Persoonia 54, 2025: 265–283
https://www.persoonia.org

RESEARCH ARTICLE

ISSN (Online) 1878-9080
https://doi.org/10.3114/persoonia.2025.54.08

A taxogenomic view of the genus *Torulaspora*: an expansion from ten to twenty-two species

M.R. Silva[1,2#], F. Paraíso[1,2#], J. Al-Oboudi[3,4#], M. Abegg[5,6], A. Aires[1,2,7], K.O. Barros[3,4,5], P.H. Brito[1,2], M. Jarzyna[3,4], K. Sylvester[3,4], Q.K. Langdon[3], D.A. Opulente[3,4], F. Carriconde[8], J.W. Fell[9], T.A. Hofmann[10], M.-A. Lachance[11], J.-L. Legras[12], D. Libkind[13], A. Pontes[1,2], P. Gonçalves[1,2], C.A. Rosa[5], M. Groenewald[14], C.T. Hittinger[3,4*], J.P. Sampaio[1,2,7*]

[1]UCIBIO, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal
[2]Associate Laboratory i4HB - Institute for Health and Bioeconomy, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal
[3]Laboratory of Genetics, Wisconsin Energy Institute, Center for Genomic Science Innovation, J. F. Crow Institute for the Study of Evolution, Microbiology Doctoral Training Program, University of Wisconsin-Madison, Madison, WI 53726, USA
[4]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI 53726, USA
[5]Departamento de Microbiologia, ICB, C.P. 486, Universidade Federal de Minas Gerais, Belo Horizonte, MG, 31270-901, Brazil
[6]Instituto de Ciências Exatas e Tecnologia - ICET, Universidade Federal do Amazonas - UFAM, Itacoatiara - AM, 69100-000, Brazil
[7]PYCC – Portuguese Yeast Culture Collection, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal
[8]Equipe "Sol & Végétation" (SolVeg), Institut Agronomique néo-Calédonien (IAC), 98000 Nouméa, New Caledonia, France
[9]Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL 33149, USA
[10]Mycological Research Center (CIMi), Herbarium UCH, Universidad Autónoma de Chiriquí - UNACHI, 0427 David, Chiriquí province, Panama
[11]Department of Biology, University of Western Ontario, London, Ontario N6A 5B7, Canada
[12]SPO, Université de Montpellier, INRAE, Institut Agro, Montpellier 34060, France
[13]Centro de Referencia en Levaduras y Tecnología Cervecera (CRELTEC), Instituto Andino Patagónico de Tecnologías Biológicas y Geoambientales (IPATEC), Universidad Nacional del Comahue, CONICET, CRUB, Quintral 1250, San Carlos de Bariloche, 8400, Río Negro, Argentina
[14]Westerdijk Fungal Biodiversity Institute, Uppsalalaan 8, 3584CT Utrecht, The Netherlands

[#]These authors contributed equally
*Corresponding authors: C.T. Hittinger, cthittinger@wisc.edu; J.P. Sampaio, jss@fct.unl.pt

Key words:
new taxa
phenotype
phylogenomics
taxogenomics
Torulaspora
yeast

Abstract: The yeast genus *Torulaspora* (subphylum *Saccharomycotina*, family *Saccharomycetaceae*) is mostly known from its type species, *T. delbrueckii*, a frequent colonizer of wine and sourdough bread fermentations. The genus currently contains 10 species that are typically found in various natural terrestrial environments in temperate and tropical climates. Here we employ taxogenomic analyses to investigate a large collection of *Torulaspora* strains obtained in multiple surveys we carried out in Asia, Australasia, North America, South America, and Europe, and to which we added several strains maintained in culture collections. Our analyses detected twelve novel species that are formally described here, thereby more than doubling the species diversity of *Torulaspora*. We also sketch a genotype-phenotype map for the genus and show how the complex relationship between key genes and the physiological traits they control both between and within species. This remarkable increase in the number of species in the genus *Torulaspora* highlights how limited the current inventory of fungal taxa is. It also shows how integrated taxogenomic approaches can foster the assessment of species circumscriptions in fungi.

Citation: Silva MR, Paraíso F, Al-Oboudi J, Abegg M, Aires A, Barros KO, Brito PH, Jarzyna M, Sylvester K, Langdon QK, Opulente DA, Carriconde F, Fell JW, Hofmann TA, Lachance M-A, Legras J-L, Libkind D, Pontes A, Gonçalves P, Rosa CA, Groenewald M, Hittinger CT, Sampaio JP (2025). A taxogenomic view of the genus *Torulaspora*: an expansion from ten to twenty-two species. *Persoonia* **54**: 265–283. doi: 10.3114/persoonia.2025.54.08
Received: 27 September 2024; **Accepted:** 2 May 2025; **Effectively published online**: 23 May 2025
Corresponding editor: J. Houbraken

INTRODUCTION

The genus *Torulaspora* currently contains 10 species that are typically found in various natural terrestrial environments, most frequently soil and plant material in temperate and tropical climates. The genus belongs to the family *Saccharomycetaceae* of the yeast subphylum *Saccharomycotina* (Groenewald *et al.* 2023) and is sister to *Zygotorulaspora* and *Zygosaccharomyces* (Shen *et al.* 2018), with *Zygotorulaspora* being intermediate between the other two genera. Salient features of *Torulaspora* are the formation of tapered protuberances resembling conjugation tubes in vegetative cells and the formation of persistent asci with up to four, but frequently two or one, ascospores (Kurtzman

2011). Physiologically, species of the genus *Torulaspora* are adapted to ferment simple sugars and exhibit the Crabtree effect; in other words, they ferment sugars when they are in high concentrations, even if oxygen is present. *Torulaspora* thus resembles *Saccharomyces*, although it is less efficient at sugar fermentation and ethanol production (Hagman *et al.* 2013). In contrast to *Saccharomyces*, their ancestor did not undergo the whole genome duplication documented in the lineage that includes *Saccharomyces* (Wolfe & Shields 1997).

Torulaspora delbrueckii is the most emblematic species in the genus because it has been frequently isolated in association with commercial fermentations, particularly those related to the production of wine and sourdough bread, but also from other sources like for example cocoa and coffee fermentations. It is also very frequently isolated from the wild and has been described as cosmopolitan (Spurley *et al.*, 2022). This species is considered an important microorganism for biotechnology due to its fermentative abilities and tolerance to various stresses such as freezing and freeze-thawing, ethanol toxicity, osmotic pressure, and acidic conditions (Pacheco *et al.*, 2012). Besides the more conventional utilizations of *T. delbrueckii* mentioned above, novel applications have been proposed for brewing (Canonico *et al.* 2016) and cider fermentations (Wei *et al.* 2019), to aroma enhancement in co-fermentations (Zhang *et al.* 2018), and utilization as starter in frozen doughs (Alves-Araújo *et al.* 2004).

Besides *T. delbrueckii*, available evidence suggests that *T. microellipsoides* might be also able to colonize anthropic niches. First, the few strains presently known were mostly isolated from artificial environments and, secondly, it has been extensively documented that wine strains of *Saccharomyces cerevisiae* horizontally acquired a 165-kb genomic region from *T. microellipsoides* (Marsit *et al.* 2015, Peter *et al.* 2018), thus suggesting co-existence in the same environment.

The remaining *Torulaspora* species have been almost exclusively found in natural environments and in relatively low frequencies. For example, *T. franciscae* was found only once, in 1955, in a sample of Spanish soil, while *T. maleeae*, described in 2007, was based mostly on strains isolated from mosses collected in Thailand and from soil samples collected in tropical Japan (Kurtzman 2011). Although some strains of *T. quercuum* have been isolated from the arboreal and soil niches in temperate climates, other strains have been isolated from humans and dairy products (Wang *et al.* 2009). At present, *T. delbrueckii* is the only species with many isolates that have been obtained both in anthropic and arboreal niches. Thus, it is the only species in the genus for which more solid inferences on ecology and population structure have been made (Albertin *et al.* 2014, Silva *et al.* 2022).

The use of genomics to evaluate yeast species circumscriptions and to inform taxonomy is gaining momentum (Libkind *et al.* 2020, Groenewald *et al.* 2023), and the number of novel species described under a taxogenomics approach is increasing (Libkind *et al.* 2011, Lopes *et al.* 2016, Haase *et al.* 2017, Morais *et al.* 2017, Čadež *et al.* 2019, David-Palma *et al.* 2020, Čadež *et al.* 2021, Santos *et al.* 2021, Brysch-Herzberg *et al.* 2023, Opulente *et al.* 2023, Santos *et al.* 2023). It is thus expected that, in the near future, taxogenomics will be the standard procedure for new species descriptions.

Over the past two decades, we have conducted various yeast surveys mostly in wild but also in anthropic environments. Several strains were molecularly identified as belonging to *Torulaspora* based on comparisons of D1/D2 and ITS rDNA sequences. Here we analyse with genomics this diverse set of *Torulaspora* isolates collected in Asia, Australasia, North America, South America, and Europe, together with several collection strains collected as far back as 1930's. Our taxogenomic analyses detected twelve novel species, thereby more than doubling the species diversity of the genus, which we propose to expand from ten to twenty-two species with formal taxonomic descriptions. We also sketch a genotype-phenotype map for the genus and show how key genes have evolved in concert with the physiological traits they control.

MATERIALS AND METHODS

Yeast isolation

Given that each novel species was obtained in a different and independent survey, the details of the isolation procedures were specific for each of them. All information available on the isolations is shown in Supplementary Data S1.

Phenotypic characterization

Morphological and physiological tests were conducted according to the methods described by Kurtzman *et al.* (2011). In brief, the evaluation of the ability to utilize different compounds as carbon or nitrogen sources was carried out in test tubes with 5 mL of liquid media, incubated at 25 °C with gentle agitation for three wk. The ability to grow at different temperatures was carried out in yeast-peptone-dextrose (YPD) broth (1 % yeast extract, 2 % peptone, 2 % glucose), in test tubes incubated in water baths at the desired temperatures. Ascospore formation was investigated on acetate agar (0.25 % yeast extract, 0.1 % glucose, 1 % potassium acetate, 1.5 % agar) and on corn meal agar (Difco). Dalmau plates for investigating formation of pseudohyphae and true hyphae were prepared with corn meal agar.

DNA barcode sequence analysis (D1/D2 domain and complete ITS region of the rDNA)

Using our dataset of assembled genomes, we retrieved using BLASTn (Altschul *et al.* 1990) the D1/D2 domain of the nuclear large subunit (LSU) 26S rDNA gene and the complete internal transcribed spacer region (ITS1–5.8S–ITS2). D1/D2 and ITS sequences from the type strains of each new species served as queries in BLASTn searches against the nucleotide database (nt) of NCBI. The top 100 sequences from each BLASTn search were retrieved, and duplicates were removed. Sequences were confirmed with ITS1 and NL4 primers and aligned using MAFFT online service (Katoh *et al.* 2019). A Maximum Likelihood (ML) tree was constructed with IQ-TREE v. 2.2.6, using automatic detection of the best-fitting model and 1000 ultrafast bootstrap replicates. Alignments and phylogenetic trees have been deposited in figshare (https://doi.org/10.6084/m9.figshare.28794995.v1).

Whole-genome sequencing, phylogenetic analyses, and gene content analyses

Paired-end Illumina NextSeq sequencing (300 cycles) was conducted at the Genomics Unit of Instituto Gulbenkian Ciência (Oeiras, Portugal) and at the University of Wisconsin Biotechnology Center. The publicly accessible reads for other *Torulaspora* species and the raw sequencing data underwent preprocessing two different ways using trimmomatic v. 0.39 (Bolger *et al.* 2014): just eliminating adapter sequences and with an optimized trimming step (LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) to eliminate low-quality bases. Subsequently, whole genome de novo assembly was carried out on all sets of pre-processed reads using SPAdes v. 3.13.1 (Bankevich *et al.* 2012). SPAdes was executed in 'careful' mode with k-mer sizes automatically determined based on the read length. The quality assessment of all resulting genome assemblies for each strain was performed using quast v. 5.0 (Gurevich *et al.* 2013), and the assembly with the highest genome size and N50 value was selected as the best. Small contigs (< 1 kb) were discarded from final assemblies.

For *T. delbrueckii* PYCC 2477 and *T. obscura* PYCC 8933, long-read data was additionally acquired using Oxford Nanopore Technology (ONT), with a MinION flowcell. For de novo assembly, Canu v. 2.2 (Koren *et al.* 2017) was utilized with default parameters, only adjusting the genome size flag to 10m. The resulting contigs were corrected with two rounds of Racon v. 1.5.0 (Vaser *et al.* 2017), one using Nanopore reads and the other using Illumina reads. Afterwards, several rounds of Pilon v. 1.24 (Walker *et al.* 2014) were executed with Illumina data until no further modifications were observed in the change file. To enhance assembly contiguity, LINKS v. 1.8.7 (Warren *et al.* 2015) was implemented.

For all genomes, including publicly available *Torulaspora* genome assemblies, ab initio prediction of protein coding genes and annotation was performed with Yeast Genome Annotation Pipeline (YGAP) (Proux-Wéra *et al.* 2012).

To reconstruct the species phylogeny, we obtained single-copy orthogroups (SCOs) using Orthofinder v. 2.5.4 (Emms & Kelly 2019) from the predicted proteomes of the *Torulaspora* genus. These SCOs were independently aligned using mafft v. 7.407 (Katoh & Standley 2013). Subsequently, the aligned SCO sequences were concatenated into a single dataset. The concatenated alignment, comprising 3766 SCOs, was then utilized to construct a Maximum Likelihood (ML) tree using IQ-TREE v. 2.2.6 (Nguyen *et al.* 2015) with a partition flag (-spp), an automatic detection of the best-fitting model of amino acid evolution (Kalyaanamoorthy *et al.* 2017) and 1000 ultrafast bootstrapping replicates (Hoang *et al.* 2018).

For the gene content analyses, protein sequences were extracted with a tBLASTN search using *T. delbrueckii* sequences as queries with an expect value (e) cutoff of 1e-50. Only full-length hits, based on the expected length of the query sequence, were considered present to exclude fragmented or partial hits and draw more direct correlations between the presence of functional genes and positive phenotypes. When no hits were returned for a tBLASTN search at e < 1e-50, the gene was assumed to be either absent or fragmented beyond standard recognition for that genome. Substrate-specificities for alpha-glucosidase genes were inferred based on previous observations in various yeast species (Viigand *et al.* 2018), including multiple *Torulaspora* species (Silva *et al.* 2023). Alignments and phylogenetic trees have been deposited in figshare (https://doi.org/10.6084/m9.figshare.28788923.v1).

Average nucleotide identities (ANI)

Average nucleotide identities (ANI) values were estimated from whole genome assemblies using OrthoANI (Lee *et al.* 2016) with an improved algorithm that employees USEARCH (OrthoANIu), as implemented in the Orthologous Average Nucleotide Identity Tool (OAT) (Yoon *et al.* 2017).

RESULTS

Detection of novel *Torulaspora* species with taxogenomics

The strains upon which the novel species descriptions are based (Table S1) were obtained from various isolation programs and surveys whose details are given in Supplementary Data S1. They were preliminarily identified using the sequences of the D1/D2 or ITS regions of the rDNA. After genome sequencing of strains suspected to represent novel species, single-copy orthogroups (SCOs) were identified from a predicted proteome dataset that included representatives of all known *Torulaspora* species. The resulting maximum likelihood phylogeny is shown in Fig. 1. The putative novel species were distributed across the genus and generally have as closest relatives already known *Torulaspora* species. To further assess discontinuities at the species level, we calculated average nucleotide identity (ANI) values within and between representatives of the main clades of Fig. 1. We took into consideration earlier studies suggesting 95 % or lower identity as a guideline for the separation of yeast (with *et al.* in italics: (Lachance *et al.* 2020, Libkind *et al.* 2020), microsporidian (de Albuquerque & Haag 2023), and prokaryotic (Jain *et al.* 2018) species, and expanded our earlier calculations for species in this genus (Silva *et al.* 2022). In the present study, all proposed novel species had ANI values that were equal or lower than 91 % when compared with their closest relatives, whereas intraspecific values were equal or higher than 96.5 % (Fig. 1 and Fig. S1). Taken together, our analyses suggested the existence of an additional twelve novel species, which would add to the ten species currently recognized in the genus *Torulaspora*, more than doubling its size.

Recognition of the novel species using DNA barcode sequences

To investigate if the new species could be recognized using DNA barcode sequences and to determine if additional representatives could be found among sequences deposited in the NCBI archive, we prepared extensive phylogenies based on the two DNA barcode regions most frequently used for the delineation of yeast species, the D1/D2 region of the LSU rDNA and the complete ITS region. The corresponding phylogeny for the D1/D2 region is shown in Fig. S2, the phylogeny based on the ITS region is shown in Fig. S3, and the phylogeny combining both regions is shown in Fig. S4. We included sequences from representative genomes used

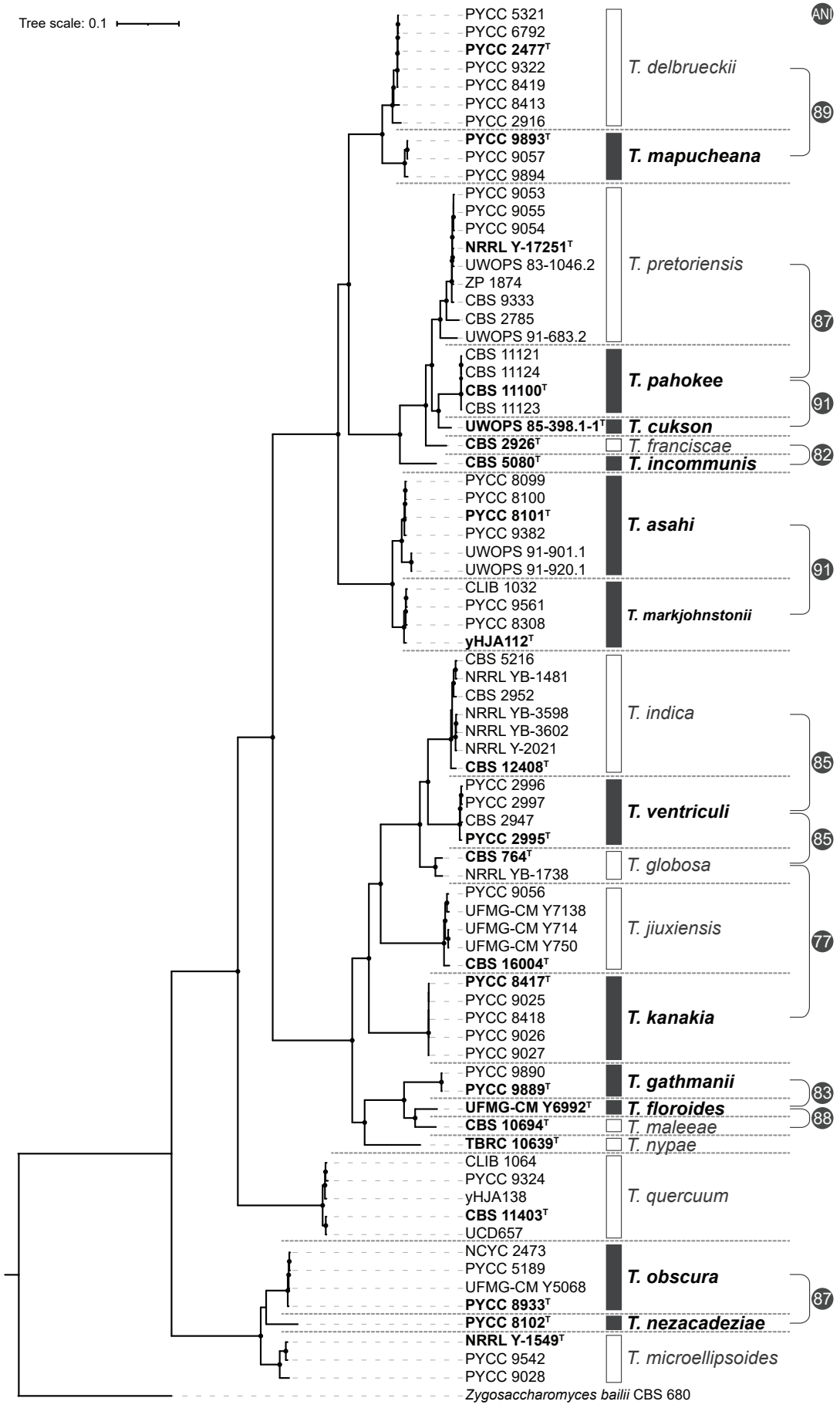

Fig. 1. Phylogenomic placement of the novel taxa within the genus *Torulaspora*. The phylogeny is based on a concatenated alignment of 3766 single-copy orthologs from 77 *Torulaspora* genomes and was rooted with *Zygosaccharomyces bailii*. The JTT+F+I+G4 model of sequence evolution and the maximum-likelihood method were employed. Black bars highlight the novel species, and white bars represent the known ones. Nodes with black dots represent ≥ 95 % bootstrap support. The scale bar corresponds to the estimated number of substitutions per site. For the proposed novel species, numerals in black circles depict ANI values to their closest relatives.

in Fig. 1, as well as sequences retrieved from the GenBank database, whose accession numbers are listed in Table S2. Although, we observed that the species delimitation obtained with whole-genome data (Fig. 1) is not completely supported by the phylogenies based on those regions, all novel species could be recognized based on the DNA barcode sequences. The D1/D2 region (Fig. S2) appeared to have less resolution than the ITS region (Fig. S3) and thus we recommend the later for species identifications in the absence of whole genome sequences. In Fig. 3, a simplified ITS phylogeny including a single representative from each species summarizes these analyses. It also depicts the number of nucleotide substitutions observed between the more closely related species. Among these, species pairs showing the lowest number of nucleotide substitutions were *T. delbrueckii* - *T. mapucheana sp. nov*. and *T. pretoriensis* – *T. cukson sp. nov*. that differed from their closest relative by five and four substitutions each, respectively (Fig. 3). These substitutions were consistently found when multiple representatives were tested and are seen as apomorphies. i.e. unique for each species.

Our extensive analyses using DNA barcode sequences shown in Figs S2, S3 and S4 allowed us to identify additional representatives of the novel species. For example, a considerable number of isolates of *T. ventriculi sp. nov.* was identified among strains isolated from soil in Cameroon and previously identified as *T. globosa* (Aljohani *et al.* 2018) (Fig. S3 and Table S2). In the case of *T. incommunis sp. nov.* for which a single isolate was initially available, we detected one D1/D2 (Fig. S2) and one ITS (Fig. S3) sequence in the pool of GenBank sequences analysed that appear to belong to two additional strains of this species (Table S2). In fact, our analyses allowed to enlarge the number of known representatives of most novel species and of all the already known species (Table S2).

We also used these analyses to ascertain if any of the currently 15 recognized synonyms of *T. delbrueckii* coincides with any of the proposed novel species. These synonyms and their molecular identification are shown in Fig. S5. All synonyms were confirmed to belong to *T. delbrueckii.* Moreover, for eight of these synonyms, whole-genome sequences also validated their identification as *T. delbrueckii.*

The genomic and phenotypic landscape of *Torulaspora*

Our updated perspective of the genus *Torulaspora*, as revealed through whole genome sequences of 77 strains from 22 species, allowed us to take a snapshot of genome composition across all currently known species. Our first observation was that the genomic landscape of *Torulaspora* was highly variable, with a marked gene content oscillation both between and within species. Genomes of *Torulaspora* spp. contain, on average, 4978 (± 87) genes with a disparity of 409 genes between the most gene-rich, *T. microellipsoides* NRRL-Y-1549 (5219 genes), and the least gene-rich, *T. asahi sp. nov*. NBRC11086 (4810 genes). Although isolates of the same species exhibit some variability in gene-richness, this value was roughly similar within species. For example, the three isolates of *T. mapucheana sp. nov.* have an average of 4863 (± 43) genes, while the three isolates of *T. microellipsoides* contain an average of 5165 (± 46) genes.

This variability was reflected in the presence or absence of genes involved in the metabolism of key carbohydrate sources and propagates to the phenotypic level (Fig. 2). For example, in *T. delbrueckii* and its closest relative, *T. mapucheana sp. nov.*, the inability to grow on galactose was observed frequently and, in those cases, two of the three genes required for canonical galactose metabolism, *GAL1*, *GAL7*, and *GAL10* (Hittinger *et al.* 2004), were absent or inactivated pseudogenes (Fig. 2). Conversely, galactose growth in *T. delbrueckii,* and in all other species in the genus, was linked to the presence of these three genes. Overall, variation across the genus was seen for maltose and melibiose metabolism, as well as for the presence or absence of the associated α-glucosidase-encoding genes (Fig. 2). Three classes of α-glucosidase-encoding genes were defined by analysing signature amino acids that correlate to substrate specificity (Table S4). Following Viigand *et al.* (2018), we tentatively grouped the α-glucosidase genes in three categories corresponding to the type of enzymes they encode: maltases, isomaltases, and mixed maltase–isomaltase activity. Our results are generally consistent with previous observations in known *Torulaspora* species (Silva *et al.* 2022) and suggest that maltose utilization is based on a widespread prevalence of α-glucosidase-encoding genes. Moreover, these genes appear to have been lost multiple times (Fig. 2, Table S4). Thus, we observed that consumption of the substrate predicted the presence of the gene coding for the canonically associated enzyme, but not vice versa. Those latter cases (i.e. gene presence but lack of the corresponding phenotype) might be explained by gene inactivation or gene expression impairments. Such variations occurred both between and within species. For example, we found that all five isolates of *T. jiuxiensis* contained the three *GAL* genes, while they were completely absent in all five isolates of *T. kanakia sp. nov.* Moreover, the intraspecific variation that we previously observed for *T. delbrueckii* (Silva *et al.* 2022) was observed here for other species, especially with respect to maltose and melezitose utilization (Fig. 2).

However, the observed gene content variation at the species level was not universal. For traits, such as sucrose and trehalose consumption, the genes most commonly associated with these phenotypes were consistently found to be present in all *Torulaspora* genomes. As above, this consistent gene presence fails to act as a perfect predictor of a positive phenotype for sucrose, raffinose, and trehalose assimilation (Fig. 2). Taken together, these results paint the picture of a genomic landscape that is highly diverse, with frequent events of gene loss that appear to be recent in terms of the ancestry of the genus *Torulaspora* (i.e. encompass a single species or a group of closely related species). This variation is not solely associated with the phylogeny. While some genes and traits are found consistently in one species but not another, others vary substantially among isolates of the same species, and the genus *Torulaspora* as a whole displays substantial variation in phenotype both within and between species (Fig. S6).

The correlation between assimilation and fermentation abilities varied between substrates and species, but as expected, assimilation was a prerequisite for fermentation. For example, all isolates that could assimilate sucrose could also ferment it; with just two exceptions, all isolates tested for galactose assimilation were also positive for galactose

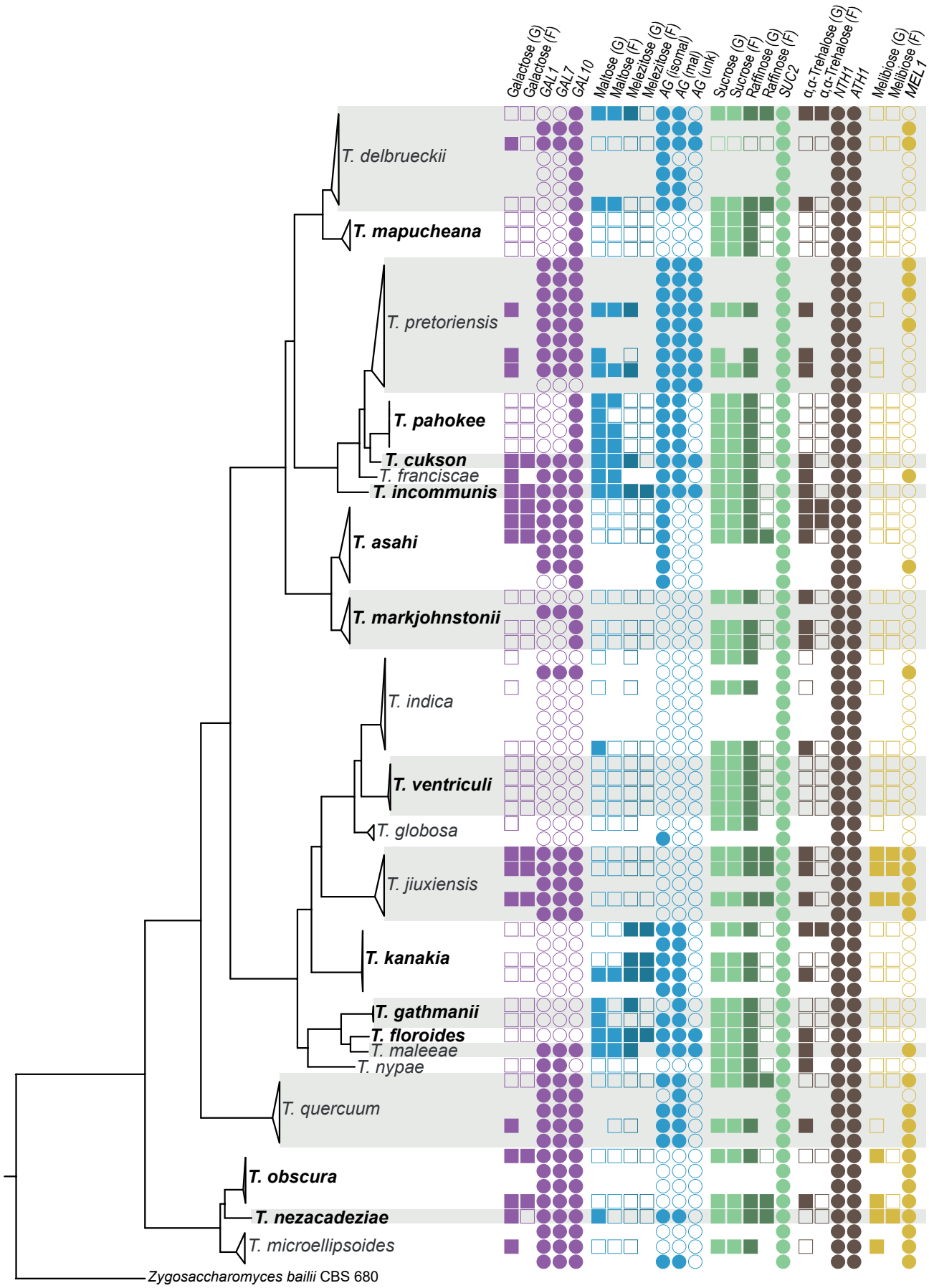

Fig. 2. Phenotype and gene content variation related to the utilization of selected carbohydrates across the genus *Torulaspora*. Squares and circles correspond to phenotypes and genotypes, respectively. A filled shape indicates a positive result (growth, G; fermentation, F; or gene presence), an empty shape indicates a negative result, and an absent shape indicates a lack of testing for that trait. Genes are designated using standard nomenclatures and capital letters corresponding to their encoded activities (isomal, isomaltase; mal, maltase; unk, unknown, encoding either for a maltase or isomaltase). Fermentation and assimilation results were obtained in this work or were taken from the literature as indicated in Table S3.

fermentation. A similarly strong correlation was found for maltose assimilation and fermentation. However, raffinose and trehalose did not exhibit this correlation; most isolates that exhibited assimilation of these carbohydrates did not also ferment them. These correlations between assimilation and fermentation sometimes persisted within species. For example, all three tested isolates of *T. jiuxiensis* both assimilated and fermented raffinose, whereas all four tested isolates of *T. ventriculi sp. nov*. only exhibited assimilation. Likewise, all three tested isolates of *T. jiuxiensis* both assimilated and fermented melibiose, whereas the two tested isolates of *T. obscura sp. nov.* only assimilated this carbohydrate. However, variation within species was common; for example, one *T. asahi sp. nov.* isolate could ferment raffinose but not trehalose, while the other two isolates could ferment trehalose but not raffinose.

Taxonomy

The formal descriptions of the new species are provided below. Species in which sexual structures were not observed are designated as *forma asexualis*, *f.a. (Lachance 2012)*.

Torulaspora asahi M. Silva, F. Paraíso, M.-A. Lachance & J.P. Sampaio, ***sp. nov.*** MB 853717. Fig. 4A–C.

Etymology: *Torulaspora asahi*. a.sa.hi', N.L. app. n. asahi, the Japanese word "asahi" meaning rising sun Japan's sobriquet ("the Land of the Rising Sun") as several strains of this species, including the ex-type strain, were isolated in that country.

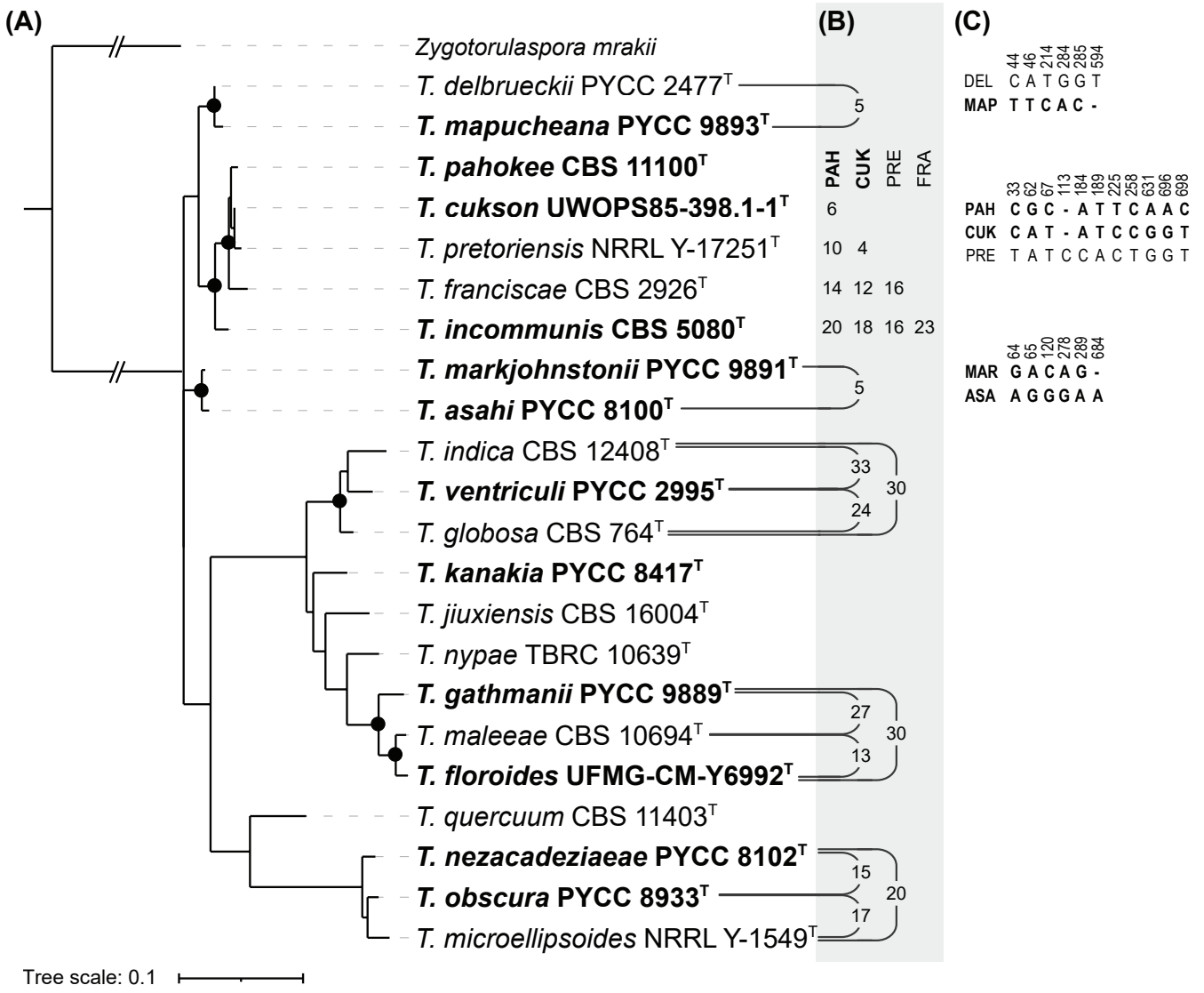

Fig. 3. Phylogeny of all the species of the genus *Torulaspora* based on the complete ITS region. **A.** Phylogenetic tree constructed with the maximum-likelihood method and the TPM2u model of sequence evolution, chosen according to the Bayesian information criterion (BIC), and rooted with *Zygotorulaspora mrakii*. The final dataset had a total of 813 positions. Names in bold indicate the novel species. The scale bar represents the estimated number of substitutions per site. Percentage bootstrap values of 1000 replicates equal or higher that 95 % are depicted as black circles at tree nodes. **B.** For the most closely related species, the number of nucleotide substitutions in the ITS alignment is given. **C.** For the species pairs exhibiting the lowest number of nucleotide substitutions the location and nature of those substitutions is given (position 1 is the first position after the primer).

Table 1. Standard physiological and biochemical profiles of the novel *Torulaspora* species.

	T. asahi	*T. cukson*	*T. floroides*	*T. gathmanii*	*T. incommunis*	*T. kanakia*	*T. mapucheana*	*T. markjohnstonii*	*T. nezacadeziaeae*	*T. obscura*	*T. pahokee*	*T. ventriculi*
Fermentation												
D-Glucose	s	+	+	+	+	+/s	+	+	+	s	+	+
D-Galactose	w	+	-	-	+	-	-	-	-	s/w	-	-
Maltose	-	+	w	-	w	v	-	-	-	-	v	-
Sucrose	s	+	+	w	+	+/w	+	+/s	+/s	s/w	+/w	+/s
α-α-Trehalose	s	-	-	-	-	v	-	-	-	-	-	-
Lactose	-	-	-	-	-	-	-	-	-	-	-	-
Melibiose	-	-	-	-	-	-	-	-	+/s	-	-	-
Melezitose	-	-	+/s	-	w	s	-	-	-	-	-	-
Cellobiose	-	-	-	-	-	-	-	-	-	-	-	-
Raffinose	s	-	-	-	-	-	-	-	+/s	v	-	-
Assimilation												
Carbon sources												
D-Glucose	+	+	+	+	+	+	+	+	+	+	+	+
D-Galactose	+/s	+	-	-	+	-	-	-	+	+	-	-
L-Sorbose	+/s	-	-	-	-	-	-	s	w	+/s	-	-
D-Glucosamine	-	-	-	-	-	-/w	-	-	-	-	-	-
D-Ribose	-	-	-	-	-	-	-	-	-	-	-	-
D-Xylose	-	-	-	-	-	-	-	-	-	-	-	-
L-Arabinose	-	-	-	-	-	-	-	-	-	-	-	-
D-Arabinose	-	-	-	-	-	-	-	-	-	-	-	-
L-Rhamnose	-	-	-	-	-	-	-	-	-	-	-	-
Sucrose	+	+	+	+	+	+	+	+	+	+	+	+
Maltose	-	+	+	+/w	+	v	-	-	s	-	+/s	-
α-α-Trehalose	+/s	+	s	-	+	v	-	+/s	-	v	-	-
Methyl α-D-Glucoside	-	+	+	+	+	v	-	v	-	-	v	-

Table 1. (Continued).

	T. asahi	*T. cukson*	*T. floroides*	*T. gathmanii*	*T. incommunis*	*T. kanakia*	*T. mapucheana*	*T. markjohnstonii*	*T. nezacadeziaeae*	*T. obscura*	*T. pahokee*	*T. ventriculi*
Cellobiose	-	-	-	-	-	-	-	-	-	-	-	-
Salicin	-	-	-	-	-	-	-	-/w	-	s/w	-	-
Melibiose	-	-	-	-	-	-	-	-	+	+	-	-
Lactose	-	-	-	-	-	-	-	-	-	-	-	-
Raffinose	+	+	+	+	+	+/s	+	+	+	+	s	+/s
Melezitose	-	+	+	v	+	+/s	-	-	-	-	-	-
Inulin	v	-	-	-	-	-	s	v	-	-	-	-
soluble Starch	-	w	w	w	w	-	w	-/w	-	+/w	w	-
Methanol	-	-	-	-	-	-	-	-	-	-	-	-
Ethanol	-/w	+	w	w	+	+/s	s/w	+/w	w	+/w	+/s	+
Glycerol	s/w	-	+	s/w	-	v	s	+/s	+	-	v	+/s
Erythritol	-	-	-	-	-	-	-	-	-	-	-	-
Ribitol	-	w	w	w	+	v	w	+/w	w	w	w	+/w
Xylitol	v	+	-	-	s	-	-	s/w	-	-	v	-
D-Glucitol	+	s	+	-	+	v	+	+	-	v	v	-
D-Mannitol	+	s	+	w	+	s/w	+	+	-	s	v	-
Galactitol	-	-	-	-	-	-	-	-	-	-	-	-
myo-Inositol	-	-	-	-	-	-	-	-	-	-	-	-
D-Glucono-1,5-lactone	-	w	+	s	s	-	s	s	s	+	s/w	-
D-Gluconate	-	-	s	w	+	-	w	v	-	+	v	-
D-Glucuronate	s	-	s	-	-	+/s	-	-	-	-	-	-
DL-Lactate	v	w	+	+	+	+	s	+	w	+/w	-	v
Succinate	-	-	-	-	-	-	-	-	-	v	-	-
Citrate	-	-	-	-	-	-	-	-	-	-	-	-
L-Malic acid	-	-	-	-	-	-	-	-	-	-	-	-
L-Tartaric acid	-	-	-	-	-	-	-	-	-	-	-	-
Protocatechuic acid	-	-	-	-	-	-	-	-	-	-	-	-

Table 1. (Continued).

	T. asahi	*T. cukson*	*T. floroides*	*T. gathmanii*	*T. incommunis*	*T. kanakia*	*T. mapucheana*	*T. markjohnstonii*	*T. nezacadeziaeae*	*T. obscura*	*T. pahokee*	*T. ventriculi*
Nitrogen sources												
Nitrate	v	-	-	-	-	-	-	-	+	v	-	-
Nitrite	-	-	-	-	-	-	-	-	-	-	-	-
Ethylamine	-	-	-	-	-	-	-	-	-	-	-	-
L-Lysine	v	+	+	+	-	v	+	-	+	+	+	v
Cadaverine	-	-	-	-	-	-	-	-	+	-	-	-
Creatine	-	-	-	-	-	-	-	-	-	-	-	-
Creatinine	+	-	-	-	-	+	-	-	-	-	-	-
Other tests												
Growth without vitamins	-	+	-	+	+	-	+	+/s	w	v	+	+
Cycloheximide 0.01 %	v	-	-	-	-	-	-	-	-	-	-	+/s
Cycloheximide 0.1 %	v	-	-	-	-	-	-	-	-	-	-	v
Starch formation	-	-	-	-	-	-	-	-	-	-	-	-
Urea hydrolysis	-	-	-	-	-	-	-	-	-	-	-	-
Growth at 30 °C	+	+	+	+	+	+	+	+	+	+	+	+
Growth at 35 °C	-	+	+	+	+	-	-	-	-	-	-	+
Growth at 37 °C	-	+	-	+	+	-	-	-	-	-	-	+
Growth at 40 °C	-	+	-	+	-	-	-	-	-	-	-	+
Growth at 42 °C	-	+	-	-	-	-	-	-	-	-	-	-
Growth at 45 °C	-	-	-	-	-	-	-	-	-	-	-	-

+, positive ; -, negative ; s, delayed ; w, weak ; v, variable.

Typus: **Japan**, Shiba Prefecture, soil underneath *Quercus acuta*, 2008, *Y. Imanishi* (**holotype** PYCC 8100H, ex-holotype cultures PYCC 8100, CBS 18509). The holotype is permanently maintained in a metabolically inactive state in the Portuguese Yeast Culture Collection, Caparica, Portugal. The genome of this species was deposited at DDBJ/ENA/GenBank under the accession GCA_931301635. The version described in this paper is v. 1.

Description: After 1 wk on YM agar at 25 °C, cultures are smooth, cream-coloured, and butyrous. After 3 d of growth on YM agar at 25 °C, cells are globose (4–5.5 µm) to subglobose (4–6 × 3–5 µm) and occur singly or in pairs, and proliferation is by multilateral budding (Fig. 4A). On Dalmau plates after 2 wk at 25 °C, no pseudohyphae nor true hyphae are formed. Sexual reproduction is observed on acetate agar after 15 d at 20 °C, and the cultures appear to be homothallic. Cellular extensions that resemble conjugation tubes are frequent and usually are not involved in conjugation (Fig. 4B). Asci are persistent and normally are formed after conjugation involving a cell and its bud. Asci produce one to two smooth and spherical ascospores, measuring 2.5–3.5 µm diam. (Fig. 4C). The physiological and biochemical profile of the species is shown in Table 1 and Table S5.

Habitat and distribution: This species was found in the arboreal niche, including soil underneath *Quercus acuta* and *Q. salicina*, bark of *Lithocarpus edulis*, root of *Clermontia* sp., exudate of *Myoporum*, and fruit of *Ficus virgata.* Currently known from collections in Japan (Chiba Prefecture and Iriomote island) and Hawaii.

Additional cultures examined: PYCC 8099, PYCC 8101, PYCC 9382, UWOPS 91-901.1, UWOPS 91-920.1 (see Table S1 for details).

Torulaspora cukson *f.a.* M. Silva, F. Paraíso, M.-A. Lachance & J.P. Sampaio, ***sp. nov.*** MB 853718. Fig. 4D.

Etymology: *Torulaspora cukson*. cuk'son, N.L. app. n. cukson, from Cuk Son, the designation in Uto-Aztecan O'odham language for Tucson, the locality in which this species was originally found.

Typus: **USA**, Arizona, *Scaptodrosophila brooksae* on *Populus fremontii*, 1985, *M.A. Lachance* (**holotype** PYCC 9174H, ex-holotype cultures PYCC 9174, CBS 18510). The holotype is permanently maintained in a metabolically inactive state in the Portuguese Yeast Culture Collection, Caparica, Portugal. The genome of this species was deposited at DDBJ/ENA/GenBank under the accession GCA_964263435. The version described in this paper is v. 1.

Description: After 1 wk on YM agar at 25 °C, *cultures* are smooth, cream-coloured, and butyrous. After 3 d of growth on YM agar at 25 °C, *cells* are ellipsoidal (3–5 × 2–3 µm) and occur singly or in pairs, and proliferation is by multilateral budding (Fig. 4D). On Dalmau plates after 2 wk at 25 °C, *pseudohyphae* and *true hyphae* are not formed. *Sexual reproduction* was not observed on acetate agar or corn meal agar after prolonged incubation (3 mo) at 18 °C. The physiological and biochemical profile of the species are shown in Table 1 and Table S5.

Habitat and distribution: This species was found only once, in an insect (*Scaptodrosophila brooksae*), Arisona, USA.

Torulaspora floroides M. Silva, F. Paraíso, K.O. Barros, C.A. Rosa & J.P. Sampaio, ***sp. nov.*** MB 853855. Fig. 4E–G.

Etymology: *Torulaspora floroides*. flo.ro'i.des, N.L. fem. adj. floroides, flower-like, referring to the peculiar budding pattern seen in this species in which a central parent cell is surrounded by multiple smaller buds, thus resembling a flower.

Typus: **Brazil**, state of Amazonas, Itacoatiara, Amazonian rainforest, Campus of the Universidade Federal do Amazonas (UFAM), rotting wood, 2019, K.O. Barros (**holotype** PYCC 9895H, ex-holotype cultures PYCC 9895, UFMG-CM-Y6992). The holotype is permanently maintained in a metabolically inactive state in the Portuguese Yeast Culture Collection, Caparica, Portugal. The genome of this species was deposited at DDBJ/ENA/GenBank under the accession GCA_964263535. The version described in this paper is v. 1.

Description: After 1 wk on YM agar at 25 °C, *cultures* are smooth, cream-coloured, and butyrous. After 3 d of growth on YM agar at 25 °C, *cells* are globose (2.5–4 µm) to subglobose (3–5 × 2.5–3 µm) and occur singly or in pairs, and proliferation is by conspicuous multilateral budding (Fig. 4E). On Dalmau plates after 2 wk at 25 °C, no *pseudohyphae* nor *true hyphae* are formed. *Sexual reproduction* is observed on acetate agar after 20 d at 20 °C, and the culture appears to be homothallic. *Asci* are persistent and normally are formed after conjugation involving a cell and its bud. Asci produce one to two smooth and spherical *ascospores*, measuring 2–3 µm diam. (Fig. 4G). Conspicuous multilateral budding resulting in a central parental cell surrounded by numerous buds with a flower-like appearance is observed in older cultures grown in sporulation medium (Fig. 4F). The physiological and biochemical profile of the species is shown in Table 1 and Table S5.

Habitat and distribution: This species was found only once, in rotting wood, Amazonas state, Brazil.

Torulaspora gathmanii J. Al-Oboudi, M. Silva, F. Paraíso, M. Jarzyna, Q.K. Langdon, D.A. Opulente, J.P. Sampaio & Hittinger, ***sp. nov.*** MB 853979. Fig. 4H–J.

Etymology: *Torulaspora gathmanii*. gath.man'i.i, N.L. gen. n. gathmanii, of Gathman, in honor of Allen C. Gathman, in recognition of his lifelong contributions to fungal genetics and genomics.

Typus: **USA**, Wisconsin, University of Wisconsin-Madison, Lakeshore Preserve, Prairie C, soil, 2016, *M. Jarzyna* (**holotype** PYCC 9889H, ex-holotype cultures PYCC 9889, CBS 18642). The holotype is permanently maintained in a metabolically inactive state in the Portuguese Yeast Culture Collection, Caparica, Portugal. The genome of this species

was deposited at DDBJ/ENA/GenBank under the accession GCA_030580195. The version described in this paper is v. 1.

Description: After 1 wk on YM agar at 25 °C, *cultures* are smooth, cream-coloured, and butyrous. After 3 d of growth on YM agar at 25 °C, *cells* are globose (3–5 µm) to sub-globose (4–6 × 3–5 µm) and occur singly or in pairs, and proliferation is by multilateral budding (Fig. 4H). On Dalmau plates after 2 wk at 25 °C, no *pseudohyphae* nor *true hyphae* are formed. *Sexual reproduction* is observed on acetate agar after 7 d at 20 °C, and the cultures appear to be homothallic. *Asci* are persistent and normally are formed after conjugation involving a cell and its bud. Asci produce one to three, possibly four, smooth and spherical *ascospores*, measuring 3–4 µm diam. (Fig. 4I, J). The physiological and biochemical profile of the species in shown in Table 1 and Table S5.

Habitat and distribution: This species was found in soil, in preserved areas of the University of Wisconsin - Madison, Wisconsin, USA.

Additional cultures examined: PYCC 9890 (see Table S1 for details).

Torulaspora incommunis M. Silva, F. Paraíso, M. Groenewald & J.P. Sampaio, ***sp. nov.*** MB 853859. Fig. 4K, L.

Etymology: *Torulaspora incommunis*. in.com'mu.nis, L. fem. adj. incommunis, uncommon, referring to the apparent rarity of this species that, albeit first isolated more than 90 years ago, is known from a very low number of isolates.

Typus: **Japan**, Wakayama, sap of an orange tree (*Citrus* sp.), 1932, *unknown collector* (**holotype** PYCC 9318H, ex-holotype cultures PYCC 9318, CBS 5080, IFO 0022). The holotype is permanently maintained in a metabolically inactive state in the Portuguese Yeast Culture Collection, Caparica, Portugal. The genome of this species was deposited at DDBJ/ENA/GenBank under the accession GCA_012851155. The version described in this paper is v. 1.

Description: After 1 wk on YM agar at 25 °C, *cultures* are smooth, cream-coloured, and butyrous. After 3 d of growth on YM agar at 25 °C, *cells* are ellipsoidal (4–5 × 2.5–4 µm) and occur singly or in pairs, and proliferation is by multilateral budding (Fig. 4K). On Dalmau plates after 2 wk at 25 °C, no *pseudohyphae* nor *true hyphae* are formed. *Sexual reproduction* is observed on acetate agar after 20 d at 20 °C, and the cultures appear to be homothallic. *Asci* are persistent and are formed after cell-to-cell conjugation frequently involving a cell and its bud and occasionally two independent cells. Asci produce one to two smooth and spherical *ascospores*, measuring 2.5–3 µm diam. (Fig. 4L). The physiological and biochemical profile of the species in shown in Table 1 and Table S5.

Habitat and distribution: This species was found in the sap of an orange tree (*Citrus* sp.), Wakayama, Japan. Two additional cultures tentatively identified as this species were found in a rose flower in Fukuyama, Hiroshima, Japan and in soil in Taiwan.

Additional cultures examined: During the investigation of D1/D2 and ITS sequences deposited at GenBank, two additional representatives of this species were presumptively detected (Table S2, Figs S2, S3): SG5S08, found in soil, Taiwan and FR 994 (NBRCN 114950), found in a rose flower, Fukuyama, Hiroshima, Japan (Hisatomi & Toyomura 2021).

Torulaspora kanakia M. Silva, F. Paraíso, A. Pontes, P. Gonçalves, F. Carriconde & J.P. Sampaio, ***sp. nov.*** MB 853861. Fig. 4M–O.

Etymology: *Torulaspora kanakia*. ka.na'ki.a, N.L. app. n. kanakia, pertaining to Kanak, the indigenous Melanesian inhabitants of New Caledonia, the island where the new species was found.

Typus: **New Caledonia**, Mont Kogis, forest soil, 2014, *J.P. Sampaio* (**holotype** PYCC 8417H, ex-holotype cultures PYCC 8417, CBS 18511). The holotype is permanently maintained in a metabolically inactive state in the Portuguese Yeast Culture Collection, Caparica, Portugal. The genome of this species was deposited at DDBJ/ENA/GenBank under the accession GCA_931302125. The version described in this paper is version one.

Description: After 1 wk on YM agar at 25 °C, *cultures* are smooth, cream-coloured, and butyrous. After 3 d of growth on YM agar at 25 °C, *cells* are globose (4–5.5 µm) to sub-globose (4–6 × 3–5 µm) and occur singly or in pairs, and proliferation is by multilateral budding (Fig. 4M). On Dalmau plates after 2 wk at 25 °C, no *pseudohyphae* nor *true hyphae* are formed. *Sexual reproduction* is observed on acetate agar after 7 d at 20 °C, and the cultures appear to be homothallic. Cellular extensions that resemble conjugation tubes are conspicuous and exceptionally long, reaching 8 µm (Fig. 4N). *Asci* are persistent and normally are formed after conjugation involving a cell and its bud. Asci produce one to two smooth and spherical *ascospores*, measuring 2.5–3.5 µm diam. (Fig. 4O). The physiological and biochemical profile of the species in shown in Table 1 and Table S5.

Habitat and distribution: This species was found in forest soil in New Caledonia.

Additional cultures examined: PYCC 9027, PYCC 8418, PYCC 9025, PYCC 9026 (see Table S1 for details).

Torulaspora mapucheana M. Silva, F. Paraíso, D. Libkind & J.P. Sampaio, ***sp. nov.*** MB 853980. Fig. 4P–R.

Etymology: *Torulaspora mapucheana*. ma.pu'che.a.na, N.L. fem. adj. mapucheana, referring to Mapuche, the group of indigenous inhabitants of present-day south-central Chile and southwestern Argentina.

Typus: **Argentina**, Lanin National Park, Yuco, bark of *Nothofagus obliqua*, 2009, *D. Libkind* (**holotype** PYCC 9893H, ex-holotype cultures PYCC 9893, CBS 18512). The holotype is permanently maintained in a metabolically inactive state in the Portuguese Yeast Culture Collection, Caparica, Portugal. The genome of this species was deposited at

DDBJ/ENA/GenBank under the accession GCA_964263445. The version described in this paper is v. 1.

Description: After 1 wk on YM agar at 25 °C, *cultures* are smooth, cream-coloured, and butyrous. After 3 d of growth on YM agar at 25 °C, *cells* are globose (1.5–4 µm) to subglobose (4–5 × 2–3 µm) and occur singly or in pairs, and proliferation is by multilateral budding (Fig. 4P). On Dalmau plates after 2 wk at 25 °C, no *pseudohyphae* nor *true hyphae* are formed. *Sexual reproduction* is observed on acetate agar after 15 d at 20 °C and the cultures appear to be homothallic. Cellular extensions that resemble conjugation tubes are frequent and usually are not involved in conjugation (Fig. 4.Q). *Asci* are persistent and normally are formed after conjugation involving a cell and its bud. Asci produce one to two smooth and spherical *ascospores*, measuring 2–3 µm diam. (Fig. 4R). The physiological and biochemical profile of the species in shown in Table 1 and Table S5.

Habitat and distribution: This species was found in the arboreal niche, in association with *Nothofagus* spp. Cultures were isolated from the bark of *Nothofagus obliqua* and the fruiting body of *Cyttaria harioti* on *N. antarctica*. Currently known from collections in Patagonia, Argentina (Yuco and Los Rápidos, Nahuel Huapi natural park).

Additional cultures examined: PYCC 9894, PYCC 9057 (see Table S1 for details).

Torulaspora markjohnstonii J. Al-Oboudi, M. Silva, F. Paraíso, K. Sylvester, Q.K. Langdon, J.-L. Legras, D.A. Opulente, J.P. Sampaio & Hittinger, ***sp. nov.*** MB 853981. Fig. 4S, T.

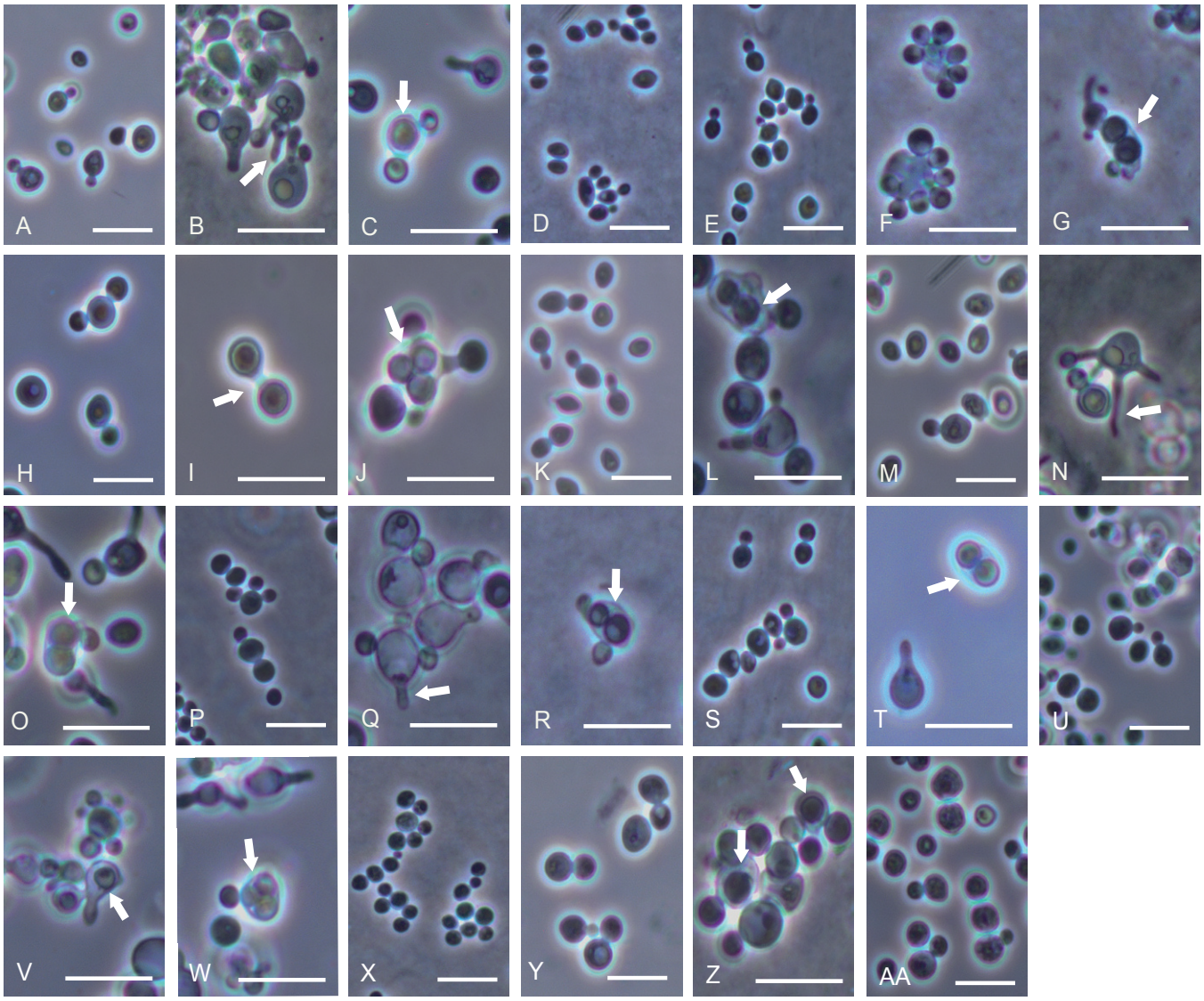

Fig. 4. Vegetative and sexual structures of the novel *Torulaspora* species. For each species, the first micrograph corresponds to budding yeast cells on YM agar after 3–5 d at 25 °C. Micrographs of asci, ascospores, or other structures were obtained from cultures grown on acetate agar for 1–3 wk at 17–20 °C. **A–C.** Yeast cells of *T. asahi* PYCC 8100^{T} (**A**), cells with tapered protuberances (**B**), and ascus with ascospore (**C**). **D.** Yeast cells of *T. cukson f.a.* PYCC 9174^{T}. **E–G.** Yeast cells of *T. floroides* UFMG-CM-Y6992^{T} (**E**), peculiar arrangement of bud cells resembling a flower (**F**), and ascus with two ascospores (**G**). **H–J.** Yeast cells of *T. gathmanii* PYCC 9889^{T} (**H**) and asci with ascospores (**I, J**). **K, L.** Yeast cells of *T. incommunis* CBS 5080^{T} (**K**) and ascus with two ascospores (**L**). **M–O.** Yeast cells of *T. kanakia* PYCC 8417^{T} (**M**), long protuberances, and asci with ascospores (**N, O**). **P–R.** Yeast cells of *T. mapucheana* PYCC 9893^{T} (**P**), cell with protuberance (**Q**), and ascus with two ascospores (**R**). **S, T.** Yeast cells of *T. markjohnstonii* PYCC 9891^{T} (**S**), a cell with a tapered protuberance, and an ascus containing two ascospores (**T**). **U–W.** Yeast cells of *T. nezacadeziae* PYCC 8102^{T} (**U**), protuberances, and asci with ascospores (**V, W**). **X–Z.** Yeast cells of *T. obscura f.a.* PYCC 8933^{T} (**X**). Yeast cells of *T. pahokee* CBS 11100^{T} (**Y**) and asci with ascospores (**Z**). **AA.** Yeast cells of *T. ventriculi f.a.* PYCC 2995^{T}. Scale bars = 10 µm.

Etymology: *Torulaspora markjohnstonii*. mark.john'sto.ni.i, N.L. gen. n. markjohnstonii, of Mark Johnston, in honor of Mark Johnston, in recognition of his lifelong contributions to yeast genetics and genomics.

Typus: **USA**, North Carolina, Winston-Salem, soil under *Acer saccharinum*, 2014, L. Shown (**holotype** PYCC 9891H, ex-holotype cultures PYCC 9891, CBS 18641). The holotype is permanently maintained in a metabolically inactive state in the Portuguese Yeast Culture Collection, Caparica, Portugal. The genome of this species was deposited at DDBJ/ENA/GenBank under the accession GCA_048593755. The version described in this paper is v. 1.

Description: After 1 wk on YM agar at 25 °C, *cultures* are smooth, cream-coloured, and butyrous. After 3 d of growth on YM agar at 25 °C, *cells* are globose (3–5.5 µm) to sub-globose (2.5–6 × 2–4 µm) and occur singly or in pairs, and proliferation is by multilateral budding (Fig. 4S). On Dalmau plates after 2 wk at 25 °C, no *pseudohyphae* nor *true hyphae* are formed. *Sexual reproduction* is observed on acetate agar after 10 d at 17 °C, and the cultures appear to be homothallic. Cellular extensions that resemble conjugation tubes are frequent and usually are not involved in conjugation. *Asci* are persistent and normally are formed after conjugation involving a cell and its bud. Asci produce one to two smooth and spherical *ascospores*, measuring 2–3 µm diam. (Fig. 4T). The physiological and biochemical profile of the species in shown in Table 1 and Table S5.

Habitat and distribution: This species was found in the bark of *Quercus pubescens* (Italy), and *Q. robur* (Portugal), soil (USA), and in a cider brewery (France). Currently known from collections in Europe and USA.

Additional cultures examined: PYCC 9561, PYCC 8308, PYCC 9892 (see Table S1 for details).

Torulaspora nezacadeziae M. Silva, F. Paraíso & J.P. Sampaio, ***sp. nov.*** MB 853982. Fig. 4U–W.

Etymology: *Torulaspora nezacadeziae*. ne.za.ca.de'zi.ae, N.L. gen n. nezacadeziae, of Neža Čadež, in honour of Neža Čadež, in recognition of her contributions to yeast taxonomy.

Typus: **Japan**, Chiba Prefecture, soil underneath *Castanopsis sieboldii*, 2008, *Y. Imanishi* (**holotype** PYCC 8102H, ex-holotype cultures PYCC 8102, CBS 18513). The holotype is permanently maintained in a metabolically inactive state in the Portuguese Yeast Culture Collection, Caparica, Portugal. The genome of this species was deposited at DDBJ/ENA/GenBank under the accession GCA_964263675. The version described in this paper is v. 1.

Description: After 1 wk on YM agar at 25 °C, *cultures* are smooth, cream-coloured, and butyrous. After 3 d of growth on YM agar at 25 °C, *cells* are globose (3–5 µm) to sub-globose (4–5 × 2–4 µm) and occur singly or in pairs, and proliferation is by multilateral budding (Fig. 4U). On Dalmau plates after 2 wk at 25 °C, no *pseudohyphae* nor *true hyphae* are formed. *Sexual reproduction* is observed on acetate agar after 15 d at 20 °C, and the cultures appear to be homothallic. Cellular extensions that resemble conjugation tubes are frequent and usually are not involved in conjugation. *Asci* are persistent and normally are formed after conjugation involving a cell and its bud. Asci produce one to three, possibly four, smooth and spherical *ascospores*, measuring 1.5–2.5 µm diam. (Fig. 4V, W). The physiological and biochemical profile of the species in shown in Table 1 and Table S5.

Habitat and distribution: This species was found only once in soil underneath *Castanopsis sieboldii*, Chiba Prefecture, Japan.

Torulaspora obscura *f.a.* M. Silva, F. Paraíso, C.A. Rosa & J.P. Sampaio, ***sp. nov.*** MB 853983. Fig. 4X.

Etymology: *Torulaspora obscura*. ob.scu'ra, L. fem. adj. obscura, obscure, referring to the unknown natural niche of this species because all known strains have been isolated from anthropic environments.

Typus: **Belgium**, Oedelem, aquakefir, 2009, *E. Vercammen* (**holotype** PYCC 8933H, ex-holotype cultures PYCC 8933, CBS 18514). The holotype is permanently maintained in a metabolically inactive state in the Portuguese Yeast Culture Collection, Caparica, Portugal. The genome of this species was deposited at DDBJ/ENA/GenBank under the accession GCA_964263565. The version described in this paper is v. 1.

Description: After 1 wk on YM agar at 25 °C, *cultures* are smooth, cream-coloured, and butyrous. After 3 d of growth on YM agar at 25 °C, *cells* are globose (2–4 µm) and occur singly or in pairs, and proliferation is by multilateral budding (Fig. 4X). On Dalmau plates after 2 wk at 25 °C, no *pseudohyphae* nor *true hyphae* are formed. *Sexual reproduction* was not observed on acetate agar or corn meal agar after prolonged incubation (3 mo) at 18 °C. The physiological and biochemical profile of the species in shown in Table 1 and Table S5.

Habitat and distribution: This species was found in anthropic environments, including aquakefir, orange soda and artificial orange juice. Currently known from collections in Belgium, UK and Brazil.

Additional cultures examined: NCYC 2473, PYCC 5189, UFMG-CM Y5068 (see Table S1 for details).

Torulaspora pahokee M. Silva, F. Paraíso, M. Groenewald, J.W. Fell & J.P. Sampaio, ***sp. nov.*** MB 853985. Fig. 4Y–Z.

Etymology: *Torulaspora pahokee*. pa.ho'kee, NL. app. n. pahokee, referring to Pahokee, meaning "Grassy Water", the name given to the Everglades by the Seminole, a Native American tribe that inhabits Florida.

Typus: **USA**, Florida, Everglades, sea water, 1996, *J. Fell* (**holotype** PYCC 9316H, ex-holotype cultures PYCC 9316, CBS 11100). The holotype is permanently maintained in a metabolically inactive state in the Portuguese Yeast Culture Collection, Caparica, Portugal. The genome of this species was deposited at DDBJ/ENA/GenBank under the accession GCA_012851095. The version described in this paper is v. 1.

Description: After 1 wk on YM agar at 25 °C, *cultures* are smooth, cream-coloured, and butyrous. After 3 d of growth on YM agar at 25 °C, *cells* are globose (3–5 µm) to sub-globose (4–5 × 2–3 µm) and occur singly or in pairs, and proliferation is by multilateral budding (Fig. 4Y). On Dalmau plates after 2 wk at 25 °C, no *pseudohyphae* nor *true hyphae* are formed. *Sexual reproduction* is observed on acetate agar after 10 d at 17 °C, and the studied strains appear to be homothallic. *Asci* are persistent and are formed after cell-to-cell conjugation involving either a cell and its bud or two independent cells. Asci produce one to two smooth and spherical *ascospores*, measuring 2.5–3 µm diam. (Fig. 4Z). Typical *Torulaspora* cellular extensions that resemble conjugation tubes are frequent but not involved in conjugation. The physiological and biochemical profile of the species in shown in Table 1 and Table S5.

Habitat and distribution: This species was found in sea water, Everglades, Florida, USA

Additional cultures examined: CBS 11121, CBS 11123, CBS 11124 (see Table S1 for details).

Torulaspora ventriculi *f.a.* M. Silva, F. Paraíso & J.P. Sampaio, ***sp. nov.*** MB 853986. Fig. 4AA.

Etymology: *Torulaspora ventriculi*. ven.tri'cu.li, L. gen. n. ventriculi, of the ventriculus, referring to the digestive tract, the habitat where several strains of this species were found.

Typus: **Mozambique**, Vila Cabral, human faeces, 1958, N. van Uden (**holotype** PYCC 2995H, ex-holotype cultures PYCC 2995, CBS 4887). The holotype is permanently maintained in a metabolically inactive state in the Portuguese Yeast Culture Collection, Caparica, Portugal. The genome of this species was deposited at DDBJ/ENA/GenBank under the accession GCA_964263485. The version described in this paper is v. 1.

Description: After 1 wk on YM agar at 25 °C, *cultures* are smooth, cream-coloured, and butyrous. After 3 d of growth on YM agar at 25 °C, *cells* are globose (2–5 µm) and occur singly or in pairs, and proliferation is by multilateral budding (Fig. 4AA). On Dalmau plates after 2 wk at 25 °C, no *pseudohyphae nor true hyphae* are formed. *Sexual reproduction* was not observed on acetate agar or corn meal agar after prolonged incubation (3 mo) at 18 °C. The physiological and biochemical profile of the species in shown in Table 1 and Table S5.

Habitat and distribution: This species was found in soil and faeces (human and hippopotamus). Currently known from collections in Mozambique (Vila Cabral and Incomati River) and Papua New Guinea.

Additional cultures examined: CBS 2947, PYCC 2996, PYCC 2997 (see Table S1 for details).

DISCUSSION

Here we take advantage of multiple surveys of yeast diversity carried out by several groups in different regions of the globe, including Asia, Australasia, Europe, Africa, North America, and South America, to employ a taxogenomic approach to delineate and describe 12 novel species in the genus *Torulaspora.* Thus, this study represents a considerable expansion of the documented diversity in this genus that previously contained only 10 species. Since the new species are found within the already known phylogenetic breadth of the genus, (i.e. between *T. delbrueckii* and *T. microellipsoides*), and no phenotypic novelties were discovered, the diagnosis of the genus is maintained. Moreover, the new species proposals integrate genomics, DNA barcode data, a detailed phenotypic characterization and, when possible, ecology and distribution. The use of overall genome relatedness indices, such as the average nucleotide identity (ANI) calculation employed here, will likely accompany the popularization of taxogenomics studies and have to be critically evaluated. Studies in prokaryotes and micro-eukaryotes suggest that distinct species have values up to 95 %. Rather than firm cutoffs, we regard this value as a guide subjected to qualitative judgment and to be integrated with other lines of evidence, such as lineage separation due to monophyly, distinct distributions or ecologies.

The novel additions strengthen the view that, from an ecological standpoint, *Torulaspora* can be seen mostly as a genus of soil and arboreal yeasts, although some *Torulaspora* spp. have been isolated from other sources (Table S1). For example, among the novel species, *T. pahokee* was found exclusively in costal seawater near the Everglades, and several representatives of *T. ventriculi* were isolated from the intestinal tracts of hippopotamuses and humans. However, even in those cases, it might be hypothesized that soil or plants cannot be ruled out as the primary reservoir of these yeasts. For most of the novel species, their isolation sources can be seen as representing wild environments and therefore their natural habitats, *T. obscura* stands apart since its various representatives were all collected from anthropic environments (Table S1). In this respect, *T. obscura* resembles *T. delbrueckii* and *T. microellipsoides*, which are also often associated with anthropic environments.

This expanded view of *Torulaspora* shows that the genus is distributed worldwide in temperate and tropical regions (Table S1). Some species have wide distributions, while other species appear to be endemic to a particular region. *Torulaspora delbrueckii*, the type species of the genus, is the most well-known example of the first group, being found in the arboreal niche in temperate regions in Europe and North America, in tropical regions in Central and South America, and in Asia (Table S1; Silva *et al.* 2022). Another frequently isolated species is *T. pretoriensis,* but its range appears not to overlap with that of *T. delbrueckii* because, in contrast to *T. delbrueckii*, it has not been found in Europe and in North America. Among the novel species, *T. asahi*, *T. markjohnstonii*, and *T. ventriculi* appear to also have broad

distributions despite the limited number of isolates collected so far, with different collection sites separated by oceans. The first species was collected in Japan and Hawaii, the second was collected in North America and Europe, and the third was found in Africa (Mozambique) and in Papua New Guinea.

Other To*rulaspora* species have very different distribution ranges. For example, *T. franciscae* is known from a limited number of soil isolates collected once in a single locality in Spain, more than 60 years ago (Capriotti 1958). Some of the novel species described here appear to be endemic to a particular region and likely have very restricted distributions, but further yeast isolation surveys are necessary to conclusively confirm or reject this possibility. For example, *T. gathmanii* was only found in Wisconsin, USA, *T. kanakia* was only collected in New Caledonia, and *T. mapucheana* was only found in temperate South America, in Patagonia.

Among the novel species described here, *T. cukson*, *T. floroides*, and *T. nezacadeziae* are based on single isolates. Although we recognize that proposing novel species based on single strains might fall short of documenting the inherent biological variability and ecological range of the species, we believe that the recognition of these novel taxa might contribute to the discovery of additional strains. *Torulaspora jiuxiensis* exemplifies one such case. The species was described in 2022 based on two strains isolated in China (Chu *et al.* 2022). During our phylogenomic analyses, we expanded the number of representatives of this species by adding several arboreal isolates collected in Brazil and in Panama (Table S1).

Among the novel species, glucose and sucrose were the sugars most frequently fermented, whereas melezitose was only fermented (weakly) by three species (*T. incommunis*, *T. floroides*, and *T. kanakia*) (Table S5). Overall, melibiose was infrequently fermented in the genus, with positive results for *T. pretoriensis* and variable results for *T. delbrueckii* (Table S6). Assimilation of galactose as sole source of carbon and energy was positive in six of the novel species and negative in the other six, with no intraspecific variable results (Table S5). This high variation is maintained when all species in the genus are considered (Table S6). Up to now, melibiose utilization had only been observed in three species*, T. jiuxiensis*, *T. maleeae*, and *T. microellipsoides*. Among the novel species, *T. gathmanii*, *T. nezacadeziae*, and *T. obscura* also assimilated this compound. Melezitose is another compound whose spectrum of utilization is markedly increased with the expansion of *Torulaspora*. It was previously only utilized by *T. delbrueckii*, *T. pretoriensis*, and *T. maleeae*, but five of the novel species can utilize it (*T. cukson*, *T. incommunis*, *T. kanakia*, *T. gathmanii*, and *T. floroides*). Finally, maximum growth temperatures were found to vary markedly among the novel species. At the lower end of the temperature spectrum, most of the novel species could grow at 30 °C, but not at 35 °C. The two exceptions were *T. ventriculi*, which could grow at 40 °C, and *T. cukson*, which could grow at 42 °C. The first species is notable in being found in the gut of some mammals, while the second one was found in the Arizona desert. These habitats might relate to the higher temperature adaptations seen in these two species. *Torulaspora ventriculi* is sister to *T. indica* that can grow even at 42 °C, while the closest relative of *T. cukson*, *T. pahokee,* has a maximum growth temperature around 30 °C.

Of the 12 novel species, nine produced a sexual morph and formed asci with ascospores. Typically, asci contained less than four ascospores, normally one or two. We speculate if this pattern of low number of ascospore formation per ascus might be evolutionary connected with the conditions of low nutrient availability in environments like soil or tree bark. This might perhaps parallel what has been observed in *S. cerevisiae*, in which nutritional scarcity leads to the formation of less than four ascospores (Taxis *et al.* 2005).

The remarkable increase in the number of species in the genus *Torulaspora* reported here highlights the limited status of the current inventory of fungi. It also shows how integrated taxogenomic approaches can foster the assessment of species circumscriptions in fungi.

ACKNOWLEDEGMENTS

This study was supported by Fundação para a Ciência e a Tecnologia (Portugal) grants 2021.08108.BD (M.R. Silva), UIDB/04378/2020 (UCIBIO), and LA/P/0140/2020 (i4HB). Research in the Hittinger Lab is funded by the National Science Foundation (DEB-2110403), USDA National Institute of Food and Agriculture (Hatch Project 7005101), in part by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE–SC0018409), and an H.I. Romnes Faculty Fellowship (Office of the Vice Chancellor for Research with funding from the Wisconsin Alumni Research Foundation). Research conducted by CA Rosa was supported by project “INCT Yeasts: Biodiversity, preservation and biotechnological innovation”, funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil, grant #406564/2022-1, and was also supported by CNPq grants 435040/2018 and 408733/2021-7. Research conducted by J.W. Fell was supported by grants from the National Science Foundation. Research conducted by M.-A. Lachance was supported by grants from the Natural Science and Engineering Research Council of Canada. The computational work was carried out with support of INCD funded by FCT and FEDER (Portugal) under grants 22153-01/SAICT/2016, and FCT Advanced Computing Projects 2023.09581.CPCA.A1 (to AP) and 2022.15871.CPCA.A2 (to PHB). We gratefully acknowledge B. Turchetti (DBVPG) and H.-M. Daniel (BCCM/MUCL) for providing cultures, K. Wolfe for making available beforehand genome data of several *T. indica* strains, K. Bensch for assistance with the nomenclature of the new species, K. Shibayama for providing complementary information on the original isolation of *T. incommunis*, N. Fernandez and Y. Imanishi, for assistance in sample collection in Yuco, Lanin National Park, Argentina, and in Chiba, Japan, respectively, and citizen scientist L. Shown for contributing the sample from which *T. markjohnstonii* was isolated.

Declaration on conflict of interest The authors declare that there is no conflict of interest.

REFERENCES

de Albuquerque NRM, Haag KL (2023). Using average nucleotide identity (ANI) to evaluate microsporidia species boundaries based on their genetic relatedness. *Journal of Eukaryotic*

Microbiology **70**(1): e12944. https://doi.org/10.1111/jeu.12944

Aljohani R, Samarasinghe H, Ashu T, *et al.* (2018). Diversity and relationships among strains of culturable yeasts in agricultural soils in Cameroon. *Scientific Reports* **8**: 1–11. https://doi.org/10.1038/s41598-018-34122-2

Albertin W, Chasseriaud L, Comte G, *et al.* (2014). Winemaking and bioprocesses strongly shaped the genetic diversity of the ubiquitous yeast *Torulaspora delbrueckii. PLoS ONE* **9**(8): e94246. https://doi.org/10.1371/journal.pone.0094246

Altschul SF, Gish W, Miller W, *et al.* (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**(3): 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Alves-Araújo C, Almeida MJ, Sousa MJ, *et al.* (2004). Freeze tolerance of the yeast *Torulaspora delbrueckii*: Cellular and biochemical basis. *FEMS Microbiology Letters* **240**(1): 7–14. https://doi.org/10.1016/j.femsle.2004.09.008

Bankevich A, Nurk S, Antipov D, *et al.* (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**(5): 455–477. https://doi.org/10.1089/cmb.2012.0021

Bolger AM, Lohse M, Usadel B (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15): 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brysch-Herzberg M, Jia GS, Sipiczki M, *et al.* (2023). *Schizosaccharomyces lindneri sp. nov.*, a fission yeast occurring in honey. *Yeast* **40**(3): 237–253. https://doi.org/10.1002/yea.3857

Čadež N, Bellora N, Ulloa R, *et al.* (2019). Genomic content of a novel yeast species *Hanseniaspora gamundiae sp. nov.* from fungal stromata (*Cyttaria*) associated with a unique fermented beverage in Andean Patagonia, Argentina. *PLoS ONE* **14**(1): e0210792. https://doi.org/10.1371/journal.pone.0210792

Čadež N, Bellora N, Ulloa R, *et al.* (2021). *Hanseniaspora smithiae sp. nov.*, a novel apiculate yeast species from Patagonian forests that lacks the typical genomic domestication signatures for fermentative environments. *Frontiers in Microbiology* **12**: 679894. https://doi.org/10.3389/fmicb.2021.679894

Canonico L, Agarbati A, Comitini F, *et al.* (2016). *Torulaspora delbrueckii* in the brewing process: A new approach to enhance bioflavour and to reduce ethanol content. *Food Microbiology* **56**: 45–51. https://doi.org/10.1016/j.fm.2015.12.005

Capriotti A (1958). Neue Hefen, isoliert aus Böden Spaniens II: *Torulaspora franciscae* nov. spec. *Archives of Microbiology* **28**(4): 338–343.

Chu SB, Hu WT, Hui FL (2022). *Torulaspora jiuxiensis sp. nov.*, a novel yeast species isolated from rotting wood. *International Journal of Systematic and Evolutionary Microbiology* **72**(11): 005629. https://doi.org/10.1099/ijsem.0.005629

Criscuolo A, Gribaldo S (2010). BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology* **10**: 210. https://doi.org/10.1186/1471-2148-10-210

David-Palma M, Libkind D, Brito PH, *et al.* (2020). The untapped Australasian diversity of astaxanthin-producing yeasts with biotechnological potential—*Phaffia australis sp. nov.* and *Phaffia tasmanica sp. nov. Microorganisms* **8**(10): 1518. https://doi.org/10.3390/microorganisms8111651

Emms DM, Kelly S (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**(1): 238. https://doi.org/10.1186/s13059-019-1832-y

Groenewald M, Hittinger CT, Bensch K, *et al.* (2023). A genome-informed higher rank classification of the biotechnologically important fungal subphylum *Saccharomycotina*. *Studies in Mycology* **101**: 1–20. https://doi.org/10.3114/sim.2023.105.01

Gurevich A, Saveliev V, Vyahhi N, *et al.* (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **29**(8): 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Haase MAB, Kominek J, Langdon QK, *et al.* (2017). Genome sequence and physiological analysis of *Yamadazyma laniorum f.a. sp. nov.* and a reevaluation of the apocryphal xylose fermentation of its sister species, *Candida tenuis*. *FEMS Yeast Research* **17**(1): fox019. https://doi.org/10.1093/femsyr/fox019

Hagman A, Säll T, Compagno C, *et al.* (2013). Yeast "Make-Accumulate-Consume" life strategy evolved as a multi-step process that predates the whole genome duplication. *PLoS ONE* **8**(5): e68734. https://doi.org/10.1371/journal.pone.0068734

Hisatomi T, Toyomura K (2021). Isolation, identification, and characterization of wild budding yeasts from rose flowers in Fukuyama City, Hiroshima, Japan, and their application in bread and wine production. *Mycoscience* **62**(6): 382–389. https://doi.org/10.47371/mycosci.2021.10.003

Hittinger CT, Rokas A, Carroll SB (2004). Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proceedings of the National Academy of Sciences* **101**(39): 14144–14149. https://doi.org/10.1073/pnas.0404319101

Jain C, Rodriguez-R LM, Phillippy AM, *et al.* (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* **9**: 5114. https://doi.org/10.1038/s41467-018-07641-9

Hoang DT, Chernomor O, Von Haeseler A, *et al.* (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35**(2): 518–522. https://doi.org/10.1093/molbev/msx281

Kalyaanamoorthy S, Minh BQ, Wong TKF, *et al.* (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**(6): 587–589. https://doi.org/10.1038/nmeth.4285

Katoh K, Standley DM (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* **30**(4): 772–780. https://doi.org/10.1093/molbev/mst010

Koren S, Walenz BP, Berlin K, *et al.* (2017). Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Research* **27**(5): 722–736. https://doi.org/10.1101/gr.215087.116

Kurtzman CP (2011). *Torulaspora* Lindner. 1904. In: *The Yeasts, A Taxonomic Study* (CP Kurtzman, JW Fell, T Boekhout, eds), 5th edn. Elsevier, Netherlands: 867–874. https://doi.org/10.1016/B978-0-444-52149-1.00007-0

Kurtzman CP, Fell JW, Boekhout T, *et al.* (2011). Methods for isolation, phenotypic characterization and maintenance of yeasts. In: *The Yeasts, A Taxonomic Study* (CP Kurtzman, JW Fell, T Boekhout, eds), 5th edn. Elsevier, Netherlands: 87–110. https://doi.org/10.1016/B978-0-444-52149-1.00007-0

Lachance MA (2012). In defense of yeast sexual life cycles: The *forma asexualis*–an informal proposal. *Yeast Newsletter* **62**: 24–25.

Lachance MA, Lee DK, Hsiang T (2020). Delineating yeast species with genome average nucleotide identity: A calibration of ANI with haplontic, heterothallic *Metschnikowia* species. *Antonie van Leeuwenhoek* **113**(12): 2097–2106. https://doi.org/10.1007/s10482-020-01480-9

Lee I, Kim YO, Park SC, *et al.* (2016). OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *International Journal of Systematic and Evolutionary*

Microbiology **66**(2): 1100–1103. https://doi.org/10.1099/ijsem.0.000760

Libkind D, Čadež N, Opulente DA, *et al.* (2020). Towards yeast taxogenomics: Lessons from novel species descriptions based on complete genome sequences. *FEMS Yeast Research* **20**(1): foaa042. https://doi.org/10.1093/femsyr/foaa042

Libkind D, Hittinger CT, Valerio E, *et al.* (2011). Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proceedings of the National Academy of Sciences* **108**(35): 14539–14544. https://doi.org/10.1073/pnas.1105430108

Lopes MR, Morais CG, Kominek J, *et al.* (2016). Genomic analysis and D-xylose fermentation of three novel *Spathaspora* species: *Spathaspora girioi sp. nov.*, *Spathaspora hagerdaliae f.a.*, *sp. nov.* and *Spathaspora gorwiae f.a.*, *sp. nov. FEMS Yeast Research* **16**(2): fow022. https://doi.org/10.1093/femsyr/fow044

Marsit S, Mena A, Bigey F, *et al.* (2015). Evolutionary advantage conferred by an eukaryote-to-eukaryote gene transfer event in wine yeasts. *Molecular Biology and Evolution* **32**(7): 1695–1707. https://doi.org/10.1093/molbev/msv057

Morais CG, Batista TM, Kominek J, *et al.* (2017). *Spathaspora boniae sp. nov*., a D-xylose-fermenting species in the *Candida albicans*/*lodderomyces* clade. *International Journal of Systematic and Evolutionary Microbiology* **67**(10): 3798 https://doi.org/10.1099/ijsem.0.002186

Opulente DA, Langdon QK, Jarzyna M, *et al.* (2023). Taxogenomic analysis of a novel yeast species isolated from soil, *Pichia galeolata sp. nov. Yeast* **40**: 608–615. https://doi.org/10.1002/yea.3905

Pacheco A, Santos J, Chaves S, *et al.* (2012). The emerging role of the yeast *Torulaspora delbrueckii* in bread and wine production: Using genetic manipulation to study molecular basis of physiological responses. In: *Structure and Function of Food Engineering* (Eissa A, ed.), IntechOpen. https://doi.org/10.5772/46024

Peter J, De Chiara M, Friedrich A, *et al.* (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* **556**: 339–344. https://doi.org/10.1038/s41586-018-0030-5

Proux-Wéra E, Armisén D, Byrne KP, *et al.* (2012). A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. *BMC Bioinformatics* **13**: 237. https://doi.org/10.1186/1471-2105-13-237

Santos ARO, Aires A, Pontes A, *et al.* (2021). *Phaffia brasiliana sp. nov.*, a yeast species isolated from soil in a cerrado–Atlantic rainforest ecotone site in Brazil. *International Journal of Systematic and Evolutionary Microbiology* **71**: 005080. https://doi.org/10.1099/ijsem.0.005080

Santos ARO, Barros KO, Batista TM, *et al.* (2023). *Saccharomycopsis praedatoria sp. nov*., a predacious yeast isolated from soil and rotten wood in an Amazonian rainforest biome. *International Journal of Systematic and Evolutionary Microbiology* **73**: 006125. https://doi.org/10.1099/ijsem.0.006125

Shen X-X, Opulente DA, Kominek J, *et al.* (2018). Tempo and mode of genome evolution in the budding yeast subphylum. *Cell* **175**: 1533–1545. https://doi.org/10.1016/j.cell.2018.10.023

Silva M, Pontes A, Franco-Duarte R, *et al.* (2022). A glimpse at an early stage of microbe domestication revealed in the variable genome of *Torulaspora delbrueckii*, an emergent industrial yeast. *Molecular Ecology* **32**: 2396–2412. https://doi.org/10.1111/mec.16428

Spurley WJ, Fisher KJ, Langdon QK, *et al.* (2022). Substrate, temperature, and geographical patterns among nearly 2000 natural yeast isolates. *Yeast* **39**: 55–68. https://doi.org/10.1002/yea.3679

Taxis C, Keller P, Kavagiou Z, *et al.* (2005). Spore number control and breeding in *Saccharomyces cerevisiae*: A key role for a self-organizing system. *Journal of Cell Biology* **171**: 627–640. https://doi.org/10.1083/jcb.200507168

Vaser R, Sović I, Nagarajan N, *et al.* (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research* **27**: 737–746. https://doi.org/10.1101/gr.214270.116

Walker BJ, Abeel T, Shea T, *et al.* (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**(11): e112963. https://doi.org/10.1371/journal.pone.0112963

Wang QM, Xu J, Wang H, *et al.* (2009). *Torulaspora quercuum sp. nov*. and *Candida pseudohumilis sp. nov.*, novel yeasts from human and forest habitats. *FEMS Yeast Research* **9**: 1322–1326. https://doi.org/10.1111/j.1567-1364.2009.00567.x

Warren RL, Yang C, Vandervalk BP, *et al.* (2015). LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. *Gigascience* **4**(1): 35. https://doi.org/10.1186/s13742-015-0076-3

Wei J, Wang S, Zhang Y, *et al.* (2019). Characterization and screening of non-*Saccharomyces* yeasts used to produce fragrant cider. *LWT - Food Science and Technology* **107**: 191–198. https://doi.org/10.3390/fermentation9090824

Wolfe KH, Shields DC (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**(6634): 708–713. https://doi.org/10.1038/42711

Yoon SH, Ha S min, Lim J, *et al.* (2017). A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek* **110**: 1281–1286. https://doi.org/10.1007/s10482-017-0844-4

Zhang BQ, Luan Y, Duan CQ, *et al.* (2018). Use of *Torulaspora delbrueckii* co-fermentation with two *Saccharomyces cerevisiae* strains with different aromatic characteristics to improve the diversity of red wine aroma profile. *Frontiers in Microbiology* **9**: 606. https://doi.org/10.3389/fmicb.2018.00606

Supplementary material

Data S1. Different yeast isolation procedures used in this study.

Fig. S1. Average nucleotide identity (ANI) matrix for the genus *Torulaspora*. The analysis used the same genomes as in Fig. 1. Inter-species ANI values are shown in white or black backgrounds and represent the averages of all possible genome comparisons between two species. Numerals in black backgrounds correspond to comparisons between novel species and their closest relatives. Pairwise averages of all relevant intraspecific ANI values are shown in coloured or grey backgrounds; these values are not available (NA) when only one genome per species was studied.

Fig. S2. Phylogenetic tree based on a sequence alignment of the D1/D2 region of the LSU rDNA of representatives of all *Torulaspora* species, including sequences retrieved from GenBank for which genome data are not available. The phylogeny tree was constructed with the maximum-likelihood method and the Tamura–Nei model of sequence evolution, with equal base frequence, and was rooted with *Zygotorulaspora mrakii*. The final dataset had a total of 588 positions. Names in black indicate sequences retrieved by us from genome assemblies (bold highlights type strains). Note that, in some cases, more than one sequence was retrieved from the genome. Names in blue represent sequences retrieved from GenBank; in these cases, the original species designations were maintained. Species

names and delimitations based on taxogenomics are indicated on the right side. The scale bar represents the estimated number of substitutions per site. Percentage bootstrap values of 1000 replicates are given at each node.

Fig. S3. Phylogenetic tree based on a sequence alignment of the complete ITS region of the rDNA of representatives of all *Torulaspora* species, including sequences retrieved from GenBank for which genome data are not available. The phylogenetic tree was constructed with the maximum-likelihood method and the Hasegawa-Kishino-Yano model of sequence evolution and was rooted with *Zygotorulaspora mrakii*. The final dataset had a total of 843 positions. Names in black indicate sequences retrieved by us from genome assemblies (bold highlights type strains). Note that, in some cases, more than one sequence was retrieved from the genome. Names in blue represent sequences retrieved from GenBank; in these cases, the original species designations were maintained. Species names and delimitations based on phylogenomics are indicated on the right side. The scale bar represents the estimated number of substitutions per site. Percentage bootstrap values of 1000 replicates are given at each node.

Fig. S4. Phylogenetic tree based on a sequence alignment combining the D1/D2 and ITS regions of the rDNA of representatives of all *Torulaspora* species, including sequences retrieved from GenBank for which genome data are not available. The phylogenetic tree was constructed with the maximum-likelihood method and the TPM2u model of sequence evolution and was rooted with *Zygotorulaspora mrakii*. Alignment positions containing gaps and missing data were eliminated before the analysis. The final dataset had a total of 1411 positions. Names in black indicate sequences retrieved by us from genome assemblies (bold highlights type strains). Note that, in some cases, more than one sequence was retrieved from the genome. Names in blue represent sequences retrieved from GenBank; in these cases, the original species designations were maintained. Species names and delimitations based on phylogenomics are indicated on the right side. The scale bar represents the estimated number of substitutions per site. Percentage bootstrap values of 1000 replicates are given at each node.

Fig. S5. Molecular validation of the synonyms of *T. delbrueckii*. (A) List of synonyms of *T. delbrueckii* and details on ITS and whole-genome sequences. (B) Phylogenetic tree based on a sequence alignment of the complete ITS region of all species in the genus *Torulaspora* (depicted in black) and the current synonyms of *T. delbrueckii* (depicted in blue). The highlighted box corresponds to the species *T. delbrueckii*. The phylogenetic tree constructed with the maximum-likelihood method and the TIM2 model of sequence evolution, chosen according to the BIC, and rooted with *Zygotorulaspora mrakii*. The final dataset had a total of 813 positions. Names in bold indicate the novel species. The scale bar represents the estimated number of substitutions per site. Percentage bootstrap values of 1000 replicates equal or higher that 95 % are depicted as black circles at tree nodes.

Fig. S6. Expanded phenotype map for the genus *Torulaspora*. The results of physiological tests including carbon fermentation, carbon assimilation, nitrogen assimilation, temperature preference, and sensitivity to select compounds for members of the *Torulaspora* genus are arranged by phylogenetic relationship, indicated by the associated tree. The presence of a square indicates a tested trait while the absence of a square indicates a trait that was not tested for that isolate. A full square indicates a positive phenotype (e.g. fermentation of a substrate or growth at a specified temperature) while an empty square indicates a negative phenotype (e.g. failure to assimilate a substrate or lack of growth in the presence of certain compounds). Phenotypic results were obtained in this work or were taken from the literature as indicated in Table S3.

Table S1. Strains and genomes analysed in this study and relevant information pertaining to them (strains are ordered as in the phylogeny of Fig. 1).

Table S2. Sources of D1/D2 and ITS rDNA sequences used to construct Figs S2, S3, and S4.

Table S3. Sources of the phenotypic information depicted in Fig. 2.

Table S4. Signature amino acids of α-glucosidase-encoding genes (strains are ordered as in Fig. 2). Signature amino acids that correlate to substrate specificity for maltases and isomaltases, numbered as in *Saccharomyces cerevisiae* IMA1, are compared for isomaltase-, maltase- and maltase/isomaltase-like sequences. Relevant amino acid signatures for maltase and isomaltase activity are highlighted in yellow and blue, respectively, and mixed maltase–isomaltase affinity is colour-coded in green.

Table S5. Standard physiological and biochemical profiles of the novel *Torulaspora* species described this study (results shown for each strain).

Table S6. Standard physiological and biochemical profiles of all species in the genus *Torulaspora* (species are ordered as in the phylogeny of Fig. 1).