International Journal of Systematic and Evolutionary Microbiology (2001), **51**, 1593–1606 Printed in Great Britain

Partial sequence analysis of the actin gene and its potential for studying the phylogeny of *Candida* species and their teleomorphs

Heide-M. Daniel,[1,2] Tania C. Sorrell[1] and Wieland Meyer[1]

[1] Molecular Mycology Laboratory, The University of Sydney/Westmead Hospital, ICPMR, Level 3, Rm 3114A, Darcy Road, Westmead, NSW 2145, Australia

[2] Technische Universität Berlin, Institut für Mikrobiologie und Genetik, Gustav-Meyer-Allee 25, D-13355 Berlin, Germany

Author for correspondence: Heide-M. Daniel. Tel: +61 2 9845 6332. Fax: +61 2 9891 5317. e-mail: heidem@westgate.wh.usyd.edu.au

The actin gene has been studied as a potential phylogenetic marker for selected members of the anamorphic genus *Candida* and seven related teleomorphic genera (*Debaryomyces*, *Issatchenkia*, *Kluyveromyces*, *Saccharomyces* and *Pichia* from the Saccharomycetaceae; *Clavispora* and *Metschnikowia* from the Metschnikowiaceae). The nucleotide sequences of 36 fungal taxa were analysed with respect to their molecular evolution and phylogenetic relationships. A total of 460 bp (47%) of the coding 979 bp were variable and 396 bp (40%) of these were found to be phylogenetically informative. Further analysis of the sequences showed that the genic G+C contents were higher than the nuclear G+C contents for most of the taxa. A strong positive correlation was found between G+C content over all codon positions and third positions. First and second codon positions were considered to be independent of the genic G+C content. The expected transition/transversion bias was detected only for third positions. Pairwise comparisons of transitional and transversional changes (substitutions) with total percentage sequence divergences revealed that the third position transitions showed no saturation for ingroup comparisons. A specific weighting scheme was set up, combining codon-position weights with change-frequency weights to enable the inclusion of distant outgroup taxa. Parsimony analyses of the investigated taxa showed four groups, three of which corresponded to major clusters that had been established previously in *Candida* by rDNA analysis. Interrelationships among the species groups in this heterogeneous anamorphic genus were determined. The polyphyletic origin of the selected *Candida* species and their close associations with several ascomycete genera were verified and known anamorph/teleomorph pairs confirmed. The actin gene was established as a valuable phylogenetic marker with the particular advantage of an unambiguous alignment.

Keywords: *Candida*, ascomycetous yeasts, phylogeny, actin, molecular evolution

INTRODUCTION

The phylogenetic results of molecular data sets for single genetic loci have to be treated as gene trees, displaying relationships among the species only for the locus that was sequenced (Fitch, 1996). As the only established phylogenies for *Candida* and related species have been based on rDNA sequences, distinguishing the overall phylogenetic relationship between the species has not been possible. Given the lack of phylogenetically useful phenotypic characters in the anamorphic genus *Candida* (Meyer *et al.*, 1998; Starmer *et al.*, 1978; Kurtzman, 1984; Fuson *et al.*, 1980; Price *et al.*, 1978), a well-based species tree should ideally be composed from several gene trees containing all relevant species. We regard this study as a contribution towards this aim. The highly conserved actin gene had the advantage of allowing an unambiguous alignment compared with rDNA se-

Abbreviations: CI, consistency index; Co-Q, coenzyme Q system or ubiquinone; RI, retention index.

The GenBank accession numbers for the nucleotide sequences reported in this paper are AJ389057–AJ389092.

quences, which may be difficult to align with confidence. It had, however, the corresponding disadvantage of yielding relatively few informative characters.

Asexual stages (anamorphs) of ascomycetous species have traditionally been classified in the genus *Candida* within the artificial form class Deuteromycetes. Their sexual counterparts (teleomorphs) are classified in the Hemiascomycetes according to their ascospores, which are not enclosed in fruiting bodies. Molecular techniques now provide an opportunity to formulate a more natural taxonomy for *Candida* species that may not be associated directly with a known teleomorph. Existing sequencing studies used to investigate the systematics and phylogeny of *Candida* species and their teleomorphs are based on the small-subunit rRNA gene (e.g. Barns *et al*., 1991; Berbee & Taylor, 1992; Cai *et al*., 1996) and the 5′ end of the large-subunit rRNA gene (e.g. Kurtzman & Robnett, 1998) or both (e.g. Mendonça-Hagler *et al*., 1993; Kurtzman & Robnett, 1994; Yamada *et al*., 1994) as well as rDNA spacer sequences (Lott *et al*., 1993; Montrocher *et al*., 1998). Because of frequently observed limitations of rDNA sequence analysis, such as alignment ambiguity and poorly resolved phylogenetic trees, the actin (*act-1*) gene was used in this study as a potentially independently evolving marker.

Actins are highly conserved structural proteins, found in all eukaryotes. They are necessary for the transformation of chemical into mechanical energy (Sheterline & Sparrow, 1994). Within the multigene family, cytoplasmic and muscular actins can be distinguished. Only the cytoplasmic actins are relevant to yeasts. They are involved in various essential cellular functions, including mitosis, and are major proteins within the cytoskeletal structure. Cytoplasmic actin genes are among the most heavily transcribed genes (Lloyd & Sharp, 1992). As many as five actin genes, encoding different isoforms, have been found in *Saccharomyces cerevisiae* (Huang *et al*., 1996). Four of these genes encode so-called 'actin-related' proteins, showing 24–47% similarity to conventional actin, encoded by the *act-1* gene. The conventional actins of *Candida albicans* and *Filobasidiella neoformans*, fungi belonging to different phyla, show a high homology of 86% (Cox *et al*., 1995). The *act-1* genes consist of a coding region of about 1200 bp and intervening sequences. In contrast to those found in other eukaryotic species, fungal *act-1* genes are found as a single copy in each haploid genome. Species in which this status has been confirmed include *Candida albicans* (Losberger & Ernst, 1989), *Saccharomyces cerevisiae* (Gallwitz & Sures, 1980; Ng & Abelson, 1980), *Schizosaccharomyces pombe* (Mertins & Gallwitz, 1987), *Kluyveromyces lactis* (Deshler *et al*., 1989) and *Filobasidiella neoformans* (Cox *et al*., 1995).

The actin gene has been used successfully to determine phylogenetic relationships among a number of eukaryotes (Bhattacharya *et al*., 1991; Bhattacharya & Ehlting, 1995; Drouin *et al*., 1995; Bouget *et al*., 1995), including certain fungi (Cox *et al*., 1995; Wery *et al*., 1996). Some of these studies demonstrated that actin gene sequences do not show an independent evolution of characters on all hierarchical levels. As this is one of the central assumptions for phylogenetic analyses, some details of the molecular evolution were investigated. Base composition, transition and transversion rates and mutational saturation were determined and the distribution of these characteristics was analysed according to the codon position, that is, first, second or the more variable third position. From these analyses, different weighting schemes were developed and were compared with an analysis of deduced amino acid sequences. The weights assigned to the codon positions were based on the premise that more conserved characters are more reliable for detection of phylogenetic relationships (Farris, 1966; Kluge & Farris, 1969).

The purpose of this study was to determine the phylogenetic relationships evidenced by yeast actins and to compare these relationships with those elucidated previously in phylogenetic comparisons of rDNA sequences. The study concentrated on commensal or free-living, opportunistic pathogenic species and was restricted to members of the anamorphic genus *Candida* and seven teleomorphic genera (*Debaryomyces*, *Issatchenkia*, *Kluyveromyces*, *Saccharomyces* and *Pichia* from the Saccharomycetaceae; *Clavispora* and *Metschnikowia* from the Metschnikowiaceae). The diversity of ascomycetous yeasts and the wide distribution of *Candida* anamorphs among them made the inclusion of very distantly related outgroup species necessary. Outgroup species chosen included *Neurospora crassa*, a member of the Euascomycetes, and *Schizosaccharomyces pombe*, a member of the Archaeascomycetes. The former fungal class is the most closely related group at this taxonomic level to the Hemiascomycetes, which contains the main group of species tested, while the latter class is basal to both of these groups (Hendriks *et al*., 1992; Liu *et al*., 1999). A significantly more distantly related group, the Basidiomycetes, was represented by *Filobasidiella neoformans*.

METHODS

Yeast strains, cultivation and identification. The strains examined in this study are listed with their sources in Table 2. All strains were grown on Sabouraud's agar (10 g mycological peptone l^{-1}, 40 g glucose l^{-1}, 15 g agar l^{-1}; pH $5{\cdot}6 \pm 0{\cdot}2$) at 27 °C with the exception of *Candida lambica* and *Pichia fermentans*, which were grown at 22 °C. The strains were maintained as water cultures at room temperature. Isolates were identified prior to DNA extraction by two carbohydrate assimilation tests: Vitek YBC and API ID 32C (both bioMérieux). The samples were also characterized by their unique PCR fingerprints (Meyer *et al*., 1997) directly after DNA extraction and again after the sequencing work was finished, to ensure that no sample transposition or contamination had occurred.

DNA sequence determination. Genomic DNA was isolated as described by Meyer *et al*. (1997). Fragments of the actin

gene were amplified using two PCR programs. The first consisted of 30 cycles of denaturation at 94 °C for 3 min in cycle 1 and 2 min in cycles 2–30, annealing at 55 °C for 2 min and extension at 72 °C for 2 min, followed by a final extension at 72 °C for 10 min (Kan, 1993). The second, developed by us for the amplification of rDNA fragments, consisted of an initial denaturation for 3 min at 97 °C, 20 cycles of denaturation at 97 °C for 35 s, annealing at 50 °C for 55 s and extension at 72 °C for 45 s (increased by 4 s per cycle), 10 cycles of denaturation at 97 °C for 45 s, annealing at 50 °C for 55 s and extension at 72 °C for 2 min (increased by 4 s per cycle), followed by a final extension at 72 °C for 6 min. Initial amplifications were performed with the following primer sequences used by Kan (1993). Numbers in parentheses indicate the position of the primer sequence relative to the *Candida albicans* sequence, GenBank accession no. X16377; the first base of the start codon is base 1. The suffix R indicates reverse primers, and all primer sequences are given in the 5′→3′ direction: CA1, GCCGGTGACGACGCTCCAAGAGCTG (721–745); CA2R, CCGTGTTCAATTGGGTATCTCAAGGTC (852–878); CA3, GACATCAAGGTATCATGGTTGGTATGGGTGC (773–802); CA4, CCATCATGAAGTGTGACATGGATGTTAG, (1499–1526) and CA5R, GTGAACAATGGATGGACCAGATTCGTCG (1743–1770). Additional primers were designed based on the generated sequence: CA7R, CCATCACCAGAATCCAAAACAATACCGG (1103–1130); CA8, TGTACTCTTCTGGTAGAACTACCGG (1082–1106); CA9R, GGTCAATACCAGCAGCTTCCAAACCT (1458–1483); CA11, AACAATGGACGGTGGTATGT (−3–16); CA13, CCAGATGGTCAAGTTATCAC (1384–1403); CA14, AACTGGGATGACATGGAGAAGATCTGGC (889–916); CA15R, TCGGTCAAATCTCTACCAGC (1198–1217); CA16, TTTACGCTGGTTTCTCCATGCCTCACGG (1151–1178) and CA17R, TTGTGGTGAACAATGGATGGACC (1753–1775). Most of the fragments were amplified with the combination of primers CA1 and CA5R. However, the use of CA3, CA8, CA9R, CA11, CA14 and CA17R was necessary to amplify specific fragments for some species. Amplification reactions were performed in 50 µl containing 25 ng genomic DNA, 10 mM Tris/HCl pH 8·3, 50 mM KCl, 1·5 mM $MgCl_2$, 0·2 mM each of dATP, dCTP, dGTP, dTTP (Boehringer), 3 mM magnesium acetate, 50 ng each of the two primers and 2·5 U AmpliTaq Polymerase (Perkin Elmer), using a Perkin Elmer thermal cycler, model 480. Amplification products were assayed by electrophoresis in 0·8 % agarose gels in Tris/borate buffer and detected by staining with ethidium bromide. These fragments were purified using the Wizard PCR preparation system (Promega). Direct cycle sequencing of PCR products was performed with an automatic sequencer 373A, version 1.2.1 (Applied Biosystems). Both DNA strands were sequenced for all strains. The sequencing reactions contained 100–150 ng amplified fragment, 10 pmol of the appropriate primer μl^{-1} and the ABI PRISM Dye Terminator cycle sequencing ready reaction kit with AmpliTaq DNA polymerase FS (Applied Biosystems). Partial sequences were assembled and aligned manually using the sequence editor SeqPup versions 0.4 and 0.6d (Gilbert, 1996). The reading frames were determined by comparison with published actin gene sequences (e.g. *Candida albicans*, X16377, and *Kluyveromyces lactis*, M25826).

Phylogenetic analyses. The numbers of variable sites and sites informative in parsimony analysis and the genic G+C content were calculated with the computer program MacClade 3.07 (Maddison & Maddison, 1997). The genic G+C contents for first, second and third codon positions were determined separately and plotted against the total G+C content of the sequences according to Bhattacharya & Ehlting (1995). Distance matrices were generated with the programs MEGA 1.01 (Kumar *et al.*, 1993) and PAUP 3.1.1 (Swofford, 1993).

The base composition and the numbers of transitions and transversions for all pairwise comparisons of taxa were calculated using MEGA. The numbers of transitions and transversions partitioned by codon positions were plotted against the total percentage difference (or '*p*-distance', the number of nucleotide differences divided by the total number of nucleotides; Kumar *et al.*, 1993) in order to determine mutational saturation (Johnson & Sorenson, 1998).

Random searches were performed using PAUP to evaluate the shape of tree length distributions, represented by their g1 statistic. The mean tree lengths and their standard deviations were recorded and compared with the length of the most parsimonious trees. PAUP did not allow random searches for the weighted data sets.

Maximum-parsimony trees were constructed for the deduced amino acid data set, the complete nucleotide data set and the data partitioned by codon positions, using heuristic searches with simple and random sequence additions in PAUP. Only the most distant outgroup taxon, *Filobasidiella neoformans*, was predefined in PAUP as outgroup, to test whether the remaining outgroup species would appear in their expected order relative to the ingroup. When more than one tree was found, the 50 % majority consensus tree was generated.

Because of the frequent occurrence of third position transitions, they were excluded by recoding third codon positions into purines and pyrimidines. Weighting according to the relative substitution frequency of nucleotide changes within each codon position was applied to the complete data set, as well as to the partitioned and the combined data sets. The weights were generated using the 'character steps/etc.' option in the chart menu of the tree window for charting character statistics in MacClade. In order to generate weights (scale 1–10) according to codon position, the number of fixed mutational steps (S) attributable to first, second and third codon positions was calculated using the trees derived from unweighted analysis of all coding characters. Weights (W) were generated using the formula $W = 1/S$. Weights of 4, 10 and 1 were obtained for first, second and third positions, respectively. Similar analysis of the pooled data for first and second positions resulted in weights of 1 for first and 10 for second positions. Weights corresponding to the relative frequency of nucleotide changes (change-frequency weighting) at each individual position were obtained by calculating the number of changes at that position. The same MacClade menu and options were used as for the assignment of weights to the codon positions. The calculation of these weights was performed using the trees derived from unweighted analyses of separate codon positions. The resulting weights ranged from 1 to 10. Combined weights incorporating both change frequency and codon position were obtained by multiplying the weights of the two separate factors for each position. For instance, a first position with a calculated change-frequency weight of 10 was given the combined weight of 40 after multiplying by 4, the weight of the corresponding (i.e. the first, as mentioned above) codon position.

PAUP was used to conduct bootstrap analyses (Felsenstein, 1985) with 100 replicates for all phylogenies. In addition,

1000 replicates were performed for selected phylogenies. Bootstrap values above 70 were considered significant in the sense of demonstrating repeatability as suggested by Hillis & Bull (1993). This was based on the premise that the weighting technique we used neutralizes the effect of unequal rates of evolution as much as possible.

The MacClade program was used to generate consistency indices (CI) as a measurement of homoplasy (Kluge & Farris, 1969) after exclusion of uninformative characters. CI is 1 for a given data set when there is no homoplasy and decreases as homoplasy increases. However, the CI is known to be negatively correlated with the number of taxa and characters (Goloboff, 1991). The retention indices (RI, with a possible range of 0 to 1) were also evaluated in MacClade, to express the data-tree fit as the amount of synapomorphy in the data sets (Farris, 1989). The RI is high when changes occur predominantly on internal branches and low when changes are concentrated on branches leading to terminal taxa. CI and RI are useful for comparing different trees generated from one data matrix (Goloboff, 1991) and were used to assess the value of the different weighting methods and schemes.

RESULTS

Sequence alignment and variation

An alignment of 39 partial sequences of the actin gene was analysed (EMBL alignment number DS: 39430). Each sequence comprised 1127 bp. The sequence of the most distant outgroup taxon, *Filobasidiella neoformans*, contained introns of 148 bp, necessitating the insertion of three gaps of 46–51 bp. In all taxa analysed, including outgroups, 460 bp (47%) of the coding 979 bp were variable and 396 bp (40%) proved to be phylogenetically informative (Table 1).

In order to gauge the effectiveness with which the taxa were separated, we looked at the range of sequence variability among the taxa analysed. The highest intergeneric variability (31.4%, data not shown; distance matrix available on request) was found between *Filobasidiella neoformans* and *Pichia/Candida norvegensis*. The lowest (20·6%) occurred between *Schizosaccharomyces pombe* and *Clavispora opuntiae*. The variability within the ingroup ranged from 22·2% for the distantly related *Candida* species *Candida zeylanoides* and *Pichia/Candida norvegensis* to 2·1% for the closely related species *Candida albicans* and *Candida dubliniensis*. The intraspecies variability ranged from 0·4% for *Metschnikowia pulcherrima* and *Candida pulcherrima* to no variability between *Kluyveromyces lactis* and *Candida sphaerica* or the two strains of *Candida glabrata*, WM 53 and WM 54.

G+C content

The genic G+C contents of the actin sequences were determined and compared with the overall nuclear G+C content (Table 2). The genic G+C contents were slightly higher than the nuclear G+C content for most taxa. The G+C content in third codon positions was positively correlated with the overall G+C content (Fig. 1). First codon position G+C contents were partially correlated with the total genic G+C content, as only the three outgroup species showed first position values higher than the mean. This is also shown by differences in the linear regression for comparisons including ($r^2 = 0{\cdot}312$) and excluding ($r^2 = 0{\cdot}141$) the outgroup species. Second codon position G+C contents were not correlated with the overall G+C content.

Base composition and transition/transversion rates

The mean base composition of the actin sequences was biased in each of the three codon positions. There was an overrepresentation of adenine (30·6%) and guanine (32·4%) in first positions, adenine (29·7%) and thymine (30·2%) in second positions and thymine (33·9%) and cytosine (28·1%) in third positions. Lower representation was recorded for cytosine (13·5%) in first positions, guanine (14·9%) in second positions and guanine (17·4%) in third positions. The pattern of variation at first and second positions was not significantly different between species, as shown by the standard deviations for first positions (0·37–2·83) and for second positions (0·4–0·6). The standard deviations for third positions (5·3–8·8) indicate significant differences between the species compared.

Transitions accounted for a mean of 70% of changes in the ingroup comparisons and 56% in the ingroup–outgroup comparisons. The portions of transitions in first, second and third codon positions for comparisons at the ingroup/outgroup levels were 4/10, 2/2 and 64/43%, respectively. About three times more transitions than transversions were detected in third positions and about twice as many transversions than transitions in first and second positions. The number of transitions and transversions continued to increase at higher divergences (*p*-distances) in the first and second positions; only transversions increased in third positions (Fig. 2). The number of third position transitions was not significantly higher in the ingroup–outgroup comparisons (mean 81.2, standard deviation 21.4) than in the ingroup–ingroup comparisons (mean 102·1, standard deviation 15·2). Closer examination of the transitions at third positions within the ingroup revealed that the largest numbers were accumulated by the comparison of the distantly related ingroup species *Saccharomycopsis capsularis* and *Pichia/Candida norvegensis*. The largest numbers of third position transitions were found in comparisons involving taxa with deviating G+C contents, e.g. *Pichia/Candida norvegensis* (lowest genic G+C content of 32·9/33%) and *Candida zeylanoides* (highest genic G+C content for all ingroup species of 50·5%) (Table 2).

Phylogenetic signal and comparative analysis

The g1 values of the tree-length distributions were determined for the unweighted data sets (Table 1). They were below the critical values proposed by Hillis & Huelsenbeck (1992) for all combinations of characters, indicating the presence of a phylogenetic signal in all data sets. The heuristic search identified par-

Table 1. Comparison of character information content for different data sets, frequency distribution of 10000 randomly sampled tree lengths and measures of fit statistics for most parsimonious trees generated by heuristic search

SD, Standard deviation; −, not evaluated; +, soft polytomies underestimate.

Characters (total of 1127 bp)	Character information content			Random search			Heuristic search			
	Total	Variable	Informative	Mean tree length	SD	g1	CI*	RI	Tree length*	Number of trees
Deduced amino acid sequence	326	94	62	399	10·5	−0·54	0·72	0·86	144+	30
All coding residues	979	460	396	3379	63·7	−0·71	0·34	0·63	1776+	3
All coding residues weighted according to:										
Codon position	979	460	396	–	–	–	0·44	0·7	2842+	1
Change frequencies	979	460	396	–	–	–	0·46	0·7	3928+	1
Codon position and change frequencies	979	460	396	–	–	–	0·57	0·79	7946+	1
1st positions	326	111	81	493	12·1	−1	0·58	0·79	189+	255
2nd positions	326	52	29	159	5·2	−0·39	0·71	0·88	49+	1460
3rd positions	327	297	286	2727	54·4	−0·74	0·31	0·61	1500+	1
3rd transversions only	327	147	127	639	15	−1·19	0·45	0·72	280+	400
1st and 2nd positions	652	163	110	653	16	−0·81	0·6	0·8	242+	22
1st and 2nd positions weighted according to:										
Codon position	652	163	110	–	–	–	0·67	0·85	688+	203
Change frequencies	652	163	110	–	–	–	0·78	0·9	8559+	24
Codon position and change frequencies	652	163	110	–	–	–	0·83	0·93	29782+	30
1st and 2nd positions and 3rd transversions	979	490	237	1292	28·2	−1·09	0·5	0·74	540+	24
1st and 2nd positions and 3rd transversions weighted according to:										
Codon position	979	490	237	–	–	–	0·65	0·84	7792+	16
Change frequencies	979	490	237	–	–	–	0·69	0·85	19622+	4
Codon position and change frequencies	979	490	237	–	–	–	0·8	0·9	28840+	4

* After exclusion of uninformative residues.

Table 2. Grouping of *Candida* species with related teleomorphs and outgroup species according to phylogenetic analysis of partial actin gene sequences

Abbreviations: NA, not available; BD, buoyant density in caesium salt gradients.

Species	Strain*	Co-Q system†	G+C content (mol%)	
			Nuclear DNA†	Partial actin gene sequence
Group A				
Candida albicans	CBS 562^{NT}	9	34·9–37·3	38·1
Candida dubliniensis	CBS 7987^{T}	NA	35·4–36·8‡	39·4
Candida tropicalis	DUMC 29-3711	9	33·2–36·1	39·9
Candida maltosa	CBS 5611^{T}	9	35·6–36·8	39·7
Candida viswanathii	CBS 4024^{T}	9[a]	45·6–46·3	44·3
Candida parapsilosis	DUMC 31-1070	9	39·3–40·8	40·2
Group B				
Pichia guilliermondii/*Candida guilliermondii*	CBS 2030^{T}/CBS 566^{T}	9	44·4 (BD); 44·6	46·6/46·7
Candida zeylanoides	CBS 619^{NT}	9	54·4–57·6	50·5
Debaryomyces etchellsii	CBS 2011^{T}	9	38·5; 40·6 (BD)	43·2
Metschnikowia pulcherrima/*Candida pulcherrima*	CBS 2255^{T}/CBS 610^{NT}	9	45·6 (BD)	47·7/47·8
Clavispora lusitaniae/*Candida lusitaniae*	CBS 6936^{T}/CBS 4413^{T}	8	45·1 (BD)–45·7 (BD)	48·1
Clavispora opuntiae	CBS 7068^{T}	8	43·1 (BD)–44·1 (BD)	44·6
Group C				
Candida glabrata	DUMC 19-1112/CBS 138^{T}	6	38·5–40·2	45
Saccharomyces kluyveri	CBS 3082^{T}	6	41·9	43·1
Kluyveromyces polysporus	CBS 2163^{T}	6[b]	35·3 (BD)	43·1
Saccharomyces cerevisiae/*Candida robusta*	CBS 1171^{NT}/CBS 1907^{T}	6	38·5–41·1	43·4/43·5
Kluyveromyces marxianus/*Candida kefyr*	CBS 712^{T}/CBS 834^{T}	6	41·1–41·5 (BD)	43/43·2
Kluyveromyces lactis/*Candida sphaerica*	CBS 683^{NT}/CBS 141^{T}	6	39·9–40·8 (BD)	44·3
Group D				
Pichia fermentans/*Candida lambica*	CBS 187^{T}/CBS 1876^{T}	7	42·2; 43·1	43·7/43·5
Pichia norvegensis/*Candida norvegensis*	CBS 6564^{T}/CBS 1922^{T}	7	35·5–37·6	33/32·9
Issatchenkia orientalis/*Candida krusei*	CBS 5147^{T}/CBS 573^{T}	7	39·7–40·3	37·1
Pichia membranifaciens/*Candida valida*	CBS 107^{T}/CBS 638^{T}	7	41·5–44·3	41·4/41·5
Ungrouped				
Pichia jadinii/*Candida utilis*	CBS 1600^{T}/CBS 621^{T}	7	45·1 (BD)–45·8 (BD)	47/46·7
Saccharomycopsis capsularis	CBS 2519^{NT}	8	43·4	50·3
Outgroup				
Neurospora crassa	–	NA	54[c]	56·2
Schizosaccharomyces pombe	L 975	10	35·1–42	45·1
Filobasidiella neoformans var. *neoformans*	H 99	10	53·2–56·7	52·4

*CBS, Centraalbureau voor Schimmelcultures, The Netherlands; DUMC, Duke University Medical Centre, USA; T, ex-type culture; NT, ex-neotype culture. Strain H 99 was used by Cox *et al.* (1995); strain L 975 was used by Mertins & Gallwitz (1997).

† For the species, reviewed by Kurtzman & Fell (1998) except as indicated by *a* (CBS database; http://www.cbs.knaw.nl), *b* (Molnár *et al.*, 1996) or *c* (Villa & Storck, 1968).

‡ Personal communication, E. O'Neill.

simonious trees that were significantly shorter than the mean length of the random trees.

Amino acid sequences were deduced from nucleotide sequences. Heuristic analysis of these sequences was considered to be the most conservative phylogenetic examination. The tree revealed four major groups of species, subsequently referred to as species groups A, B, C and D (Table 2). Groups A and B formed the significantly supported cluster A/B.

The data for individual codon positions were considered separately in further analyses. The ability of these analyses to detect the groups delineated in amino acid analysis is summarized in Table 3. The comparison of data sets for the first, second and third positions showed that the high variability within the third position was coupled with the lowest CI and RI in all subsets analysed (Table 1). The single tree resulting from the analysis of the third position placed one of the outgroup taxa (*Schizosaccharomyces pombe*)

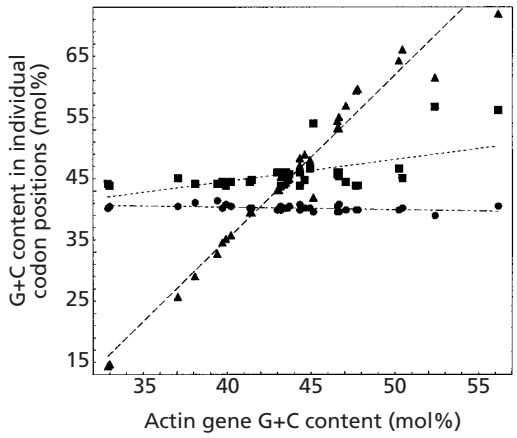

Fig. 1. Correlation of G+C content over all positions with first (■), second (●) and third (▲) positions of actin codons for 39 fungal taxa. The ranges of G+C content at each codon position were 43·0–56·7 mol% (first), 39·6–41·4 mol% (second) and 14·4–71·9 mol% (third). The r^2 values from linear regression for comparisons including/excluding the outgroup species were 0·312/0·141 (first), 0·202/0·222 (second) and 0·965/0·995 (third).

within the ingroup (data not shown), as sister group to *Pichia/Candida guilliermondii*, taxa with similar genic G+C contents. Separate analyses of the first and third positions showed different interrelationships among groups A, C and D (data not shown). Analysis of the second codon positions divided each of groups B and C into two neighbouring clades. When the first and second codon positions were analysed together, a lack of informative characters was reflected in short branches and polytomies in groups C and D.

Phylogeny reconstruction and weighting

The influences of different weighting schemes on phylogeny reconstruction were investigated. Four types of weights were used: (i) exclusion of third position transitions based on observed transitional bias, (ii) change-frequency weighting based on the assumption that fewer required steps in a character indicate stronger support for a tree, (iii) codon-position weighting as weighting of first, second and third positions in relation to redundancy of the genetic code and (iv) combinations of these. The character information content and statistical analyses for the different data sets under different weighting schemes are summarized in Table 1.

The analyses of third-position transversions alone and in combination with first and second positions detected groups A, C and D (Table 3). However, the interrelationships among these groups differed. These analyses failed to assign some species to any of the groups, but clustered the *Clavispora* and *Metschnikowia* species together. The CI and RI values were increased compared with the corresponding analyses involving the complete third positions (Table 1).

Change-frequency weights led to increased CIs and RIs as well as significantly increased bootstrap values in the heuristic searches on all combined data sets compared with their unweighted alternatives. However, the outgroup taxa were not separated in the expected order.

Codon-position weighting led to slightly lower values for CI and RI compared with the change-frequency weighting. Bootstrap values were increased compared with those of the unweighted analyses, but lower when compared with those of the change-frequency-weighted analyses. Analysis of the complete data set transformed by codon-position weighting distinguished group B, while analysis of the unweighted alternative failed to do so (Table 3). This type of weighting applied to the two recombined data sets (first plus second and first plus second plus recoded third positions) separated the outgroup taxa in the expected order.

Analysis of the two recombined data sets using the combined values incorporating both change-frequency and codon-position weights led to similar tree topologies. In the analysis of combined first and second positions, one taxon, *Kluyveromyces polysporus*, was separated from group C and appeared as a sister group to it (data not shown). The highest values for CI and RI were detected for the data sets of combined first and second positions and combined first, second and recoded third positions (Table 1). Bootstrap values were decreased compared with those assigned to change-frequency-weighted analyses. Analyses of these data sets showed identical interrelationships of the groups compared with the analysis of the deduced amino acids. However, the tree derived from the three combined positions was chosen for Fig. 3 to represent the complete data set.

Molecular analysis and phylogeny of *Candida* species

The basidiomycetous yeast *Filobasidiella neoformans*, the ascomycetous yeast *Schizosaccharomyces pombe* and the filamentous ascomycete *Neurospora crassa* were included as the outgroup taxa to root the trees obtained. Within the four major groups detected in the ingroup (Fig. 3; Table 2), all anamorph/teleomorph relationships were confirmed. Opportunistic pathogens did not cluster into a single group.

The tree including the highly variable third positions was used in order to detect the intragroup relationships (Fig. 3). Within group A, *Candida dubliniensis* was most closely related to *Candida albicans*. *Candida tropicalis*, *Candida maltosa* and *Candida viswanathii* were the next most closely related taxa to *Candida albicans*. *Candida parapsilosis* appeared basal (closer to the root) to the other species of this group. The *Clavispora* species formed one clade in group B. The relationships of other members of group B, *Pichia/Candida guilliermondii*, *Debaryomyces etchellsii*, *Metschnikowia/Candida pulcherrima* and *Candida zeylanoides* could not be resolved. Group C contained a cluster of *Kluyveromyces marxianus/Candida kefyr* with *Kluyveromyces lactis/*

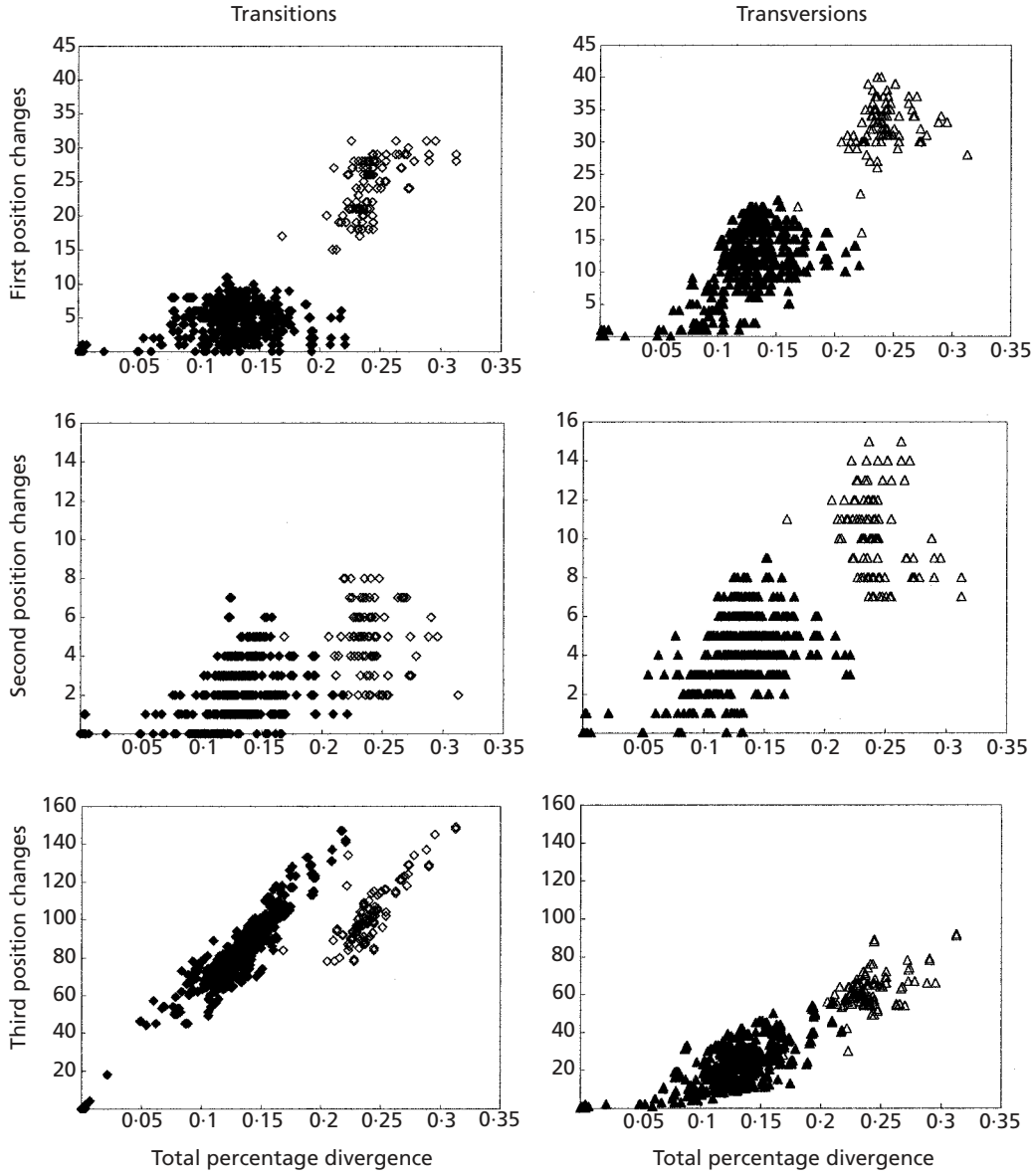

Fig. 2. Plots of pairwise transitional and transversional substitutions against total percentage divergence (*p*-distance; Kumar *et al.*, 1993) partitioned by codon position. Filled symbols refer to ingroup comparisons, open symbols refer to comparisons of outgroup species as well as to comparisons between ingroup and outgroup species. Diamonds refer to transitions, triangles to transversions.

Candida sphaerica. *Kluyveromyces polysporus* was basal to *Saccharomyces cerevisiae*/*Candida robusta* followed by *Candida glabrata* and *Saccharomyces kluyveri*. Group D showed a close relationship between *Pichia*/*Candida norvegensis* and *Issatchenkia krusei*/*Candida orientalis*. The relationships within this group could not be resolved further.

The sister-group relationship of groups A and B was resolved by several analyses, including the data set of the first and second codon positions and by the analysis of deduced amino acids (Table 3). Groups C and D were more closely related to the cluster A/B than were *Pichia jadinii* and *Candida utilis*. However, the order of these groups relative to the A/B cluster could not be resolved. Groups C and D were sister groups in the weighted and unweighted analyses of the complete data set and in the analysis weighted according to the change frequency of first and second positions. Other analyses showed that either species group C or D was closer to the A/B cluster, but none of these formations was favoured. *Saccharomycopsis capsularis* was placed basal to the remainder of the ingroup species.

Tree support

Bootstrap analyses revealed high support for the anamorph/teleomorph pairs in all analyses, except for

Table 3. Frequency of recovery of groups A, B, C and D and cluster A/B in 100 bootstrap replicates

Characters included	A	B	C	D	A/B	Outgroup*
Deduced amino acid sequence	84	50	45	69	95	82
All coding residues	73	–	57	91	–	–
All coding residues weighted according to:						
Codon position	100	71	66	96	76	–
Change frequencies	99	51	84	100	43	–
Codon position and change frequencies	100	85	73	98	87	69
1st positions	83	43	52	53	35	–
2nd positions	75	–	–	48	–	95
3rd positions	13	–	8	47	–	–
3rd transversions only	63	–	17	81	–	–
1st and 2nd positions	95	49	52	70	78	–
1st and 2nd positions weighted according to:						
Codon position	98	51	41	70	83	93
Change frequencies	96	63	83	92	93	–
Codon position and change frequencies	94	53	23	81	90	89
1st and 2nd positions and 3rd transversions	97	–	70	95	–	–
1st and 2nd positions and 3rd transversions weighted according to:						
Codon position	100	60	40	85	72	83
Change frequencies	100	–	91	98	–	–
Codon position and change frequencies	95	73	47	96	90	94

* Frequency with which the outgroup was detected as expected, that is *Neurospora crassa* closest to the ingroup while *Filobasidiella neoformans* was defined as the outgroup.

the two *Kluyveromyces* pairs, which were insufficiently resolved unless the complete data set, including the third position transitions, was applied (Fig. 3). The separation of the outgroup taxa from the ingroup was supported by significant bootstrap values (generally 70–100). Table 3 details bootstrap support for the groups detected, including the cluster A/B. The connecting node of groups A and B was supported significantly in most of the analyses that detected group B. The position of groups C and D could not be resolved at a significant bootstrap level.

DISCUSSION

Consideration of variable and informative characters at each of the different codon positions gave insight into the relative contribution of each position to the resolution of the trees that were generated (Table 1). The range of variability allowed us to distinguish intraspecific from interspecific variation. The overlap in levels of genetic divergence for outgroup–ingroup comparisons and some comparisons among ingroup species illustrated the breadth of diversity within the artificial form genus *Candida*. As most of this sequence variation was concentrated in the third codon position, several tests were carried out separately for this position as well as the other codon positions. Based on analysis of the genic G+C content (Fig. 1), the second position evolved independently from the overall G+C content and can be used to infer phylogenies for all species. The first position is informative for all ingroup species but not for ingroup–outgroup comparisons. The third position contains useful information for comparisons of species with similar G+C content. However, it would be misleading in the detection of phylogenetic relationships for species with deviating G+C contents. The detected correlation of the genic G+C content with the overall G+C content was observed previously in an analysis of actin gene sequences in eukaryotes (Bhattacharya & Ehlting, 1995). This, together with the biased base composition, is the basis for the application of weights according to codon position in our analysis. The biases detected in base composition at all three positions showed that one of the central assumptions of phylogenetic analysis, the independence of characters, was violated. Even though these positions are biased towards certain bases, the low species-to-species variation seen within this bias permits a valid phylogenetic analysis. Therefore, the compositional biases observed for first and second positions should not interfere with the phylogenetic analysis (Ducroz *et al.*, 1998).

The general bias of transitions over transversions (Brown *et al.*, 1982) was found only for the third position. The more frequent occurrence of transitions has been attributed to misincorporation of bases during replication rather than to other types of mutations. The conserved positions are unlikely to be strongly influenced by misincorporation, because any resulting amino acid changes are highly likely to eliminate functionality, leading to a lethal mutation. This could explain why first and second positions did

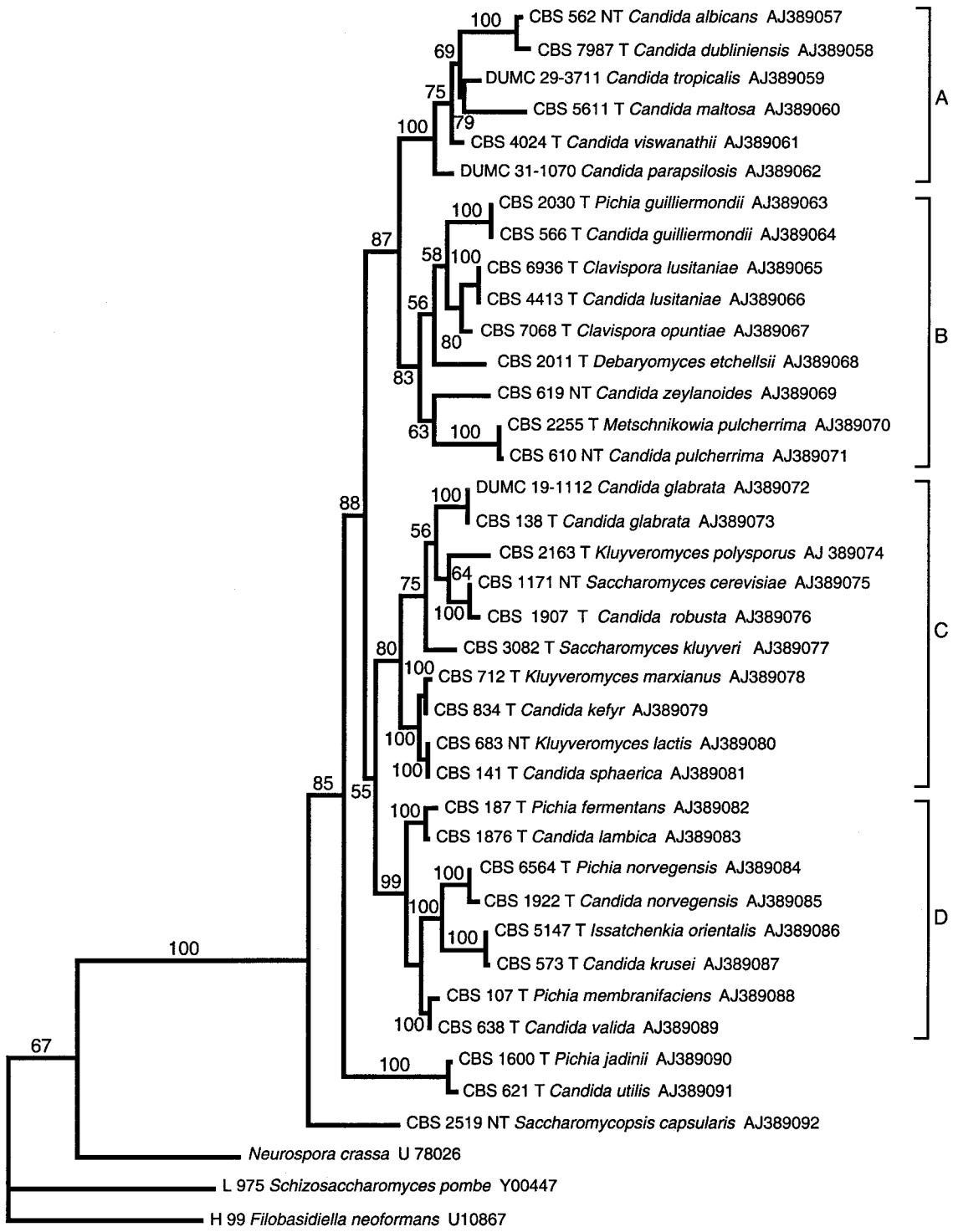

Fig. 3. Phylogenetic tree produced by weighted parsimony analysis of partial sequences of the actin gene from 39 yeast taxa. The tree was derived by a heuristic search (PAUP 3.1.1) from the complete data set of all three codon positions (979 bp), under application of combined weights for the codon positions and for character changes according to their frequency (MACCLADE 3.07). *Filobasidiella neoformans* was defined as outgroup. Numbers on branches indicate bootstrap values greater than 50 after 1000 replications. This single most parsimonious tree consists of 7946+steps (soft polytomies underestimate) and shows the indices CI = 0·57 (excluding uninformative characters) and RI = 0·79.

not follow the predicted transition–transversion bias. However, the reason was not clarified in this study. A level of transitions near 50% was proposed as an indication that sequences had become saturated with a maximal number of substitutions (Knight & Mindell, 1993). According to this, the third position transitions are not saturated for ingroup comparisons (64%), but reached a percentage that indicated saturation for outgroup comparisons (43%).

This difference between the more closely and distantly related clade was also observed in the plots shown in Fig. 2 of transitions and transversions against divergence distances (*p*-distances). In these plots, the numbers of transitions and transversions that have occurred over the evolutionary distance between all possible pairs of yeast species in the data set are mapped against *p*-distances. The approximately equal number of third position changes for ingroup and outgroup comparisons indicated saturation only for this class of changes. After determining which species comparisons accounted for the most divergent data points in this transition plot, it was found that the ingroup comparisons showed a distribution concordant with presumed evolutionary distances among the species (distant species showed the largest number of changes). Therefore, it was concluded that third position transitions were not saturated at the more recent evolutionary divergences (ingroup comparisons). However, they are not useful for the evaluation of more distant relationships (outgroup comparisons).

As it was shown that the three codon positions contain very different levels of useful phylogenetic information, separate phylogenetic analyses of each position were performed to test the congruence of the resulting trees with each other and with trees derived from amino acid sequences. The analyses of first and second positions gave generally compatible results. The analysis of third position data resulted in an obvious misplacement of an outgroup species, *Schizosaccharomyces pombe*. As an Archaeascomycete, this species is expected to be more distantly related to the ingroup than is the ascomycetous *Neurospora crassa*. However, third position analysis resolved closer relationships, as was expected, after these data confirmed empirically their relevance and importance (Kälersjö *et al*., 1999). The highly variable third position characters were shown to be influenced by compositional biases, such as the overall G+C content. The G+C content, which is not specific to single groups of fungi, tends to bring a degree of spurious similarity into comparisons among widely separated fungi that share this bias. The observed compositional bias (G+C content) confers structure on the data, especially potential homoplastic data. Therefore, they become misleading the more the G+C content departs from the mean range. In contrast, unstructured homoplasy, which can be understood as 'noise', hardly affects the probability of retrieving the best possible phylogeny (Wenzel & Siddall, 1999). In comparisons of very distantly related taxa, however, as had to be done in our study to include appropriate outgroups, frequently changing sites approached probable structured randomness and were not analytically useable without down-weighting the most frequent changes.

The codon-position weighting scheme and the exclusion of third positions or third position transitions were required so that codon positions other than the second position could be used to analyse the outgroup and ingroup species together. In addition, the same salvage of useful information could be accomplished by application of combined codon-position weight and change-frequency weights to the complete data set. The latter procedure results in only one tree, as opposed to the 203 and 16 trees offered by the alternative techniques. The results obtained by the weighted analyses were highly similar to those obtained with inferred amino acid sequences. Minor differences were observed in the branching order within the groups detected. Cautious weighting proved to be effective in optimizing the useable information that could be extracted from a data set that combined unequal rates of evolution, as is typically found in sequence analyses of genes that encode proteins subject to strong selection.

The actin gene tree showed that pathogenic *Candida* species were concentrated in, but not confined to, group A. The recent divergence of *Candida albicans* and *Candida dubliniensis* (Sullivan *et al*., 1995) was confirmed. A more distant relationship between *Candida albicans* and *Candida parapsilosis*, as reflected in actin sequences, is consistent with analyses of the large-subunit rRNA gene (Kurtzman & Robnett, 1998). Our analyses did not resolve details of the relationships between *Candida tropicalis*, *Candida maltosa* and *Candida viswanathii*, which clustered closely together within group A.

The clustering of *Clavispora* and *Metschnikowia* species with *Pichia/Candida guilliermondii*, *Debaryomyces etchellsii* and *Candida zeylanoides* in group B is different from the findings of most rDNA analyses, which place *Clavispora* and *Metschnikowia* species in a single clade, distantly related to any others. Their distinct relationship to other Hemiascomycetes is supported by the presence of large deletions in their rDNA sequences. However, no deletions are found in the actin genes of selected *Metschnikowia* and *Clavispora* species, due to strong selective forces operating on these sequences that lead to the observed high degree of conservation. Only one early analysis of the small-subunit rRNA gene (Barns *et al*., 1991) places *Candida lusitaniae* next to our group A, with its closest neighbour being *Candida guilliermondii*. The other three species of group B are reportedly close to *Candida albicans* and related species (Ohkuma *et al*., 1993; Cai *et al*., 1996; Mannarelli & Kurtzman, 1998). More comprehensive studies (Kurtzman & Robnett, 1997, 1998; Kurtzman & Blanz, 1998), showed several species that were not included in this study incor-

porated in clusters that included *Pichia/Candida guilliermondii*, *Debaryomyces etchellsii* and *Candida zeylanoides*. The difference between our well-defined group B and the more complex grouping seen in rDNA analyses may result from different evolution of the rDNA cluster and the actin gene. The relatively close relationship between our group A and the cluster containing *Clavispora* and *Metschnikowia* was confirmed by the significant bootstrap support for the link between groups A and B.

Within *Kluyveromyces*, two or three clusters have been identified by phenotypic and genotypic analyses (reviewed by Lachance, 1993) and by sequence analysis of the small subunit rDNA gene (Cai *et al*., 1996). The closely related *Kluyveromyces marxianus* and *Kluyveromyces lactis* belong to one of these previously recognized clusters that is considered to be monophyletic and well demarcated, worthy of recognition as a distinct genus as proposed by Cai *et al*. (1996). This cluster was found to be a sister group to other coenzyme-Q_6 (Co-Q_6) yeasts in our analysis, including two *Saccharomyces* species, *Candida glabrata* and *Kluyveromyces polysporus*. Our group C, which encompassed the *Kluyveromyces* species and their relatives, was the least robust of our four groups. It was split in the analysis of second codon positions and *Kluyveromyces polysporus* was separated from it in the analysis of the first and second codon positions, weighted according to codon position and change frequency. This was most likely due to the lack of informative characters in these positions, as group C was supported by significant bootstrap values in the weighted analysis of all three codon positions. The whole group C represents more than one genus, but all species within it are unified in the Co-Q_6 system (Cai *et al*., 1996).

Pichia is a very diverse genus that includes yeasts with Co-Q systems 7, 8 and 9. They are dispersed in six major clades throughout the ascomycetous yeasts, according to rDNA analyses (Kurtzman & Robnett, 1998). The *Pichia* species that were found in species group D appeared to be homogeneous. They are known to be closely related to cactophilic *Pichia* species (Shen & Lachance, 1993) and certain *Issatchenkia* species, e.g. *Issatchenkia orientalis*. *Pichia jadinii* constitutes a taxon separate from the others (Yamada *et al*., 1994, 1995; Kurtzman & Robnett, 1998).

Saccharomycopsis capsularis is classified in the Saccharomycopsidaceae, a distinct family with abundant development of true mycelium in some species and coenzyme Co-Q_8. The species of this family are known to form a separate clade at the generic level based on the large rRNA subunit (Kurtzman & Smith, 1998). Based on our analysis of the actin gene, *Saccharomycopsis capsularis* evolved separately from the other budding yeasts that we investigated.

Co-Q data and G+C contents (reviewed by Kurtzman & Fell, 1998; Molnár *et al*., 1996; Villa & Storck, 1968) provide useful characteristics in yeast taxonomy and were compared with the species groups that we detected (Table 2). The major Co-Q system is generally homogeneous among species within a monophyletic genus (Yamada *et al*., 1976). Nuclear G+C contents usually vary by no more than 10% within a monophyletic genus. Overlaps between unrelated taxa are known for both characters. Therefore, their homogeneity may or may not indicate monophyly, but their heterogeneity supports polyphyly at the genus level (Kurtzman, 1998). The G+C contents of the partial actin gene sequences were slightly higher than the nuclear G+C contents, when the latter was detected by buoyant density in caesium salt gradients. However, there were some species for which the nuclear G+C contents, determined by thermal denaturation, were higher than the genic G+C contents (e.g. *Candida viswanathii* and *Candida zeylanoides*). Group A (Co-Q_9) appeared to be homogeneous in its Co-Q system, although this was not determined for two species. However, the range of the nuclear G+C content in this group increased significantly from 7·6% in the monophyletic lineage excluding *Candida viswanathii* to 13·1% in the same lineage including *Candida viswanathii*. The genera *Metschnikowia* and *Clavispora*, combined in group B, are known to be only weakly associated (no significant bootstrap support in rDNA analyses, different Co-Q systems) (Kurtzman & Robnett, 1998). Therefore, the apparent close relationship between them could be an artefact of underestimation of their evolutionary divergence. Again, the range of the nuclear G+C contents increased from 7·2% in the main group to 19·1% after inclusion of *Candida zeylanoides*. Groups C (Co-Q_6) and D (Co-Q_7) were homogeneous in their Co-Q systems and in their ranges of nuclear G+C content (8·8 and 6·2%, respectively). These data support the monophyly of each of these groups. The actin tree showed coherence with the Co-Q systems and G+C levels within each of the groups.

We conclude that the actin tree is compatible with the phylogenies generated from rDNA sequences, other than in the positioning of the *Metschnikowia/Clavispora* clade. The identification of four potentially monophyletic groups within the diversity of the taxa investigated is a noteworthy advance in yeast systematics, as even the currently recognized teleomorph genera of ascomycetous yeasts are not monophyletic. The actin gene proved to be a valuable phylogenetic marker, sufficiently conserved to allow an unambiguous alignment and variable enough to resolve basal lineages significantly for the entire ingroup. It could be useful for the assignment of ascomycetous species to groups that should subsequently be resolved further using sequences of higher variability. Some species relationships within the detected groups could not be supported by significant bootstrap values. However, all anamorph/teleomorph pairs included were validated by high bootstrap values. The investigation of more species using actin gene sequences may reduce

the heterogeneity in such highly polyphyletic genera as *Pichia* and *Kluyveromyces* and may facilitate the search for as-yet unknown anamorph/teleomorph connections. The investigation of more genes (e.g. elongation factors or tubulin) is required for the further development of a natural classification of the Hemiascomycetes.

ACKNOWLEDGEMENTS

We are grateful to David Yarrow for providing type strains and advice on yeast taxonomy, Richard Summerbell and an anonymous referee for productive comments on the manuscript, Cletus Kurtzman for discussion of our results and Ok Cha Lee for technical support on the biochemical identification of yeast cultures. The work was supported by a project grant from the National Health & Medical Research Council, Australia (#960792), to W.M., a grant from the German Research Society (Me 1393/1-1) to W.M., two Grants-in-Aid from the Westmead Research Foundation, Australia, to W.M., an initiation grant from the Westmead Research Foundation to H.M.D. and a PhD scholarship from the Centre for Infectious Diseases and Microbiology, Australia, to H.M.D.

REFERENCES

Barns, S. M., Lane, D. J., Sogin, M. L., Bibeau, C. & Weisburg, W. G. (1991). Evolutionary relationships among pathogenic *Candida* species and relatives. *J Bacteriol* **173**, 2250–2255.

Berbee, M. L. & Taylor, J. W. (1992). Detecting morphological convergence in true fungi, using 18S rRNA gene sequence data. *Biosystems* **28**, 117–125.

Bhattacharya, D. & Ehlting, J. (1995). Actin coding regions: gene family evolution and use as a phylogenetic marker. *Arch Protistenkd* **145**, 155–164.

Bhattacharya, D., Stickel, S. K. & Sogin, M. L. (1991). Molecular phylogenetic analysis of actin genic regions from *Achlya bisexualis* (Oomycota) and *Costaria costata* (Chromophyta). *J Mol Evol* **33**, 525–536.

Bouget, F.-Y., Kerbourc'h, C., Liaud, M.-F., Loiseaux de Goer, S., Quatrano, R. S., Cerff, R. & Kloareg, B. (1995). Structural features and phylogeny of the actin gene of *Chondrus crispus* (Gigartinales, Rhodophyta). *Curr Genet* **28**, 164–172.

Brown, W. M., Prager, E. M., Wang, A. & Wilson, A. C. (1982). Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J Mol Evol* **18**, 225–239.

Cai, J., Roberts, I. N. & Collins, M. 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.

Cox, G. M., Rude, T. H., Dykstra, C. C. & Perfect, J. R. (1995). The actin gene from *Cryptococcus neoformans*: structure and phylogenetic analysis. *J Med Vet Mycol* **33**, 261–266.

Deshler, J. O., Larson, G. P. & Rossi, J. J. (1989). *Kluyveromyces lactis* maintains *Saccharomyces cerevisiae* intron-encoded splicing signals. *Mol Cell Biol* **9**, 2208–2213.

Drouin, G., Moniz de Sa, M. & Zuker, M. (1995). The *Giardia lamblia* actin gene and the phylogeny of eukaryotes. *J Mol Evol* **41**, 841–849.

Ducroz, J.-F., Volobouev, V. & Granjon, L. (1998). A molecular perspective on the systematics and evolution of the genus *Arvicanthis* (Rodentia, Muridae): inferences from complete cytochrome b gene sequences. *Mol Phylogenet Evol* **10**, 104–117.

Farris, J. S. (1966). Estimation of conservatism of characters by constancy within biological populations. *Evolution* **20**, 587–591.

Farris, J. S. (1989). The retention index and the rescaled consistency index. *Cladistics* **5**, 417–419.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Fitch, W. M. (1996). Uses for evolutionary trees. In *New Uses for New Phylogenies*, pp. 116–133. Edited by P. H. Harvey, A. J. L. Brown, J. M. Smith & S. Nee. Oxford: Oxford University Press.

Fuson, G. B., Price, C. W. & Phaff, H. J. (1980). Deoxyribonucleic acid base sequence relatedness among strains of *Pichia ohmeri* that produce dimorphic ascospores. *Int J Syst Bacteriol* **30**, 217–219.

Gallwitz, D. & Sures, I. (1980). Structure of a split yeast gene: complete nucleotide sequence of the actin gene in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **77**, 2546–2550.

Gilbert, D. (1996). SeqPup version 0.6d. Bloomington, IN: Indiana University.

Goloboff, P. A. (1991). Homoplasy and the choice among cladograms. *Cladistics* **7**, 215–232.

Hendriks, L., Goris, A., van de Peer, Y., Neefs, J. M., Vancanneyt, M., Kersters, K., Berny, J. F., Hennebert, G. L. & de Wachter, R. (1992). Phylogenetic relationships among ascomycetes and ascomycete-like yeasts as deduced from small ribosomal subunit RNA sequences. *Syst Appl Microbiol* **15**, 98–104.

Hillis, D. H. & Bull, J. J. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol* **42**, 182–192.

Hillis, D. M. & Huelsenbeck, J. P. (1992). Signal, noise, and reliability in molecular phylogenetic analyses. *J Hered* **83**, 189–195.

Huang, M.-E., Souciet, J.-L., Chuat, J.-C. & Galibert, F. (1996). Identification of act4, a novel essential actin-related gene in the yeast *Saccharomyces cerevisiae*. *Yeast* **12**, 839–848.

Johnson, K. P. & Sorenson, M. D. (1998). Comparing molecular evolution in two mitochondrial protein coding genes (cytochrome b and ND2) in the dabbling ducks (Tribe: Anatini). *Mol Phylogenet Evol* **10**, 82–94.

Kälersjö, M., Albert, V. A. & Farris, J. S. (1999). Homoplasy increases phylogenetic structure. *Cladistics* **15**, 91–93.

Kan, V. L. (1993). Polymerase chain reaction for the diagnosis of candidemia. *J Infect Dis* **168**, 779–783.

Kluge, A. G. & Farris, J. S. (1969). Quantitative phyletics and the evolution of anurans. *Syst Zool* **18**, 1–32.

Knight, A. & Mindell, D. P. (1993). Substitution bias, weighting of DNA sequence evolution, and the phylogenetic position of fea's viper. *Syst Biol* **42**, 18–31.

Kumar, S., Tamura, K. & Nei, M. (1993). MEGA: Molecular Evolutionary Genetics Analysis 1.01. Pennsylvania, PA: Pennsylvania State University.

Kurtzman, C. P. (1984). Synonymy of the yeast genera *Hansenula* and *Pichia* demonstrated through comparisons of deoxyribonucleic acid relatedness. *Antonie Leeuwenhoek* **50**, 209–217.

Kurtzman, C. P. (1998). Nuclear DNA hybridization: quantitation of close genetic relationships. In *The Yeasts, a Taxonomic Study*, 4th edn, pp. 63–68. Edited by C. P. Kurtzman & J. W. Fell. Amsterdam: Elsevier.

Kurtzman, C. P. & Blanz, P. A. (1998). Ribosomal RNA/DNA sequence comparisons for assessing phylogenetic relationships. In *The Yeasts, a Taxonomic Study*, 4th edn, pp. 69–74. Edited by C. P. Kurtzman & J. W. Fell. Amsterdam: Elsevier.

Kurtzman, C. P. & Fell, J. W. (editors) (1998). *The Yeasts, a Taxonomic Study*, 4th edn. Amsterdam: Elsevier

Kurtzman, C. P. & Robnett, C. J. (1994). Synonymy of the yeast genera *Wingea* and *Debaryomyces*. *Antonie Leeuwenhoek* **66**, 337–342.

Kurtzman, C. P. & Robnett, C. J. (1997). Identification of clinically important ascomycetous yeasts based on nucleotide divergence in the 5′ end of the large-subunit (26S) ribosomal DNA gene. *J Clin Microbiol* **35**, 1216–1223.

Kurtzman, C. P. & Robnett, C. J. (1998). Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Antonie Leeuwenhoek* **73**, 331–371.

Kurtzman, C. P. & Smith, M. T. (1998). *Saccharomycopsis* Schiönning. In *The Yeasts, a Taxonomic Study*, 4th edn, pp. 374–386. Edited by C. P. Kurtzman & J. W. Fell. Amsterdam: Elsevier.

Lachance, M.-A. (1993). *Kluyveromyces*: systematics since 1970. *Antonie Leeuwenhoek* **63**, 95–104.

Liu, Y. J., Whelen, S. & Hall, B. D. (1999). Phylogenetic relationships among ascomycetes: evidence from an RNA polymerase II subunit. *Mol Biol Evol* **16**, 1799–1808.

Lloyd, A. T. & Sharp, P. M. (1992). Evolution of codon usage patterns: the extent and nature of divergence between *Candida albicans* and *Saccharomyces cerevisiae*. *Nucleic Acids Res* **20**, 5289–5295.

Losberger, C. & Ernst, J. F. (1989). Sequence of the *Candida albicans* gene encoding actin. *Nucleic Acids Res* **17**, 9488.

Lott, T. J., Kuykendall, R. J. & Reiss, E. (1993). Nucleotide sequence analysis of the 5·8S rDNA and adjacent ITS2 region of *Candida albicans* and related species. *Yeast* **9**, 1199–1206.

Maddison, W. P. & Maddison, D. R. (1997). MacClade. Analysis of Phylogeny and Character Evolution, version 3.07. Sunderland, MA: Sinauer Associates.

Mannarelli, B. M. & Kurtzman, C. P. (1998). Rapid identification of *Candida albicans* and other human pathogenic yeasts by using short oligonucleotides in a PCR. *J Clin Microbiol* **36**, 1634–1641.

Mendonça-Hagler, L. C., Hagler, A. N. & Kurtzman, C. P. (1993). Phylogeny of *Metschnikowia* species estimated from partial rRNA sequences. *Int J Syst Bacteriol* **43**, 368–373.

Mertins, P. & Gallwitz, D. (1987). A single intronless actin gene in the fission yeast *Schizosaccharomyces pombe*: nucleotide sequence and transcripts formed in homologous and heterologous yeast. *Nucleic Acids Res* **15**, 7369–7379.

Meyer, W., Latouche, G. N., Daniel, H.-M., Thanos, M., Mitchell, T. G., Yarrow, D., Schönian, G. & Sorrell, T. C. (1997). Identification of pathogenic yeasts of the imperfect genus *Candida* by polymerase chain reaction fingerprinting. *Electrophoresis* **18**, 1548–1559.

Meyer, S. A., Payne, R. W. & Yarrow, D. (1998). *Candida* Berkout. In *The Yeasts, a Taxonomic Study*, 4th edn, pp. 454–573. Edited by C. P. Kurtzman & J. W. Fell. Amsterdam: Elsevier.

Molnár, O., Prillinger, H., Lopandic, K., Weigang, F. & Staudacher, E. (1996). Analysis of coenzyme Q systems, monosaccharide patterns of purified cell walls, and RAPD-PCR patterns in the genus *Kluyveromyces*. *Antonie Leeuwenhoek* **70**, 67–78.

Montrocher, R., Verner, M.-C., Briolay, J., Gautier, C. & Marmeisse, R. (1998). Phylogenetic analysis of the *Saccharomyces cerevisiae* group based on polymorphisms of rDNA spacer sequences. *Int J Syst Bacteriol* **48**, 295–303.

Ng, R. & Abelson, J. (1980). Isolation and sequence of the gene for actin in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **77**, 3912–3916.

Ohkuma, M., Hwang, C. W., Masuda, Y., Nishida, H., Sugiyama, J., Ohta, A. & Takagi, M. (1993). Evolutionary position of *n*-alkane-assimilating yeast *Candida maltosa* shown by nucleotide sequence of small-subunit ribosomal RNA gene. *Biosci Biotechnol Biochem* **57**, 1793–1794.

Price, C. W., Fuson, G. B. & Phaff, H. J. (1978). Genome comparison in yeast systematics: delimitation of species within the genera *Schwanniomyces*, *Saccharomyces*, *Debaryomyces*, and *Pichia*. *Microbiol Rev* **42**, 161–193.

Shen, R. & Lachance, M.-A. (1993). Phylogenetic study of ribosomal DNA of cactophilic *Pichia* species by restriction mapping. *Yeast* **9**, 315–330.

Sheterline, P. & Sparrow, J. C. (1994). Actin. *Protein Profile* **1**, 1–121.

Starmer, W. T., Phaff, H. J., Miranda, M. & Miller, M. W. (1978). *Pichia amethionina*, a new heterothallic yeast associated with the decaying stems of cereoid cacti. *Int J Syst Bacteriol* **28**, 433–441.

Sullivan, D. J., Westerneng, T. J., Haynes, K. A., Bennett, D. E. & Coleman, D. C. (1995). *Candida dubliniensis* sp. nov.: phenotypic and molecular characterization of a novel species associated with oral candidosis in HIV-infected individuals. *Microbiology* **141**, 1507–1521.

Swofford, D. L. (1993). PAUP, Phylogenetic Analysis Using Parsimony, version 3.1.1Champaign, IL: Illinois Natural History Survey.

Villa, V. D. & Storck, R. (1968). Nucleotide composition of nuclear and mitochondrial deoxyribonucleic acid of fungi. *J Bacteriol* **96**, 184–190.

Wenzel, J. W. & Siddall, M. E. (1999). Noise. *Cladistics* **15**, 51–64.

Wery, J., Dalderup, M. J. M., ter Linde, J., Boekhout, T. & van Ooyen, A. J. J. (1996). Structural and phylogenetic analysis of the actin gene from the yeast *Phaffia rhodozyma*. *Yeast* **12**, 641–651.

Yamada, Y., Nojiri, M., Matsuyama, M. & Kondo, K. (1976). Coenzyme Q system in the classification of the ascosporogenous yeast genera *Debaryomyces*, *Saccharomyces*, *Kluyveromyces*, and *Endomycopsis*. *J Gen Appl Microbiol* **22**, 325–337.

Yamada, Y., Maeda, K. & Mikata, K. (1994). The phylogenetic relationships of the hat-shaped ascospore-forming, nitrate-assimilating *Pichia* species, formerly classified in the genus *Hansenula* Sydow et Sydow, based on the partial sequences of 18S and 26S ribosomal RNAs (Saccharomycetaceae): the proposals of three new genera, *Ogataea*, *Kuraishia*, and *Nakazawaea*. *Biosci Biotechnol Biochem* **58**, 1245–1257.

Yamada, Y., Matsuda, M. & Mikata, K. (1995). The phylogenetic relationships of *Pichia jadinii*, formerly classified in the genus *Hansenula*, and related species based on the partial sequences of 18S and 26S ribosomal RNAs (Saccharomycetaceae). *Biosci Biotechnol Biochem* **59**, 518–520.