 26S D1/D2 = domains 1 and 2, large-subunit (26S) ribosomal DNA (rDNA); ITS1, ITS2 = internal transcribed spacers 1 and 2, rDNA; EF-1 α = translation elongation factor 1 α ; mtSm = mitochondrial small-subunit rDNA; COX II = cytochrome oxidase II.

^e Numbers indicate nucleotide differences with the type strain.

^f Indels that are comprised of two or more contiguous nucleotides are treated as single events; e.g., NRRL Y-1375, ITS1 is 2(4) and indicates two indel sites, one of which has three contiguous insertions for a total of four nucleotide differences with the type strain.

(F), NL-5AR (5'-CCAGCTACTAGATGGTTCGATTAGTC) (R), NL-7AR (5'-CCGACTTCCATGGCCACCGTCC) (R), NL-8AF, NL-9 (5'-GGAGACGTCGCGCGGAGCCCTGG) (F), NL-11 (5'-ACTTAG-AACTGGTACGGACAAGG) (F), NL-1611AR (5'-CACCTTGGA-GACCTGCTGCGG) (R), NL-8/11AF (5'-GTAAGTTCGGGATAAGGATTGG) (F), NL-8/11BR (5'-GAGCCAATCCTTATCCCGAAG) (R), and NL-11BR. GenBank accession numbers for this region are AF399761–AF399820.

Region 2963–3341. The following primers were used in various combinations for symmetrical amplifications: NL-1611F (5'-CCGCAGCAGGTCTCCAA) (F), NL-G19B (5'-GACCGTCGTGAGACAGGTTAG) (F), ETS-2AR (5'-GATCGTAACAACAAGGCTACTCTA) (R), ETS-2DR (5'-CTGCTTACAATACCCCGTTGTAC) (R), and ETS-2FR (5'-CTGGCTTAGAGGCGTTTCAGCC) (R). Sequencing primers were NL-E27 (5'-GGTAGCCAAATGCCTCGTCA) (F), NL-G19A (5'-GGGAACGTGAGCTGGGTTTAGACCG) (F), NL-G19B, NL-1611AF (5'-CCGCAGCAGGTCTCCAAGGTG) (F), NL-11R (5'-CCTTGTCCGTACCAGTTCTAAGT) (R), NL-13R (5'-GCGTTATCGTTTAAACAGATGTGCCG) (R), ETS-2DR, and ETS-2FR. GenBank accession numbers for this region are AY046086–AY046145.

26S rRNA gene, nucleotides 63–3341. The nearly entire gene was sequenced for 18 species and was amplified in two halves using the primer pairs NL-1+NL-E27R and NL-1611F+ETS-2AR. Sequencing primers were: NL-3A, NL-5A, NL-5AR, NL-7AR, NL-9, NL-9A (5'-GGA-GACGTCGCGCGAGCCCTGGGAGG) (F), NL-1611AR, NL-11, NL-E27R, NL-1611AF, NL-11R, NL-E27 (5'-GGTAGCCAAATGCCTCGTCA) (F), NL-13R, NL-G19A, and ETS-2FR. GenBank accession numbers for these sequences are AY048154–AY048171.

ITS1–5.8S–ITS2 rDNA. Primers for symmetrical amplifications were various combinations of the following: ITS-5 (5'-GGAAGTAAAAGTCGTAACAAGG) (F), ITS-4 (5'-TCCTCCGCTTATTGATATGC) (R), NS-7A, NL-2A, NL-4. Temperatures for PCR were 52°C annealing and 72°C extension or 39°C annealing and 60°C extension. Sequencing primers were ITS-5, ITS-2 (5'-GCTGC-GTTCTTCATCGATGC) (R), ITS-3 (5'-GCATCGATGAAGAACGCAGC) (F) and ITS-4. GenBank accession numbers for these sequences are AY046146–AY046221 and AY130303–AY130313.

Mitochondrial small-subunit rDNA. Primers for symmetrical amplifications were various combinations of the following, which also were used for the sequencing reactions. MS-1 (5'-CAGCAGTCAAGAATATTAGTCAATG) (F),

MS-1A (5'-CGAAAGATTGATCCAGTTA) (F), MS-1B (5'-GATTGATCCAGTTACTTATTAG) (F), MS-1D (5'-GTGCCAGCAGTCGCGGT) (F), MS-2 (5'-GCGGAT-TATCGAATTAAATAAC) (R), MS-2A (5'-CCGTCT-ATTGTTTTGAGTTTCA) (R), and MS-2B (5'-ACTAC-TCGGGTATCGAAT) (R). Temperatures for symmetrical amplifications were either 50°C annealing and 72°C extension or 39°C annealing and 60°C extension. A few of the sequencing reactions used 42°C annealing and 45°C extension, in contrast to the standard temperatures of 50°/60°C. GenBank accession numbers for these sequences are AF442281–AF442356 and AY130314–AY130324. Sequences were not obtained for *S. bulderi*, *S. rosinii*, all species of *Hanseniaspora*, *Kloeckera* and *Saccharomyces ludwigii*.

COX II. Primers for symmetrical amplifications were various combinations of the following, which were also used for the sequencing reactions. The sequences for COII-3 and COII-5 were given by Belloch et al. [8]; sequences for the other primers listed were developed in the present work. COII-5 (5'-GGTATTTTAGAATTACATGA) (F), COII-5A (5'-GTTTTATTTATTAGTTATTTTAGG) (F), COII-5B (5'-ATTAGTTATTTTAGGTTTAG) (F), COII-3 (5'-ATTTATTGTTCTTTAATCA) (R), COII-3A (5'-GCAGAACTTGATTTAATCTACC) (R), and COII-3B (5'-CCTTCTCTTTGAATTAATGC) (R). Temperatures for symmetrical amplification were either 45°C annealing and 72°C extension or 39°C annealing and 60°C extension. GenBank accession numbers for these sequences are AF442206–AF442280 and AY130325–AY130335. Sequences were not obtained for *S. bulderi*, *H. guilliermondii* and *H. uvarum*.

Translation EF-1 α A, nucleotides 64–1190 [13]. Primers for symmetrical amplifications were the following, which were used in various combinations: YTEF-1 (5'-GGTCAYGTYGAYKCTGGTAAGT) (F), YTEF-1G (5'-GGTAAGGGTTCTTTCAAGTACGCTTGGG) (F), YTEF-6A (5'-GGTABTCRSTGAARGYYTCAACRGACA) (R), and YTEF-6G (5'-CGTTCTTGAGTCACCA-CAGACGTTACCTC) (R). Primers for sequencing reactions were the following: YTEF-1, YTEF-1C (5'-CAAGTGCGGTGGTATTGACAAGCGTAC) (F), YTEF-1E (5'-CAAGTGTTGGTGGTATTGACAAGAGAACY-ATC) (F), YTEF-1G (5'-GGTAAGGGTTCTTTCAAGTACGCTTGGG) (F), YTEF-2B (5'-CCTTCTTGATRAAGTTGGAGGTYTCCTT) (R), YTEF-2C (5'-CAAT-CATGTTGTCACCGTTCCAACC) (R), YTEF-2E (5'-CAATCATGTTGTCACCGTTCC) (R), YTEF-3A (5'-CCAACTTTATCAAGAAGGTTGGTT) (F), YTEF-3B (5'-GGTTGGAACGGTGACAACATGATTG) (F), YTEF-4 (5'-TCRACGGACTTGACTTCAGTGGTVACACC) (R), YTEF-4B (5'-GTCACCACAGACGTTACCTCTTC) (R), YTEF-5D (5'-CCGGTGTCA-TCAAGCCAGGTATG) (F), YTEF-5E (5'-CCAGG-TGACAACGTTGGTTTCAACG) (F), YTEF-5F (5'-CCCGTCGGTCGTGTCGAGACTGGTGTGCATC) (F),

YTEF-6A, YTEF-6D (5'-CCGGACTTCAAGAACTTTGGATGGTC) (R), and YTEF-6G (5'-CGTTCTTGAGTCACCACAGACGTTACCTC) (R). GenBank accession numbers for these sequences are AF402004–AF402094, AY130808, and AY130810–AY130813.

Actin-1, nucleotides 783–1680. Initial primers were from the study of Daniel et al. [14] followed by development of group-specific primers for the species in the present study. The following are the primers for symmetrical amplifications, which were used in various combinations. Temperature protocols for PCR ranged from 52, 42, and 39°C for annealing, to 72, 45, and 42°C for extension. CA1 (5'-GCGGTGACGACGCTCCAAGAGCTG) (F), CA2R (5'-CCGTGTTCAATTGGGTATCTCAAGGTC) (R), and ACT3R (5'-GAACCACCAATCCAGACGGAG) (R). Sequencing primers were: CA1, CA2 (5'-GACCTTGAGATACCAATTGAACACGG) (F), CA2R, ACT1BF (5'-CAAGGTATCATGGTCGGTATGGG) (F), ACT2F (5'-GAACGTGGTTACTCTTTCTC) (F), ACT2R (5'-GAGAAAGAGTAACCACGTC) (R), ACT2AF (5'-CAAGAAATGCAAACCGCTGCTCAATC) (F), ACT2AR (5'-GATTGAGCAGCGGTTTGCATTCTTG) (R), ACT3R, and ACT3AR (5'-GGAGCAATGATCTTGACCTTC) (R). Temperatures for sequencing reactions were either 50°C annealing and 60°C extension or 42°C annealing and 45°C extension. Actin sequences were determined only for Clade 2 species and outgroup taxa. GenBank accession numbers for the sequences are AF527913–AF527941.

RNA polymerase II (RPB2), nucleotides 2447–3127 [15]. Initial primers were from the work of Liu et al. [16] followed by development of additional primers specific for species in the present study. The following are primers for symmetrical amplifications and were used in various combinations. Primer designations ending in F are forward and those with an R are reverse. Thermal cycling temperatures were 39 or 42°C for annealing and 42 or 45°C for extension. RPB2-6F (5'-TGGGGKWTGGTYTGYCCTGC), RPB2-6AF (5'-GTTGGTACAGATCCGATGCC), fRPB2-7cR (5'-CCCATRGCCTTGTTTCCAT), and RPB2-7cAR (5'-CACCCATAGCTTGC-TTACC). Sequencing primers included the following, which were cycled at 42°C annealing and 45°C extension. RPB2-6F, RPB2-6AF, RPB2-6CF (5'-GGTGATATCAATCCGGAAG), RPB2-6CR (5'-CTTCCGGATTGATATCACC), RPB2-6DF (5'-GATGCCGGTAGAGTTTATAGACC), RPB2-6DR (5'-GGTCTATAAACTCTACCGGCATC), fRPB2-7cR, and RPB2-7cAR. Temperatures for sequencing reactions were 42°C annealing and 45°C extension. RPB2 sequences were determined only for Clade 2 species and outgroup taxa. GenBank accession numbers for the sequences are AF527884–AF527912.

2.4. Phylogenetic analyses of gene sequences

Estimates of phylogenetic relatedness among species

were determined using the maximum parsimony (MP), neighbor-joining (NJ) and maximum likelihood (ML) programs of PAUP* 4.063a [17]. Because of computational constraints, ML analyses were conducted only for species in Clade 2. The Kimura-2 parameter distance correction was used for NJ analyses, which gave phylogenetic trees similar to all other distance correction matrices available in the PAUP package. Maximum likelihood was used with the defaults provided as well as with the gamma correction determined from the Clade 2 dataset.

Phylogenetic analyses of individual gene sequences are presented as bootstrap consensus trees to illustrate the statistical strength of each gene tree. Bootstrap support for all phylogenetic trees was determined from 1000 replications. Third-position nucleotides for codons of the protein-encoding genes EF-1 α , actin-1 and RNA polymerase II appeared saturated when viewed across the entire dataset, and inclusion of third-position nucleotides in analyses sometimes resulted in non-congruence for divergent species. An alternative was to translate the gene sequences into amino acid sequences, but this resulted in significant loss of resolution. To resolve this problem, the third position of each codon was deleted, resulting in stronger congruence among gene trees. One exception to deleting the third position was the EF-1 α comparison of species in the *S. cerevisiae* clade. Genetic distances are small and there was no indication of third-position saturation. *Pichia anomala* (hemiascomycete clade), *Neurospora crassa* (euascomycete clade), and *Schizosaccharomyces pombe* (archiascomycete clade) were tested as outgroup species, but choice of outgroup had little effect on most phylogenetic trees.

Analyses were conducted on datasets that included all nucleotides of each gene sequence as well as on datasets in which nucleotide regions of possible uncertain alignments were removed. While there were relatively few topological differences between the two treatments, the impact of deleting these areas was to reduce the 18S dataset from 1806 to 1647 positions and the 26S D1/D2 dataset from 624 to 452 positions, resulting in some loss of resolution between closely related taxa, but increased stability for deeper lineages. Consequently, for single-gene analyses, entire sequences were included because closely related species offered no alignment uncertainties, but for analyses of combined genes that examined deeper lineages, the pruned datasets were used.

Individual gene sequence datasets were combined to increase phylogenetic signal. Combining data is expected to increase phylogenetic accuracy by increasing signal and dispersing noise [18], and any informational conflicts between genes would not be expected to increase statistical support for affected nodes [19]. Datasets combined in the present study gave essentially the same terminal relationships as when analyzed individually, and progressive addition of gene sequences resulted in a progressive increase in

bootstrap support across the entire phylogenetic tree, indicating no significant internal conflicts.

3. Results and discussion

3.1. Intraspecific gene divergence

Table 1 lists the species examined and includes a comparison of intraspecific nucleotide divergence among genes of selected species. For most species, strains showing no nucleotide differences with the type strain in the D1/D2 domain of 26S rDNA also had few or no nucleotide differences in the other genes compared. However, there are several exceptions. Some strains of *S. cerevisiae* and *Zygosaccharomyces rouxii* had short deletions in their ITS genes (Table 1), which could result in incorrect species designations if only ITS sequences were used for identification. Another disparity concerns NRRL Y-7245, the type strain of *Candida milleri*, which appears conspecific with *C. humilis* on the basis of little or no divergence in 26S D1/D2, mitochondrial small-subunit rDNA and COX II, but shows what appear to be species level differences in ITS1, ITS2 and EF-1 α . One explanation is that this strain may represent a partial hybrid derived from *C. humilis* and an as yet unrecognized sister species. In view of this, *C. milleri* may represent a distinct species.

Other intraspecific differences of note are the large number of nucleotide substitutions in the COX II gene among strains of *S. cerevisiae*, *S. kudriavzevii* and *S. mikatae* (Fig. 1). These three species, along with *S. bayanus*, *S. cariocanus* and *S. paradoxus*, are heterothallic and closely related, but Naumov et al. [20] have reported that they are reproductively isolated from one another. An additional species of this clade, *S. pastorianus*, was considered a hybrid of *S. cerevisiae* and *S. bayanus* by Naumov et al. [20], as proposed by Vaughan Martini and Kurtzman [21]. Despite the apparent reproductive isolation of the six biological species of the *S. cerevisiae* complex, present data suggest a relatively high rate of recombination in the COX II gene for *S. cerevisiae*, *S. paradoxus* and *S. mikatae*. Diversity in the COX II gene has been shown by Fischer et al. [22] for *S. kudriavzevii* mating types and here it is seen for additional genes as well (Fig. 1). Groth et al. [23,24] have demonstrated from nuclear and mitochondrial gene sequences that interspecific hybrids are produced among *Saccharomyces* species and that they survive in nature. In other molecular comparisons, Nguyen et al. [25] have reported that the type strain of *S. bayanus* appears to be a natural hybrid of *S. cerevisiae* and *S. uvarum*, the latter species often being regarded as conspecific with *S. bayanus*. Our results are consistent with the observations that genetically intermediate forms are found among species of the *S. cerevisiae* complex and that some strains of these biological species might themselves be hybrids as suggested by Pedersen [26].

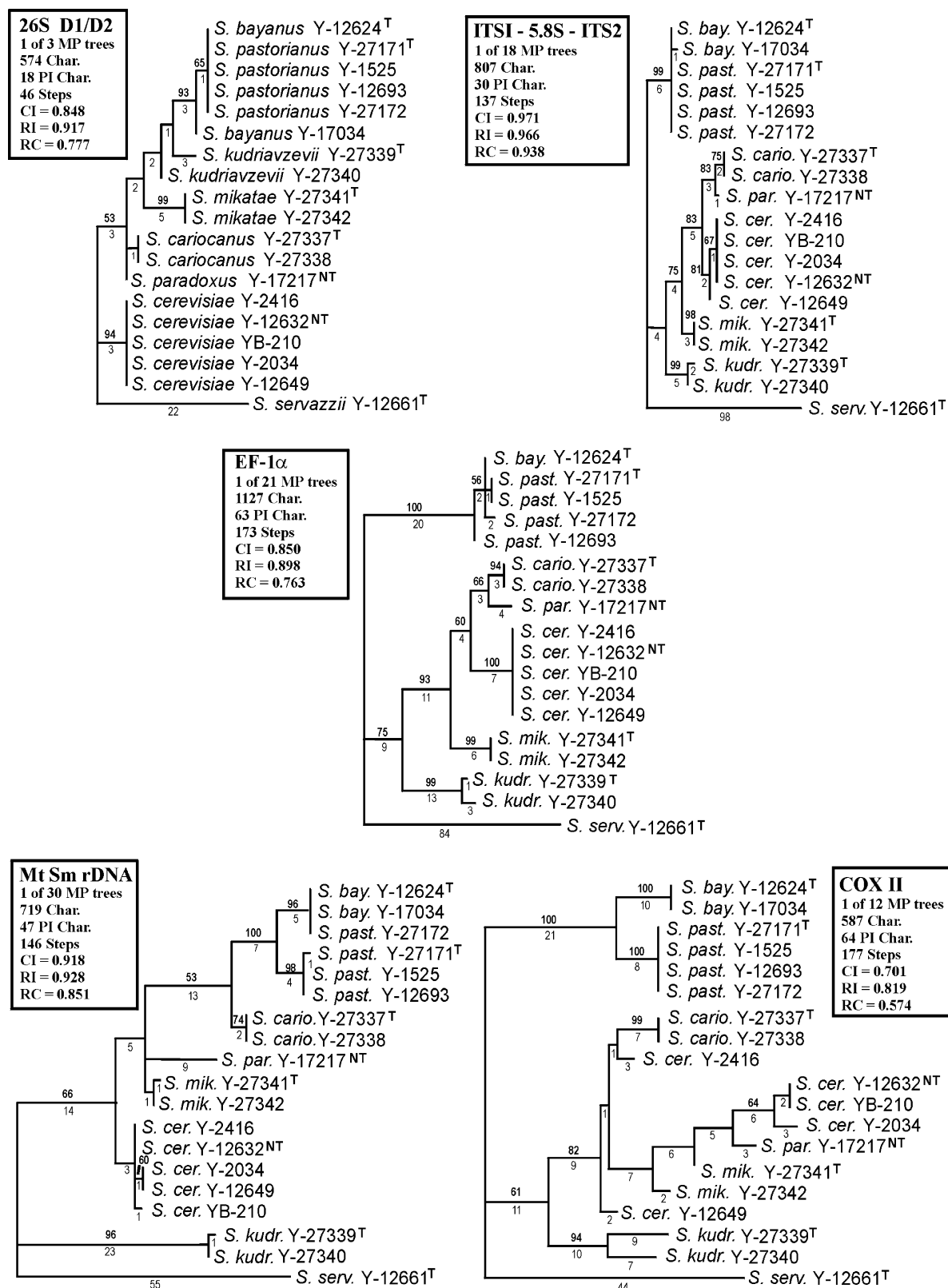


Fig. 1. MP analyses of species in the *S. cerevisiae* clade from nucleotide sequences of 26S rDNA domains 1 and 2, ITS1–5.8S–ITS2, the nuclear protein encoding gene EF-1α, and the mitochondrially encoded small-subunit rDNA and COX II genes. Numbers above branches are bootstrap values, shown when greater than 50%, and numbers below the branches are nucleotide differences. PI Char. = phylogenetically informative characters, CI = consistency index, RI = retention index, RC = rescaled consistency index. Type strains for each species are designated with a superscript T. *Saccharomyces servazzii* was the outgroup species in all analyses.

ITS1-5.8S-ITS2

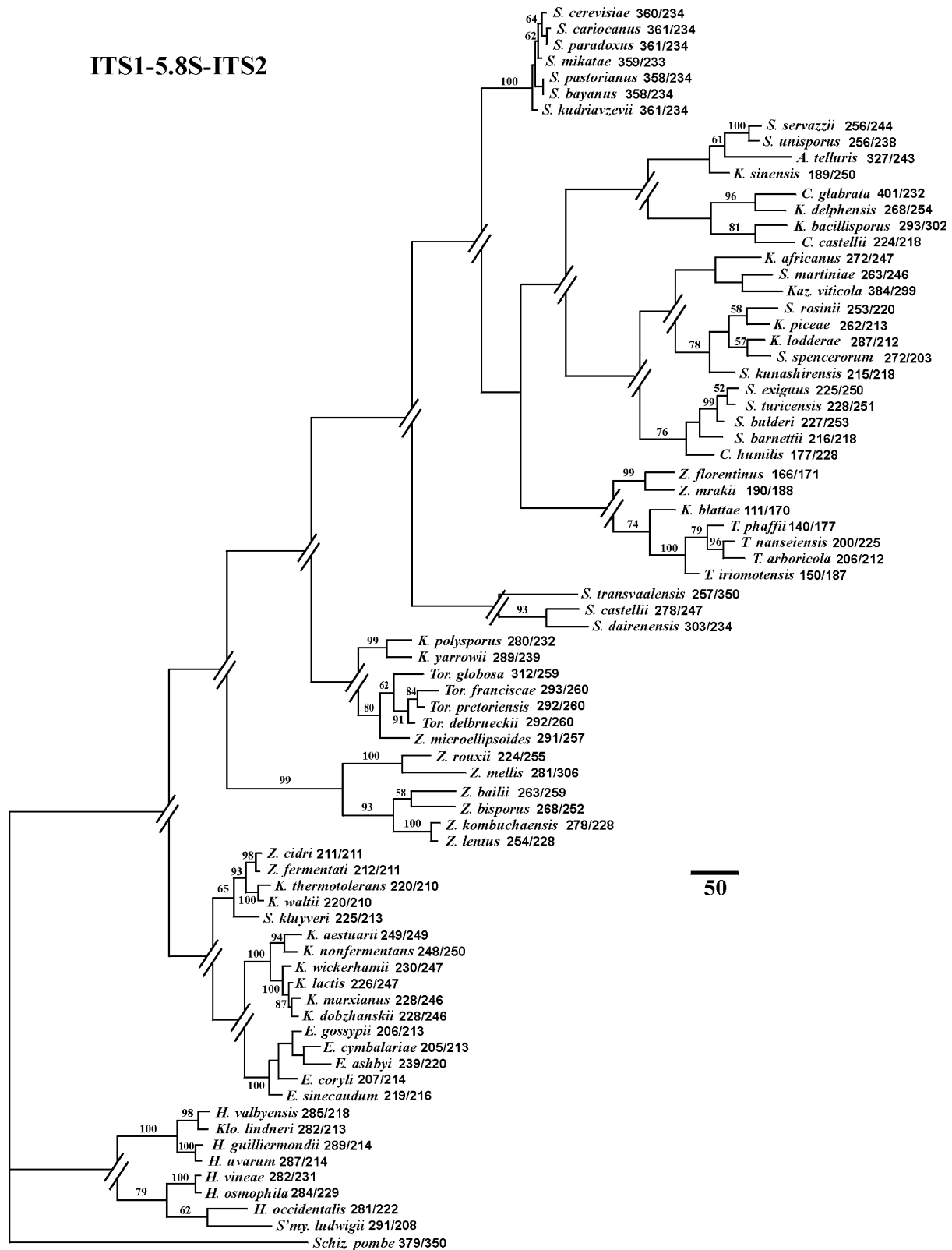


Fig. 2. MP analysis of species in the 'Saccharomyces complex' from ITS1–5.8S–ITS2 sequences. Branches with broken lines indicate that adjacent clades differ from one another in sequence length or that interclade alignment is ambiguous and therefore phylogenetic positions between clades are uncertain. ITS1/ITS2 nucleotide lengths follow each species name and are based on determinations from type strains. Bootstrap values > 50% are given for terminal nodes. The bar indicates branch lengths as nucleotide substitutions. *Schizosaccharomyces pombe* is the outgroup species. Abbreviations for figures: A. = *Arxiozyma*, C. = *Candida*, E. = *Eremothecium*, H. = *Hanseniaspora*, K. = *Kluyveromyces*, Kaz. = *Kazachstania*, Klo. = *Kloeckera*, S. = *Saccharomyces*, Schiz. = *Schizosaccharomyces*, S'my. = *Saccharomycodes*, T. = *Tetrapispora*, Tor. = *Torulaspora*, Z. = *Zygosaccharomyces*.

3.2. Ribosomal rDNA genes

3.2.1. ITS1–5.8S–ITS2

The length of the internal transcribed spacer genes (ITS1/ITS2) differed considerably among some of the species compared, with *Kluyveromyces blattae* having the shortest (111/170) and *Kazachstania viticola* the longest (383/299) (Fig. 2). For many closely related species, such as members of the *S. cerevisiae* clade, there are few inter-specific length differences, but in some instances such unrelated species as *Saccharomyces turicensis* (228/251) and *Kluyveromyces marxianus* (228/246) have nearly identical ITS lengths. Consequently, use of ITS length for species identification will often be unreliable. Sequences of closely related species are easily aligned, but accurate alignment of sequences between more distantly related groups is limited to a few short regions on either side of the 5.8S gene, resulting in an unambiguously aligned database of just 203 nucleotides, including the 157 nucleotides of the highly conserved 5.8S gene. As a result, basal lineages of the clades depicted in Fig. 2, although similar to the combined gene trees to be discussed, are exceptionally weak when derived from the well-aligned regions of ITS, and have been presented with broken branch lines to emphasize lack of support. Paradoxically, bootstrap support for the basal lineages, when determined from all nucleotides (data not shown), is quite strong because the lack of alignment between clades is treated as divergence by the treeing algorithms. For terminal taxa, species relationships from ITS are congruent with those from other gene trees, but there is no greater resolution.

3.2.2. Small-subunit (18S) and large-subunit (26S) rDNAs

Species relationships determined from 18S rDNA (Fig. 3) and the ca. 600-nucleotide domains D1 and D2 of 26S rDNA (Fig. 4) are congruent for species with strong bootstrap values. Neither dataset provides strong support of basal lineages, and there are some differences between the two genes in extent of support. The *Z. rouxii* clade is well supported in the 18S tree but not in the 26S D1/D2 tree, whereas the *Torulaspora* clade is more strongly supported in the D1/D2 tree than in the 18S tree. Combining the two datasets resulted in stronger support for some basal lineages, but many were still only weakly resolved.

In an effort to increase phylogenetic signal, the entire 26S rDNA gene was sequenced for a subset comprised of the following 18 species: *S. cerevisiae*, *S. paradoxus*, *S. bayanus*, *S. servazzii*, *S. unisporus*, *S. rosinii*, *S. spencerorum*, *S. exiguus*, *S. barnettii*, *S. castellii*, *S. dairenensis*, *Kluyveromyces africanus*, *K. lodderae*, *K. blattae*, *K. polysporus*, *K. yarrowii*, *Tetrapisispora phaffii* and *S. pombe*. An analysis of this dataset showed 71% of the phylogenetic signal to be derived from three regions, i.e., nucleotides 63–642 (domains D1/D2), 1498–2152, and 2963–3341, which comprise less than half of the entire 26S gene. Sequences from these three regions were determined

18S
>100 MP Trees
1806 Char.
241 PI Char.

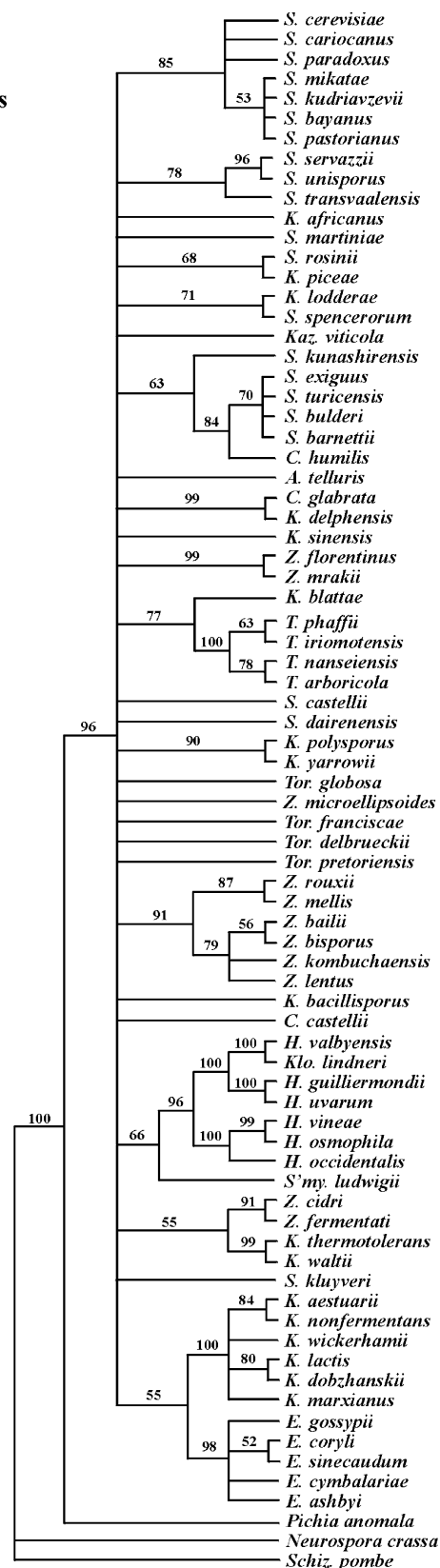


Fig. 3. Bootstrap consensus tree of species in the 'Saccharomyces complex' from MP analysis of 18S rDNA. Bootstrap values > 50% are given. Tree length = 1018, CI = 0.509, RI = 0.739, RC = 0.376. *S. pombe* is the outgroup species, and all species are represented by type strains.

26S D1/D2
>100 MP Trees
624 Char.
233 PI Char.

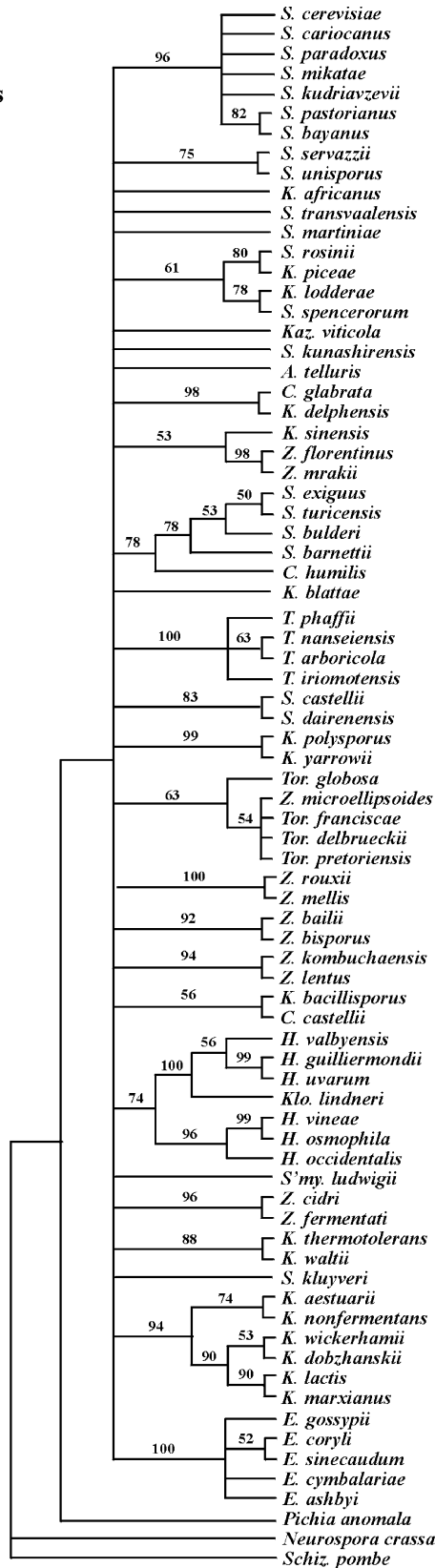


Fig. 4. Bootstrap consensus tree of species in the 'Saccharomyces complex' from MP analysis of 26S D1/D2 rDNA. Bootstrap values > 50% are given. Tree length=1377, CI=0.361, RI=0.623, RC=0.225. *S. pombe* is the outgroup species, and all species are represented by type strains.

18S-5.8S-26S
12 MP Trees
3767 Char.
833 PI Char.

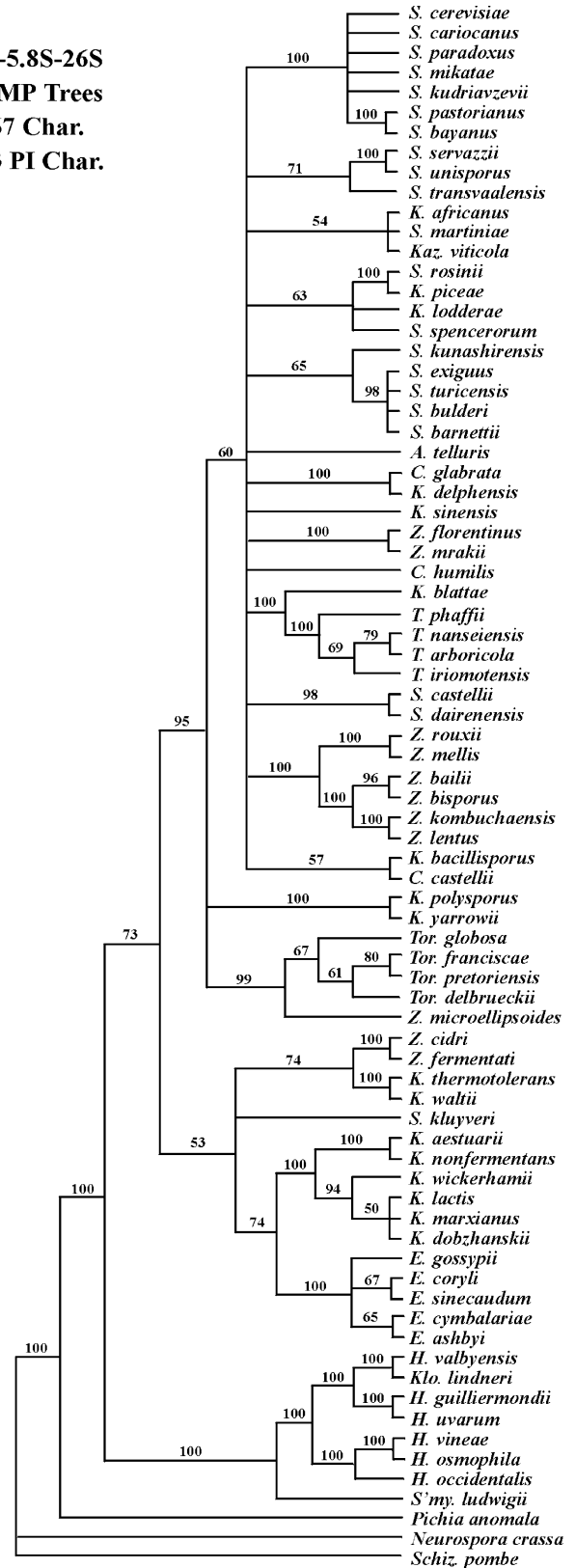


Fig. 5. Bootstrap consensus tree of species in the 'Saccharomyces complex' from MP analysis of combined sequences from 18S, 5.8S, and 26S (nucleotides 63–642, 1498–2152, 2963–3341) rDNAs. Bootstrap values > 50% are given. Tree length=4256, CI=0.418, RI=0.673, RC=0.281. *S. pombe* is the outgroup species, and all species are represented by type strains.

for all species in the study and combined with 5.8S and 18S sequences. Although this markedly strengthened the dataset, numerous lineages ended in polytomies on bootstrap consensus trees (Fig. 5).

3.3. Protein-encoding nuclear genes

Three nuclear protein-encoding genes were selected for comparison: EF-1 α , actin and RPB2. The latter two genes were sequenced only for species in Clade 2, and will be discussed relative to the combined gene analysis. As noted in Section 2, third-position nucleotides were removed from the analysis. Phylogenetic signal from EF-1 α is weaker than that of 26S D1/D2 or 18S rDNA (Fig. 6). Species relationships with greater than 50% bootstrap support were congruent with the rDNA trees, but as with the rDNA trees, certain clades were disproportionately supported. In single-gene rDNA trees, the *K. marxianus* clade was strongly supported, but not in the EF-1 α tree. Nonetheless, the EF-1 α tree supported the concept that the Saccharomycetales (hemiascomycetes) are monophyletic, as did the other gene trees.

3.4. Mitochondrial genes

COX II has been demonstrated by Belloch et al. [8] to be useful for examining relationships among species of *Kluyveromyces*. Although a protein-encoding gene, third-position nucleotides appear not to be saturated and, in contrast to EF-1 α , actin and RPB2, all nucleotides in the sequence were included in the analysis. Phylogenetic relationships determined from this gene are congruent with those of the rDNA trees (Fig. 7). Mitochondrial small-subunit rDNA (Fig. 8) also gave a phylogenetic tree congruent with nuclear rDNA trees, and with the exception of the *Eremothecium* clade, bootstrap support was more robust than in the D1/D2 tree.

3.5. Combined gene analyses

To more strongly resolve basal lineages, a dataset comprised of the previously described regions of 26S, 5.8S/alignable ITS, 18S, EF-1 α , mitochondrial small-subunit rDNA, and COX II was analyzed using MP and NJ with *P. anomala*, *N. crassa* and *S. pombe* as outgroups, respectively. MP and NJ trees were similar and selection of the outgroup species had little effect on relationships among clades with strong basal support. In this combined dataset, species resolved into 14 clades, most of which show strong basal support (Fig. 9). Clade 8 (*Zygosaccharomyces florentinus*, *Z. mrakii*) is basal to and weakly associated with *Torulaspora* (Clade 9) when the outgroup is *P. anomala* as shown in Fig. 9. Designation of *S. pombe* as outgroup causes Clade 8 to become basal to and weakly associated with Clade 7. Many of the species in Clade 2 are only weakly resolved. With *P. anomala* as outgroup,

EF-1 α
>100 MP Trees
748 Char.
136 PI Char.

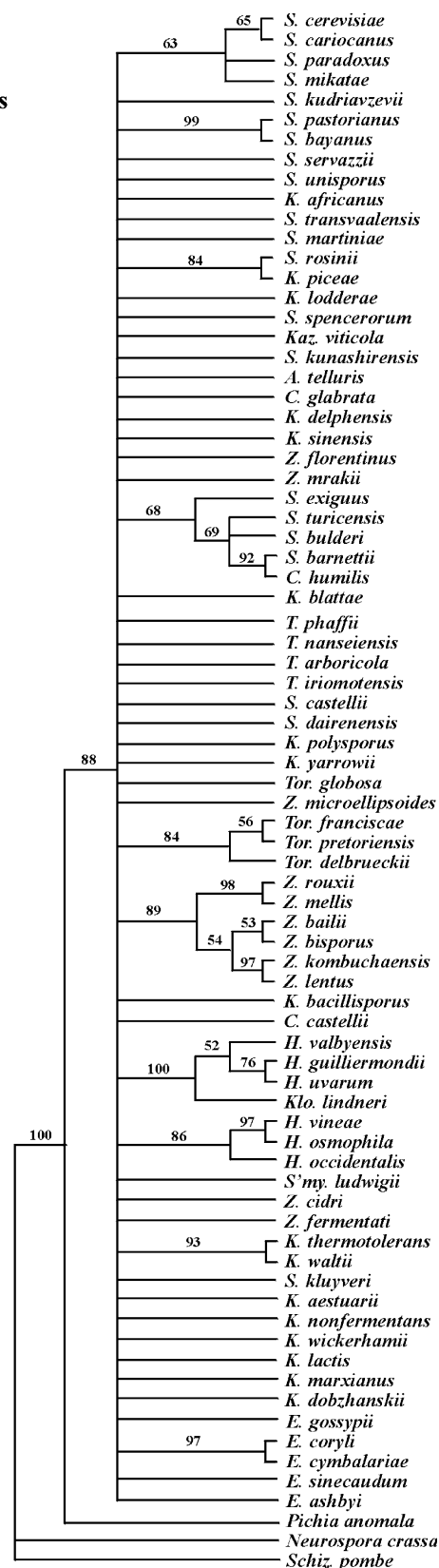


Fig. 6. Bootstrap consensus tree of species in the 'Saccharomyces complex' from MP analysis of codon positions 1 and 2 of the EF-1 α DNA sequence. Bootstrap values >50% are given. Tree length = 738, CI = 0.358, RI = 0.583, RC = 0.208. *S. pombe* is the outgroup species, and all species are represented by type strains.

COX II
4 MP Trees
579 Char.
286 PI Char.

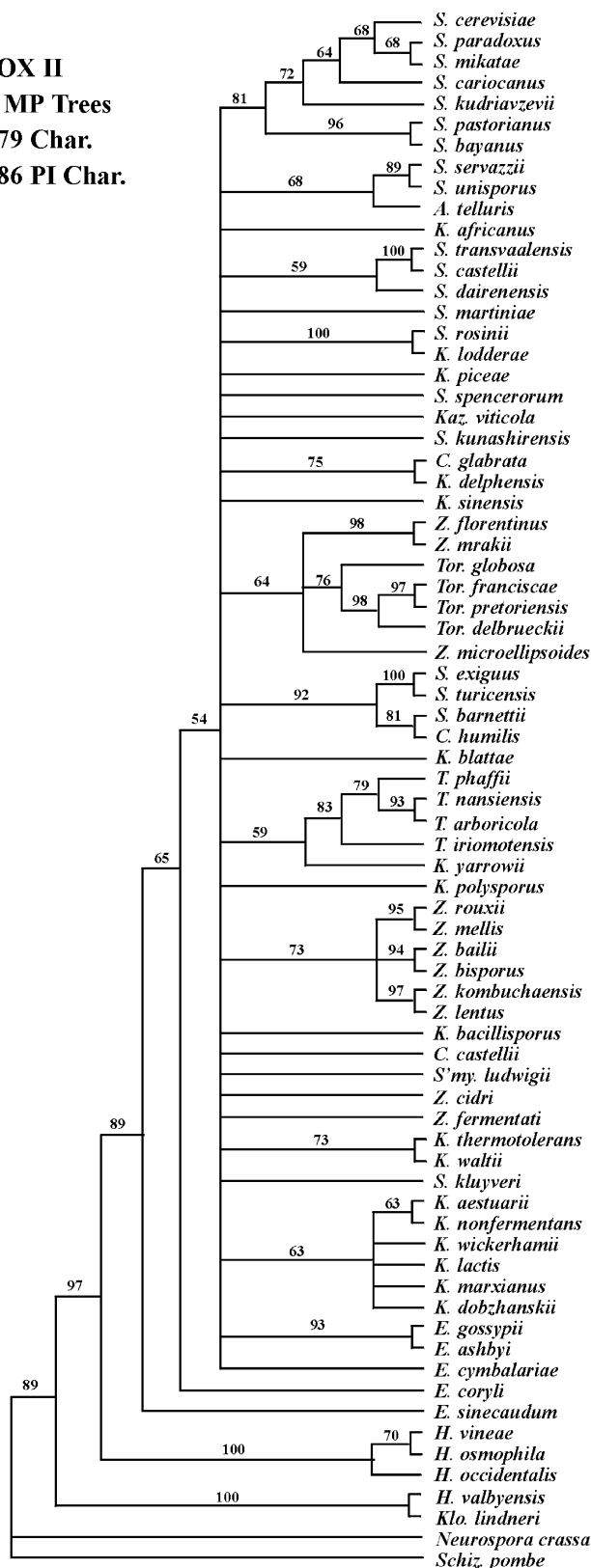


Fig. 7. Bootstrap consensus tree of species in the 'Saccharomyces complex' from MP analysis of COX II DNA sequences. Bootstrap values > 50% are given. Tree length = 1974, CI = 0.286, RI = 0.556, RC = 0.159. *Saccharomyces bulderi*, *Hanseniaspora guilliermondii*, *H. uvarum* and *Pichia anomala* were not included in the analysis. *S. pombe* is the outgroup species, and all species analyzed are represented by the type strains.

Mt Sm rDNA
>100 MP Trees
457 Char.
220 PI Char.

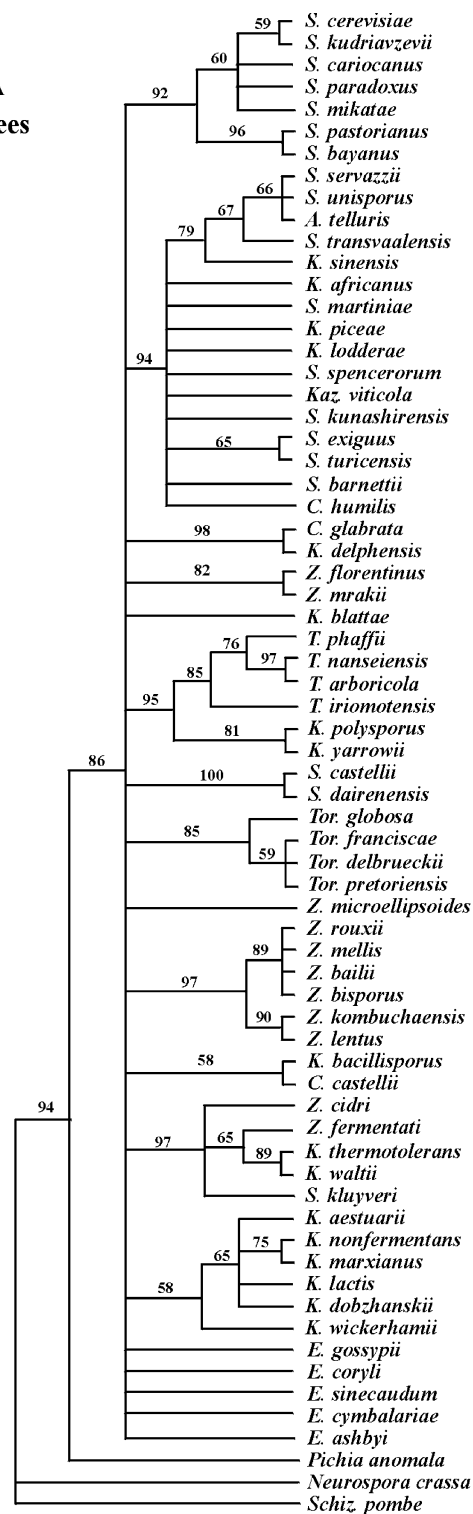


Fig. 8. Bootstrap consensus tree of species in the 'Saccharomyces complex' from MP analysis of mitochondrial small-subunit rDNA sequences. Bootstrap values > 50% are given. Tree length = 1079, CI = 0.437, RI = 0.669, RC = 0.292. *Saccharomyces bulderi*, *S. rosinii*, all *Hanseniaspora/Kloeckera* spp., and *Saccharomycodes ludwigii* were not included in the analysis. *S. pombe* is the outgroup species, and all species analyzed are represented by type strains.

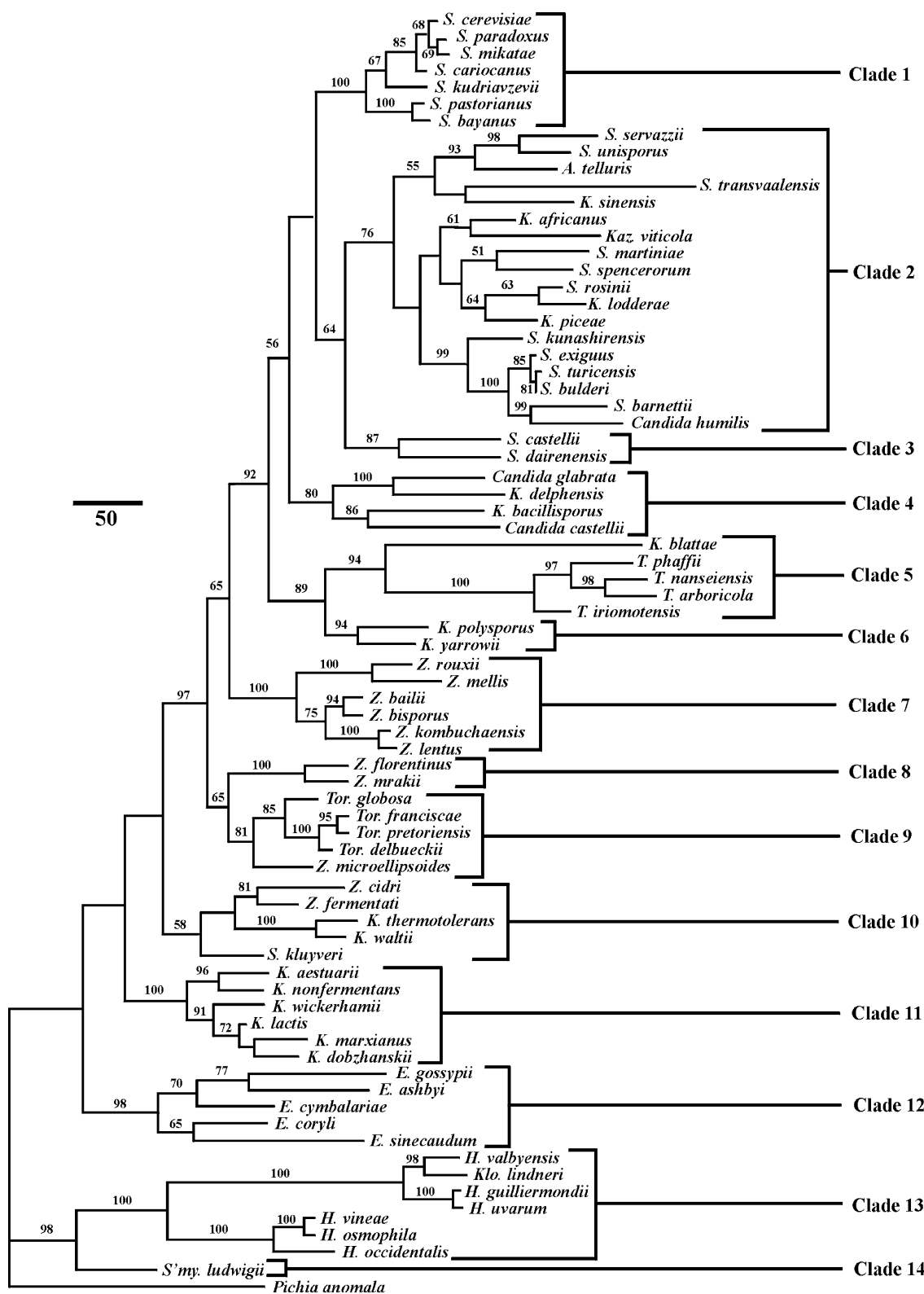


Fig. 9. Phylogenetic tree resolving species of the 'Saccharomyces complex' into 14 clades as represented by one of three most parsimonious trees derived from MP analysis of a dataset comprised of nucleotide sequences from 18S, 5.8S/alignable ITS, and 26S (three regions) rDNAs, EF-1 α , mitochondrial small-subunit rDNA and COX II. Tree length = 5135, CI = 0.329, RI = 0.631, RC = 0.208. The dataset, pruned of all potentially ambiguously aligned regions, has 4962 characters of which 929 are parsimony-informative. Bootstrap values > 50% are given. *P. anomala* is the outgroup species, and all species are analyzed from type strains.

the species separated into three subclades delimited by *S. servazzii*/*K. sinensis*, *K. africanus*/*K. piceae* and *S. kunashirensis*/*S. barnettii*. When *S. pombe* was used as outgroup, *K. africanus*, *Kaz. viticola* and *S. martiniae* moved to a basal position in the clade, but other species relationships were unchanged.

Two additional genes, actin and RPD2, were sequenced for members of Clade 2, as well as for *S. cerevisiae* and *K. polysporus*, with the latter taxon used as outgroup species for analyses. Actin provided little phylogenetic signal after removal of third-position nucleotides (598 nucleotides, 27 of which were phylogenetically informative), but the signal from RPD2 was moderately strong (639 nucleotides, 204 of which were phylogenetically informative). With or without the actin and RPD2 genes in the dataset, the clade had 96% bootstrap support, which was ca. 30% higher than when this subset of species was included in the whole dataset (Fig. 9). Including actin and RPD2 increased bootstrap support of terminal lineages up to 10%, but provided no additional support for internal nodes. Trees from this dataset placed *K. africanus*, *Kaz. viticola* and *S. martiniae* basal to the other species in Clade 2, but bootstrap support was less than 50% for each of the three species. The Clade 2 datasets were analyzed by MP, ML and NJ, and species relationships were the same in all three treatments.

4. Conclusions

The multigene sequence analysis presented has resolved 75 species of the ‘*Saccharomyces* complex’ into 14 clades. In many cases, these clades do not correspond to currently circumscribed genera [27]. The *S. cerevisiae* clade (Clade 1) comprised of the *Saccharomyces* sensu stricto species is phylogenetically separate from the *Saccharomyces* sensu lato species. Species of *Kluyveromyces* are found in six clades indicating the polyphyly of the genus as presently defined, and the inadequacy of ascus deliquescence as a phylogenetic descriptor for *Kluyveromyces*. Members of the genus *Zygosaccharomyces* are distributed among four clades.

The species assigned to *Eremothecium* from earlier molecular analyses [28] remain as a well-supported clade following multigene analysis. *Hanseniaspora* had been shown from other comparisons to be divided into two closely related lineages [29,30], and that earlier work is confirmed here. The present analysis placed *Saccharomycodes ludwigii* basal to the *Hanseniaspora* clade, confirming a close relationship that had been predicted from the similarity of vegetative cell division by bipolar budding. Resolution of lineages basal to each of the 14 clades (Fig. 9) detected in this study is generally weak, precluding assignment of genera to families solely from the phylogenetic analysis presented.

Results from the present study show that many of the

genera in the ‘*Saccharomyces* complex’ are not phylogenetically circumscribed, and this misclassification is particularly acute for species now assigned to *Saccharomyces*, *Kluyveromyces* and *Zygosaccharomyces*. There is presently no widely accepted definition for a phylogenetically circumscribed genus. Clades 1 (*Saccharomyces* sensu stricto), 7 (*Zygosaccharomyces* sensu stricto), 9 (*Torulaspora*) and 12 (*Eremothecium*) may serve as models for this concept because they represent isolated and well-supported monophyletic groups of species that have overall phenotypic similarity. Consequently, in a forthcoming paper, species will be classified into genera corresponding to the 14 clades that were resolved in the present study.

Acknowledgements

Larry W. Tjarks is gratefully acknowledged for skillful operation of the nucleic acid sequencer and oligonucleotide synthesizer, and Raymond F. Sylvester for preparation of final figures. We thank Todd J. Ward for helpful advice on maximum likelihood analysis. The mention of firm names or trade products does not imply that they are endorsed or recommended by the U.S. Department of Agriculture over other firms or similar products not mentioned.

References

- [1] Kurtzman, C.P. and Robnett, C.J. (1994) Synonymy of the yeast genera *Wingea* and *Debaryomyces*. *Antonie van Leeuwenhoek* 66, 337–342.
- [2] Cai, J., Roberts, I.N. and Collins, D. (1996) Phylogenetic relationships among members of the ascomycetous yeast genera *Brettanomyces*, *Debaryomyces*, *Dekkera*, and *Kluyveromyces* deduced by small-subunit rRNA gene sequences. *Int. J. Syst. Bacteriol.* 46, 542–549.
- [3] James, S.A., Collins, M.D. and Roberts, I.N. (1994) Genetic interrelationships among species of the genus *Zygosaccharomyces* as revealed by small-subunit rRNA gene sequences. *Yeast* 10, 871–881.
- [4] James, S.A., Cai, J., Roberts, I.N. and Collins, M.D. (1997) A phylogenetic analysis of the genus *Saccharomyces* based on 18S rRNA gene sequences: description of *Saccharomyces kunashirensis* sp. nov. and *Saccharomyces martiniae* sp. nov. *Int. J. Syst. Bacteriol.* 47, 453–460.
- [5] James, S.A., Collins, M.D. and Roberts, I.N. (1996) Use of an rRNA internal transcribed spacer region to distinguish phylogenetically closely related species of the genera *Zygosaccharomyces* and *Torulaspora*. *Int. J. Syst. Bacteriol.* 46, 189–194.
- [6] Kurtzman, C.P. and Robnett, C.J. (1995) Molecular relationships among hyphal ascomycetous yeasts and yeastlike taxa. *Can. J. Bot.* 73, S824–S830.
- [7] Kurtzman, C.P. and Robnett, C.J. (1998) Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Antonie van Leeuwenhoek* 73, 331–371.
- [8] Belloch, C., Querol, A., Garcia, M.D. and Barrio, E. (2000) Phylogeny of the genus *Kluyveromyces* inferred from the mitochondrial cytochrome-c oxidase II gene. *Int. J. Syst. Evol. Microbiol.* 50, 405–416.
- [9] Mikata, K., Ueda-Nishimura, K. and Hisatomi, T. (2001) Three new

- species of *Saccharomyces* sensu lato van der Walt from Yaku Island in Japan: *Saccharomyces naganishii* sp. nov., *Saccharomyces humaticus* sp. nov. and *Saccharomyces yakushimaensis* sp. nov. *Int. J. Syst. Evol. Microbiol.* 51, 2189–2198.
- [10] Wickerham, L.J. (1951) Taxonomy of yeasts. USDA Tech. Bull. 1029, Washington, DC.
- [11] Raeder, U. and Broda, P. (1985) Rapid preparation of DNA from filamentous fungi. *Lett. Appl. Microbiol.* 1, 17–20.
- [12] Kurtzman, C.P., Robnett, C.J. and Basehoar-Powers, E. (2001) *Zygosaccharomyces kombuchaensis*, a new ascosporeogenous yeast from 'Kombucha tea'. *FEMS Yeast Res.* 1, 133–138.
- [13] Nagata, S., Nagashima, K., Tsunetsugu-Yokota, Y., Fujimura, K., Miyazaki, M. and Kaziro, Y. (1984) Polypeptide chain elongation factor 1 α (EF-1 α) from yeast: nucleotide sequence of one of two genes for EF-1 α from *Saccharomyces cerevisiae*. *EMBO J.* 3, 1825–1830.
- [14] Daniel, H.-M., Sorrell, T.C. and Meyer, W. (2001) Partial sequence analysis of the actin gene and its potential for studying the phylogeny of *Candida* species and their teleomorphs. *Int. J. Syst. Evol. Microbiol.* 51, 1593–1606.
- [15] Sweetser, D., Nonet, M. and Young, R.A. (1987) Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. *Proc. Natl. Acad. Sci. USA* 84, 1192–1196.
- [16] Liu, Y.J., Whelen, S. and Hall, B.D. (1999) Phylogenetic relationships among ascomycetes: evidence from an RNA polymerase II subunit. *Mol. Biol. Evol.* 16, 1799–1808.
- [17] Swofford, D.L. (1998) PAUP* 4.0: Phylogenetic Analysis using Parsimony. Sinauer Associates, Sunderland, MA.
- [18] de Queiroz, A., Donoghue, M.J. and Kim, J. (1995) Separate versus combined analysis of phylogenetic evidence. *Annu. Rev. Ecol. Syst.* 26, 657–681.
- [19] Sullivan, J. (1996) Combining data with different distributions of among-site rate variation. *Syst. Biol.* 45, 375–380.
- [20] Naumov, G.I., James, S.A., Naumova, E.S., Louis, E.J. and Roberts, I.N. (2000) Three new species in the *Saccharomyces sensu stricto* complex: *Saccharomyces cariocamus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*. *Int. J. Syst. Evol. Microbiol.* 50, 1931–1942.
- [21] Vaughan Martini, A. and Kurtzman, C.P. (1985) Deoxyribonucleic acid relatedness among species of the genus *Saccharomyces* sensu stricto. *Int. J. Syst. Bacteriol.* 35, 508–511.
- [22] Fischer, G., James, S.A., Roberts, I.N., Oliver, S.G. and Louis, E.J. (2000) Chromosomal evolution in *Saccharomyces*. *Nature* 405, 451–454.
- [23] Groth, C., Hansen, J. and Piskur, J. (1999) A natural chimeric yeast containing genetic material from three species. *Int. J. Syst. Bacteriol.* 49, 1933–1938.
- [24] Groth, C., Petersen, R.F. and Piskur, J. (2000) Diversity in organization and the origin of gene orders in the mitochondrial DNA molecules of the genus *Saccharomyces*. *Mol. Biol. Evol.* 17, 1833–1841.
- [25] Nguyen, H.-V., Lepingle, A. and Gaillardin, C. (2000) Molecular typing demonstrates homogeneity of *Saccharomyces uvarum* strains and reveals the existence of hybrids between *S. uvarum* and *S. cerevisiae*, including the *S. bayanus* type strain CBS 380. *Syst. Appl. Microbiol.* 23, 71–85.
- [26] Pedersen, M.B. (1986) DNA sequence polymorphisms in the genus *Saccharomyces*. IV. Homoeologous chromosomes, *S. carlsbergensis* and *S. uvarum*. *Carlsberg Res. Commun.* 51, 185–202.
- [27] Kurtzman, C.P. and Fell, J.W. (1998) The Yeasts, A Taxonomic Study, 4th edn. Elsevier Science, Amsterdam.
- [28] Kurtzman, C.P. (1995) Relationships among the genera *Ashbya*, *Eremothecium*, *Holleya* and *Nematospora* determined from rDNA sequence divergence. *J. Ind. Microbiol.* 14, 523–530.
- [29] Boekhout, T., Kurtzman, C.P., O'Donnell, K. and Smith, M.T. (1994) Phylogeny of the yeast genera *Hanseniaspora* (anamorph *Kloeckera*), *Dekkera* (anamorph *Brettanomyces*), and *Eeniella* as inferred from partial 26S ribosomal DNA nucleotide sequences. *Int. J. Syst. Bacteriol.* 44, 781–786.
- [30] Yamada, Y., Maeda, K. and Banno, I. (1992) An emendation of *Kloeckeraspora* Niehaus with the type species *Kloeckeraspora osmophila* Niehaus, and the proposals of two new combinations, *Kloeckeraspora occidentalis* and *Kloeckeraspora vineae* (Saccharomycetaceae). *Bull. Jpn. Fed. Cult. Collect.* 8, 79–85.