

The use of parsimony network analysis for the formal delineation of phylogenetic species of yeasts: *Candida apicola*, *Candida azyma*, and *Candida parazyza* sp. nov., cosmopolitan yeasts associated with floricolous insects

Marc-André Lachance · Jessica Dobson ·
Dilini N. Wijayanayaka · Allison M. E. Smith

Received: 18 August 2009 / Accepted: 12 November 2009 / Published online: 26 November 2009
© Springer Science+Business Media B.V. 2009

Abstract Parsimony network analysis of rDNA sequences was used to delimit phylogenetic species of yeasts in an objective, formal manner. Many strains assigned to *Candida apicola* (*Starmerella* clade), when compared to the type, fell outside the inclusion limits proposed by Kurtzman and Robnett (1998) based on a pair-wise comparison of the large subunit rRNA gene D1/D2 domains. However, when these sequences were analyzed jointly with ITS rDNA sequences by parsimony network analysis, 28 of the 30 strains formed a cohesive set. Two strains, MUCL 45721 and CBS 4353, were excluded from the species, but there was no evident justification to subdivide the rest. A similar analysis of 81 isolates originally assigned to *Candida azyma* (*Wickerhamiella* clade) yielded dramatically different results, giving rise to six independent networks corresponding to *Candida azyma* sensu stricto (18 strains), *Candida azymoides* (2 strains), a pair of isolates from Australian hibiscus flowers, a single isolate from the same substrate, a single isolate from Malaysian bertam palm nectar, and 57 isolates that are assigned to the new species *Candida parazyza* (type = UWOPS 91-652.1^T = CBS 11563^T = NRRL Y-48669^T). The strains retained in *C. azyma* sensu stricto differed from one another by

up to four substitutions in their D1/D2 sequences, but their polymorphism at the level of the ITS was considerable and suggested a history of divergence resulting from dispersal. Strains of *C. parazyza* fell into seven variant haplotypes based on sequences of the rDNA ITS and D1/D2 regions. The most abundant haplotype occurred across the global range of the species. Others were either endemic to Belize, Costa Rica, Rarotonga, or Tennessee, suggestive of vicariance, or occurred across remote localities, offering partial support to the notion of rapid dispersal.

Keywords *Candida apicola* · *Candida azyma* · *Candida parazyza* · Parsimony networks · Phylogenetic species delineation

Introduction

DNA sequencing has become a central tool in yeast systematics. The earliest and most widespread application has been in yeast identification, an approach which is now applied to a vast array of organisms, including plants and animals, and referred to as “DNA barcoding”. The use of sequences in yeast identification is made possible through complete databases of sequences for the large nuclear ribosomal RNA gene (D1/D2 domains), initially available due to the efforts of Kurtzman and Robnett (1998) and Fell et al. (2000). The databases have been subsequently updated by authors of novel

M.-A. Lachance (✉) · J. Dobson · D. N. Wijayanayaka ·
A. M. E. Smith
Department of Biology, University of Western Ontario,
London, ON N6A 5B7, Canada
e-mail: lachance@uwo.ca

species. D1/D2 sequences also serve to generate provisional phylogenies that can be later refined by the inclusion of well-selected sequences that represent many independent, orthologous genes.

Relevant to the present study, D1/D2 sequences have been used extensively as a means of delimiting species. Kurtzman and Robnett (1998) observed that ascomycetous yeast strains that were considered conspecific based on reproductive compatibility or DNA/DNA reassociation rarely differed by more than three substitutions and that strains known to represent different species generally differed by six or more substitutions. On that basis, the authors stated (our italics): “it is *predicted* that strains showing greater than 1% substitutions in the ca. 600-nucleotide D1/D2 domain *are likely to be* different species and that strains with 0–3 nucleotide differences are either conspecific or sister species.” It would appear, however, that the prediction may have in some cases acquired a more axiomatic spirit, as exemplified by a later account by Kurtzman and Droby (2001, our italics): “Kurtzman and Robnett (1998) *demonstrated* for ascomycetous yeasts that strains differing by more than 1% substitutions in the D1/D2 domain *represent* separate species.” This notion has in many circles been accepted as a stand-alone criterion upon which to delimit species.

Whereas species delineated from DNA sequence divergence may indeed, in the majority of cases, match taxa defined on the basis of reproductive isolation, the approach suffers from a number of unresolved problems. For example, the species studied by Kurtzman and Robnett (1998) were often represented by very few isolates and in most cases, a single isolate, precluding an appraisal of the distribution of polymorphic sites within species. Yet, polymorphisms do occur at the level of barcode sequences. Kurtzman and Robnett (1998) did report a divergence of five substitutions between mating strains of a species (*Metschnikowia agaves*), although only the sequence of the type was deposited in GenBank. Later, Lachance et al. (2003) observed that mating strains of *Clavispora lusitaniae* that give rise to sporulating asci possess two divergent families of D2 sequences. The variants differed by as many as 32 substitutions, were found in strains of either mating type, and were uncorrelated with the few growth characteristics that are polymorphic. Most importantly, a few strains possessed mixtures of both D2 variants, confirming the existence of recombination

among species members. At the other extreme, some closely related but nonetheless reproductively isolated species (Lachance 1993) may have fewer than three substitutions, as in the case of *Kluyveromyces marxianus* and *Kluyveromyces lactis*, with only two substitutions. A similar difficulty was noted for basidiomycetous yeasts by Fell et al. (2000), who recommended that sequences of the internal transcribed spacers and the 5.8S rRNA gene be considered also in the separation of certain species.

The extent of divergence in barcoding sequences has enormous potential to assist in the delineation of species, particularly in yeasts with diplontic or automictic sexual cycles, or no sex at all. However, the significance of polymorphism in those sequences has not been properly assessed. For example, six species have been described, in the last decade, on the basis that they differ from *Metschnikowia pulcherrima* by the required number of D1/D2 substitutions (Kurtzman and Droby 2001; Molnár and Prillinger 2005; Suh et al. 2004a; Xue et al. 2006). These yeasts are otherwise virtually indistinguishable from *M. pulcherrima*. All are diplontic and form ascospores from chlamydospores without conjugation. When many available sequences of *M. pulcherrima* relatives were compared, as done to some extent by Molnár and Prillinger (2005) or Sipiczki (2006), sharp discontinuities were not so easily observed. The distinct status of the six new species therefore remains debatable, because it is not clear whether the variant sequences should be treated as orthologs (same ancestry, different species) or simply as alleles of a polymorphic locus. At the other extreme, three recent studies (Groenewald et al. 2008; Nguyen et al. 2009; Jacques et al. 2009) focused on automictic yeasts that would, based solely on Kurtzman and Robnett’s (1998) criterion, be assigned to the single species *Debaryomyces hansenii*. Using an eclectic palette of approaches, the various authors assigned up to 170 strains to several distinct species.

Yeasts of ascomycetous affinity but for which a sexual cycle has not been identified are usually assigned to the genus *Candida*. A notable consequence of Kurtzman and Robnett’s (1998) publication has been the subsequent description of a phenomenal number of putatively new *Candida* species, with very little attention to potential polymorphisms in barcoding sequences. We suggest that an analysis of multiple, independent isolates is particularly important when

dealing with asexual forms, for several reasons. First, descriptions based on single isolates or isolates that may not be genetically distinct preclude the observation of a sexual cycle in heterothallic haplonts. Second, such descriptions overlook the fact that, like most biological phenomena, the attributes of a species are characterized by a norm and a range. Single isolates have no range and consequently no norm. The third reason is more complex and requires a more detailed exposition.

The Darwinian model of common descent holds that species are the result of tree-like divergence and predicts that species, by loss of intermediate types, should form discontinuous clusters in sequence divergence space. As Darwin himself stated: “the only distinction between species and well-marked varieties is, that the latter are known, or believed, to be connected at the present day by intermediate gradations, whereas species were formerly thus connected.” (Darwin 1859). Our task as systematists is therefore to identify such discontinuities and use them as species boundaries. One should not, however, confuse discontinuities that are artefacts of insufficient sampling with those that arose as a consequence of the speciation process. And although well sampled yeasts sometimes do form discontinuous clades that conform to Kurtzman and Robnett’s (1998) criterion (Suh et al. 2004b), this is not always the case. Barcoding sequences such as D1/D2 and ITS have the potential to assist systematists in delimiting phylogenetic species in a non-arbitrary fashion. This requires a formal, statistically-based means of sorting out isolates whose sequences do not exhibit clear hierarchical patterns.

In phylogenetic trees derived from sequence data, extant taxa are positioned at terminal nodes and the internal nodes are meant to represent ancestral taxa that no longer exist. Whether the extant taxa in a particular clade should be assigned to a species remains a matter to be adjudicated by a taxonomist. Although there is a consensus that phylogenetically defined species should be monophyletic, there is little agreement as to the degree of inclusivity required (Taylor et al. 2000). Parsimony networks, by contrast, are grounded on the assumption that the sequences in a network represent alleles of a locus within a single species (Posada and Crandall 2001). The network is expected to contain both derived and ancestral sequences, and the most ancestral sequence is normally the most abundant. Only sequences that

are deemed, based on coalescent theory, to represent alleles in a conspecific population are accepted into the network. Sequences interpreted as orthologs from distinct species are treated as members of distinct networks. In the light of these principles, Hart and Sunday (2007) reviewed the application of parsimony networks to the analysis of barcoding sequences in species of a wide range of plants and animals, and reached the conclusion that the method correctly identifies members of a species in a majority of cases. This paper examines parsimony network analysis as a means of delineating asexual yeast species by unravelling significant discontinuities in strains once assigned to *Candida apicola* and *Candida azyma*.

C. apicola was until recently known from just a handful of strains of diverse origins (Meyer et al. 1998). Examination by sequencing (Lachance, unpublished) of some strains deposited by various authors as *C. apicola* in the CBS showed that misidentification based on growth tests had been frequent. Strains CBS 5710 and CBS 6366 were re-identified as *Starmerella bombicola* and strain CBS 7444 was reassigned to *Candida sorbosivorans* (M.A. Lachance, unpublished data). The correct identification of strains CBS 4353 (AY574387) and CBS 8413 (AY574388), which differ from the type of *C. apicola* by 13 and 5 substitutions in D1/D2 sequences, respectively, was more problematic, especially in the context of the isolation of several new strains from tropical bees (Rosa et al. 2003). It is possible that members of the *Starmerella* clade form unusually long branches in phylogenetic trees constructed from rRNA gene sequences because of high speciation rates, as suggested more generally by Webster et al. (2003). However, the relationship between sequence divergence and species density is not simple (Wright et al. 2006) and a theoretical framework linking the two in yeasts is lacking. In consequence, a naïve application of Kurtzman and Robnett’s (1998) criterion to species delineation in *C. apicola* could be misleading. We shall address the question by evaluating the depth of sequence discontinuities through parsimony network analysis.

C. azyma was until recently known (Meyer et al. 1998) mostly from a pair of isolates recovered from lichen in South Africa. However, intensive isolations from Convolvulaceae flowers and associated insects have yielded large numbers of strains that were provisionally assigned to the species on the basis of growth responses (Lachance et al. 2001). Rosa et al.

(2006) described the close relative *Candida azymoides* to accommodate four isolates from tropical fruit and associated fly larvae, based on a divergence from the type of *C. azyma* of seven substitutions in D1/D2 sequences. At the time, it was noted that strain ST19 (DQ400365, S. Jindamorakot, S. Limtong, T. Nakase) differed from the type of *C. azyma* by four substitutions. Further sequence analyses indicated that *C. azyma* may in fact be a cryptic complex of strains that in some cases should be treated as members of separate species. We now report on these analyses and their interpretation by parsimony network analysis. We also propose the new species *Candida parazyama* to accommodate a large number of strains that exhibit major sequence differences from *C. azyma*.

Methods

Strains were obtained through studies conducted in this laboratory or from other collections, as specified in Tables 1, 2, and 3. They were characterized by standard growth tests (Yarrow 1998). Attempts to induce ascus formation made use of several of the usual media, including McClary's acetate agar and 5% malt agar. In addition, strains were mixed in pairs on YCBAS (Yeast Carbon Base, Difco supplemented with 0.01% ammonium sulphate) agar and GY (glucose 1%, yeast extract 0.01%) agar to detect conjugation. The ITS–D1/D2 rDNA segment was amplified directly from whole cells as specified by Marinoni and Lachance (2004) and sequenced at the Robarts Research Institute, London, Ontario. Sequences were edited and aligned with the programs Chromas (Technelysium Pty Ltd) and DNAMAN (Lynnon Biosoft). Parsimony networks were constructed from aligned sequences with the program TCS 1.21 (Clement et al. 2000). Importantly, gapped positions were excluded from the analyses.

Results

Candida apicola

Thirty strains at one time assigned to *C. apicola* were considered. The parsimony network based on D1/D2 sequences (Fig. 1) excluded only strain CBS 4353 at

the 95% connection limit. Nine strains differed by three or fewer substitutions from the type (Fig. 1, subset A) and thus fit Kurtzman and Robnett's (1998) criterion for conspecificity. Another six strains (Fig. 1, subset B) differed from the type by at least five substitutions, but they were connected to one or more members of subset A by three or fewer substitutions. The remaining strains in the network differed from the type by six or fewer substitutions, except for strain MUCL 45721, with 12 substitutions. The sequence for strain UWOPS 01-663b2, identified as ancestral by the program TCS, was found in only one strain, but its putative ancestral nature is ostensibly linked to the fact that it is connected by single steps to three other sequences and multiple steps to two more.

Inclusion of ITS sequences (Fig. 2) caused no major changes in overall topology. However, strain MUCL 45721 was excluded from the network. Three additional strains joined subset B to form subset C, within which the ITS/5.8S region was invariant. Here, the sequence considered ancestral did coincide with the most abundant rDNA haplotype in the sample. In both networks (Figs. 1, 2), three adjacent haplotypes (subset D) had reticulate connections to the rest of the network. Inspection of the sequence alignment revealed that the haplotypes in subset D have in common only two D1/D2 and five ITS substitutions that differentiate them unambiguously from any haplotype in the other subsets. The remaining variation consists of homoplastic polymorphic sites that do not follow that pattern, resulting in two equally parsimonious paths by which to connect subset D to the rest.

Careful examination of the growth characteristics (not shown) of the strains in the light of the sequence-based analyses indicated that the little variation observed in growth abilities was independent of the sequence patterns, with three exceptions. Strain CBS 4353, which was excluded from both networks, failed to grow at 30°C and was the only strain to exhibit strong growth in the presence of 50% glucose. Strain MUCL 45721, excluded from the network with combined sequences, did not utilize sucrose or raffinose and did not grow on YM agar in the presence of 6% ethanol. Interestingly, strain CBS 8413, which can be linked to the type through a chain of intermediates that differ by no more than two substitutions in the D1/D2 region, also failed to grow at 30°C or in the presence of 6% ethanol. To place these results in perspective, all

Table 1 Selected characteristics of strains of *Candida apicola* and relatives

Strain number	GenBank accession number(s) or sequence identity	Source
CBS 1887	Identical to UFMG Jat 128.2	Fermenting cucumbers, USA
CBS 1888	Identical to UFMG Jat 128.2	Fermenting cucumbers, USA
CBS 2222	Identical to UFMG Jat 128.2	Sputum of patient with tuberculosis, Norway
CBS 2650	Identical to CBS 2868 ^T	Beehive mite biocontrol preparation, France
CBS 2868 ^T = NRRL Y-2481 ^T	EU926482, U45703 ^a	Gut of bee, Croatia
CBS 2869	Identical to CBS 2868 ^T	Gut of bee, Croatia
CBS 4078	EU926481	Grape must, Japan
CBS 4353	EU926486, AY574387	Sugar, United Kingdom
CBS 8413	EU926483, AY574388	Brine bath of cheese factory, The Netherlands
MUCL 45721/CRGF 96	FM209493, FM209492	Commercial bee pollen, Cuba
UFMG 07-MC1	EU926479	<i>Melipona</i> sp., Brazil
UFMG Jat 35.1	Identical to UFMG Jat 128.2	<i>Melipona quadrifasciata</i> hive, Brazil
UFMG Jat 128.2	EU926484, AY257054	<i>Melipona rufiventris</i> , Brazil
UFMG MP448	EU926477	<i>Melipona rufiventris</i> pollen, Brazil
UFMG Jat 515.1	EU926478	<i>Melipona rufiventris</i> pollen, Brazil
UWOPS 01-108b2	EU926474, AY257052	Beetle (unidentified) in <i>Ipomoea batatoides</i> , Costa Rica
UWOPS 01-125.4	Identical to UWOPS 01-193.1	<i>Conotelus</i> sp. in <i>Ipomoea batatas</i> , Costa Rica
UWOPS 01-131.1	Identical to UWOPS 01-193.1	<i>Trigona</i> sp. in Bignoniaceae flower, Costa Rica
UWOPS 01-132.1	EU926485, AY257055	<i>Trigona</i> sp. in Bignoniaceae flower, Costa Rica
UWOPS 01-133.1	Identical to UWOPS 01-132.1	<i>Trigona</i> sp. in Bignoniaceae flower, Costa Rica
UWOPS 01-135a2	Identical to UWOPS 01-132.1	<i>Trigona</i> sp. in Bignoniaceae flower, Costa Rica
UWOPS 01-135b3	EU926475, AY257053	Wasp (unidentified) in Bignoniaceae flower, Costa Rica
UWOPS 01-135c3	Identical to UWOPS 01-135b3	<i>Trigona</i> sp. in Bignoniaceae flower, Costa Rica
UWOPS 01-153b1	Identical to UWOPS 01-108b2	<i>Diabrotica</i> sp. in <i>Merremia tuberosa</i> flower, Costa Rica
UWOPS 01-193.1	EU926476, AY257051	<i>Trigona</i> sp. in flower of unidentified tree, Costa Rica
UWOPS 01-194.1	Identical to UWOPS 01-193.1	<i>Trigona</i> sp. in flower of unidentified tree, Costa Rica
UWOPS 01-663b2	EU926480	<i>Merremia tuberosa</i> flower, Costa Rica
UWOPS 01-677b2	Identical to UWOPS 01-108b2	<i>Conotelus</i> sp. in <i>Merremia tuberosa</i> , Costa Rica
UWOPS 05-219.3	Identical to CBS 4078	<i>Trigona</i> sp. on flower bud of <i>Eugeissona tristis</i> , Malaysia
UWOPS 05-243.4	Identical to CBS 4078	<i>Trigona</i> sp. on flower bud of <i>Eugeissona tristis</i> , Malaysia

Accession numbers for the ITS/5.8S rRNA gene region and the LSU rRNA D1/D2 domains are given separately or for the complete segment, as appropriate

CBS Centraalbureau voor Schimmelcultures, The Netherlands; NRRL Agricultural Research Service Culture Collection, National Center for Agricultural Utilization Research, Peoria, IL, USA; MUCL Mycothèque de l'Université catholique de Louvain, Belgium; CRGF Colección de Recursos Genéticos Fungicos; UFMG culture collection of the Universidade Federal de Minas Gerais, Brazil; UWOPS culture collection of the Department of Biology (formerly Plant Sciences), University of Western Ontario

^a Kurtzman and Robnett (2003)

^T Type strain

other strains utilized sucrose, raffinose, mannitol, and glucitol consistently, but gave variable responses for L-sorbose, D-ribose, glycerol, succinic acid, citric acid, ethylamine, and growth in the presence of 10% NaCl.

Candida azyma

Eighty-one strains at one time assigned to *Candida azyma* on the basis of growth tests were examined. As

Table 2 Selected properties of strains of *Candida azyma* and related species

Strain number	Identification	D1/D2 divergence from type	GenBank accession number	Locality	Source
CBS 6826 ^T	<i>Candida azyma</i>	(0)	EF536346	South Africa	Lichen
UWOPS 95-764.4	<i>Candida azyma</i>	2		Queensland, Australia	<i>Ipomoea indica</i> flower
UWOPS 98-119.5	<i>Candida azyma</i>	2	EF533997	Northern Territory, Australia	<i>Hibiscus meraukensis</i> flower
UWOPS 05-260.1	<i>Candida azyma</i>	2	EF533996	Cameron Highlands, Malaysia	<i>Ipomoea cairica</i> nitidulid
UWOPS 05-264.3	<i>Candida azyma</i>	2		Cameron Highlands, Malaysia	<i>Ipomoea cairica</i> <i>Drosophila</i> sp.
UWOPS 05-265.2	<i>Candida azyma</i>	2		Cameron Highlands, Malaysia	<i>Ipomoea cairica</i> nitidulid
UWOPS 05-270.2	<i>Candida azyma</i>	2		Cameron Highlands, Malaysia	<i>Ipomoea indica</i> nitidulid
UWOPS 05-271.2	<i>Candida azyma</i>	2		Cameron Highlands, Malaysia	<i>Ipomoea indica</i> nitidulid
UWOPS 05-272.1	<i>Candida azyma</i>	2		Cameron Highlands, Malaysia	<i>Ipomoea indica</i> nitidulid
UWOPS 05-261.1	<i>Candida azyma</i>	2		Cameron Highlands, Malaysia	<i>Ipomoea cairica</i> nitidulid
UWOPS 05-267.1	<i>Candida azyma</i>	2	EU203671	Cameron Highlands, Malaysia	<i>Ipomoea cairica</i> nitidulid
UWOPS 04-111.2	<i>Candida azyma</i>	3	EF536347	Molokai, Hawai'i	<i>Clermontia</i> sp. flower
UWOPS 04-157.5	<i>Candida azyma</i>	3		Maui, Hawai'i	<i>Clermontia arborescens</i> flower
UWOPS 04-158.3	<i>Candida azyma</i>	3		Maui, Hawai'i	<i>Clermontia arborescens</i> flower
UWOPS 95-693.4	<i>Candida azyma</i>	3	EF601042	Queensland, Australia	<i>Hibiscus heterophyllus</i> flower
UWOPS 95-766.5	<i>Candida azyma</i>	3		Queensland, Australia	<i>Ipomoea indica</i> flower
UWOPS 04-160.6	<i>Candida azyma</i>	3		Maui, Hawai'i	<i>Clermontia arborescens</i> flower
ST-19 ^a	<i>Candida azyma</i>	3		Thailand	Unknown
UFMG R287	<i>Candida azymoides</i>	7	DQ985171	Sergipe, Brazil	<i>Eugenia uniflora</i> fruit
UFMG 05-201.2	<i>Candida azymoides</i>	7		Tocantins, Brazil	<i>Anastrepha mucronota</i> larvae in <i>Peritassa campestris</i> fruit
UWOPS 95-805.2	<i>Candida</i> sp.	9	EF601043	Australia	<i>Hibiscus heterophyllus</i> flower
UWOPS 95-813.3	<i>Candida</i> sp.	9		Australia	<i>Hibiscus heterophyllus</i> flower
UWOPS 95-863.2	<i>Candida</i> sp.	10	EF601041	Australia	<i>Hibiscus heterophyllus</i> flower
UWOPS 03-446.4	<i>Candida</i> sp.	62	EU443389	Malaysia	<i>Eugeissona tristis</i> nectar
UWOPS 91-652.1 ^T	<i>Candida parazyza</i>	70	EF601044	Hawaii	<i>Drosophila floricola</i> in <i>Ipomoea indica</i> flower
CBS 11563 ^T NRRL Y-48669 ^T					

The number of substitutions separating the rRNA LSU gene D1/D2 domains from that of the type of *C. azyma* is given, along with the accession number for the entire ITS-D1/D2 rDNA segment. Horizontal lines separate strains that had identical sequences

^T Type strain

^a BIOTEC Culture Collection, National Center for Genetic Engineering and Biotechnology (BIOTEC), 113 Paholyothin Road, Klong 1, Klong Luang, Pathumthani 12120, Thailand. See Table 1 for other abbreviations

indicated in Table 2, 17 strains diverged from the type by three or fewer D1/D2 substitutions. The maximum divergence between any two of these

strains was four substitutions, observed between two Australian isolates and seven isolates of diverse origins that differed from the type by three

Table 3 Selected characteristics of strains assigned to *Candida parazyima*. ITS-D1/D2 rDNA haplotypes (A–G) are given along GenBank accession numbers of voucher deposits

UWOPS number	rDNA	Locality	Insect/plant
91-652.1 ^T	A EF601044	Hawaii Island	<i>Drosophila floricola</i> on <i>Ipomoea indica</i>
91-681.2	A	Hawaii Island	<i>Ipomoea indica</i> flower
91-697.2	A	Hawaii Island	<i>Ipomoea indica</i> flower
91-703.2	A	Hawaii Island	<i>Conotelus mexicanus</i> on <i>Ipomoea indica</i>
91-704.2	A	Hawaii Island	<i>Ipomoea indica</i> flower
91-711.2	A	Hawaii Island	<i>Ipomoea indica</i> flower
91-720.2	A	Hawaii Island	<i>Ipomoea indica</i> flower
91-727.2	A	Hawaii Island	<i>Ipomoea pes-caprae</i> flower
91-751.2	A	Hawaii Island	<i>Drosophila floricola</i> on <i>Ipomoea indica</i>
91-760.2	A	Hawaii Island	<i>Drosophila floricola</i> on <i>Ipomoea indica</i>
91-764.2	A	Hawaii Island	<i>Drosophila floricola</i> on <i>Ipomoea indica</i>
91-882.2	A	Hawaii Island	<i>Ipomoea indica</i> flower
91-894.2	A	Hawaii Island	<i>Ipomoea pes-caprae</i> flower
95-821.3	A	Eastern Australia	<i>Ipomoea indica</i> flower
95-964.2	A	Eastern Australia	<i>Aethina concolor</i> on <i>Hibiscus splendens</i>
95-1002.2	A	Eastern Australia	<i>Ipomoea indica</i> flower
95-1058.3	A	Eastern Australia	<i>Aethina</i> sp. on <i>Ipomoea indica</i>
99-333.1	A EF533998	Costa Rica	<i>Conotelus</i> sp. on <i>Ipomoea carnea</i>
99-669.2	A	Viti Levu	<i>Aethina concolor</i> on <i>Hibiscus</i> sp.
99-694.2	A	Mana Island	<i>Aethina concolor</i> on <i>Ipomoea tuboides</i>
00-132.3	A	Costa Rica	<i>Conotelus</i> sp. on <i>Merremia quinquefolia</i>
00-135.3	A	Costa Rica	<i>Conotelus</i> sp. on <i>Tabebuia rosea</i>
00-313.3	A	Maui	<i>Conotelus mexicanus</i> on <i>Pandorea jasminoides</i> var. <i>rosea</i>
00-348.2	A	Kauai	<i>Conotelus mexicanus</i> on <i>Thevetia peruviana</i>
00-601a2	A	Oahu	<i>Conotelus mexicanus</i> on Bignoniaceae
00-629b1	A	Oahu	<i>Drosophila</i> sp. on <i>Ipomoea indica</i>
00-660.3	A	Hawaii Island	<i>Conotelus mexicanus</i> on <i>Ipomoea indica</i>
00-722a1	A	Molokai	<i>Drosophila floricola</i> on <i>Ipomoea indica</i>
00-739.2	A	Molokai	<i>Aethina concolor</i> on <i>Ipomoea indica</i>
04-370.3	A	Hawaii Island	<i>Prosopeus subaeneus</i> on <i>Ipomoea indica</i>
S99-201.1*	A	Tennessee	<i>Conotelus obscurus</i> on <i>Ipomoea pandurata</i>
S99-204.1*	A	Tennessee	<i>Conotelus obscurus</i> on <i>Ipomoea pandurata</i>
S99-205.3*	A	Tennessee	<i>Conotelus obscurus</i> on <i>Ipomoea pandurata</i>
S99-212.2*	A	Tennessee	<i>Conotelus obscurus</i> on <i>Ipomoea pandurata</i>
S99-213.1	B EU443390	Tennessee	<i>Conotelus obscurus</i> on <i>Ipomoea</i> sp.
S99-215.1	B	Tennessee	<i>Conotelus obscurus</i> on <i>Calystegia sepium</i>
S99-205.1*	C	Tennessee	<i>Conotelus obscurus</i> on <i>Ipomoea pandurata</i>
00-150b2	C EU273475	Costa Rica	<i>Conotelus obscurus</i> on <i>Ipomoea batatas</i>
95-403.1	C	Brazil	<i>Conotelus</i> sp. on <i>Ipomoea</i> sp.
S99-206.1*	C	Tennessee	<i>Conotelus obscurus</i> on <i>Ipomoea pandurata</i>
09-101.1	D	Belize	<i>Conotelus</i> sp. on <i>Ipomoea</i> sp. 1
00-154.2	D	Costa Rica	<i>Conotelus</i> sp. on <i>Merremia tuberosa</i>
01-113a1	D EF601045	Costa Rica	Nitidulidae on Bignoniaceae
99-724.2	E EF601046	Rarotonga	<i>Aethina concolor</i> on <i>Ipomoea carnea</i>

Table 3 continued

UWOPS number	rDNA	Locality	Insect/plant
09-102a2	F	Belize	<i>Conotelus</i> sp. on <i>Ipomoea</i> sp. 1
09-107a3	F	Belize	<i>Conotelus</i> sp. on <i>Ipomoea</i> sp. 2
09-107d2	F	Belize	<i>Conotelus</i> sp. on <i>Ipomoea</i> sp. 2
09-110.4	F GQ411401	Belize	<i>Conotelus</i> sp. on <i>Ipomoea</i> sp. 2
09-113.3	F	Belize	<i>Conotelus</i> sp. on <i>Ipomoea</i> sp. 2
09-114.1	F	Belize	<i>Conotelus</i> sp. on <i>Ipomoea</i> sp. 2
09-115.6	F	Belize	<i>Conotelus</i> sp. on <i>Ipomoea</i> sp. 2
09-116.3	F	Belize	<i>Conotelus</i> sp. on <i>Ipomoea</i> sp. 2
09-104a2	F	Belize	<i>Conotelus</i> sp. on <i>Ipomoea</i> sp. 2
09-105.2	F	Belize	<i>Conotelus</i> sp. on <i>Ipomoea</i> sp. 2
09-106.3	F	Belize	<i>Conotelus</i> sp. on <i>Ipomoea</i> sp. 2
09-112.4	F	Belize	<i>Conotelus</i> sp. on <i>Ipomoea</i> sp. 2
09-117.3	G GQ411402	Belize	<i>Conotelus</i> sp. on <i>Ipomoea</i> sp. 1

Horizontal lines separate strains with identical sequences

* Number in the yeast culture collection of the Department of Biology, Syracuse University (SUB)

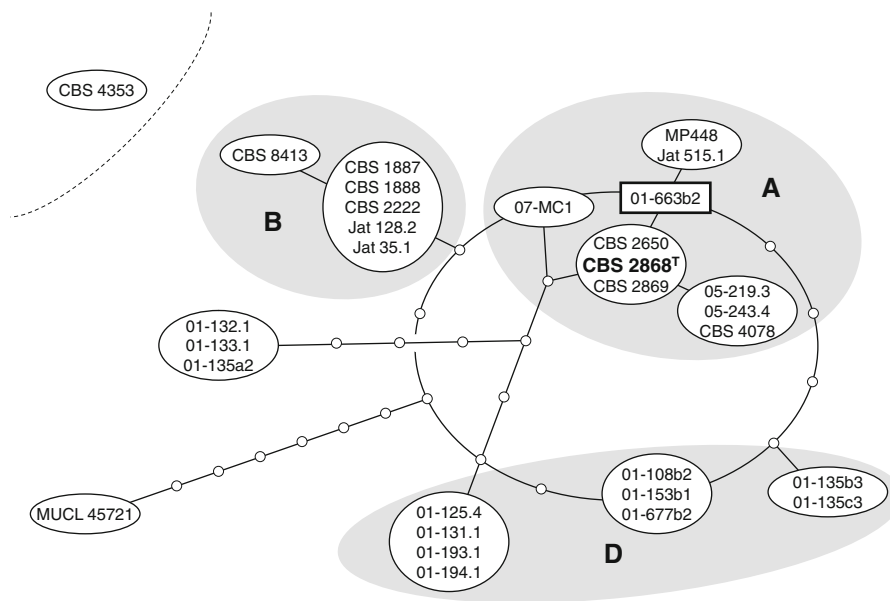


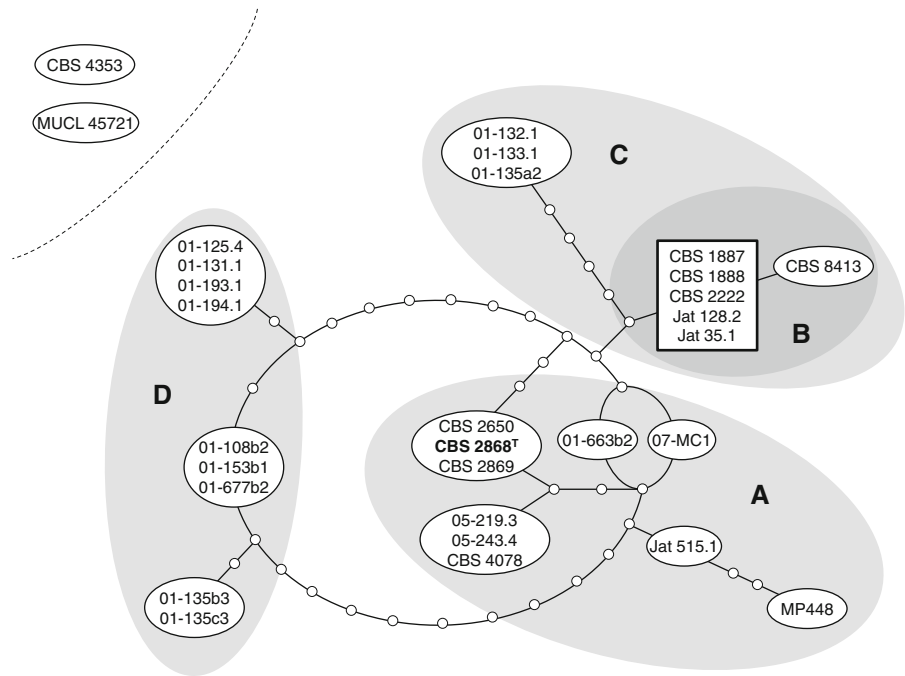
Fig. 1 Parsimony network analysis of the LSU rRNA gene D1/D2 domains of strains of *C. apicola* and relatives. Each connecting line represents one substitution and each *small circle* represents a missing intermediate sequence. A *rectangle* identifies the sequence identified as ancestral by the analysis. The *shaded area A* shows a subset of strains that differ from

the type by three or fewer substitutions. *Area B* shows strains that are connected to *subset A* by three or fewer substitutions. The significance of *area D* is explained in the text. The *dashed line* shows that strain CBS 4353 was excluded from the network

substitutions (Table 2). However, when the ITS region was taken into account as well (Fig. 3), considerably more divergence was observed. The added divergence was not hierarchical but sequential.

The analysis further confirmed the discontinuities elicited in Table 2 by assigning the remaining isolates to five separate sets outside the *C. azyma* network. The sets include two singletons, a pair of

Fig. 2 Parsimony network analysis of the ITS–D1/D2 rDNA haplotypes of strains of *C. apicola* and relatives. Each connecting line represents one substitution and each *small circle* represents a missing intermediate sequence. A *rectangle* identifies the haplotype identified as ancestral by the analysis. *Shaded areas A and B* correspond to those in Fig. 1. *Shaded area C* shows a subset of strains that have identical ITS/5.8S rRNA haplotypes. The significance of shaded area **D** is explained in the text. The *dashed line* shows that strains CBS 4353 and MUCL 45721 were excluded from the main network



isolates, the recognized species *Candida azymoides* (of which four isolates are known), and the 43 isolates now reassigned to *Candida parazyama* sp. nov. Note that a parsimony analysis of D1/D2 sequences only (not shown) yielded the same network memberships.

Most strains in the *C. azyma* complex utilize sucrose, the common α -glucosides, L-sorbose, xylitol, mannitol, glucitol, succinic acid, and 2-keto-D-gluconic acid, grow at 30°C or higher, and can obtain nitrogen from ethylamine, lysine, and cadaverine, but not nitrate or nitrite. All are strongly resistant to cycloheximide and able to hydrolyze Tween 80, and none is fermentative. Variation occurs primarily in the utilization of galactose, D-xylose, L-arabinose, ethanol, glycerol, ribitol, galactitol, and citric acid, hydrolysis of gelatin or casein, and growth in the presence of 10% NaCl or 50% glucose. The variation is generally uncorrelated with affinities elicited by sequence analyses, with the following exceptions. Strains UWOPS 95-805.2 and UWOPS 95-813.3 do not utilize galactose. Strain UWOPS 03-446.4 exhibits a weak utilization of α -glucosides, no growth on glycerol or in the presence of 10% NaCl, but can grow in the presence of 10 mg L⁻¹ CTAB or 6% ethanol. *C. azymoides* consistently exhibits strong growth on citric acid, which is normally negative in

other strains. The seven strains that differ from the type by three substitutions in the D1/D2 region (shaded area in Fig. 3) utilize 1-propanol weakly, whereas other strains do not assimilate that carbon source. In all remaining cases, the variation in growth responses exhibits no correlation whatsoever with the relationships depicted in Fig. 3.

Candida parazyama Lachance sp. nov

Phylogenetic placement

The phylogram in Fig. 4 shows that *C. parazyama* lies in a basal position with respect to *C. azyma*, *C. azymoides*, and other closely related species as delineated in Fig. 3. The exception was *Candida* sp. 03-446.4, which formed a clade with the three species selected to serve as outgroup.

Standard description

On YM agar after 3 days at 25°C, the cells are ovoid to ellipsoidal (2–3 × 2–4 μ m). Budding is multilateral. A ring is formed in fermentation medium. Colonies are white, convex or umbonate, smooth, often with concentric circles. In Dalmau plates after 2 weeks on cornmeal agar, neither pseudohyphae nor true hyphae

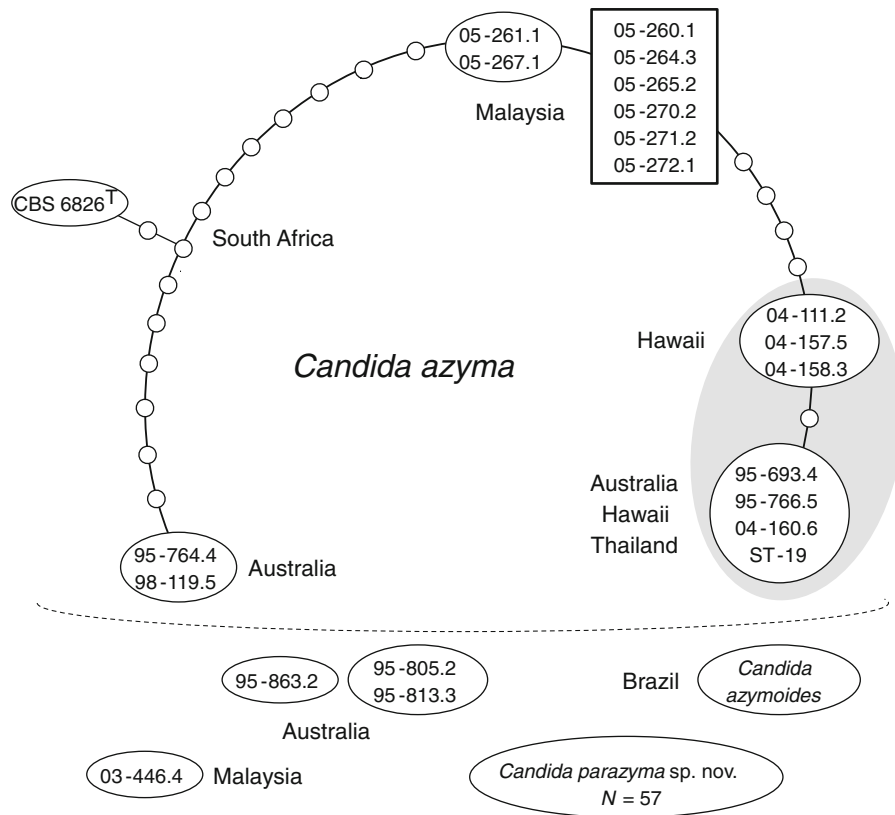
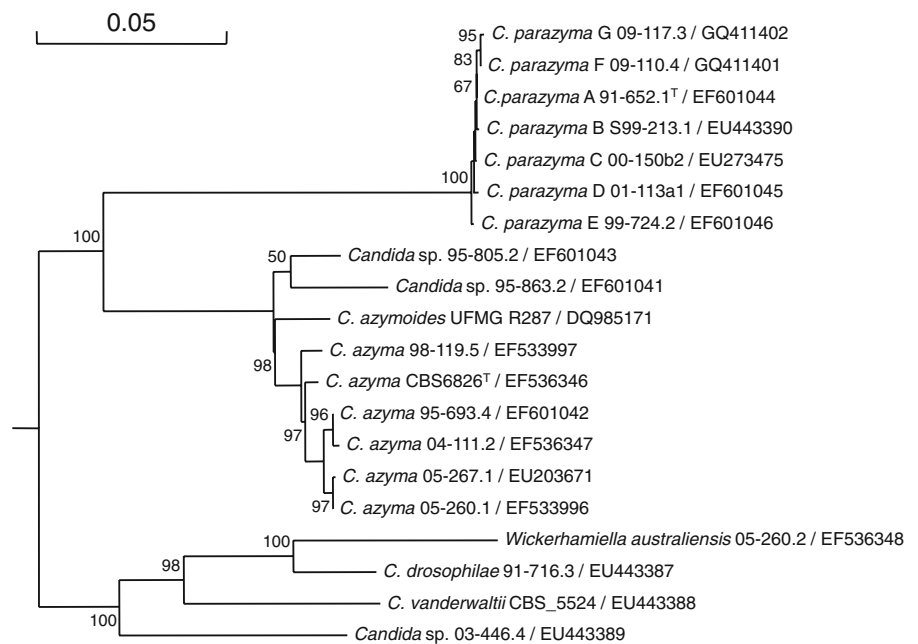


Fig. 3 Parsimony network analysis of the ITS–D1/D2 rDNA haplotypes of strains of *C. azyma* and relatives. Each connecting line represents one substitution and each small circle represents a missing intermediate sequence. The shaded

area shows strains that differ from the type by three substitutions in the D1/D2 region and assimilate 1-propanol weakly. The dashed line shows the exclusion of five sets of strains from the main network

Fig. 4 Neighbour-joining phylogram of the ITS–D1/D2 rDNA sequences of selected strains showing the phylogenetic placement of *C. parazyma*. The scale bar shows the degree of sequence divergence (K_{nuc}). Bootstrap values of 50% or greater are shown; they were determined from 1,000 iterations



are formed. Asci have not been observed in pure or mixed cultures on common sporulation media. Glucose is not fermented. Glucose, sucrose, galactose (sometimes slow), trehalose, maltose, melezitose, α -methyl-D-glucoside (variable), L-sorbose, D-xylose (slow), ethanol (slow), glycerol (variable), xylitol, mannitol (slow), D-glucitol, succinic acid, and 2-keto-D-gluconate are assimilated. No growth occurs on inulin, raffinose, melibiose, lactose, soluble starch, cellobiose, salicin, L-rhamnose, L-arabinose (sometimes slow), D-arabinose, D-ribose, methanol, 1-propanol, 2-propanol, 1-butanol, erythritol, ribitol (sometimes weak), galactitol (sometimes weak), myo-inositol, lactic acid, citric acid, gluconic acid, glucosamine, *N*-acetyl-glucosamine, acetone, ethyl acetate, or hexadecane. Assimilation of nitrogen compounds: positive for lysine, ethylamine-HCl and cadaverine, and negative for nitrate and nitrite. Growth in vitamin-free medium is negative. Growth in amino-acid-free medium is positive. The maximum growth temperature is 34–36°C. Growth on YM agar with 5% sodium chloride is positive, and at 10% sodium chloride is variable. Growth in the presence of 50% glucose is variable. Growth in the presence of 1,000 mg L⁻¹ cycloheximide growth is positive. Starch-like compounds are not produced. The Diazonium Blue B reaction is negative. The habitat is flowers and associated insects in warmer climates worldwide. The type strain of *Candida parazyza* is strain UWOPS 91-652.1^T isolated from *Drosophila floricola* collected in a flower of *Ipomoea indica* in Kipuka Puau, Island of Hawaii. It has been deposited in the collection of the Yeast Division of the Centraalbureau voor Schimmelcultures, Utrecht, the Netherlands, as strain CBS 11563^T (NRRL Y-48669^T).

Eymology: the epithet *parazyza* (pa.ra.zy'ma) N.L. nom. f. sing. n., from the Greek παρα (para) meaning near and the epithet *azyza*, referring to the similarity and relatedness of the new species to *Candida azyza*.

Latin diagnosis

In agaro YM post dies tres cellulae singulae aut binae, ovoidae (2–3 × 2–4 μm). In medio liquid annulus formatur. Cultura convexa, glabra et candida. In agaro farinae Zea mays post dies 14 mycelium nec pseudo-mycelium non formantur. Glucosum non fermentatur. Glucosum, sucrosus, galactosus (aliquando lente),

trehalosus, maltosus, melezitosus, α -methyl-D-glucosidus (variabilis), L-sorbosus, D-xylosus (lente), ethanolus (lente), glycerolus (variabilis), xylitolus, mannitolus (lente), glucitolus, acidum succinicum et 2-keto-gluconatum assimilantur, at non inulinum, raffinosus, melibiosus, lactosus, amyllum solubile, cellobiosus, salicinus, L-rhamnosus, L-arabiosus (aliquando lente), D-arabiosus, D-ribosus, methanolus, 1-propanolus, 2-propanolus, 1-butanolus, erythritolus, ribitolus, galactitolus (aliquando exigue), meso-inositolus, acidum lacticum, acidum citricum, acidum gluconicum, glucosaminus, *N*-acetyl-D-glucosaminus, acetonus, ethyl acetatus, nec hexadecanus. Ethylaminus, lysinus et cadaverinus assimilantur at non natrium nitricum nec natrium nitrosus. Ad crescentiam vitamina externa necessaria sunt. Augmentum in 34–36°C. Habitat flores et insectis junctis. Typus UWOPS 91-652.1^T *Drosophila floricola* e flore *Ipomoea indica* in Hawaii isolatus est. In collectione zymotica Centraalbureau voor Schimmelcultures, Trajectum ad Rhenum, sub no. CBS 11563^T depositus est.

As indicated in Table 3, *Candida parazyza* has been isolated in most localities where floricolous nitidulid beetles and other insects have been sampled, worldwide, but not at higher latitudes. The ITS and D1/D2 rDNA sequences comprised seven haplotypes represented by the letters A to G, in order of their discovery (Table 3 and Figs. 4, 5). Haplotype B is unique in that it entails a substitution in the D1/D2 region. All other rDNA polymorphisms are confined to the internal transcribed spacers. The maximum amount of divergence was observed between the Central American haplotypes D and G, which differ by five substitutions in total. The distribution of rDNA haplotypes had no bearing on the growth characteristics of *C. parazyza*, which is less extensive than that of the *C. azyza* complex as a whole.

Discussion

Parsimony networks in systematics

Much of the forthcoming discussion hinges on the interpretation of patterns identified by parsimony network analysis and for this reason, it is fitting to review briefly the underlying theory and properties of the method. Clement et al. (2000), in the description

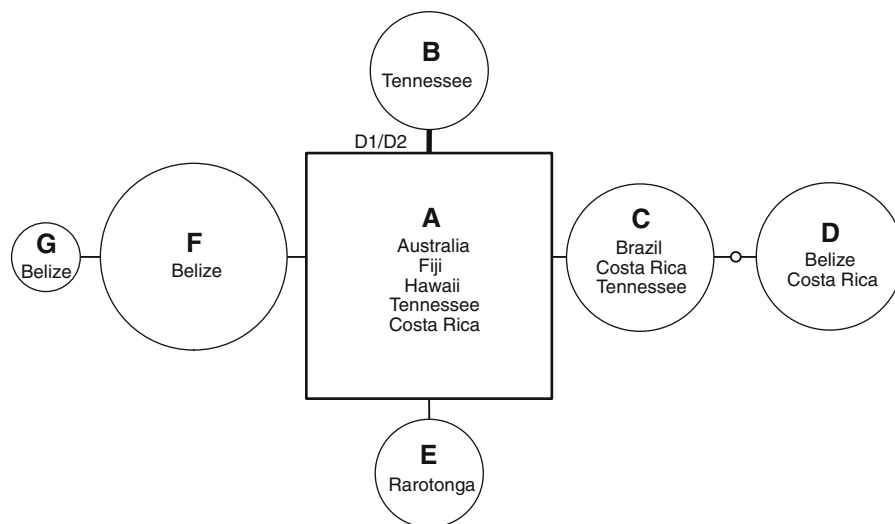


Fig. 5 Parsimony network analyses of the ITS–D1/D2 rDNA haplotypes of strains of *C. parazyama*. Each connecting line represents one substitution and the *small circle* represents a missing intermediate sequence

of the computer program TCS, pointed out that phylogenetic trees are inappropriate models of the genealogy of alleles within a species. Posada and Crandall (2001) developed the idea further: phylogenetic trees are based on the assumption that speciation causes ancestral sequences to be replaced by new sequences through lineage sorting and phyletic divergence. Thus, the internal nodes of a tree represent the hypothetical positions of ancestral sequences that no longer exist. By contrast, coalescent theory posits that within a population, the most ancestral allele should tend to be the most abundant, provided that the different alleles are selectively neutral. The relationships between alleles within a species are tokogenetic (parent-offspring kinship) rather than phylogenetic (sibling or cousin kinship). Inter-allele relationships are not hierarchical and instead follow patterns of stepwise accumulation of mutations. Moreover, most tree construction methods inadequately deal with sequences or haplotypes that have arisen by recombination or homoplasy. Parsimony networks depict such events in the form of reticulation or loops. The networks are hypothetical reconstructions of the most parsimonious path connecting alleles via intermediates that differ each by a single substitution. Importantly, membership in a network is based on the probability that DNA sequences share parsimonious relationships, in other words, that they can be connected by single steps.

The connection limit is the probability that the steps depicted in a network correspond to single, and not multiple, substitutions. Sequences that do not pass that test are excluded from the network. Parsimony network analysis also attempts to identify one allele or haplotype as the most likely to be ancestral. This is done independently of allele frequencies, although one expects the allele identified as ancestral from coalescent theory to be the most abundant. A more complete discussion of the relevant mathematics has been offered by Templeton et al. (1992).

Given that parsimony networks are meant to identify tokogenetic relationships, it is reasonable to expect sequences that represent homologous alleles within a species to join the same parsimony network and those that represent orthologues (from different species) to be assigned to distinct networks. This is in fact what was observed by Hart and Sunday (2007) in a meta-analysis of hundreds of reports where parsimony network analyses have been applied to barcoding sequences determined for members of well-defined species of animals and plants. Separate biological species tend to form distinct networks at the 95% parsimony connection limit. Less well defined cases were often accountable to hybridization. The authors concluded that the formation of parsimony networks from barcoding sequences constitutes an objective means of circumscribing phylogenetic species.

C. apicola is a polymorphic, cosmopolitan, asexual species

Based on the strictest application of Kurtzman and Robnett's (1998) criterion, one would retain in *C. apicola* only those strains that differ by no more than three substitutions from the type in the D1/D2 domains. These are represented by set A in Fig. 1. Assignment of the strains of set B would be problematic. Based on the approach used in many recently published descriptions of yeast species, where sampling is sometimes poor and even occasionally limited to a single strain, it is not inconceivable that the 30 strains included in the present study could be assigned to as many as seven separate species. The parsimony network generated from D1/D2 sequences suggests on the contrary that all strains but CBS 4353 could be viewed as member of a cohesive evolutionary unit, although the inclusion of strain MUCL 45721 might appear dubious. Although internal transcribed spacers are a poor source of phylogenetic signal, the ITS region has been useful in discriminating between members of distinct biological species that have nearly identical D1/D2 regions (e.g., Lachance et al. 2005). Indeed, inclusion of ITS sequences in a barcode system was seen as desirable by Fell et al. (2000) for basidiomycetous species and the ITS is currently favored as candidate to serve as official barcoding sequence for fungi (Seifert 2009). Inclusion of ITS sequences in the parsimony network analysis (Fig. 2) reinforced the notion that most strains should be treated as members of a single evolutionary population but caused strain MUCL 45721 to be excluded at the 95% limit. The dependence of these observations on adequate sampling further suggests that the description of new species to accommodate strains CBS 4353 and MUCL 45721 should await the isolation of a sufficient number of additional strains.

Addition of ITS sequences to the analysis resulted in the designation of a different strain's haplotype as ancestral (Figs. 1, 2). This is manifestly due to the fact that in the first instance, the D1/D2 sequence of strain UWOPS 01-663b2 could be connected to three others each by a single step, whereas in the combined analysis the haplotype shared by five strains was the only with multiple connections that included at least one single-step connection. Whether this should or should not be taken to mean that the ITS and D1/D2

regions have different ancestral sequences should be appraised in the light of the relatively high polymorphism to sample size ratio of the *C. apicola* isolates. This is in sharp contrast to what was found for *C. parazyza* (Fig. 5), where the ancestral status of haplotype A is firmly grounded and represented by a clear majority of isolates.

Our attempts to observe inter-strain matings have been arduous and exhaustive, leading us to reaffirm the conclusion that *C. apicola* is an asexual species. The reticulation observed in the networks presented in Figs. 1 and 2 is unlikely to be the result of recombination, given that the polymorphic sites responsible for reticulation are dispersed within contiguous sequences that span just over a thousand nucleotides. They must therefore be attributed to homoplasmy, which further reinforces the notion that an arbitrary number of substitutions cannot be applied broadly to delineate species across a broad taxonomic range. Membership in a parsimony network of sequences spanning the ITS-D1/D2 rDNA region would seem to be a suitable surrogate for mating compatibility in the delineation of asexual yeasts, in view of its stronger theoretical foundation compared to DNA/DNA reassociation or a specified number of substitutions.

Examination of the source data in Table 1 shows that strains from Costa Rica exhibited the largest degree of haplotype diversity ($n = 5$), with representatives in subsets A, C, and D (Fig. 2). Not enough is known on the phylogenetic systematics of the bees from which most yeast strains originated to speculate on possible species-specific associations, but it may be safe to infer that the very large bee diversity of Costa Rica accounts for the diversity of *C. apicola* in the bees of that region. The sampling effort in Brazil is comparable in intensity (Rosa et al. 2006) and this surely bears on the finding of three distinct haplotypes (in subsets A and B) in Brazilian bees. Although the available data do not allow one to conclude much on the population structuring of *C. apicola*, it is safe to state that the species as a whole has a global distribution.

The morphospecies *C. azyma* is part of a complex of at least six, mostly cryptic species

Identification based on growth properties initially would have led to the assignment of the 81 strains

considered here to *C. azyma*. Small differences in growth responses prompted the examination of some strains by rDNA sequencing and confirmed the distinct status of strains now assigned to *C. azymoides*, of strains UWOPS 95-805.2 and UWOPS 95-813.3, and of strain UWOPS 03-446.4. Strain UWO 95-863.2 as well as those included in *C. parazyza*, however, were not properly distinguishable from *C. azyma* by growth characteristics and would require identification by sequence analysis. Species assignment of the strains retained in the *C. azyma* parsimony network (Fig. 3) purely on the basis of number of substitutions would be problematic, depending of what strains are available. For example, strains UWOPS 95-764.4 and UWOPS 95-766.5, recovered from morning glory flowers at the same Australian collection site, differ by 23 substitutions in their combined rDNA haplotypes. Taken in isolation, this information could have led some authors to describe them as separate species, even though the analysis presented here argues that they are part of the *C. azyma* continuum. Furthermore, two testable predictions can be made: first, that intensified sampling will yield *C. azyma* isolates with rDNA haplotypes that are intermediate between those of known strains, and second, that additional sampling is unlikely to yield haplotypes that are linear intermediates between those of the six distinct networks in Fig. 3. These predictions arise from the expectation that allele genealogies within a species proceed by small successive steps along a continuum from which isolates of a species are sampled, whereas species genealogies proceed principally in a tree-like fashion by divergence from ancestors that are no longer accessible. Increased sampling is therefore likely to fill in missing intermediates within a species, but not between species. Pursuant to this, three of the putatively new species elicited by this study are represented by only one or two isolates and therefore remain to be properly sampled prior to proposing formal descriptions.

Contrasting biogeographies in the *C. azyma* species complex

The biogeography of the *C. azyma* species complex would seem to be a mixture of endemism (*C. azymoides*) and broad dispersal. Member species appear to be mostly confined to tropical and subtropical

regions, with no confirmed reports of isolations at latitudes exceeding 40°N. Of particular interest is the mostly Australasian distribution of *C. azyma sensu stricto* (with the exception of the type, from South Africa). It is hoped that focused sampling combined with sequence-based identification over larger portions of the old world will clarify the intriguing pattern presented in Fig. 3.

C. azyma was once reported to be the most widespread species to be isolated globally from floricolous insects (Lachance et al. 2001). As it were, the majority of isolates, which had been identified from growth characteristics, are now known instead to be representatives of the cryptic relative *C. parazyza*. Very few other ascomycetous yeast species have been isolated from an ecosystem found across such a broad geographic range and so it is of interest to see if the distribution of genetic variants within the species follows biogeographic rules. Fenchel and Finlay (2004) proposed that microorganisms, by virtue of their small size, disperse so rapidly that they do not follow biogeographically meaningful distributions as do larger organisms. In response to criticisms that morphological identification of microorganisms does not allow such generalizations to be drawn, Finlay et al. (2006) characterized free-living protozoa by sequencing and concluded that the distribution of genetic variants is dictated by local niche characteristics and not by historical or spatial factors. Greig (2007) pointed out that yeasts are probably not free-living because of their complex interactions with vector insects and suggested that *Saccharomyces* species probably do have a biogeography, a prediction that seems to have received strong support recently from phylogenomic studies (Liti et al. 2009). So far, yeasts found in association with floricolous insects have followed strong biogeographic patterns, be it at the supraspecific (Lachance et al. 2001, 2005) or the infraspecific (Lachance et al. 2008; Wardlaw et al. 2009) levels. The analysis depicted in Fig. 5 can be explained in part by a vicariance model, whereby strains with the broadly dominant haplotype A have dispersed globally followed by the local generation of genetic variants. As is often the case, the highest degree of polymorphism was observed in Central America with the identification of most of the haplotypes in Belize and Costa Rica. However, partial support for the ubiquity model arises from the observation of the combined

haplotype C in such remote localities as Brazil, Costa Rica, and Tennessee.

Asexuality in *C. parazyza*

As was the case for *C. apicola*, our efforts to observe sexual interaction in the entire *C. azyza* complex have been extensive and unsuccessful. It should be noted that at least one precedent exists in the broader *Wickerhamiella* clade for sexuality to be very cryptic (Lachance et al. 2000). All known species in the teleomorphic genus have a haplontic sexual cycle, four of the five are heterothallic, and the last discovered species, *Wickerhamiella lipophila*, was first described as a *Candida* species because of fastidious and unpredictable mating and sporulation habits. A preliminary study comparing the pattern reported here for ribosomal sequences with alleles of a polymorphic marker with homology with *Essential Nuclear Protein 1* and *imidazole glycerol phosphate synthase* genes of *Yarrowia lipolytica* (A.E. Smith and M.A. Lachance, unpublished), suggests that the two loci may have experienced considerable recombination and consequently that the species may have a hitherto cryptic sexual cycle.

Concluding remarks on parsimony networks

Parsimony network analysis provides a statistical, theory-laden approach to the delineation of phylogenetic yeast species from barcode sequence data. The method aims at distinguishing, within sequence space, variation that can be regarded as polymorphism (within species) from differences that are the consequence of speciation. We have found it useful in determining which strains should be retained in *C. apicola* or *C. azyza* in spite of sequence polymorphism. The approach is not meant to replace or even to augment species delineation based on reproductive isolation when such information is available, nor is it meant to invalidate simpler interpretations of sequence divergence when clear continuities exist, as in the case of *C. parazyza*. A future reinterpretation of the polymorphisms present in *C. apicola* in the light of new information may very well favour a subdivision of the species. Importantly, we hope that this study will contribute to a better appreciation for the importance of sufficient taxon sampling in the exploration of yeast biodiversity.

Acknowledgments This work was funded by the Natural Science and Engineering Research Council of Canada. We acknowledge the gift of strains from the Centraalbureau voor Schimmelcultures, The Netherlands (V. Robert, M. Groenewald), the Mycothèque de l'Université catholique de Louvain, Belgium (H.M. Daniel), the BIOTEC Culture collection, Thailand (S. Jindamorakot), as well as the research collections of C.A. Rosa and W.T. Starmer. We also benefited from the field collaboration of numerous individuals, including J.S.F. Barker, J.M. Bowles, D.H. Janzen, J. McNeil, M. Polak, and others. Special thanks are owed to C. Harding for the Belize collection.

References

- Clement M, Posada D, Crandall K (2000) TCS: a computer program to estimate gene genealogies. *Mol Ecol* 9: 1657–1660
- Darwin C (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. 1st edn. John Murray, London, p 502
- Fell JW, Boekhout T, Fonseca A, Scorzetti G, Stätzell-Tallman A (2000) Biodiversity and systematics of basidiomycetous yeasts as determined by large-subunit rDNA D1/D2 domain sequence analysis. *Int J Syst Evol Microbiol* 50:1351–1371
- Fenchel T, Finlay BJ (2004) The ubiquity of small species: patterns of local and global diversity. *Bioscience* 54: 777–784
- Finlay BJ, Esteban GF, Brown S, Fenchel T, Hoef-Emden K (2006) Multiple cosmopolitan ecotypes within a microbial eukaryote morphospecies. *Protist* 157:377–390
- Greig D (2007) Population biology: wild origins of a model yeast. *Curr Biol* 17:R251–R253
- Groenewald M, Daniel HM, Robert V, Poot GA, MTh Smith (2008) Polyphasic re-examination of *Debaryomyces hansenii* strains and reinstatement of *D. hansenii*, *D. fabryi* and *D. subglobosus*. *Persoonia* 21:17–27
- Hart M, Sunday J (2007) Things fall apart: biological species form unconnected parsimony networks. *Biol Lett* 3: 509–512
- Jacques N, Mallet S, Casaregola S (2009) Delimitation of the species of the *Debaryomyces hansenii* complex by intron sequence analysis. *Int J Syst Evol Microbiol* 59:1242–1251
- Kurtzman CP, Droby S (2001) *Metschnikowia fructicola*, a new ascospore yeast with potential for biocontrol of postharvest fruit rots. *Syst Appl Microbiol* 24:395–399
- Kurtzman CP, Robnett CJ (1998) Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Antonie van Leeuwenhoek* 73:331–371
- Kurtzman CP, Robnett CJ (2003) 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
- Lachance MA (1993) *Kluyveromyces*: systematics since 1970. *Antonie van Leeuwenhoek* 63:95–104
- Lachance MA, Bowles JM, Mueller C, Starmer WT (2000) On the biogeography of yeasts in the *Wickerhamiella* clade

- and description of *Wickerhamiella lipophila* sp. nov., the teleomorph of *Candida lipophila*. *Can J Microbiol* 46:1145–1148
- Lachance MA, Bowles JM, Kwon S, Marinoni G, Starmer WT, Janzen DH (2001) *Metschnikowia lochheadii* and *Metschnikowia drosophilae*, two new yeast species isolated from insects associated with flowers. *Can J Microbiol* 47:103–109
- Lachance MA, Daniel HM, Meyer W, Prasad GS, Gautam SP, Boundy-Mills K (2003) The D1/D2 domain of the large subunit rDNA of the yeast species *Clavispora lusitaniae* is unusually polymorphic. *FEMS Yeast Res* 4:253–258
- Lachance MA, Ewing CP, Bowles JM, Starmer WT (2005) *Metschnikowia hamakuensis* sp. nov., *Metschnikowia kamakouana* sp. nov., and *Metschnikowia mauiiiana* sp. nov., three endemic yeasts from Hawaiian nitidulid beetles. *Int J Syst Evol Microbiol* 55:1369–1377
- Lachance MA, Lawrie D, Dobson J, Piggott J (2008) Biogeography and population structure of the Neotropical endemic yeast species *Metschnikowia lochheadii*. *Antonie van Leeuwenhoek* 94:403–414
- Liti G, Carter DM, Moses AM et al (2009) Population genomics of domestic and wild yeasts. *Nature* 458:337–341
- Marinoni G, Lachance MA (2004) Speciation in the large-spored *Metschnikowia* clade and establishment of a new species, *Metschnikowia borealis* comb. nov. *FEMS Yeast Res* 4:587–596
- Meyer SA, Payne RW, Yarrow D (1998) *Candida* Berkhout. In: Kurtzman CP, Fell JW (eds) *The yeasts, a taxonomic study*, 4th edn. Elsevier, Amsterdam, pp 454–573
- Molnár O, Prillinger H (2005) Analysis of yeast isolates related to *Metschnikowia pulcherrima* using the partial sequences of the large subunit rDNA and the actin gene; description of *Metschnikowia andauensis* sp. nov. *Syst Appl Microbiol* 28:717–726
- Nguyen HV, Gaillardin C, Neuvéglise C (2009) Differentiation of *Debaryomyces hansenii* and *Candida famata* by rRNA gene intergenic spacer fingerprinting and reassessment of phylogenetic relationships among *D. hansenii*, *C. famata*, *D. fabryi*, *C. flareri* (= *D. subglobosus*) and *D. prosopidis*: description of *D. vietnamensis* sp. nov. closely related to *D. nepalensis*. *FEMS Yeast Res* 9:641–662
- Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol Evol* 16:37–45
- Rosa CA, Lachance MA, Silva J, Teixeira A, Marini MM, Antonini Y, Martins RP (2003) Yeast communities associated with stingless bees. *FEMS Yeast Res* 4:271–275
- Rosa CA, Morais PB, Lachance MA, Pimenta RS, Santos RO, Trindade RC, Figueroa DL, Resende MA, Bragança MAL (2006) *Candida azymoides* sp. n., a yeast species from tropical fruits and larva (Ascomycota) of *Anastrepha mucronota* (Diptera: Tephritidae). *Lundiana* 7:83–86
- Seifert KA (2009) Barcoding fungi. Progress towards DNA barcoding of fungi. *Mol Ecol Resour* 9(Suppl 1):83–89
- Sipiczki M (2006) *Metschnikowia* strains isolated from botrytized grapes antagonize fungal and bacterial growth by iron depletion. *Appl Environ Microbiol* 72:6716–6724
- Suh SO, Gibson CM, Blackwell M (2004a) *Metschnikowia chrysoperlae* sp. nov., *Candida picachoensis* sp. nov. and *Candida pimensis* sp. nov., isolated from the green lacewings *Chrysoperla comanche* and *Chrysoperla carnea* (Neuroptera: Chrysopidae). *Int J Syst Evol Microbiol* 54:1883–1890
- Suh SO, McHugh JV, Blackwell M (2004b) Expansion of the *Candida tanzawaensis* yeast clade: 16 novel *Candida* species from basidiocarp-feeding beetles. *Int J Syst Evol Microbiol* 54:2409–2429
- Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, Fisher MC (2000) Phylogenetic species recognition and species concepts in Fungi. *Fungal Genet Biol* 31:21–32
- Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132:619–633
- Wardlaw AM, Berkers TE, Man KC, Lachance MA (2009) Population structure of two beetle-associated yeasts: comparison of a New World asexual and an endemic Nearctic sexual species in the *Metschnikowia* clade. *Antonie van Leeuwenhoek* 96:1–15
- Webster AJ, Payne RJ, Pagel M (2003) Molecular phylogenies link rates of evolution and speciation. *Science* 301:478
- Wright S, Keeling J, Gillman L (2006) The road from Santa Rosalia: a faster tempo of evolution in tropical climates. *Proc Natl Acad Sci USA* 103:7718–7722
- Xue ML, Zhang LQ, Wang QM, Zhang JS, Bai FY (2006) *Metschnikowia sinensis* sp. nov., *Metschnikowia zizyphicola* sp. nov. and *Metschnikowia shanxiensis* sp. nov., novel yeast species from jujube fruit. *Int J Syst Evol Microbiol* 56:2245–2250
- Yarrow D (1998) Methods for the isolation and identification of yeasts. In: Kurtzman CP, Fell JW (eds) *The yeasts, a taxonomic study*, 4th edn. Elsevier, Amsterdam, pp 77–100