International Journal of Systematic and Evolutionary Microbiology (2006), **56**, 1157–1165 **DOI** 10.1099/ijs.0.64052-0

Phylogenetic placement of *Hanseniaspora–Kloeckera* species using multigene sequence analysis with taxonomic implications: descriptions of *Hanseniaspora pseudoguilliermondii* sp. nov. and *Hanseniaspora occidentalis* var. *citrica* var. nov.

Neza Cadez,[1] Peter Raspor[1] and Maudy Th. Smith[2]

[1]University of Ljubljana, Biotechnical Faculty, Department of Food Science and Technology, Jamnikarjeva 101, 1000 Ljubljana, Slovenia

[2]Centraalbureau voor Schimmelcultures, PO Box 85167, 3508 AD Utrecht, The Netherlands

Correspondence
Neza Cadez
neza.cadez@bf.uni-lj.si

Two protein-coding genes, actin and translation elongation factor-1α (EF-1α), as well as two ribosomal gene regions, D1/D2 domains of the large subunit and both internal transcribed spacers including the 5·8S gene region, were evaluated regarding their usefulness for reconstruction of phylogenetic relationships in the *Hanseniaspora–Kloeckera* species group. This included analyses of sequence divergence values, heterogeneity of evolutionary rates and the reliability of the inferred trees. Both protein-coding genes showed greater capacities to resolve at the strain level and between the closely related species of *Hanseniaspora–Kloeckera*, compared with the ribosomal gene regions. However, to obtain a fully resolved and reliable phylogenetic tree that reflected the biological relationships it was necessary to combine three congruent sequence datasets. The novel species *Hanseniaspora pseudoguilliermondii* sp. nov. (type strain CBS 8772^T) is described as a result of the application of various molecular approaches to delimit species. Furthermore, incongruent gene genealogies of genetically divergent strains of *Hanseniaspora occidentalis*, as determined by amplified fragment length polymorphism analysis and DNA–DNA reassociation measurements, indicated the presence of two novel varieties, *H. occidentalis* var. *occidentalis* (type strain CBS 2592^T) and *H. occidentalis* var. *citrica* var. nov. (type strain CBS 6783^T), which could be distinguished by habitat preference.

INTRODUCTION

The name *Hanseniaspora* was proposed for ascosporogenous apiculate yeasts by Zikes (1911), although their common presence in fruit juices and fermenting musts had already led to the description of the apiculate species *Saccharomyces apiculatus* (Reess, 1870). The unstable ability of these yeasts to form ascospores contributed to several changes in nomenclature and definition of the genus (Lodder & Kreger-van Rij, 1952). With the introduction of molecular taxonomy, relationships between the species of *Hanseniaspora* and of the anamorphic genus *Kloeckera* were established on the basis of DNA base composition (Nakase & Komagata, 1970) and DNA–DNA relatedness (Meyer *et al.*, 1978). The latter study represents the basis for the current classification, which was confirmed by several phylogenetic studies based on different parts of the ribosomal gene and two protein-coding genes (Yamada *et al.*, 1992a; Boekhout *et al.*, 1994; Esteve-Zarzoso *et al.*, 2001; Kurtzman & Robnett, 2003). The genus *Hanseniaspora* currently comprises ten species, four of which (*Hanseniaspora meyeri*, *Hanseniaspora clermontiae*, *Hanseniaspora lachancei* and *Hanseniaspora opuntiae*) were recently circumscribed on the basis of DNA–DNA reassociation data (Smith, 1998; Cadez *et al.*, 2003).

Abbreviations: AFLP, amplified fragment length polymorphism; EF-1α, translation elongation factor-1α; ITS, internal transcribed spacer; LSU, large subunit; RAPD, randomly amplified polymorphic DNA.

The GenBank/EMBL/DDBJ accession numbers for the gene sequences determined in this study are listed in Table 1.

Figures showing saturation plots and phylogenetic analyses of sequences of actin and EF-1α genes and the D1/D2 regions of the LSU rDNA and ITS regions of *Hanseniaspora–Kloeckera* strains, and tables comparing dataset characteristics and characteristics inferred from the datasets using maximum-parsimony analyses and DNA–DNA relatedness values among strains of *Hanseniaspora occidentalis* are available as supplementary material in IJSEM Online.

Phylogenetic placement of members of *Hanseniaspora* within the *Saccharomyces* clade revealed that the genus is divided into two clusters (Kurtzman & Robnett, 1998, 2003). The first cluster contains the species *Hanseniaspora vineae*, *Hanseniaspora osmophila* and *Hanseniaspora occidentalis.* Strains belonging to the latter species were found to be heterogeneous with respect to their randomly amplified polymorphic DNA (RAPD) profiles, electrophoretic karyotypes and restriction patterns of the internal transcribed spacer (ITS) regions (Cadez *et al.*, 2002). The second cluster comprises *Hanseniaspora valbyensis* and four recently described species that are closely related to *Hanseniaspora uvarum* and *Hanseniaspora guilliermondii* (Cadez *et al.*, 2003). The poorly resolved rDNA-based phylogenetic tree suggested that there were very close relationships among species of the *H. uvarum–H. guilliermondii* complex, which was not in agreement with the genetic relatedness determined by DNA–DNA reassociation analyses (Cadez *et al.*, 2003). This supports the observation that the relationships between recently diverged species complexes are difficult to resolve on the basis of a single gene only (Rokas *et al.*, 2003).

Several studies have demonstrated the phylogenetic usefulness of protein-coding genes for inferring relationships among different taxonomic units (Baldauf *et al.*, 2000; Daniel *et al.*, 2001; Belloch *et al.*, 2000; O'Donnell *et al.*, 1998). Housekeeping genes such as translation elongation factor-1α (EF-1α) and actin are frequently used as independently evolving markers because they do not appear to have undergone horizontal transfer or gene duplication. Moreover, the rates of sequence substitutions in exons are constrained by codon positions, which may determine the taxonomic level of analysis (Baldauf & Palmer, 1993; Kretzer & Bruns, 1999).

Because the species of *Hanseniaspora–Kloeckera* fall into several groups of closely related taxa separated by large phylogenetic distances (Cadez *et al.*, 2003), the phylogenetic marker for resolving *Hanseniaspora–Kloeckera* species should have both the capacity to resolve the sister species and the potential to realistically reconstruct the phylogeny between the divergent lineages. Consequently, the aim of the study was to assess the phylogenetic usefulness of four phylogenetic markers by describing sequence divergence and by comparing the inter- and intraspecific phylogenetic relationships with those derived from genomic DNA–DNA relatedness and physiological properties.

METHODS

Yeast strains. The strains examined and their origins and accession numbers are listed in Table 1.

Strains of *H. occidentalis* were characterized physiologically by using standard methods (Yarrow, 1998). Fermentation ability was determined after 3 and 7 days of incubation. Utilization of carbon compounds was tested in liquid media in tubes, with reciprocal shaking at 30 r.p.m. at 25 °C for 3 weeks. Assimilation of nitrogen compounds was examined by using the auxoanographic method after 1 week (Barnett *et al.*, 2000).

DNA extraction, amplification and sequencing. DNA was isolated from cultures grown on yeast-malt agar (0·3 % yeast extract, 0·3 % malt extract, 0·5 % peptone, 1 % glucose, 2 % agar), according to the method of Möller *et al.* (1992).

PCR primers for the amplification of actin and EF-1α protein-coding regions were designed based on sequences of *Saccharomyces cerevisiae* (GenBank accession numbers NC_001138 and U51033, respectively). Additional primers were designed from generated sequences of the actin gene of *H. vineae* (Act-HV) and the EF-1α gene of *H. meyeri* (EF1-HM). Primer sequences, annealing temperatures and annealing sites and references are given in Table 2.

Amplifications were performed under the following conditions: an initial denaturing step of 5 min at 94 °C, followed by 35 cycles of 30 s at 94 °C, 30 s at primer-pair-specific annealing temperatures (Table 2), 2 min at 72 °C and termination with a final extension step of 7 min at 72 °C. The PCR products were purified using GenElute PCR Clean-up (Sigma–Aldrich), according to the manufacturer's instructions. The pGEM-T vector system was used to clone the PCR products of the protein-coding genes, following the procedure described in the manual (Promega). Plasmid DNA was isolated and purified for subsequent sequencing by using GenElute Plasmid Minipreps (Sigma–Aldrich). Nucleotide sequences were determined in both directions by using a Thermo Sequenase Primer Cycle sequencing kit (Amersham Biosciences) and labelled universal primers M13(−10) or M14R with an ALFexpress II Automatic DNA Sequencer (Amersham Pharmacia Biotech). The D1/D2 regions of the large subunit (LSU) rDNA and ITS regions were sequenced as described previously (Cadez *et al.*, 2003). Sequences of both strands were aligned and edited with the ALFwin Sequence Analyser 2.10.06 software (Amersham Pharmacia Biotech).

Phylogenetic analyses. The number of variable and parsimony-informative sites, the genomic DNA G + C contents and the mean uncorrected sequence divergences between the groups of taxa were calculated using the computer program MEGA 2.1 (Kumar *et al.*, 2001).

Sequences were aligned automatically using the multiple sequence alignment program CLUSTAL X version 1.83 (Thompson *et al.*, 1997). Positions where gaps existed in one of the sequences were manually excluded using BioEdit (Hall, 1999). All subsequent phylogenetic analyses employed the PAUP* 4.0b10 software package (Swofford, 2002). Most-parsimonious trees were generated by a heuristic search procedure with 1000 random addition replicates and tree bisection–reconnection branch swapping. Nucleotide sites were equally weighted. The stability of the branches was assessed by bootstrap analysis (Felsenstein, 1985) in which 1000 replicates were used. The homogeneity of the signal from the four datasets was tested with partition-homogeneity tests (Farris *et al.*, 1995) implemented in PAUP* 4.0b10 with 1000 heuristic replicate searches and exclusion of the invariant characters.

For each dataset, a saturation plot was drawn to evaluate the degree of homoplasy and the heterogeneity of evolutionary rates and to choose an optimal outgroup species (Bush & Everett, 2001; Misof *et al.*, 2001). The corrected evolutionary rates of nucleotide substitutions per site were plotted against the proportion of nucleotide changes per site (p distance). The corrected distances were calculated under the optimality criteria of maximum-likelihood using PAUP* 4.0b10 with corrections for multiple substitutions at the same site, substitutional rate biases and differences in evolutionary rates among sites for each dataset. These corrections were included in substitution models selected using the likelihood-ratio test implemented in MODELTEST 3.5 (Posada & Crandall, 1998). The general time reversible model with a gamma distribution parameter for among-site rate variations ($\mathrm{GTR}+\Gamma$) was found to be of adequate complexity for the actin

Table 1. Strains studied, their origin and GenBank/EMBL/DDBJ accession numbers

CBS, Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands; DBVPG, Industrial Yeast Collection, Perugia, Italy; NCAIM, National Collection of Agricultural and Industrial Microorganisms, Budapest, Hungary.

Species and strain designations	Origin	Accession number			
		Actin	EF-1α	D1/D2	ITS
Hanseniaspora clermontiae					
CBS 8821$^{\mathrm{T}}$	Stem rot, *Clermontia* sp., Hawaii	AM039472	AM039516	AJ512452	AJ512441
CBS 8822	Stem rot, *Clermontia* sp., Hawaii	AM039473	AM039517	AJ512456	AJ512442
Hanseniaspora guilliermondii					
CBS 95	Fermenting bottled tomatoes, The Netherlands	AM039450	AM039494	–	AJ512427
CBS 465$^{\mathrm{T}}$	Infected nail, South Africa	AM039457	AM039501	U84230	AJ512433
Hanseniaspora lachancei					
CBS 8818$^{\mathrm{T}}$	Fermenting agave juice, Mexico	AM039469	AM039513	AJ512457	AJ512439
CBS 8819	*Drosophila* sp., fermenting agave juice, Mexico	AM039470	AM039514	–	–
Hanseniaspora meyeri					
CBS 8734$^{\mathrm{T}}$	Fruit of *Sapindus* sp., Hawaii	AM039466	AM039510	AJ512454	AJ512436
CBS 8815	Drosophilid from *Sapindus* sp. berries, Hawaii	AM039468	AM039512	AJ512458	AJ512438
Hanseniaspora occidentalis var. *occidentalis*					
CBS 280*	Soil, Java	AM039451	AM039495	AJ973097	AJ973087
CBS 282*	Soil, Java	AM039452	AM039496	AJ973098	AJ973091
CBS 283*	Soil, Java	AM039453	AM039497	AJ973099	AJ973089
CBS 284*	Soil, Java	–	–	–	AJ973090
CBS 2569*	*Drosophila* sp., Brazil	AM039460	AM039504	AJ973100	AJ973088
CBS 2578*	Soil, West Indies (St Thomas)	AM039461	AM039505	–	–
CBS 2592$^{\mathrm{T}}$*	Soil, West Indies (St Croix)	AM039463	AM039507	U84225	AJ512429
CBS 8835	Unknown	–	–	–	AJ973093
Hanseniaspora occidentalis var. *citrica* var. nov.					
CBS 6782	Orange juice, Italy	–	–	–	–
CBS 6783$^{\mathrm{T}}$*	Orange juice, Italy	AM039464	AM039508	AJ973101	AJ973092
CBS 9921 (DBVPG 4654)	Rotten orange, Argentina	AM039474	AM039518	AJ973102	AJ973094
CBS 9922 (DBVPG 4656)	Rotten orange, Argentina	AM039475	AM039519	AJ973096	AJ973095
Isolate no. 568	Rotten orange, Argentina	–	–	–	–
Isolate no. 666	Rotten orange, Argentina	–	–	–	–
Hanseniaspora opuntiae					
CBS 8733$^{\mathrm{T}}$	*Opuntia ficus-indica* rot, Hawaii	AM039465	AM039509	AJ512453	AJ512435
CBS 8820	*Opuntia ficus-indica* rot, Hawaii	AM039471	AM039515	AJ512451	AJ512440
Hanseniaspora osmophila					
CBS 313$^{\mathrm{T}}$	Ripe Riesling grape, Germany	AM039455	AM039499	U84228	AJ512431
Hanseniaspora pseudoguilliermondii sp. nov.					
CBS 8772$^{\mathrm{T}}$ (NCAIM Y.741$^{\mathrm{T}}$)	Orange juice concentrate, GA, USA	AM039467	AM039511	AJ512455	AJ512437
Hanseniaspora uvarum					
CBS 314$^{\mathrm{T}}$	Muscatel grape, Crimea, Ukraine	AM039456	AM039500	U84229	AJ512432
CBS 2584	Unknown	AM039462	AM039506	–	AJ512428
Hanseniaspora valbyensis					
CBS 479$^{\mathrm{T}}$	Soil, Germany	AM039458	AM039502	U73596	AJ512434
Hanseniaspora vineae					
CBS 2171$^{\mathrm{T}}$	Soil of vineyard, South Africa	AM039459	AM039503	U84224	AJ512443
Kloeckera lindneri					
CBS 285$^{\mathrm{T}}$	Soil, Java	AM039454	AM039498	U84226	AJ512430

*Used for DNA–DNA reassociation.

Table 2. Primers used

T_{anneal}, Annealing temperature. Position refers to *S. cerevisiae* sequence.

Designation	Primer sequence	T_{anneal} (°C)	Position	Reference
Actin gene				
Act1-F	5′-CTCGTGCTGTCTTCCCATCT-3′	56	388–407	
Act1-R	5′-ACGACGAAAGTGGTCCATCT-3′	56	1393–1412	
Act-HV-F	5′-TGGTGGAGCAATGATTTTGA-3′	53	468–487	
Act-HV-R	5′-GTTGTTGATGAAGCCCAATC-3′	53	1288–1307	
EF-1α gene				
EF20-F	5′-GGTCATGTCGATTTCGGTAAGT-3′	53	40–61	Kurtzman & Robnett (2003)
EF400-R	5′-GGTGGAGGTTTCCTTGACGA-3′	53	503–522	
EF1Y-F	5′-GGCTTTCACCTTGGGTGTTA-3′	53	417–436	
EF1Y-R	5′-CGTCTGTGGTGACGCTAAGA-3′	53	966–985	
EF1-HM-F	5′-CCGTTATTGATGCTCCAGGT-3′	52·8	263–282	
EF1-HM-R	5′-AGGAAATCCGAAGAGGTAACG-3′	52·8	947–967	
D1/D2 domains of LSU rDNA				
NL-1	5′-GCATATCAATAAGCGGAGGAAAAG-3′	53·5	63–86	Kurtzman & Robnett (1998)
NL-4	5′-GGTCCGTGTTTCAAGACGG-3′	53·5	624–642	Kurtzman & Robnett (1998)
Internal transcribed spacer				
ITS1-F	5′-CTTGGTCATTTAGAGGAACTAA-3′	55	–	Gardes & Bruns (1993)
ITS4	5′-TCCTCCGCTTATTGATATGC-3′	55	–	White *et al.* (1990)

gene dataset. The Tamura–Nei model was chosen for the EF-1α and D1/D2 datasets, with only the gamma distribution parameter (TN + Γ) for the former dataset and the addition of the estimate of invariant sites (TN + I + Γ) for the latter. The Hasegawa–Kishino–Yano 85 model with the gamma distribution parameter and the estimate of invariant sites (HKY + I + Γ) was proposed for the ITS dataset. When sequences of closely related taxa are compared, a strong correlation between phyletic and *p* distances exists. With increasing distances between taxa, the degree of sequence divergence is underestimated due to parallel and backward substitutions and the regression curve starts to decline. When the slope of the curve approaches zero, mutational saturation is reached and the phylogenetic reconstruction is more likely to suffer from artefacts such as long-branch attraction (Felsenstein, 1978).

S. cerevisiae, *Lachancea kluyveri* and *Pichia anomala* were tested as outgroup species using the saturation plots. Because the levels of sequence divergence between *S. cerevisiae*, *L. kluyveri* and ingroup species overlapped, *P. anomala* was chosen as the outgroup.

AFLP assay. Amplified fragment length polymorphism (AFLP) analysis was performed according to the procedure described by Vos *et al.* (1995) and adapted for fluorescence-based detection by Bandelj *et al.* (2004). Genomic DNA was digested with the restriction enzymes *Mse*I (1 U) and *Eco*RI (5 U) and linked to *Mse*I and *Eco*RI adapters (50 and 5 pmol, respectively). Restricted and ligated DNA was pre-amplified using MseI-core and EcoRI-core primers. The PCR conditions for the pre-amplification step were 2 min at 72 °C, followed by 20 cycles of 30 s at 94 °C, 30 s at 56 °C and 2 min at 72 °C. Selective amplifications were performed with diluted (1 : 10, by vol.) pre-amplification mixture using three selective primer combinations (MseC/EcoAC, MseC/EcoCG and MseAC/EcoCG). The EcoAC and EcoCG primers were labelled for fluorescence detection with Cy5 dye. The PCR conditions were 2 min at 94 °C, 10 cycles of 30 s at 94 °C, 30 s at 66 °C (decreasing by 1 °C each step of the cycle) and 2 min at 72 °C, followed by 20 cycles of 30 s at 94 °C, 30 s at 56 °C and 2 min at 72 °C. The PCR products were prepared for electrophoresis by addition of an equal volume of formamide loading dye (98 % formamide, 10 mg blue dextran ml^{-1}, 50 mM EDTA) and denatured by heating at 94 °C for 3 min. AFLP fragments were separated on 7·5 % polyacrylamide gels containing 7 M urea on ALF Express II (Amersham Pharmacia Biotech). ALF-Express sizer 50-500 (Amersham Pharmacia Biotech) was used as an external standard.

The AFLP profiles were combined in composite fingerprints using the Bionumerics software package (version 4.0; Applied Maths). Similarities between profiles were calculated using Pearson's correlation coefficient and cluster analysis was performed using the UPGMA algorithm.

Genomic DNA analysis. For DNA extraction, the strains were grown for 2 days at 25 °C on a rotary shaker at 125 r.p.m. in 0·5 l yeast extract/malt extract broth (YM; Difco) using 1 l flat-bottom flasks. Isolation and purification of the DNA, determination of DNA base composition and DNA–DNA reassociation were performed according to procedures described previously (Golubev *et al.*, 1989). Strains used for DNA–DNA reassociation are indicated in Table 1.

RESULTS AND DISCUSSION

Comparison of phylogenetic information in different datasets for resolving *Hanseniaspora–Kloeckera* species

Partial sequences of two nuclear protein-coding genes, actin and EF-1α, were generated and analysed phylogenetically for 26 strains of *Hanseniaspora–Kloeckera* species. The newly generated sequences were compared with the D1/D2 sequences of the LSU rDNA and ITS1–5·8S rDNA–ITS2 sequences. Sequences for the two regions of the ribosomal gene complex were obtained from previous studies in which they were used as phylogenetic markers for inferring relationships between *Hanseniaspora–Kloeckera* species (Kurtzman & Robnett, 1998; Cadez *et al.*, 2003). We compared sequence divergence values, the heterogeneity of evolutionary rates and the reliability of the inferred trees from the four datasets of *Hanseniaspora–Kloeckera* strains. The results are presented in Supplementary Table S1 and Fig. S1 in IJSEM Online. The four datasets are influenced by functional and structural constraints and therefore differed in the amount and form of variation. For example, the ITS regions are under more relaxed structural constraints, which was evident from the highest mean sequence divergence between *Hanseniaspora–Kloeckera* species (0·253; Supplementary Table S1). On the other hand, the protein-coding datasets resolved better at the strain level (0·005; Supplementary Table S1) and between the closely related species of *Hanseniaspora–Kloeckera* (0·03; Supplementary Table S1). These results are consistent with the observations of Daniel & Meyer (2003) that synonymous substitutions in codons account for the high degree of sequence divergence between closely related taxa. Furthermore, the saturation plots (Supplementary Fig. S1 in IJSEM Online) indicated heterogeneity in the evolutionary rates for the ITS dataset (i.e. mean p distance, 0·12; SD = 0·11; Supplementary Fig. S1, ITS, point 3). This had been already predicted from the fact that the number of substitutions in the ITS regions was incongruent with the DNA–DNA reassociation values for closely related species of *Hanseniaspora* (Cadez *et al.*, 2003).

For all four datasets, we detected an overlap in the levels of sequence divergence between the distantly related *Hanseniaspora–Kloeckera* species and the levels of sequence divergence between the *Hanseniaspora–Kloeckera* species and the outgroup species of the Saccharomycetaceae (Supplementary Fig. S1 in IJSEM Online). Kurtzman (2003) observed a similar overlap of intra- and intergeneric distances for *Hanseniaspora* and four related genera calculated from a dataset of six genes. In view of these results, the reinstatement of the genus *Kloeckeraspora* to accommodate *H. occidentalis*, *H. osmophila* and *H. vineae*, as proposed by Yamada *et al.* (1992b), could be justified. Nevertheless, we advocate the argument of Boekhout *et al.* (1994) against splitting the genus on the basis of genetic divergence only, because the two species groups share many similarities in morphology, physiology and ecology.

The phylogenetic usefulness of each dataset could only be determined after cladistic analyses (Damgaard & Cognato, 2003). Most-parsimonious trees based on single datasets and their characteristics are available as Supplementary Fig. S2 and in Supplementary Table S1, respectively, in IJSEM Online. In general, the trees showed some degree of congruence. However, the lack of divergence between the recently diverged *H. uvarum–H. meyeri–H. clermontiae* and *H. guilliermondii–H. lachancei–H. opuntiae–Hanseniaspora* sp. CBS 8772^T species complexes resulted in a lack of statistical support for these clades in single genealogies. Therefore, we concatenated the datasets in order to increase the phylogenetic information. Prior to concatenation, we tested the congruence of the datasets with the partition homogeneity test (Farris *et al.*, 1995). When all four datasets were combined, significant incongruence ($P = 0·016$) was found and this resulted in a weakly supported phylogenetic tree. Therefore we tested different combinations of the concatenated datasets that resulted in the highest P values for the combined dataset without either ITS or EF-1α sequence data ($P = 0·35$ and 0·24, respectively). Finally, parsimony analysis of the combined dataset of actin, D1/D2 and ITS sequences without EF-1α data produced a completely resolved tree with high statistical support (Fig. 1).

With the aim of evaluating whether the inferred phylogenetic tree reflects the biological relationships, we compared the relationships with key physiological characteristics and DNA–DNA reassociation data of *Hanseniaspora–Kloeckera* species reported in previous studies (Meyer *et al.*, 1978; Cadez *et al.*, 2003). Fig. 1 shows that clustering of the species is consistent with the current taxonomy. Clusters of closely related species (*H. vineae–H. osmophila*, *H. guilliermondii–H. lachancei–H. opuntiae–Hanseniaspora* sp. CBS 8772^T, *H. uvarum–H. meyeri–H. clermontiae* and *H. valbyensis–Kloeckera lindneri*) generally share the same nutritional characteristics with two exceptions, *H. lachancei* and *H. clermontiae*. Furthermore, the congruence between the relationships determined from the concatenated datasets and DNA–DNA reassociation improved, in particular for *H. lachancei*, *H. opuntiae* and *Hanseniaspora* sp. CBS 8772^T, which were placed inconsistently in the single genealogies (bootstrap support <50 %; Supplementary Fig. S2 in IJSEM Online). The latter species was not named in our previous study (Cadez *et al.*, 2003) because its description would rest on a single isolate. However, based on additional molecular evidence of species delimitation we propose that strain CBS 8772^T represents a novel species with the name *Hanseniaspora pseudoguilliermondii* sp. nov.

Conflicting gene genealogies within *H. occidentalis*

A comparison of multiple gene genealogies can be used to detect reproductive or geographical isolation between distinct populations. Incongruences between different gene genealogies may be indicative of genetic exchange among individuals within a species (Taylor *et al.*, 2000). In order to clarify the species concept for *H. occidentalis*, we examined the consistency of phylogenetic relationships among *H. occidentalis* strains using three phylogenetic markers. In a previous study (Cadez *et al.*, 2002), the six strains of *H. occidentalis* were found to be genetically heterogeneous, as shown by their chromosomal make-up, RAPD-PCR profiles and restriction patterns of the ITS regions.

First, we determined the genetic variability of 14 *H. occidentalis* strains by cluster analysis of the combined AFLP fingerprints. Fig. 2 shows that the strains segregated into two major groups at a similarity level of 32 %. Interestingly, this grouping coincided with the source of isolation: the strains from group I were isolated from soil whereas the strains from group II were isolated from oranges or orange

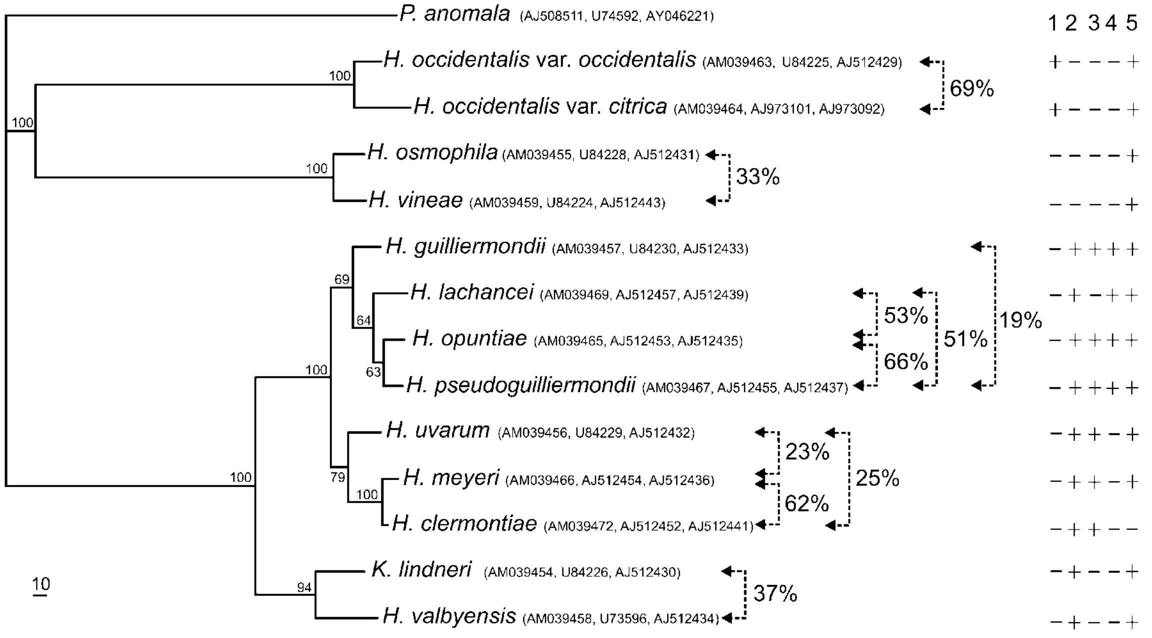

Fig. 1. Phylogeny of *Hanseniaspora–Kloeckera* species based on concatenated datasets of actin, D1/D2 and ITS gene sequences. Each species is represented by its type strain. The single most-parsimonious tree (tree length, 1169; consistency index = 0·7707; retention index = 0·8319) was constructed by heuristic search procedure under the parsimony criterion in PAUP*. Bootstrap percentages from 1000 replicates are shown. *Pichia anomala* was used as the outgroup. Bar, number of nucleotide substitutions. Interspecies DNA–DNA reassociation values are indicated by arrows. +/− show responses to key physiological tests for *Hanseniaspora–Kloeckera* species identification: 1, fermentation of sucrose; 2, growth with 0·01 % cycloheximide; 3, assimilation of 2-keto-D-gluconate; 4, growth at 37 °C; 5, growth at 30 °C.

juices. An exception was strain CBS 2569, which was isolated from a fruit fly. Nevertheless, this strain showed a moderate AFLP similarity (42 %) to the rest of the strains in group I.

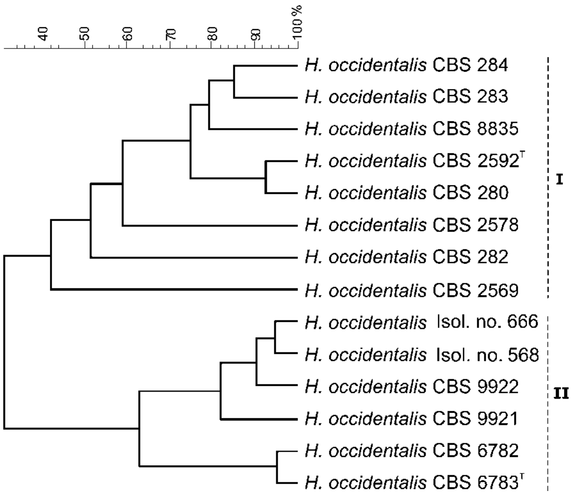

Fig. 2. UPGMA dendrogram of *H. occidentalis* strains based on combined AFLP fingerprints obtained with three primer combinations (MseC/EcoAC, MseC/EcoCG and MseAC/EcoCG). The distance between the strains was calculated as the Pearson's correlation coefficient.

The existence of two major groups was confirmed by DNA–DNA reassociation (Supplementary Table S2 in IJSEM Online). The reassociation values obtained when group I strains (CBS 2592^T, CBS 280, CBS 282, CBS 283, CBS 284, CBS 2569 and CBS 2578) were paired with group II strains (CBS 6783^T and CBS 6782) ranged from 60 to 78 %. The DNA–DNA relatedness values within group I, which included the type strain, ranged from 88 to 100 %. The mutual DNA–DNA relatedness of strains of group II was 99 %. Within group I, the lowest relatedness value (88 %) was observed between the type strain and strains CBS 2569 and CBS 282, which is in agreement with the AFLP fingerprinting results (Fig. 2). However, the interpretation of DNA–DNA reassociation values of 60–78 % determined between the major groups is ambiguous without further evidence of genetic isolation between the strains.

Phylogenetic trees inferred from the sequences of the actin gene, the EF-1α gene and the ITS regions of *H. occidentalis* strains are presented in Supplementary Fig. S3 in IJSEM Online. Only the topology of the actin tree (Supplementary Fig. S3a) recovered the relationships obtained from the AFLP fingerprints (Fig. 2), whereas the positions of strains CBS 2569 and CBS 9921 in the EF-1α and ITS trees were incongruent. For example, strain CBS 9921, a member of group II in the actin and ITS trees (Supplementary Fig. S3a, c), was placed in group I in the EF-1α tree (Supplementary

Fig. S3b). On the other hand, strain CBS 2569 was basal to group I in the actin tree but was basal to group II in the EF-1α and ITS trees. The sequence divergence in the D1/D2 regions was too low (0–1 substitutions) for phylogenetic reconstruction.

From the results presented, we might predict that the two groups of *H. occidentalis* are isolated by habitat preference and may present an example of an ongoing speciation event. Therefore, we propose that the two divergent groups within *H. occidentalis* should be recognized as varieties, with the names *H. occidentalis* var. *occidentalis* for group I and *H. occidentalis* var. *citrica* for group II.

Taxonomy

Latin diagnosis of *Hanseniaspora pseudoguilliermondii* N. Cadez, P. Raspor *et* M. Th. Smith sp. nov.

In medio liquido post 48 *horas* 25 °C *cellulae apiculatae, ovoideae vel elongatae*, 2·2–8·7 × 1·6–4·2 μm, *singulae vel binae*; *gemmatione bipolari repruducentes. Post unum mensem annulus tenuis et sedimentum formantur. In agaro farina Zeae maydis confecto pseudomycelium rudimentarium. In quoque asco* 4, *ascosporae petasiformes. Glucosum et cellobiosum fermentantur. Glucosum*, *cellobiosum*, *salicinum*, *arbutinum*, *glucono-δ-lactonum*, 2-*ketogluconatum*, *acidum gluconicum*, *ethylaminum*, *lysinum et cadaverinum assimilantur. Non assimilantur galactosum*, L-*sorbosum*, D-*glucosaminum*, D-*ribosum*, D-*xylosum*, L-*arabinosum*, D-*arabinosum*, L-*rhamnosum*, *sucrosum*, *maltosum*, *trehalosum*, *methyl* α-D-*glucosidum*, *melibiosum*, *lactosum*, *raffinosum*, *melezitosum*, *amylum solubile*, *glycerolum*, *erythritolum*, *ribitolum*, *xylitolum*, L-*arabinitolum*, D-*glucitolum*, D-*mannitolum*, *galactitolum*, *inositolum*, *acidum* D-*glucuronicum*, *acidum* D-*galacturonicum*, *acidum* DL-*lacticum*, *acidum succinicum*, *acidum citricum*, *methanolum et ethanolum. Maxima temperatura crescentiae* 37 °C. *Crescit in medio addito* 10 % NaCl, 50 % *glucoso et* 0·1 % *cycloheximido*. G + C *acidi deoxyribonucleati* 31·5 mol%. *Typus* CBS 8772^T (= NCAIM Y.741^T) *in collectione zymotica* Centraalbureau voor Schimmelcultures, *Trajectum ad Rhenum lyophilus depositus*.

Description of *Hanseniaspora pseudoguilliermondii* N. Cadez, P. Raspor & M. Th. Smith sp. nov.

Etymology: the epithet *pseudoguilliermondii* is chosen because the species is similar to *H. guilliermondii*.

In YM liquid medium after 48 h at 25 °C cells are apiculate, ovoid to elongate, 2·2–8·7 × 1·6–4·2 μm, single or in pairs. Budding is bipolar. Sediment is present. After 1 month, a very thin ring and a sediment are formed. After 1 month at 25 °C a streak culture on malt agar is cream coloured, butyrous, smooth, glossy, flat to slightly raised at the centre, with an entire to slightly undulate margin. On cornmeal agar a rudimentary pseudomycelium is formed. Asci containing four hat-shaped ascospores are observed on 5 % Difco malt extract agar at 25 °C. Glucose and cellobiose are fermented. The following carbon compounds are assimilated: glucose, cellobiose, salicin, arbutin, glucono-δ-lactone, 2-ketogluconate and D-gluconate. The following are not assimilated: galactose, L-sorbose, D-glucosamine, D-ribose, D-xylose, L-arabinose, D-arabinose, L-rhamnose, sucrose, maltose, trehalose, methyl α-D-glucoside, melibiose, lactose, raffinose, melezitose, starch, glycerol, erythritol, ribitol, xylitol, L-arabinitol, D-glucitol, D-mannitol, galactitol, *myo*-inositol, D-glucuronate, D-galacturonate, DL-lactate, succinate, citrate, methanol and ethanol. Assimilation of nitrogen compounds is positive for ethylamine, lysine and cadaverine but negative for sodium nitrate and sodium nitrite. Growth at 37 °C is positive; growth at 40 °C is negative. Growth on YM agar with 10 % NaCl is positive. Growth on 50 % glucose (w/w) yeast extract agar is weak. Growth in the presence of 0·1 % cycloheximide is positive. Diazonium blue B reaction is negative. G + C content of DNA (T_m) is 31·5 mol%.

The type strain, CBS 8772^T (= NCAIM Y.741^T), was isolated from orange juice concentrate, GA, USA. It has been deposited at the Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands.

Hanseniaspora occidentalis Smith var. *occidentalis*

Antonie van Leeuwenhoek (1974) **40**, 441–444.

Type strain: ex-type isolate CBS 2592^T (= ATCC 32053^T) is a living strain in the CBS yeast collection.

Hanseniaspora occidentalis var. *citrica* N. Cadez, P. Raspor & M. Th. Smith var. nov.

Etymology: the epithet *citrica* of *Citrus*, the genus name of the host plant.

Type strain: ex-type isolate CBS 6783^T is a living strain in the CBS yeast collection. *Varietas a Hanseniaspora occidentalis var. citrica differt*: *trehalosum assimilantur (inferme)*. *Typus*: CBS 6783^T *in collectione zymotica* Centraalbureau voor Schimmelcultures, *Trajectum ad Rhenum lyophilus depositus*.

This variety differs from *Hanseniaspora occidentalis* var. *occidentalis* by assimilation of trehalose (weak). The type strain, CBS 6783^T, was isolated from orange juice from Italy. It has been deposited at the Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands.

ACKNOWLEDGEMENTS

We thank Ann Vaughan-Martini for providing cultures. Vincent Robert is acknowledged for valuable comments on the manuscript.

REFERENCES

Baldauf, S. L. & Palmer, J. D. (1993). Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc Natl Acad Sci U S A* **90**, 11558–11562.

Baldauf, S. L., Roger, A. J., Wenk-Siefert, I. & Doolittle, W. F. (2000). A kindom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**, 972–977.

Bandelj, D., Jakše, J. & Javornik, B. (2004). Assessment of genetic variability of olive varieties by microsatellite and AFLP markers. *Euphytica* **136**, 93–102.

Barnett, J. A., Payne, R. W. & Yarrow, D. (2000). *Yeasts: Characteristics and Identification.* Cambridge: Cambridge University Press.

Belloch, C., Querol, A., Garcia, M. D. & Barrio, E. (2000). Phylogeny of the genus *Kluyveromyces* inferred from the mitochondrial cytochrome-*c* oxidase II gene. *Int J Syst Evol Microbiol* **50**, 405–416.

Boekhout, T., Kurtzman, C. P. O'Donnell K. & Smith, M. T. (1994). Phylogeny of the yeast genera *Hanseniaspora* (anamorph *Kloeckera*), *Dekkera* (anamorph *Brettanomyces*), and *Eeniella* as inferred from partial 26S ribosomal DNA nucleotide sequences. *Int J Syst Bacteriol* **44**, 781–786.

Bush, R. M. & Everett, K. D. E. (2001). Molecular evolution of the *Chlamydiaceae*. *Int J Syst Evol Microbiol* **51**, 203–220.

Cadez, N., Raspor, P., de Cock, A. W. A. M., Boekhout, T. & Smith, M. Th. (2002). Molecular identification and genetic diversity within species of the genera *Hanseniaspora* and *Kloeckera*. *FEMS Yeast Res* **1**, 279–289.

Cadez, N., Poot, G. A., Raspor, P. & Smith, M. Th. (2003). *Hanseniaspora meyeri* sp. nov., *Hanseniaspora clermontiae* sp. nov., *Hanseniaspora lachancei* sp. nov. and *Hanseniaspora opuntiae* sp. nov., novel apiculate yeast species. *Int J Syst Evol Microbiol* **53**, 1671–1680.

Damgaard, J. & Cognato, A. I. (2003). Sources of character conflict in a clade of water strides (Heteroptera: Gerridae). *Cladistics* **19**, 512–526.

Daniel, H.-M. & Meyer, W. (2003). Evaluation of ribosomal RNA and actin gene sequences for the identification of ascomycetous yeasts. *Int J Food Microbiol* **86**, 61–78.

Daniel, H.-M., Sorrell, T. C. & Meyer, W. (2001). Partial sequence analysis of the actin gene and its potential for studying the phylogeny of *Candida* species and their teleomorphs. *Int J Syst Evol Microbiol* **51**, 1593–1606.

Esteve-Zarzoso, B., Peris-Torán, M. J., Ramón, D. & Querol, A. (2001). Molecular characterisation of *Hanseniaspora* species. *Antonie van Leeuwenhoek* **80**, 85–92.

Farris, J. S., Kallersjo, M., Kluge, A. G. & Bult, C. (1995). Testing significance of incongruence. *Cladistics* **10**, 315–319.

Felsenstein, J. (1978). Cases in which parsimony and compatibility methods will be positively misleading. *Syst Zool* **27**, 401–410.

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

Gardes, M. & Bruns, T. D. (1993). ITS primers with enhanced specificity for basidiomycetes – application to the identification of mycorrhizae and rusts. *Mol Ecol* **2**, 113–118.

Golubev, W. I., Smith, M. Th., Poot, G. A. & Kock, J. L. F. (1989). Species delineation in the genus *Nadsonia* Sydow. *Antonie van Leeuwenhoek* **55**, 369–382.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**, 95–98.

Kretzer, A. M. & Bruns, T. D. (1999). Use of atp6 in fungal phylogenetics: an example from the Boletales. *Mol Phylogenet Evol* **13**, 483–492.

Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244–1245.

Kurtzman, C. P. (2003). Phylogenetic circumscription of *Saccharomyces*, *Kluyveromyces* and other members of the Saccharomycetaceae, and the proposal of the new genera *Lachancea*, *Nakaseomyces*, *Naumovia*, *Vanderwaltozyma* and *Zygotorulaspora*. *FEMS Yeast Res* **4**, 233–245.

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 van Leeuwenhoek* **73**, 331–371.

Kurtzman, C. P. & Robnett, C. J. (2003). Phylogenetic relationships among yeasts of the '*Saccharomyces* complex' determined from multigene sequence analyses. *FEMS Yeast Res* **3**, 417–432.

Lodder, J. & Kreger-van Rij, N. J. W. (1952). *The Yeasts, a Taxonomic Study*. Amsterdam: North Holland.

Meyer, S. A., Smith, M. Th. & Simione, F. P., Jr (1978). Systematics of *Hanseniaspora* Zikes and *Kloeckera* Janke. *Antonie van Leeuwenhoek* **44**, 79–96.

Misof, B., Rickert, A. M., Buckley, T. R., Fleck, G. & Sauer, K. P. (2001). Phylogenetic signal and its decay in mitochondrial SSU and LSU rRNA gene fragments of Anisoptera. *Mol Biol Evol* **18**, 27–37.

Möller, E. M., Bahnweg, G., Sandermann, H. & Geiger, H. H. (1992). A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues. *Nucleic Acids Res* **20**, 6115–6116.

Nakase, T. & Komagata, K. (1970). Significance of DNA base composition in the classification of yeast genera *Hanseniaspora* and *Kloeckera*. *J Gen Appl Microbiol* **16**, 241–250.

O'Donnell, K., Cigelnik, E. & Nirenberg, H. I. (1998). Molecular systematics and phylogeography of the *Gibberella fujikuroi* species complex. *Mycologia* **90**, 465–493.

Posada, D. & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.

Reess, M. (1870). *Botanische Untersuchungen über die Alkoholgährungspilze*. Leipzig: A. Felix (in German).

Rokas, A., Williams, B. L., King, N. & Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**, 798–804.

Smith, M. Th. (1998). *Hanseniaspora* Zikes. In *The Yeasts, a Taxonomic Study*, 4th edn, pp. 214–220. Edited by C. P. Kurtzman & J. W. Fell. Amsterdam: Elsevier.

Swofford, D. L. (2002). PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods), version 4.0b10. Sunderland, MA: Sinauer Associates.

Taylor, J. W., Jacobson, D. J., Kroken, S., Kasuga, T., Geiser, D. M., Hibbett, D. S. & Fisher, M. C. (2000). Phylogenetic species recognition and species concepts in fungi. *Fungal Genet Biol* **31**, 21–32.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.

Vos, P., Hogers, R., Bleeker, M. & 7 other authors (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* **23**, 4407–4414.

White, T. J., Bruns, T., Lee, S. & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR Protocols: a Guide to Methods and Applications*, pp. 315–322. Edited by N. Innis, D. Gelfand, J. Sninsky & T. White. London: Academic Press.

Yamada, Y., Maeda, K. & Banno, I. (1992a). The phylogenetic relationships of the Q_6-equipped species in the teleomorphic apiculate yeast genera *Hanseniaspora*, *Nadsonia*, and *Saccharomycodes* based on the partial sequences of 18S and 26S ribosomal ribonucleic acids. *J Gen Appl Microbiol* **38**, 585–596.

Yamada, Y., Maeda, K. & Banno, I. (1992b). An emendation of *Kloeckeraspora* Niehaus with the type species, *Kloeckeraspora osmophila* Niehaus and the proposals of two new combinations, *Kloeckeraspora occidentalis* and *Kloeckeraspora vineae* (Saccharomycetaceae). *Bull Jpn Fed Cult Collect* **8**, 79–85.

Yarrow, D. (1998). Methods for the isolation, maintenance and identification of yeasts. In *The Yeasts, a Taxonomic Study*, 4th edn, pp. 77–100. Edited by C. P. Kurtzman & J. W. Fell. Amsterdam: Elsevier.

Zikes, H. (1911). Zur Nomenclaturfrage der Apiculatushefen. *Zentralbl Bakteriol Parasitenkd Abt II* **30**, 145–149 (in German).