

A yeast love triangle: multiple hybridizations shape genome evolution in the *Pichia cactophila* species complex

Mixão V^{1,2,3}, Saus E^{1,2}, Hagen F^{4,5,6}, Boekhout T⁷, Stevens RW⁸, and Gabaldón T^{1,2,9,10*}

¹ Life Sciences Department, Barcelona Supercomputing Center (BSC), Jordi Girona, 29, 08034 Barcelona, Spain

² Institute for Research in Biomedicine (IRB), The Barcelona Institute of Science and Technology, Barcelona, Spain

³ Bioinformatics Unit, Infectious Diseases Department, National Institute of Health Dr. Ricardo Jorge, Av. Padre Cruz, 1649-016, Lisbon, Portugal

⁴ Westerdijk Fungal Biodiversity Institute, Utrecht, Netherlands

⁵ Institute of Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, Netherlands

⁶ Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, Netherlands

⁷ College of Sciences, King Saud University, 11451 Riyadh, Saudi Arabia

⁸ Department of Pharmacy Services, Mayo Clinic, 201 1st St. SW Rochester, MN 55902, USA

⁹ ICREA, Pg. Lluís Companys 23, Barcelona 08010, Spain

¹⁰ Centro de Investigación Biomédica En Red de Enfermedades Infecciosas, Barcelona, Spain

Citation – Mixão V, Saus E, Hagen F, Boekhout T, Stevens RW, Gabaldón T 2024 – A yeast love triangle: multiple hybridizations shape genome evolution in the *Pichia cactophila* species complex. *Mycosphere* 15(1), 1083–1109, Doi 10.5943/mycosphere/15/1/9

Abstract

Hybrids are chimeric organisms carrying genetic material from at least two divergent parental lineages. Hybridization can contribute to the emergence of novel lineages with unique phenotypic traits that may facilitate their adaptation to new environments. In recent years, genomic analyses have revealed the hybrid nature of several opportunistic human pathogenic yeasts. One of them is *Pichia inconspicua*, a member of the *Pichia cactophila* species complex, for which all sequenced strains isolated from Europe are hybrids with so far unknown parentals. It has been recently proposed that *P. inconspicua* and *P. cactophila* s.s. should be ranked as the same species due to their genetic similarity. To obtain a better understanding of the evolution of this clade, we sequenced the genomes of the *P. cactophila* type strain, of its close-relative *Pichia pseudocactophila*, and of a putative *P. inconspicua* clinical isolate from Alaska and compared them with the previously sequenced genomes of *Pichia norvegensis*, *P. inconspicua* and the recently described *Pichia galeolata*. Our results show evidence for the existence of distinct lineages with a hybrid origin within this clade and suggest an intricate scenario of recurrent hybridizations in this species complex, some of them giving rise to lineages with the ability to infect humans. Given their different hybridization histories, we propose that *P. inconspicua*, *P. cactophila*, and the new clinical isolate from Alaska should represent three distinct species and suggest the name *Pichia alaskaensis* for the new taxon. Our results clarify the evolutionary relationships within the *P. cactophila* species complex and underscore the importance of non-vertical evolution.

Keywords – *C. inconspicua* – comparative genomic – *P. alaskaensis* – *P. cactophila* – *P. galeolata* – *P. norvegensis* – *P. pseudocactophila*

INTRODUCTION

Hybridization, i.e. the cross of two diverged lineages, gives rise to organisms with highly heterozygous genomes, which can be a source of negative epistatic interactions, and often lead to hybrid unviability (McClintock 1984, Mixão & Gabaldón 2018). Although some authors limit the

Submitted: 28 January 2024; **Accepted:** 25 May 2024; **Published:** 27 August 2024

***Corresponding Author:** Toni Gabaldón – e-mail – toni.gabaldon@bsc.es

Accepted by: Dhanushka N. Wanasinghe

definition of “hybrid” to the F1 offspring generation (Sipiczki 2018), the standard practice in yeast evolutionary biology, and the one we apply here, considers also the clonal progeny of F1 hybrids as hybrids, constituting a hybrid lineage (Morales & Dujon 2012, Gabaldón 2020, Boekhout et al. 2021). Different processes can shape hybrid genomes leading to loss of heterozygosity (LOH) and consequent genome stabilization (Payseur & Rieseberg 2016, Mixão et al. 2023), favoring the emergence of lineages with advantageous combinations of genomic and phenotypic traits (Gladieux et al. 2014, Mixão & Gabaldón 2018). Over the last years, a growing number of yeast hybrid lineages have been described (Gabaldón 2020), and hybridization has been hypothesized as a possible mechanism leading to the emergence of lineages with the ability to grow in new environments, such as the human body, as it is the case of *Candida orthopsilosis*, *Candida metapsilosis*, *Pichia inconspicua* (also known as *Candida inconspicua*), *Candida tropicalis* or *Candida albicans* (Pryszcz et al. 2014, 2015, Mixão & Gabaldón 2018, Mixão et al. 2019, Mixão & Gabaldón 2020, Mixão et al. 2021a, Mixão et al. 2021b, O’Brien et al. 2021, O’Brien et al. 2022, Del Olmo et al. 2023).

Pichia inconspicua (syn. = *Candida inconspicua*) is an emerging yeast pathogen, recently transferred to the *Pichia* genus (Zhu et al. 2024), for which all sequenced strains were isolated in Europe and correspond to hybrids with so far unknown parental lineages (Pfaller et al. 2010, Mixão et al. 2019). The analysis of the genomes of multiple clinical isolates revealed the existence of two distinct clades which differ in terms of genomic variability and ploidy levels. Despite these differences, it remained unclear whether the two identified clades resulted from one or two independent hybridization events (Mixão et al. 2019). *P. inconspicua* belongs to the *Pichia cactophila* species complex, which also includes *P. cactophila* s.s., *Pichia pseudocactophila*, *Pichia norvegensis* (syn. = *Candida norvegensis*) and the recently described *Pichia galeolata*, among others (Kurtzman et al. 2008, Ganter et al. 2010, Opulente et al. 2023) (Fig. 1). While *P. inconspicua* and *P. norvegensis* are considered emerging pathogens, *P. cactophila* and *P. pseudocactophila* are more frequently found on cacti and *P. galeolata* was isolated from soil (Starmer et al. 1978, Sandven et al. 1997, Moraes et al. 2005, Guitard et al. 2013, Czaika et al. 2013, Gonçalves et al. 2023, Opulente et al. 2023). Despite their differences in the ability to sporulate, which is almost nonexistent in the case of *P. inconspicua*, ITS-based phylogenetic analysis and identification through *EF-1 α* sequencing and MALDI-TOF revealed a high similarity between *P. inconspicua* and *P. cactophila* (Guitard et al. 2015). For this reason, the same authors suggested that they correspond to the same species (Guitard et al. 2015). As a result, some studies performed since 2015 do not make any distinction between these lineages (Douglass et al. 2018, Borman et al. 2019) and proposed the transfer of the name *P. inconspicua* in the genus *Pichia* as *P. cactophila* (Borman & Johnson 2020).

Considering the hybrid nature of *P. inconspicua*, and its apparent genetic similarity to *P. cactophila*, we hypothesized that *P. cactophila* could correspond to one of its missing parental lineages. To test this, as well as to obtain further insights on the evolution of this clade, we sequenced the type strains of *P. cactophila* (CBS 6926) and *P. pseudocactophila* (CBS 6929) and performed a comparative genomics analysis with *P. inconspicua*. We have also analyzed the publicly available data of the closely related species *P. norvegensis* and *P. galeolata* (Krause et al. 2018, Shen et al. 2018, Opulente et al. 2023), and that of a putative *P. inconspicua* clinical strain from Alaska (CBS 16747), which we propose to be ranked as a new species named *Pichia alaskaensis*.

MATERIALS AND METHODS

Genome sequencing

Genomic DNA sequencing was performed for *P. cactophila* (CBS 6926) and *P. pseudocactophila* (CBS 6929) type-strains, and for the putative *P. inconspicua* (CBS 16747) isolated in a hospital in Alaska (USA) (Caldwell et al. 2024). A modified protocol from the MasterPure™ Yeast DNA Purification Kit was used to extract the DNA. In brief, samples were

grown overnight in liquid YPD at 30 °C. Cells were pelleted and lysed with RNase treatment at 65°C for 15min. After 5min of cooling down on ice, samples were purified by the kit reagent by mixing, centrifugation and removal of the debris as described in the kit protocol. Further, samples were left at -20 °C with absolute ethanol for at least 2h after which the DNA was precipitated for 30min at 4 °C. The pellet was washed in 70% ethanol and left to dry. TE buffer was used to resuspend the DNA. Genomic DNA Clean & Concentrator kit (ZymoResearch, Irvine, CA, USA) was used for the final purification.

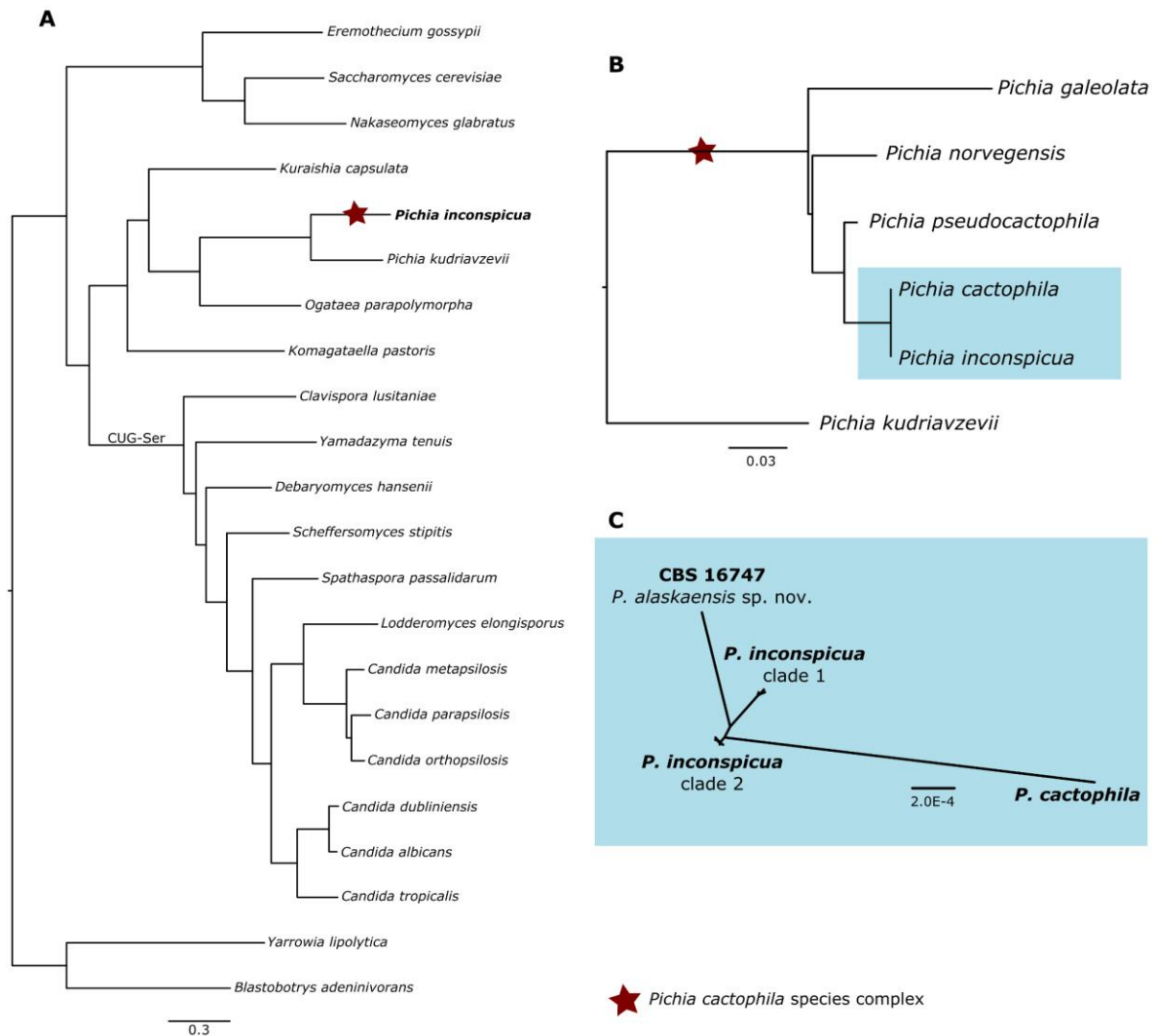


Figure 1 – Phylogenetic analysis of the members of the *P. cactophila* species complex. A) Phylogenetic tree reconstruction of the concatenated sequence alignment of the *P. inconspicua* phylome 498 (Mixão et al. 2019), available at PhylomeDB (Fuentes et al. 2022), showing the phylogenetic placement of *P. cactophila* species complex (marked with a star) within Ascomycota. B) Phylogenetic tree reconstruction of the alignment of the ITS region of the species analyzed in this study. C) Phylogenetic tree reconstruction of the alignment comprising all the homozygous SNPs resultant from read mapping of *P. inconspicua* clade 1 strains, *P. inconspicua* clade 2 strains, CBS 1647 and *P. cactophila* s.s. on *P. inconspicua* reference genome (Mixão et al. 2019).

Libraries were prepared using the NEBNext® DNA Library Prep Reagent Set for Illumina® kit (New England Biolabs, Ipswich, MA, USA) according to the manufacturer’s protocol. Briefly, 1 µg of gDNA was fragmented by nebulization in g-tubes (Covaris, Woburn, MA, USA) to approximately 600 bp and subjected to end repair, addition of adenine bases to the 3’-ends and ligation of Truseq adapters. All purification steps were performed using Qiagen PCR purification

columns (Qiagen, Hilden, Germany). Library size selection was done with 2% low-range agarose gels. Fragments with average insert size of 700 bp were cut from the gel, and DNA was extracted using QIAquick Gel extraction kit (Qiagen) and eluted in 30 µl EB. Ten µl of adapter-ligated size-selected DNA were used for library amplification by PCR using the Truseq Illumina primers. Final libraries were analyzed using Agilent DNA 1000 chip to estimate the quantity and check size distribution and were then quantified by qPCR using the KAPA Library Quantification Kit (ref. KK4835; Kapa Biosystems, Wilmington, MA, USA) prior to amplification with Illumina's cBot. Libraries were loaded at a concentration of 2 pM onto the flow cell and were sequenced 2 x 125bp on Illumina's HiSeq 2500.

Public data

For a comparative genomics analysis with *P. inconspicua*, the sequencing library of *P. inconspicua* CBS 180 (SRR8506592) and the respective genome assembly (ASM493185v1) were downloaded from NCBI database (Mixão et al. 2019). Moreover, the concatenated sequence alignment of the phylome 498, which uses this strain as seed, was downloaded from PhylomeDB (Fuentes et al. 2022). For the analysis of *P. norvegensis* and *P. galeolata*, publicly available data was retrieved from NCBI. Specifically, we downloaded the genome assemblies under the accession numbers ASM370546v1 and ASM3055572v1, and the sequencing libraries under the sequencing runs SRR6476040 and SRR16974481, respectively (Krause et al. 2018; Shen et al. 2018; Opulente et al. 2023).

Read mapping and Variant calling

Next-Generation Sequencing data was inspected with FastQC v0.11.5 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Paired end reads were filtered for quality below 10 or size below 31 bp and for the presence of adapters with Trimmomatic v0.36 (Bolger et al. 2014). Read mapping and variant calling were performed using HaploTypo pipeline v1.0.1 with default parameters (Pegueroles et al. 2020). Briefly, BWA-MEM v0.7.15 (Li 2013) was used for read mapping on the respective genome assemblies, and GATK v4 (McKenna et al. 2010) was used to call variants with the tool HaplotypeCaller followed by VariantFiltration. The read alignment was inspected with IGV v2.0.30 (Thorvaldsdóttir et al. 2013). Mapping coverage was determined with Samtools v1.9 (Li et al. 2009). Only positions in the reference with 20 or more reads were considered for the analysis, and these were determined with bedtools genomecov v2.26.0 (Quinlan & Hall 2010). Ploidy estimation was performed with nQuire (Weiß et al. 2018).

LOH block definition

To determine for each heterozygous strain the presence of LOH blocks, heterozygous and homozygous variants were separated. Then, the procedure applied and validated by Prysycz and colleagues and later refined by the same group was used (Prysycz et al. 2015, Mixão & Gabaldón 2020). Briefly, bedtools merge v2.26.0 (Quinlan & Hall 2010) with a distance of 100 bp was used to define heterozygous regions, and by opposite, LOH blocks would be all non-heterozygous regions in the genome (Mixão & Gabaldón 2020). Moreover, the minimum LOH and heterozygous blocks sizes were established at 100 bp. Divergence between homeologous chromosomes was calculated by dividing the number of variants present in heterozygous regions by the size of those same regions.

Analysis of four universal marker genes

Previous analysis of markers traditionally used for species identification revealed a high similarity between species of *P. cactophila* complex. Therefore, for the reconstruction of the phylogenetic relationships between the different hybrid lineages, we used four marker genes previously shown to be optimal phylogenetic markers for Ascomycota (Capella-Gutierrez et al. 2014), namely, *KOG1*, *CLU1*, *VPS53* and *RFA1*, and phased in each hybrid strain using HapCUT2 (Edge et al. 2017). The different haplotypes of each gene were aligned with MAFFT v7 (Katoh &

Standley 2013) and trimmed with trimAL v1.4.rev15 (Capella-Gutiérrez et al. 2009). RAxML v8 (Stamatakis 2014) was used to reconstruct the Maximum Likelihood phylogenetic tree of each of the multi-sequence alignments, using the GTRCAT model.

Genome assembly

The K-mer Analysis Toolkit v2.4.1 (KAT, (Mapleson et al. 2017)) was used to count k -mer frequency and estimate the expected genome size using default parameters ($k = 27$). SOAPdenovo v2.04 (Luo et al. 2012) and SPAdes v3.9 in both SPAdes and dipSPAdes modes (Bankevich et al. 2012; Safonova et al. 2015) were used separately to perform the genome assembly. Afterwards, redundant contigs were removed from each assembly with Redundans v0.13c (Pryszcz & Gabaldón 2016). The quality of the different assemblies was inspected with Quast v4.5 (Gurevich et al. 2013). Genome annotation was performed with Augustus v3.5 using *C. albicans* as model organism (Stanke & Morgenstern 2005). The assembly completeness was estimated with KAT and BUSCO v3 using the Ascomycota database (Mapleson et al. 2017, Waterhouse et al. 2019). The best assembly for each species was chosen based on the assembly completeness, genome size, N50 and number of scaffolds.

Phylogenetic tree reconstruction

For the phylogenetic analysis of the ITS region, the ITS sequences of *P. cactophila*, *P. pseudocactophila* and strain CBS 16747 were retrieved from their respective genome assemblies using the ITS sequence of *P. inconspicua* type strain (NR_111116.1) as query. Additionally, ITS sequences of *P. norvegensis*, *P. galeolata* and *P. kudriavzevii* were retrieved from the NCBI database (accession numbers OP800188.1, OL583853.1 and NR_131315.1, respectively). The multiple sequence alignment of these sequences was generated with MAFFT v7 (Katoh & Standley 2013) and trimmed with trimAL v1.4.rev15 (Capella-Gutiérrez et al. 2009).

For the whole-genome scale phylogenetic analysis of *P. cactophila* s.s., *P. inconspicua* and CBS 16747, the concatenated alignment of all the homozygous SNPs observed in at least one of these strains when mapped onto the *P. inconspicua* genome assembly (Mixão et al. 2019) was used. Phylogenetic tree reconstruction was performed with RAxML v8 (Stamatakis 2014).

Phenotypic growth test

The morphology and nutritional growth patterns of the strain CBS 16747, the putative new species *P. alaskaensis*, was investigated according to standard methods used in yeast taxonomy (Kurtzman et al. 2011). Fermentation and carbon utilization was tested in liquid media, whereas nitrogen compounds were tested on solid media using the auxanogram method. The morphology was investigated using MEA, yeast morphology agar (YMA), potato dextrose agar (PDA) and glucose yeast peptone agar (GYPA), 5% glucose in yeast nitrogen broth and Dalmau plates on yeast morphology agar. Formation of asci and ascospores was investigated for two months on the following media for which the recipes can be found in Kurtzman et al. (2011) (Kurtzman et al. 2011): YMoA, CMA, GYPA, MEA, 1/10 YMA, Fowell acetate agar, V8-agar and McLarry agar.

Antifungal susceptibility testing

The type-strains of *P. cactophila* (CBS 6926), *P. pseudocactophila* (CBS 6929), *P. inconspicua* (CBS 180), *P. norvegensis* (CBS 6564) and CBS 16747 (here proposed as the newly described species *P. alaskaensis*) were subjected to antifungal susceptibility testing according to the EUCAST broth dilution method for yeasts E.DEF 7.3.2.; <http://www.eucast.org/>). Antifungals tested were amphotericin B, 5-flucytosine, fluconazole, itraconazole, posaconazole, voriconazole, isavuconazole, anidulafungin and micafungin (all from Merck). The growth inhibition was checked spectrophotometrically as well as visually.

Data availability

The clinical isolate from Alaska, which is here proposed to be ranked as a new species

(*P. alaskaensis*), was deposited at the CBS yeast collection under the accession number CBS 16747 and at the IHEM/BCCM culture collection with the accession number IHEM 29050. Sequencing data, and respective genome assemblies and annotation, are available at the NCBI database under the BioProject PRJNA694915.

RESULTS

***P. cactophila* type strain and CBS 16747 represent two additional hybridization events in the *P. cactophila* species complex**

For a better understanding of the evolution of the hybrid opportunistic pathogen *P. inconspicua*, we sequenced the type strain of *P. cactophila* (CBS 6926) and the genome of a putative *P. inconspicua* clinical isolate from Alaska (CBS 16747). Given the expected genetic proximity between these new strains and *P. inconspicua* type strain, our initial strategy was to perform a *k*-mer comparison and a read mapping approach between these new libraries and the reference genome of *P. inconspicua* (Mixão et al. 2019) (details in Materials and Methods section). The *k*-mer-based analysis of CBS 16747 revealed that this strain has more than one peak of coverage, and part of the *k*-mers present in the sequencing data are absent from the *P. inconspicua* genome assembly (Supplementary Fig. 1). These patterns indicate that CBS 16747 has more than one haplotype, and only one of them is represented in the assembly, suggesting that this strain may also have a hybrid origin (Mixão et al. 2019). In addition, the presence of multiple peaks of coverage suggests a ploidy different from 2, and, indeed, computational estimations point that this strain has a triploid genome (nQuire histotest $r^2 = 0.98$). After read-mapping and variant-calling on *P. inconspicua* genome assembly, we detected 36.39 variants/kb from which 29.04 correspond to heterozygous positions (Table 1). These levels of heterozygosity are surprisingly high for a putative *P. inconspicua* strain, when compared to previously analyzed *P. inconspicua* strains (minimum 14, and maximum 19.76 heterozygous variants/kb, (Mixão et al. 2019)). Nevertheless, their distribution along the genome forming blocks of heterozygosity with similar haplotype divergence separated by blocks of LOH confirms that CBS 16747 is also a hybrid (Supplementary Fig. 2). In this regard, 37.94% of this genome corresponds to LOH regions, a value consistent with a higher level of heterozygosity in this strain when compared to the previously analyzed isolates of *P. inconspicua* (minimum 51% LOH, (Mixão et al. 2019)). Moreover, the current nucleotide divergence between the homeologous chromosomes of this strain is 4.46%, ~1% higher than the previous *P. inconspicua* strains (Table 1). These differences in the levels of LOH and parental sequence divergence are unexpected for a *P. inconspicua* strain, but in accordance with the observations of the *k*-mer analysis. Altogether, these results suggest that CBS 16747 derives from a hybridization event unrelated to any of the two previously described clades of *P. inconspicua* (Mixão et al. 2019).

Similar to the analysis of CBS 16747, *P. cactophila* was analyzed based on a read mapping approach on the *P. inconspicua* reference genome. The *k*-mer comparison between these strains revealed that only part of the genome of *P. cactophila* is represented in *P. inconspicua* genome assembly (Supplementary Fig. 1), and our estimations suggest that *P. cactophila* has a diploid genome (nQuire histotest $r^2 = 0.98$). Indeed, this species has 47.52 variants/kb when compared to *P. inconspicua* genome assembly, from which 39.52 are heterozygous (Table 1). Interestingly, once again, the heterozygous SNPs are not homogeneously distributed along the genome. Instead, they form blocks of heterozygosity with an estimated current nucleotide divergence of 6.41% (Supplementary Fig. 2), ~3% higher than *P. inconspicua* strains (Mixão et al. 2019), and ~2% higher than CBS 16747 (Table 1), supporting a scenario of yet an additional hybridization event in the clade. Of note, *P. cactophila* has 40.84% of the genome in LOH regions. The higher sequence divergence observed in *P. cactophila* indicates that the two parental lineages of this species are not the same as the parentals of *P. inconspicua* nor CBS 16747, but the observation of shared *k*-mers between the two species suggests that they possibly share at least one of the parental lineages of their respective hybridizations. As seen in Fig. 1c, the putative *P. inconspicua* strains (including

CBS 16747) form different hybridization clades, and *P. cactophila* is distantly related to all of them. Importantly, these results are at odds with the previously proposed scenario that *P. cactophila* and *P. inconspicua* correspond to the same species (Guitard et al. 2015).

Table 1 Summary of genomic variability in CBS 16747, *P. cactophila* type strain, *P. pseudocactophila* type strain, *P. norvegensis* type strain and *P. galeolata* type strain, with indication of the reference genome in which their sequencing data was aligned, ploidy, number of variants per kilo–base (kb), heterozygous variants per kb, percentage of the genome in LOH blocks, and the estimated haplotype divergence in heterozygous regions. *Ploidy estimated with nQuire (Weiß et al. 2018).

Lineage	Reference	Ploidy*	Variants/ kb	Heterozygous variants/kb	LOH (>100bp)	Haplotype divergence
CBS 16747	<i>P. inconspicua</i>	Triploid	36.39	29.04	37.94%	4.46%
<i>P. cactophila</i>	<i>P. inconspicua</i>	Diploid	47.52	39.52	40.84%	6.41%
<i>P. pseudocactophila</i>	<i>P. pseudocactophila</i>	Diploid	0.14	0.14	-	-
<i>P. norvegensis</i>	<i>P. norvegensis</i>	Diploid	4.03	4.03	91.04%	3.02%
<i>P. galeolata</i>	<i>P. galeolata</i>	Diploid	0.1	0.1	-	-

CBS 16747 and *P. inconspicua* share a parental lineage

To get a better understanding of the number of lineages involved in the origin of the new hybrids described here, we analyzed four previously proposed universal fungal phylogenetic marker genes (*CLUI*, *KOG1*, *RFA1*, and *VPS53*) (Capella–Gutierrez et al. 2014) and the *MAT* locus of these strains. Briefly, we phased these genes to recover the two haplotypes in each heterozygous strain and then reconstructed their phylogenetic relationships (see Materials and Methods). Due to the absence of known parental lineages, it was not possible to concatenate haplotypes from different genes and the analysis was performed separately for each of them. The reconstruction of the different haplotypes of the four marker genes in the *P. cactophila* clade revealed at least three putative sub–clades (Supplementary Fig. 3). Considering the existence of at least three different hybridization events, this suggests that some of these hybrid lineages share at least one of their parents. Indeed, for each gene and for both CBS 16747 and *P. cactophila*, one of the haplotypes is shared with *P. inconspicua* and the other one is not (Supplementary Fig. 3).

In the previous study in which the genomes of multiple *P. inconspicua* strains were investigated (Mixão et al. 2019), it was suggested that the polymorphisms between the two *P. inconspicua* clades observed in the *MAT* locus could be the consequence of accumulation of mutations after a shared hybridization event, and not differences between their parental lineages. These results were used to suggest that the two *P. inconspicua* clades possibly correspond to diverged lineages of the same hybridization event. However, our results show that CBS 16747 is a result of an independent hybridization event, and, even so, it shares many of the polymorphisms present in *P. inconspicua* clade 1 *MAT* α (Supplementary Fig. 4). Thus, these polymorphisms were possibly present before both hybridization events occurred (CBS 16747 and *P. inconspicua* clade 1), and therefore *P. inconspicua* clades 1 and 2 represent two different hybridizations. Furthermore, if most of these polymorphisms are ancestral, these results suggest that the CBS 16747 and *P. inconspicua* clade 1 share the parental lineage carrying the *MAT* α , and differ in the donor of *MAT* a . This scenario is in accordance with the observation of a heterozygous ITS region in CBS 16747, with one of the haplotypes being similar to *P. inconspicua* clade 1 (Supplementary Fig. 5).

P. cactophila derives from a secondary hybridization involving a *P. inconspicua* hybrid and another lineage

As mentioned above, the reconstruction of the different haplotypes in the set of phylogenetic marker genes revealed that one of the haplotypes of *P. cactophila* is close to *P. inconspicua*, but the alternative one is distantly related to any of its lineages (Supplementary Fig. 3). This result supports the hypothesis that *P. cactophila* and *P. inconspicua* share a parental lineage but differ in the other

one. Nevertheless, further inspection of the *MAT* locus revealed a much more complex scenario. Indeed, *P. cactophila* *MAT a* seems close to that present in the *P. inconspicua* clade 1. However, the *MAT a* allele has a *MAT α1* that is different from *P. inconspicua* and CBS 16747, and a *MAT α2* that is heterozygous (Supplementary Fig. 4). To confirm this observation, i.e., *P. cactophila* has two *MAT a* alleles, we assembled the genome of *P. cactophila* type strain (see Materials and Methods, and Supplementary Table 1). From this genome assembly, we recovered the full *MAT a* and *MAT α* alleles of *P. cactophila*. Our results show that one of the haplotypes of *MAT alpha2* has recombined in the *MAT a* and corresponds to the *MAT α2* of *P. inconspicua* (Fig. 2a, Supplementary Fig. 6). The existence of this recombination event supports a scenario in which one of the parents of *P. cactophila* was a *P. inconspicua* hybrid that underwent a recombination event in the *MAT* locus. This is in accordance with the ITS similarity observed between this species and *P. inconspicua* clade 1 (Supplementary Fig. 5). It is important to note that, as mentioned before, our estimations point to a diploid state of *P. cactophila*, which is not compatible with the idea of a hybrid parent. Therefore, we suggest that, after the recombination event in the *MAT* locus, the *P. inconspicua* hybrid that gave origin to *P. cactophila* experienced a ploidy reduction before the subsequent hybridization that originated *P. cactophila* (Fig. 2b). Alternatively, a *P. inconspicua* hybrid carrying the recombined *MAT* alleles crossed with the alternative parental of *P. cactophila*, and the ploidy reduction occurred afterwards.

***P. norvegensis* is not a parental lineage, but rather another hybrid**

With all the lineages analyzed so far in the *P. cactophila* species complex being hybrids, we decided to extend the analysis to *P. pseudocactophila*, *P. galeolata* and *P. norvegensis*, to assess whether one of them could be the alternative parental of *P. cactophila*. While for *P. pseudocactophila* we sequenced the type strain, for *P. galeolata* and *P. norvegensis* we retrieved the publicly available sequencing library and genome assembly from NCBI (details in the Materials and Methods section). A *k*-mer comparison showed that none of these three species is the parent of any of the hybrids identified in the clade. An inspection of *P. pseudocactophila* and *P. galeolata* revealed that both these species have a highly homozygous diploid genome (Table 1, Supplementary Fig. 1, and Supplementary Table 1). In contrast, similarly to other members of the clade, *P. norvegensis* presents two peaks of *k*-mer coverage (Supplementary Fig. 1), a pattern that, as mentioned before, is a good indicator of a highly heterozygous genome. *P. norvegensis* has 4.03 heterozygous variants/kb when mapped to its own reference genome (Table 1), thus presenting a lower heterozygosity than the one observed in *P. inconspicua* and *P. cactophila*. Interestingly, similarly to these other two species, the variants of *P. norvegensis* are not spread across the genome, but rather form blocks of heterozygosity, and the sequence divergence in these blocks has a single density peak (Supplementary Fig. 7), indicating that the heterozygosity of *P. norvegensis* was acquired at a single time point, and that it may have a hybrid nature (Mixão & Gabaldón 2020). The current sequence divergence in heterozygous blocks is 3%, and 91% of the genome is homozygous. This indicates that, despite its possible hybrid ancestry, *P. norvegensis* is much more homozygous than any other hybrid analyzed so far in the clade, pointing to an older hybridization event or to the occurrence of a massive LOH through chromosome recombination (as LOH blocks are interspersed with heterozygous genomes and it does not present aneuploidies as assessed by changes in coverage), similarly to what was previously described for the *C. albicans* clade (Mixão & Gabaldón 2020, Mixão et al. 2021b).

Proposal to rank CBS 16747 as a new species: *Pichia alaskaensis*

The species concept has evolved over time and different criteria have been applied to define yeast species (Boekhout et al. 2021). With recent technological advances, which, for example, nowadays provide the possibility to differentiate specimens at the nucleotide level, yeast species are undergoing significant nomenclature changes aiming to correct past practices (Kidd et al. 2023). In the case of the *P. cactophila* species complex, there is an important nomenclature uncertainty which is the distinction between *P. inconspicua* and its close relative *P. cactophila* s.s.. Indeed,

although they were both initially described as different species given their morphological and phenotypic differences, the finding of their genetic similarity in the ITS, D1/D2 and large ribosomal subunit region (shown in Supplementary Fig. 5) led to the ranking of both of them as the same species (Guitard et al. 2015, Borman & Johnson 2020). Nevertheless, the comparative genomics analysis performed in this study suggests they are different species. Although the two species are very similar in a conserved region commonly used to classify yeast species, namely the rDNA ITS region (Fig. 1b, Supplementary Fig. 5) (Boekhout et al. 2021), our results based on whole genome information clearly show that *P. cactophila* s.s. has a different genotype from *P. inconspicua* in almost 5% of its nucleotide positions. For this reason, we consider that it is important to maintain their classification as two separate species.

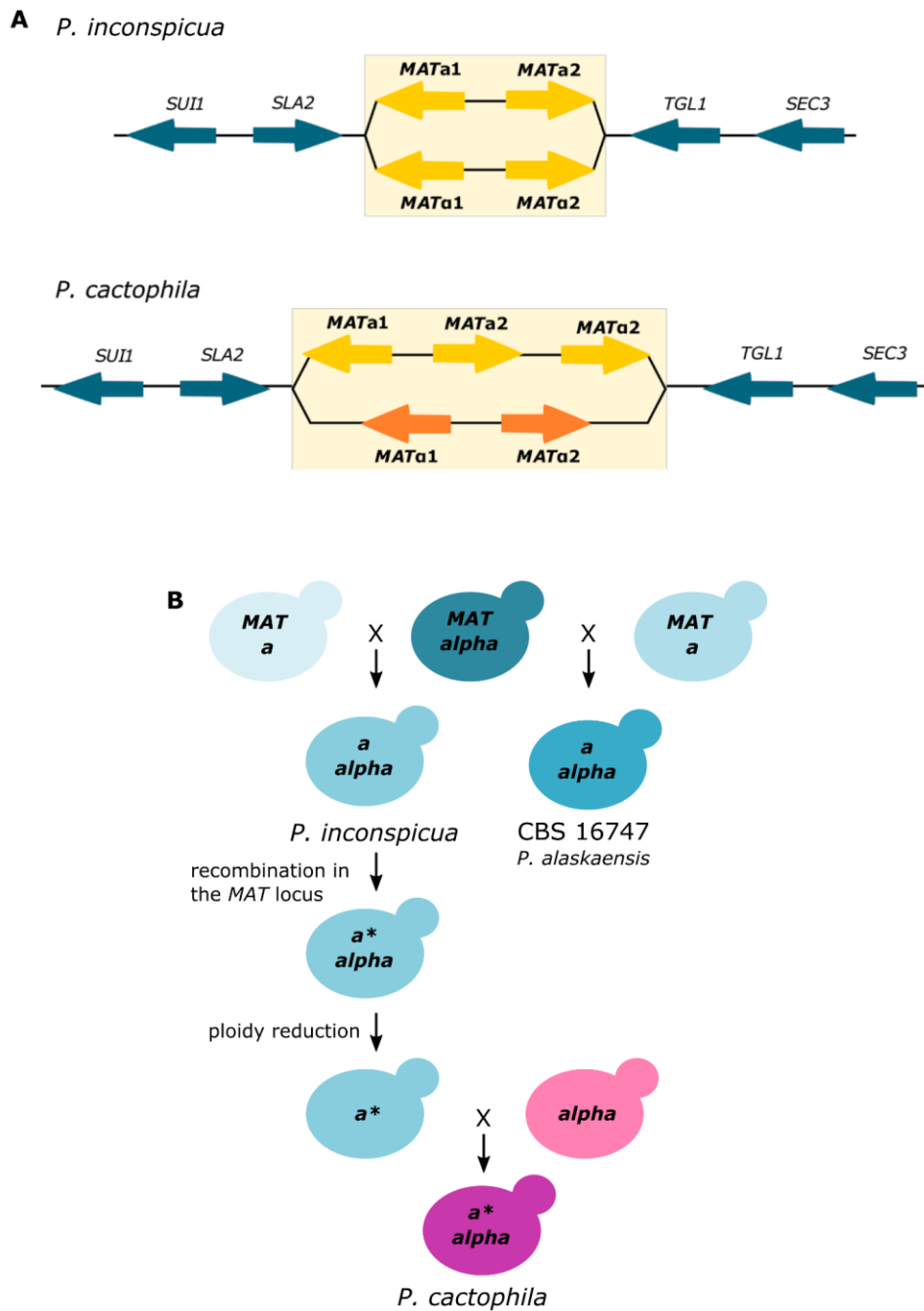


Figure 2 – Schematic representation of *P. cactophila* MAT locus and proposed evolutionary scenario. A) MAT locus representation of *P. inconspicua* and *P. cactophila* type strains. Yellow (*P. inconspicua*) and orange (*P. cactophila* alternative parental) arrows represent the distinct genes

of the *MAT a* and α idiomorphs. The direction of the arrow represents the DNA strand coding the gene. *P. cactophila* *MAT* locus comprises two different alleles, one of them including both *MAT a* and a single *MAT a* gene of *P. inconspicua*. B) Schematic representation of the evolutionary scenario proposed for the analyzed strains of the *P. cactophila* species complex.

This study uncovered another hybrid lineage within the species complex represented by strain CBS 16747. This strain was initially considered to represent *P. inconspicua* given its genetic similarity with this species in the ribosomal DNA genes, but we here showed that it does not belong to any of the previously described clades of this species (Mixão et al. 2019). As shown in Supplementary Fig. 5, CBS 16747 is heterozygous in the ITS region, with one ITS allele identical to that of *P. inconspicua*, likely leading to its erroneous initial classification as *P. inconspicua*, and the other ITS allele differing to the ITS of *P. inconspicua* type strain in 6 positions, suggesting it is a different species according to standard species discrimination criteria (Boekhout et al. 2021). The comparative genomics analysis clarifies this conundrum and shows that CBS 16747 results from the cross of one of the parents of *P. inconspicua* clade 1 and a yet unknown, more distantly related lineage (Table 1), which makes it phylogenetically distant from any of the known species of the complex (shown in Fig. 1c and in Supplementary Fig. 4). This genetic distance is also reflected phenotypically, as CBS 16747 presents different abilities not only regarding fermentation and metabolism of different compounds, but also regarding growth at very high temperatures and in the antifungal susceptibility profiles (Fig. 3, Tables 2, 3). Hybrid speciation is a known mechanism for the origin of novel species that is characterized by the new lineage being reproductively isolated from the two parental lineages (Arnold 1997). The presence of large blocks of heterozygosity across the genome indicates reproductive isolation with discrete loss of heterozygosity, as the existence of common crosses with any of the parental lineages would lead to introgression patterns characterized by the dominance of one of the haplotypes, which we do not observe. Thus, considering genotypic, phenotypic, and phylogenetic evidence presented here we propose that CBS 16747 represents a new taxon, different from *P. inconspicua*, for which we here propose the name *Pichia alaskaensis* (see below).

Description of *Pichia alaskaensis* sp. nov.

Pichia alaskaensis Gabaldón, Mixão, Hagen, Boekhout, spec. nov.

Mycobank number: MB851200

Etymology: alaskaensis, from Alaska, the state, USA, where the species was found

Morphology

Growth in glucose fermentation broth with a thin whitish film, somewhat adhering to the glass of the test tube, and with white sediment. Colony on yeast morphology agar after 7 days at 25 °C is 6–10 mm width, flat, smooth, shiny, off–white, butyrous, reverse whitish, with straight and entire margin (Fig. 3). On malt extract agar, it is 17 mm width, somewhat striate near margin; Dalmau culture on malt extract agar and potato dextrose agar showed patches of pseudohyphae. Yeast cells in glucose fermentation broth are ellipsoidal to cylindrical, 6.0–8.5 × 3.0–5.0 µm, with polar, sympodial to multilateral budding with the buds sessile on a rather broad base, single, in pairs or in short, branched chains that from small clusters of cells. Yeast cells on yeast morphology agar and malt extract agar similar, 6.0–10.5 × 2.8–5.5 µm, with cylindrical cells forming pseudohyphae that form laterally blastoconidia. After two months at YMoA larger cells occur, 8.0–11.0 × 2.0–5.0 µm, with polar and sympodial budding and that adhere into short pseudohyphae. No sexual state has been observed on YMoA, CMA, GYPA, MEA, 1/10 YMA, Fowell acetate agar, V8–agar and McLarry agar after up to two months of observations.

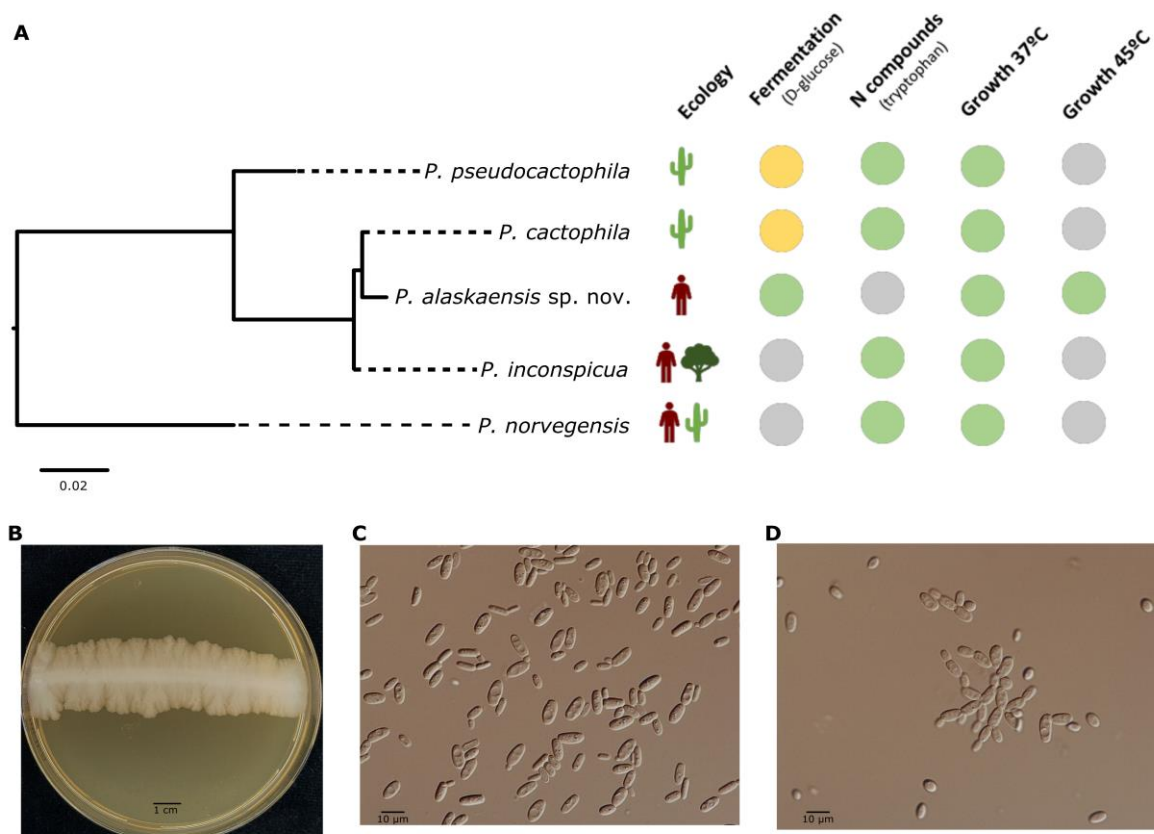


Figure 3 – Description of *P. alaskaensis* sp. nov. (CBS 16747) genetic and phenotypic features. A) Phylogenetic tree reconstruction of the nucleotide sequence alignment of the ITS region of the species of *P. cactophila* species complex analyzed in this study, with indication of their ecological niche (until the date of this study) and some phenotypic differences. Green circles indicate presence of phenotype, yellow circles indicate variable phenotype and gray circles indicate absence of phenotype. B) Colonies of *P. alaskaensis* on MEA and Yeast Morphology agar (YMoA) at 25 °C for 5 days. C) *P. alaskaensis* cells on YYMoA at 25 °C for 5 days. D) *P. alaskaensis* cells on glucose broth at 25 °C for 5 days.

Table 2 Morphological and growth features of *P. alaskaensis* and related species. Data on ecology and reproduction are taken from (Kurtzman et al. 2011) and (Gueho & Phaff 1983). Data on growth profiles were taken from <https://wi.knaw.nl/page/Collection>, except those for *P. alaskaensis* that were generated in this research, as described in the methods section. Legends: –, no growth; + growth; v, variable growth; d, delayed growth (see Kurtzman et al. 2011 for explanation of scores).

	<i>Pichia alaskaensis</i>	<i>Pichia inconspicua</i>	<i>Pichia cactophila</i>	<i>Pichia pseudocactophila</i>	<i>Pichia norvegensis</i>
Number of strains	1	6	12	5	17
Ascosporeulation*	Not known	Not known	Present	Present	Present
Asci*	Not known	Not known	Unconjugated	Heterothallic	Unconjugated
Ascospores*	Not known	Not known	2 per ascus, hat-shaped	4 per ascus, hat-shaped	1-4 per ascus, hat-shaped
Ecology*	Clinical	Sputum, tree fluxes oak tree	Various cactus species	Cardon cactus, hecho cactus	Sputum, vagina; roots of <i>Opuntia</i>
Fermentation					
D-glucose	+	-	v	v	wd,-
D-galactose	-	?	?	?	?
maltose	-	?	?	-	-
sucrose	-	?	?	-	-

Table 2 Continued.

	<i>Pichia alaskaensis</i>	<i>Pichia inconspicua</i>	<i>Pichia cactophila</i>	<i>Pichia pseudocactophila</i>	<i>Pichia norvegensis</i>
α,α trehalose	-	?	?	?	?
lactose	-	?	?	?	?
raffinose	-	?	?	?	?
D-xylose	-	?	?	?	?
Carbon compounds					
D-glucose	+	+	+	+	+
D-galactose	w	-	-	-	-
L-sorbose	wd,-	-	-	-	-
D-glucosamine	+	+	+	+	+
D-ribose	wd,-	-	-	-	-
D-xylose	wd	d,-	v	-	d,-
L-arabinose	-	-	-	-	-
D-arabinose	wd,-	-	-	-	-
L-rhamnose	w,-	-	-	-	-
sucrose	-	-	-	-	-
maltose	w,-	-	-	-	-
α,α trehalose	-	-	-	-	-
methyl α -glucoside	wd,-	-	-	-	-
cellobiose	w,-	-	-	-	+
salicin	-	-	-	-	+,d
arbutin	-	-	-	-	v
melibiose	w,-	-	-	-	-
lactose	-	-	-	-	-
raffinose	-	-	-	-	-
melezitose	-	-	-	-	-
inuline	w	-	-	-	-
soluble starch	-	-	-	-	-
glycerol	+	+	v	+	+
<i>Meso</i> erythritol	-	-	-	-	-
ribitol	-	-	-	-	-
xylitol	-	-	-	-	-
L-arabinitol	wd,-	-	-	-	-
D-glucitol	-	-	-	-	-
D-mannitol	-	-	-	-	-
galactitol	-	-	-	-	-
<i>myo</i> -inositol	-	-	-	-	-
glucono D-lactone	w,-	d,-	d,-	-	-
2-keto-D-gluconate	-	-	-	-	-
D-gluconate	-	-	-	-	-
D-glucuronate	-	-	-	-	-
D-galacturonate	-	-	-	-	-
DL-lactate	+	+	+	+	+
succinate	d,+	+	+	+	+
citrate	d,-	+	+	d,+	+
methanol	-	-	-	-	-
ethanol	+	+	+	+	+
propane 1,2 diol	w	-	-	-	?
butane 2,3 diol	-	-	-	-	-
quinic acid	-	-	-	-	-
saccharate	-	-	?	?	?
N compounds					
nitrate	-	-	-	-	-
nitrite	-	-	-	-	-
ethylamine	+	+	+	+	+

Table 2 Continued.

	<i>Pichia alaskaensis</i>	<i>Pichia inconspicua</i>	<i>Pichia cactophila</i>	<i>Pichia pseudocactophila</i>	<i>Pichia norvegensis</i>
L-lysine	+	+	+	+	+
cadaverine	+	+	+	+	+
creatine	-	-	-	-	-
creatinine	-	-	-	-	-
glucosamine	+	+	+	+	+
imidazole	-	-	-	-	-
proline	-	?	?	?	?
tryptophan	-	+	+	+	+
w/o inositol	?	+	+	+	+
w/o panthotenate	?	+,d	+	+	+
w/o biotin	?	-	-	v	v
w/o thiamine	?	-	-	-	-
w/o bio + thia	?	-	-	-	-
w/o pyridoxine	?	-	-	-	-
w/o pyr + thia	?	-	-	-	-
w/o niacin	?	+	+	+	v
w/o PABA	?	+	+	+	+
Other tests					
0.01% cycloheximide	-	-	-	-	-
1% acetic acid	?	-	-	-	-
50% glucose	-	-	-	-	-
starch production	-	-	-	-	-
acetic acid production	?	-	-	-	-
urea test	-	-	-	-	-
DBB reaction	-	-	-	-	-
Growth at different temperatures					
4°C	w	?	?	?	?
10°C	+	?	?	?	?
18°C	+	?	?	?	?
21°C	+	?	?	?	?
25°C	+	+	+	+	+
30°C	+	+	+	+	+
35°C	+	+	+	+	+
37°C	+	+	+	+	+
40°C	+	+	+	+	+
42°C	+	+	+	+	+
45°C	+	-	-	-	-
50°C	-	-	-	-	-

Table 3 Overview of antifungal susceptibility testing results using the EUCAST protocol.

	<i>P. alaskaensis</i> CBS 16747	<i>P. inconspicua</i> CBS 180	<i>P. cactophila</i> CBS 6926	<i>P. pseudocactophila</i> CBS 6929	<i>P. norvegensis</i> CBS 6564
Amphotericin B	0.5	0.125	0.25	0.5	0.5
5-Flucytosine	0.125	0.06/0.125 [#]	0.25	0.03	0.05
Fluconazole	>64	64	>64	32	>64
Itraconazole	0.25	0.03	0.06/0.125 [#]	0.03	0.06
Voriconazole	0.03/0.025 [#]	0.03	0.03/0.06 [#]	0.03	0.03/0.125 [#]
Posaconazole	0.25	0.06	0.25	0.06	0.125
Isavuconazole	0.125	0.03	0.125	0.03	0.06/0.125 [#]
Anidulafungin	0.03	0.03	0.03	0.03	0.03
Micafungin	0.03	0.03	0.03	0.03	0.06/0.125 [#]

* Antifungal susceptibility data is given in µg/ml; [#] indicates MIC-values that differ between 24h and 48h of incubation

For the growth profile of *P. alaskaensis*, see Table 2. Glucose is fermented, carbon utilization is similar to that of the related species and thus does not distinguish the species. The same holds for the utilization of nitrogen compounds. Contrary to the related species, *P. alaskaensis* can grow at 45 °C. The MIC values of *P. alaskaensis* for the antifungal drugs tested are presented in Table 3. For all antifungal drugs tested, the strain yielded low MIC values, except for fluconazole to which *Pichia* species are intrinsically resistant (Jamiu et al. 2021). The MIC-values for amphotericin B, itraconazole, voriconazole, posaconazole, anidulafungin and micafungin for CBS 16747 were below that of the epidemiological cut-off values (ECOFFs) for the clinically relevant species *P. kudriavzevii* (= *Candida krusei*) as outlined by Astvad et al (2022) (Astavd et al. 2022). For 5-flucytosine and isavuconazole there are no ECOFFs defined.

Origin of strain(s)

Holotype: CBS 16747, preserved metabolically inactive in the Westerdijk Fungal Biodiversity Institute, Utrecht, the Netherlands. CBS 16747 was, in 2018, isolated from the blood of a patient in Anchorage, Alaska, USA

Ex-type isolates: TG20210326 (collection Toni Gabaldón) = 2MG-A1202-16 (working collection Medical Mycology research group at WI) = IHEM 29050 (IHEM/BCCM culture collection).

Note – *Pichia alaskaensis* can be phenotypically distinguished from its close relatives by its ability to grow at 45 °C, to grow with D-galactose and inuline as sole carbon sources, lack of growth with D-tryptophane as a nitrogen source, and higher MIC values for itraconazole (Tables 2, 3).

DISCUSSION

In recent years, the emergence of new pathogenic lineages has shifted the epidemiology of yeast infections (Pfaller et al. 2010). Species that so far were not considered as medically relevant are now emerging as new opportunists (Pfaller et al. 2010a, 2010b, Mixão & Gabaldón 2018). For this reason, it is important to understand their evolution. Recent studies have pointed to a role of hybridization on the emergence of lineages with the ability to survive in new environments, and an increasing number of hybrid lineages has been described from clinical strains (Pryszcz et al. 2015, Mixão & Gabaldón 2018, Mixão et al. 2019, 2020, Gabaldón 2020). These comprise *P. inconspicua*, a member of the *P. cactophila* species complex, for which all clinical strains were identified as hybrids with so far unknown parental lineages (Mixão et al. 2019). The absence of known parental lineages among clinical isolates led to the hypothesis that they were possibly environmental, and hybridization possibly played a role in the emergence of this opportunistic pathogen (Mixão et al. 2019). Therefore, to get a better understanding of the evolution of *P. inconspicua*, the major goal of this study was to analyze other isolates of the *P. cactophila* species complex, namely the *P. cactophila* neotype and a putative isolate of *P. inconspicua* from Alaska, USA. Our results revealed that none of these species represents the parental of the previously described hybrids, and it remains still unclear what was the role of hybridization in the emergence of *P. inconspicua* pathogenicity.

Despite our unsuccessful results on the identification of the putative parents of *P. inconspicua*, a much more complex scenario than initially thought was uncovered, involving at least four different hybrid lineages occurring in the *P. cactophila* species complex. Our results suggest that the previously described clades of *P. inconspicua* (Mixão et al. 2019) correspond to two different hybridization events. This resembles the case of *C. orthopsilosis* in which the multiple crosses of the same two lineages originated multiple pathogenic hybrids (Schröder et al. 2016). More interesting, one of the parental lineages of *P. inconspicua* clades, specifically the one providing *MAT* α , has crossed with an alternative lineage with a higher sequence divergence than that of the alternative parent of *P. inconspicua*, giving rise to the lineage to which CBS 16747, isolated from patient from Alaska (Caldwell et al. 2024), belongs. Besides the possible relevance of hybridization for the ability of these lineages to cause opportunistic infections, this result raises the

question of the taxonomic classification of hybrid lineages, particularly those lacking known parentals. For instance, this clinical isolate from Alaska was originally identified as *P. inconspicua* (Caldwell et al 2024), but it does not result from the cross of the same two parental lineages as the *P. inconspicua* type strain. Therefore, it is uncertain whether they should have the same taxonomic classification, especially if we take into consideration their different phenotypes. This question may also be generalized to hybrids originated from a cross of the same parental lineages because the process of genome shaping may generate lineages with different genetic information, and possibly different phenotypes, despite their shared ancestry. Hybridization is a known source of genetic isolation and speciation in yeasts (Gabaldón 2020), and the case uncovered here underscores the potential complexity of resulting scenarios when the same species can originate hybrids with different partners.

The analysis of the *P. cactophila* type strain revealed that this lineage is the result of a hybridization event of a *P. inconspicua* hybrid and a yet unknown lineage. This shows that, contrary to what was previously suggested (Guitard et al. 2015), *P. inconspicua* and *P. cactophila* are not the same species. The conclusion regarding the origin of *P. cactophila* was taken based on a whole-genome analysis and on a recombination event found in the *MAT* locus of this species, as it presents a recombinant *MAT* $\alpha 2$ of *P. inconspicua* in the *MAT a* allele of the same species. Of note, such an event has previously been described in *C. metapsilosis* hybrids (Pryszcz et al. 2015), thus suggesting that gene conversion in the *MAT* locus may represent an important advantage for hybrid lineages. Indeed, changes in the *MAT* locus of hybrid yeasts has been shown to be important to restore fertility. For instance, the loss of the entire chromosome III (harboring the *MAT* locus) of one of the parental subgenomes of *Saccharomyces* hybrids was shown to break the sterility barrier (Sipiczki et al. 2020), while in *Zygosaccharomyces* the disruption of the *MAT* locus restores hybrid's fertility (Ortiz-Merino et al. 2017). Therefore, we hypothesize that, although only covering the *MAT* locus, the recombination event we found allowed a *P. inconspicua* strain to restore fertility and mate with *P. cactophila* alternative parent. Given the diploid state of *P. cactophila* we also hypothesize that after the recombination in the *MAT* locus, the *P. inconspicua* parent of *P. cactophila* restored the haploid state. Nevertheless, we cannot exclude the possibility that ploidy reduction occurred after the hybridization event giving origin to *P. cactophila*. The contribution of such a recombination in the *MAT* locus for the possible restoration of fertility and the haploid state before hybridization is still unknown. Thus, addressing the contribution of such an event for hybrids' fertility and the emergence of new lineages is an important future step to understand their impact on genetic isolation, and, hence, speciation.

As an attempt to determine the alternative parental lineage of *P. cactophila*, we analyzed the genome of *P. norvegensis*. This analysis revealed that this species is not the parent of *P. cactophila*, and instead it represents an additional hybrid lineage. However, despite its possible hybrid nature, *P. norvegensis* has possibly undergone a massive LOH through chromosome recombination (and not chromosome loss), resembling the parallel and independent massive losses of heterozygosity described for *Candida africana* and *Candida stellatoidea*, two lineages which share a hybrid ancestor with *C. albicans* (Mixão & Gabaldón 2020, Mixão et al. 2021b). Alternatively, if LOH is taken as a proxy for the relative age of hybrid lineages (Schröder et al. 2016), *P. norvegensis* hybridization is likely much older than any of the other reported hybridization events of the *P. cactophila* species complex.

Altogether, these results show a high propensity of species of the *P. cactophila* species complex to hybridize and originate lineages with medical relevance. From the fourteen isolates of the complex which were analyzed in the past or in this study all of them are hybrids, showing the important role of hybridization on the evolution of the clade. The three new hybrid lineages here described are part of a growing number of yeast hybrids that are described using next-generation sequencing data, even in the absence of known parental lineages (Pryszcz et al. 2014, 2015, Hagen et al. 2015, Mixão & Gabaldón 2018, Mixão et al. 2019, Gabaldón 2020, Mixão & Gabaldón 2020, Mixão et al. 2021b). This is only possible due to the characteristic genomic patterns of hybrid genomes, which are only identified when looking at the whole genome level, and not at a single

gene. For instance, the analysis of ITS showed a high genetic similarity between *P. inconspicua* and *P. cactophila*, leading to the proposal that they should be considered as the same species (Borman & Johnson 2020, Guitard et al. 2015). Although the species concept is difficult to be applied in a context of non-vertical evolution, we consider of extreme relevance the distinction between these lineages, as they present different phenotypic traits (Fig. 3, Table 2). Indeed, both lineages were thought to be associated with different environments (Starmer et al. 1978, Sandven et al. 1997, Moraes et al. 2005, Czaika et al. 2013, Guitard et al. 2013), and perhaps this is the reality. However, if a similar species name is used, we will lose the power to distinguish their geographical and host distribution. A similar scenario is found for the CBS 16747 strain, which due to its high similarity to *P. inconspicua* in highly conserved genetic regions, such as ITS, was classified as a member of this species. Nevertheless, both the genomic analysis and the phenotypic assays performed in this study showed that this isolate is significantly different from *P. inconspicua*, which led us to propose its ranking as a new species named *Pichia alaskaensis*. The distinction between the species of this complex is particularly relevant in the clinical setting, as for instance it would be relevant to distinguish opportunistic infections caused by *P. inconspicua* and *P. cactophila* for the assessment of their potential as emerging pathogens. Therefore, there is an urgent need for proper guidelines for lineage classification in a context of hybridization.

ACKNOWLEDGEMENTS

This work was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement N° H2020-MSCA-ITN-2014-642095. TG group also acknowledges support from the Spanish Ministry of Economy, Industry, and Competitiveness (MEIC) for grant PGC2018-099921-B-I00 co-founded by European Regional Development Fund (ERDF); from the CERCA Programme/Generalitat de Catalunya; from the Catalan Research Agency (AGAUR) SGR857; and grants from the European Union's Horizon 2020 research and innovation programme under the grant agreements ERC-2016-724173, and MSCA-747607. TG also receives support from an INB Grant (PT17/0009/0023 – ISCIII-SGEFI/ERDF). TB acknowledges support of the Distinguished Scientist Fellow program of King Saud University, Riyadh, Saudi Arabia. We would like to thank Elaine Francisco for generating the antifungal susceptibility data, and Dr. Benjamin Westley for his help in coordinating shipment of the clinical strain.

REFERENCES

- Arnold ML. 1997 – Natural Hybridization and Evolution. Oxford University Press, USA.
- Astvad KMT, Thyssen, Arikan-Akdagli S, Arendrup MC. 2022 – A Pragmatic Approach to Susceptibility Classification of Yeasts without EUCAST Clinical Breakpoints. *Journal of Fungi* (Basel, Switzerland) 8(2): 141. Doi 10.3390/jof8020141.
- Bankevich A, Nurk S, Antipov D, Gurevich AA et al. 2012 – SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 19(5): 455–77.
- Boekhout T, Aime MC, Begerow D, Gabaldón T et al. 2021 – The Evolving Species Concepts Used for Yeasts: From Phenotypes and Genomes to Speciation Networks. *Fungal Diversity* 109 (1): 27–55.
- Bolger AM, Lohse M, Usadel B. 2014 – Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30 (15): 2114–20.
- Borman AM, Johnson EM. 2020 – Name Changes for Fungi of Medical Importance, 2018–2019. *Journal of Clinical Microbiology* 59(2): e01811-20. Doi 10.1128/JCM.01811-20
- Borman AM, Muller J, Walsh-Quantick J, Szekely A et al. 2019 – Fluconazole Resistance in Isolates of Uncommon Pathogenic Yeast Species from the United Kingdom. *Antimicrobial Agents and Chemotherapy* 63(8): e00211-19. Doi 10.1128/AAC.00211-19.

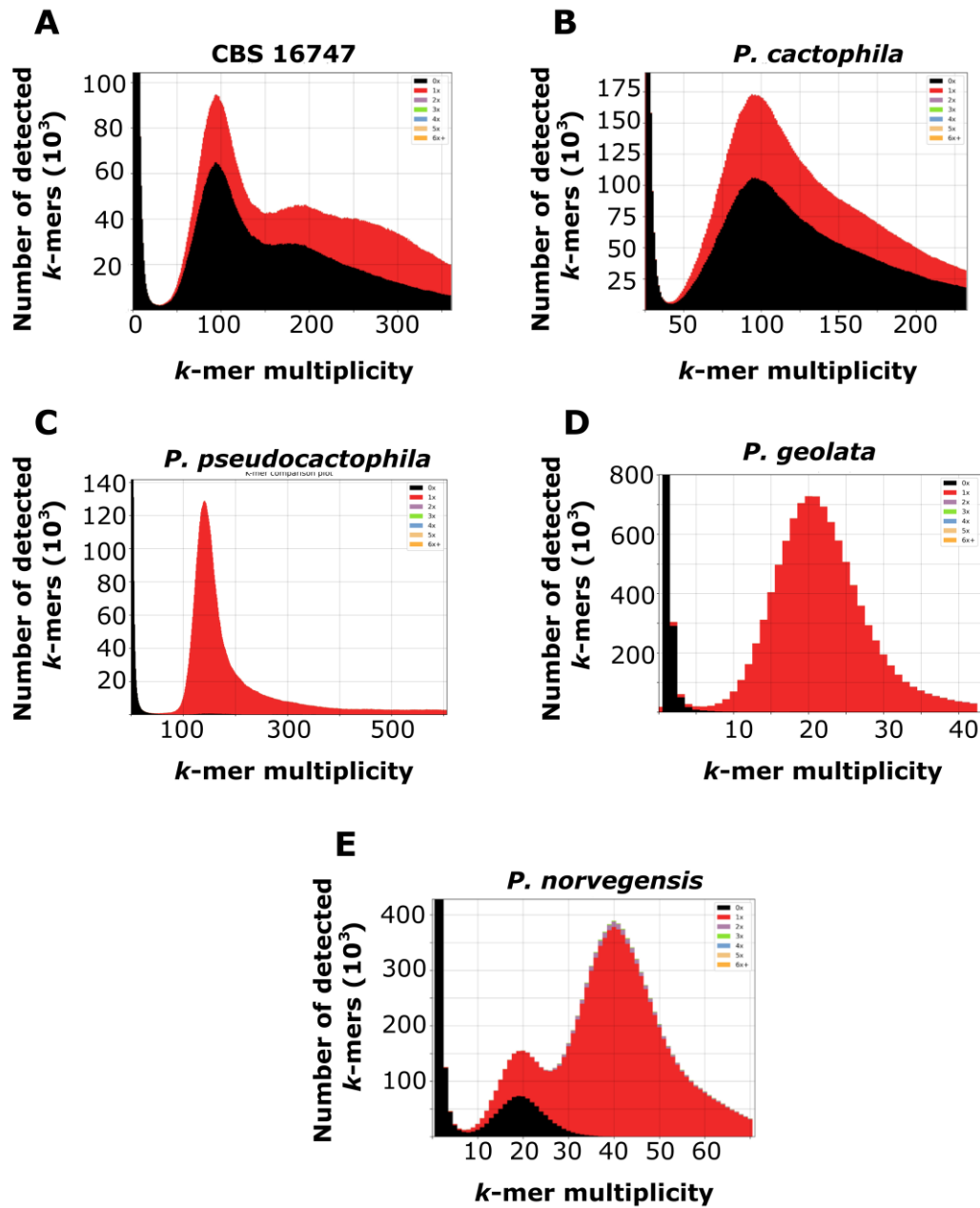
- Caldwell AT, Gabaldón T, Mixão V, Doherty J et al. 2024 – An Inconspicuous Identification: Isolation and Identification of a Novel *Pichia* Species Presenting as Fungemia Following Cardiac Surgery. *International Journal of Infectious Diseases* 143: 107040.
- Capella-Gutierrez S, Kauff F, Toni T. 2014 – A Phylogenomics Approach for Selecting Robust Sets of Phylogenetic Markers. *Nucleic Acids Research* 42(7): e54.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009 – trimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses. *Bioinformatics* 25 (15): 1972–73.
- Czaika V, Nenoff P, Glöckner A, Fegeler W et al. 2013 – Epidemiology and Changes in Patient – Related Factors from 1997 to 2009 in Clinical Yeast Isolates Related to Dermatology, Gynaecology, and Paediatrics. *International Journal of Microbiology* 2013 (December): 703905.
- Del Olmo V, Mixão V, Fotedar R, Saus E et al. 2023 – Origin of Fungal Hybrids with Pathogenic Potential from Warm Seawater Environments. *Nature Communications* 14(1): 6919.
- Douglass AP, Offei B, Braun-Galleani S, Coughlan AY et al. 2018 – Population Genomics Shows No Distinction between Pathogenic *Candida krusei* and Environmental *Pichia kudriavzevii*: One Species, Four Names. *PLoS Pathogens* 14(7): e1007138.
- Edge P, Bafna V, Bansal V. 2017 – HapCUT2: Robust and Accurate Haplotype Assembly for Diverse Sequencing Technologies. *Genome Research* 27 (5): 801–812. Doi 10.1101/gr.213462.116.
- Fuentes D, Molina M, Chorostecki U, Capella-Gutiérrez S et al. 2022 – PhylomeDB V5: An Expanding Repository for Genome-Wide Catalogues of Annotated Gene Phylogenies. *Nucleic Acids Research* 50 (D1): D1062–68.
- Gabaldón T. 2020 – Hybridization and the Origin of New Yeast Lineages. *FEMS Yeast Research* 20 (5): foaa040. Doi 10.1093/femsyr/foaa040
- Ganter PF, Cardinali G, Boundy-Mills K. 2010 – *Pichia Insulana* sp. nov., a Novel Cactophilic Yeast from the Caribbean. *International Journal of Systematic and Evolutionary Microbiology* 60(Pt 4): 1001–7.
- Gladieux P, Ropars J, Badouin H, Branca A et al. 2014 – Fungal Evolutionary Genomics Provides Insight into the Mechanisms of Adaptive Divergence in Eukaryotes. *Molecular Ecology* 23(4): 753–73.
- Gonçalves C, Harrison M-C, Steenwyk JL, Opulente DA et al. 2023 – Diverse signatures of convergent evolution in cacti-associated yeasts. *bioRxiv* 2023.09.14.557833.
- Gueho E, Phaff HJ. 1983 – *Pichia norvegensis*, Teleomorph of *Candida norvegensis*: Validation by Chromosomal Homology. *Bulletin de La Societe Francaise de Mycologie Medicale* 12 (2): 203–7.
- Guitard JRA, Brossas JY, Meyer I, Gits M et al. 2015 – *Candida inconspicua* and *Candida norvegensis*: New Insights into Identification in Relation to Sexual Reproduction and Genome Organization. *Journal of Clinical Microbiology* 53 (5): 1655–61.
- Guitard J, Angoulvant A, Letscher-Bru V, L'Ollivier C et al. 2013 – Invasive Infections due to *Candida norvegensis* and *Candida inconspicua*: Report of 12 Cases and Review of the Literature. *Medical Mycology: Official Publication of the International Society for Human and Animal Mycology* 51(8): 795–99.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013 – QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* 29(8): 1072–75.
- Hagen F, Khayhan K, Theelen B, Kolecka A et al. 2015 – Recognition of Seven Species in the *Cryptococcus gattii/Cryptococcus neoformans* Species Complex. *Fungal Genetics and Biology: FG & B* 78(May): 16–48.
- Jamiu AT, Albertyn J, Sebolai OM, Pohl CH. 2021 – Update on *Candida krusei*, a Potential Multidrug-Resistant Pathogen. *Medical Mycology: Official Publication of the International Society for Human and Animal Mycology* 59(1): 14–30.

- Katoh K, Standley DM. 2013 – MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 30(4): 772–80.
- Kidd SE, Abdolrasouli A, Hagen F. 2023 – Fungal Nomenclature: Managing Change Is the Name of the Game. *Open Forum Infectious Diseases* 10(1): ofac559.
- Krause DJ, Kominek J, Opulente DA, Shen X-X et al. 2018 – Functional and Evolutionary Characterization of a Secondary Metabolite Gene Cluster in Budding Yeasts. *Proceedings of the National Academy of Sciences of the United States of America* 115(43): 11030–35.
- Kurtzman C, Fell JW, Boekhout T. 2011 – *Pichia* E.C. Hansen (1904). In *The Yeasts: A Taxonomic Study*, 2: 685–707. Amsterdam: Elsevier.
- Kurtzman CP, Fell JW, Boekhout T, Robert V. 2011 – Methods for Isolation, Phenotypic Characterization and Maintenance of Yeasts. In *The Yeasts*, 87–110. Elsevier.
- Kurtzman CP, Robnett CJ, Basehoar-Powers E. 2008 – Phylogenetic Relationships among Species of *Pichia*, *Issatchenkia* and *Williopsis* Determined from Multigene Sequence Analysis, and the Proposal of *Barnettozyma* gen. nov., *Lindnera* gen. nov. and *Wickerhamomyces* gen. nov. *FEMS Yeast Research* 8(6): 939–54.
- Li H. 2013 – Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. arXiv 1303.3997. <https://arxiv.org/abs/1303.3997>
- Li H, Handsaker B, Wysoker A, Fennell T et al. 2009 – The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* 25(16): 2078–79.
- Luo R, Liu B, Xie Y, Li Z et al. 2012 – SOAPdenovo2: An Empirically Improved Memory-Efficient Short-Read de Novo Assembler. *GigaScience* 1(1): 18.
- Mapleson D, Accinelli GD, Kettleborough G, Wright J, Clavijo BJ. 2017 – KAT: A K-Mer Analysis Toolkit to Quality Control NGS Datasets and Genome Assemblies. *Bioinformatics* 33(4): 574–76.
- McClintock B. 1984 – The Significance of Responses of the Genome to Challenge. *Science* 226 (4676): 792–801.
- McKenna A, Hanna M, Banks E, Sivachenko A et al. 2010 – The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data. *Genome Research* 20(9): 1297–1303.
- Mixão V, Gabaldón T. 2018 – Hybridization and Emergence of Virulence in Opportunistic Human Yeast Pathogens. *Yeast* 35(1): 5–20.
- Mixão V, Gabaldón T. 2020 – Genomic Evidence for a Hybrid Origin of the Yeast Opportunistic Pathogen *Candida albicans*. *BMC Biology* 18(1): 48.
- Mixão V, Hansen AP, Saus E, Boekhout T et al. 2019 – Whole-Genome Sequencing of the Opportunistic Yeast Pathogen *Candida inconspicua* Uncovers Its Hybrid Origin. *Frontiers in Genetics* 10: 383. Doi 10.3389/fgene.2019.00383
- Mixão V, Hegedúsová E, Saus E, Prysycz LP, et al. 2021a – Genome Analysis of *Candida subhashii* Reveals Its Hybrid Nature and Dual Mitochondrial Genome Conformations. *DNA Research* 28(3): dsab006. Doi 10.1093/dnares/dsab006
- Mixão V, Nunez-Rodriguez JC, Del Olmo V, Ksiezopolska E et al. 2023 – Evolution of Loss of Heterozygosity Patterns in Hybrid Genomes of *Candida* Yeast Pathogens. *BMC Biology* 21(1): 105.
- Mixão V, Saus E, Boekhout T, Gabaldón T. 2021b – Extreme Diversification Driven by Parallel Events of Massive Loss of Heterozygosity in the Hybrid Lineage of *Candida albicans*. *Genetics* 217 (2): iyaa004. Doi 10.1093/genetics/iyaa004
- Moraes ME, Rosa CA, Sene FM. 2005 – Preliminary Notes on Yeasts Associated with Necrotic Cactus Stems from Different Localities in Brazil. *Brazilian Journal of Biology = Revista Brasileira de Biologia* 65(2): 299–304.
- Morales L, Dujon B. 2012 – Evolutionary role of interspecies hybridization and genetic exchanges in yeasts. *Microbiology and Molecular Biology Reviews* 76(4): 721–39.

- O'Brien CE, Oliveira-Pacheco J, Cinnéide EO, Haase MAB et al. 2021 – Population Genomics of the Pathogenic Yeast *Candida tropicalis* Identifies Hybrid Isolates in Environmental Samples. *PLoS Pathogens* 17(3): e1009138.
- O'Brien CE, Zhai B, Ola M, Bergin SA et al. 2022 – Identification of a Novel *Candida metapsilosis* Isolate Reveals Multiple Hybridization Events. *G3* 12(1): jkab367. Doi 10.1093/g3journal/jkab367.
- Opulente DA, Langdon QK, Jarzyna M, Buh KV et al. 2023 – Taxogenomic Analysis of a Novel Yeast Species Isolated from Soil, *Pichia galeolata* sp. nov. *Yeast* 40(12): 608–615. Doi 10.1002/yea.3905
- Ortiz-Merino RA, Kuanyshev N, Braun-Galleani S, Byrne KP et al. 2017 – Evolutionary Restoration of Fertility in an Interspecies Hybrid Yeast, by Whole-Genome Duplication after a Failed Mating-Type Switch. *PLoS Biology* 15 (5): e2002128. Doi 10.1371/journal.pbio.2002128
- Payseur BA, Rieseberg LH. 2016 – A Genomic Perspective on Hybridization and Speciation. *Molecular Ecology* 25(11): 2337–60.
- Pegueroles C, Mixão V, Carreté L, Molina M, Gabaldón T. 2020 – HaploTypo: A Variant-Calling Pipeline for Phased Genomes. *Bioinformatics* 36 (8): 2569–71.
- Pfaller MA, Diekema DJ, Gibbs DL, Newell VA et al. 2010 – Results from the ARTEMIS DISK Global Antifungal Surveillance Study, 1997 to 2007: A 10.5-Year Analysis of Susceptibilities of *Candida* Species to Fluconazole and Voriconazole as Determined by CLSI Standardized Disk Diffusion. *Journal of Clinical Microbiology* 48(4): 1366–77.
- Pfaller MA, Diekema DJ, Gibbs DL, Newell VA et al. 2010 – Geographic Variation in the Frequency of Isolation and Fluconazole and Voriconazole Susceptibilities of *Candida glabrata*: An Assessment from the ARTEMIS DISK Global Antifungal Surveillance Program. *Diagnostic Microbiology and Infectious Disease* 67(2): 162–71.
- Pryszcz LP, Gabaldón T. 2016 – Redundans: An Assembly Pipeline for Highly Heterozygous Genomes. *Nucleic Acids Research* 44(12): e113.
- Pryszcz LP, Németh T, Gácsér A, Gabaldón T. 2014 –Genome Comparison of *Candida orthopsilosis* Clinical Strains Reveals the Existence of Hybrids between Two Distinct Subspecies. *Genome Biology and Evolution* 6(5): 1069–78.
- Pryszcz LP, Németh T, Saus E, Ksiezopolska E et al. 2015 – The Genomic Aftermath of Hybridization in the Opportunistic Pathogen *Candida metapsilosis*. *PLoS Genetics* 11(10): e1005626.
- Quinlan AR, Hall IM. 2010 – BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* 26(6): 841–42.
- Safonova Y, Bankevich A, Pevzner PA. 2015 – dipSPAdes: Assembler for Highly Polymorphic Diploid Genomes. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 22(6): 528–45.
- Sandven PKN, Digranes A, Tjade T, Lassen J. 1997 – *Candida norvegensis*: A Fluconazole-Resistant Species. *Antimicrobial Agents and Chemotherapy* 41(6): 1375–76.
- Schröder MS, Martínez de San Vicente K, Prandini THR, Hammel S et al. 2016 – Multiple Origins of the Pathogenic Yeast *Candida orthopsilosis* by Separate Hybridizations between Two Parental Species. *PLoS Genetics* 12(11): e1006404.
- Shen X-X, Opulente DA, Kominek J, Zhou X et al. 2018 – Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell* 175(6): 1533–45.e20.
- Sipiczki M. 2018 – Interspecies Hybridisation and Genome Chimerisation in *Saccharomyces*: Combining of Gene Pools of Species and Its Biotechnological Perspectives. *Frontiers in Microbiology* 9: 3071.
- Sipiczki M, Antunovics Z, Szabo A. 2018 – *MAT* heterozygosity and the second sterility barrier in the reproductive isolation of *Saccharomyces* species. *Current Genetics* 66(5): 957–969.
- Stamatakis A. 2014 – RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* 30(9): 1312–13.

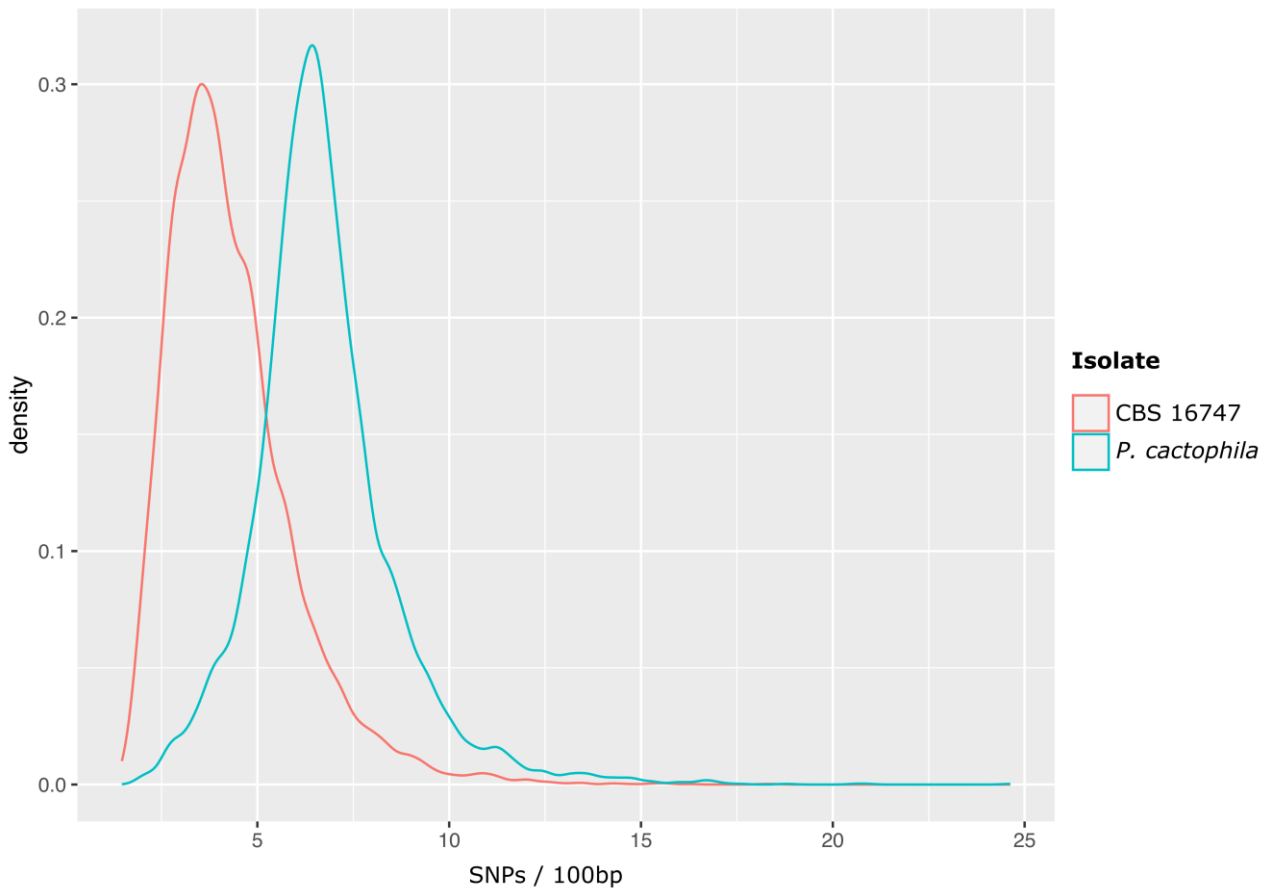
- Stanke M, Morgenstern B. 2005 – AUGUSTUS: A Web Server for Gene Prediction in Eukaryotes That Allows User-Defined Constraints. *Nucleic Acids Research* 33 (Web Server issue): W465–7. Doi 10.1093/nar/gki458
- Starmer WT, Phaff HJ, Miranda M, Miller MW. 1978 – *Pichia cactophila*, a New Species of Yeast Found in Decaying Tissue of Cacti. *International Journal of Systematic Bacteriology* 28(2): 318–25.
- Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013 – Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration. *Briefings in Bioinformatics* 14(2): 178–92.
- Waterhouse RM, Seppey M, Simão FA, Zdobnov EM. 2019 – Using BUSCO to Assess Insect Genomic Resources. *Methods in Molecular Biology* 1858: 59–74.
- Weiß CL, Pais M, Cano LM, Kamoun S, Burbano HA. 2018 – nQuire: A Statistical Framework for Ploidy Estimation Using next Generation Sequencing. *BMC Bioinformatics* 19(1): 122.
- Zhu H-Y, Huo L-C, Hu S, Wei Y-H et al. 2024 – *Pichia kurtzmaniana* f.a. sp. nov., with the transfer of eight *Candida* species to *Pichia*. *International Journal of Systematic and Evolutionary Microbiology* 74: 006306.

Supplementary materials

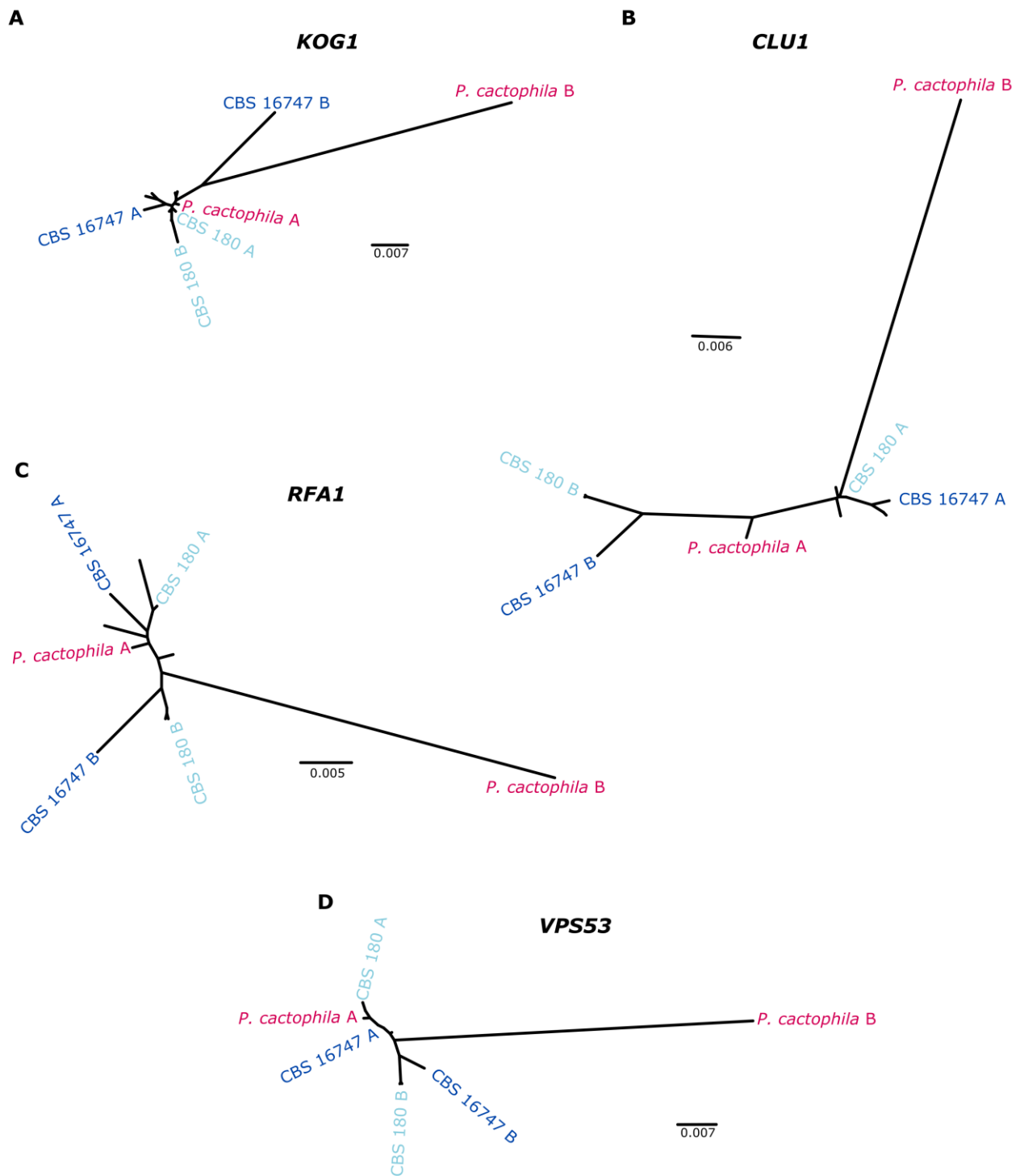


Supplementary Figure 1 – K-mer plots of the newly sequenced isolates, indicating the number of different k -mers across the genome coverage. Red represents the k -mers also observed in *P. inconspicua* reference genome and black represents the absent ones. A) CBS 16747 (*P. alaskaensis* sp. nov.). B) *P. cactophila*. C) *P. pseudocactophila*. D) *P. galeolata*. E) *P. norvegensis*.

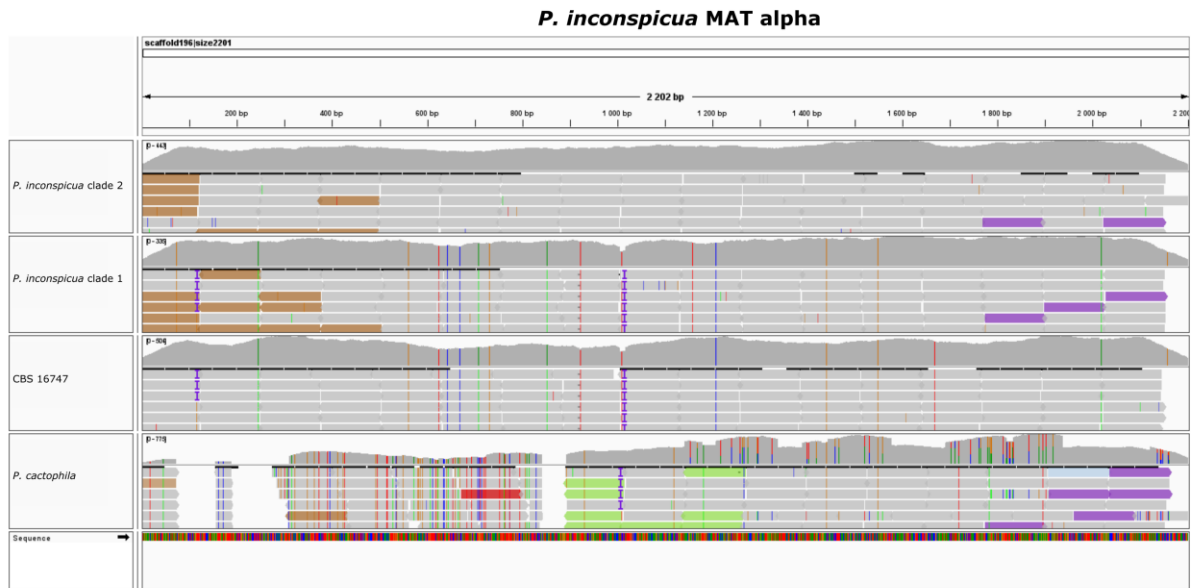
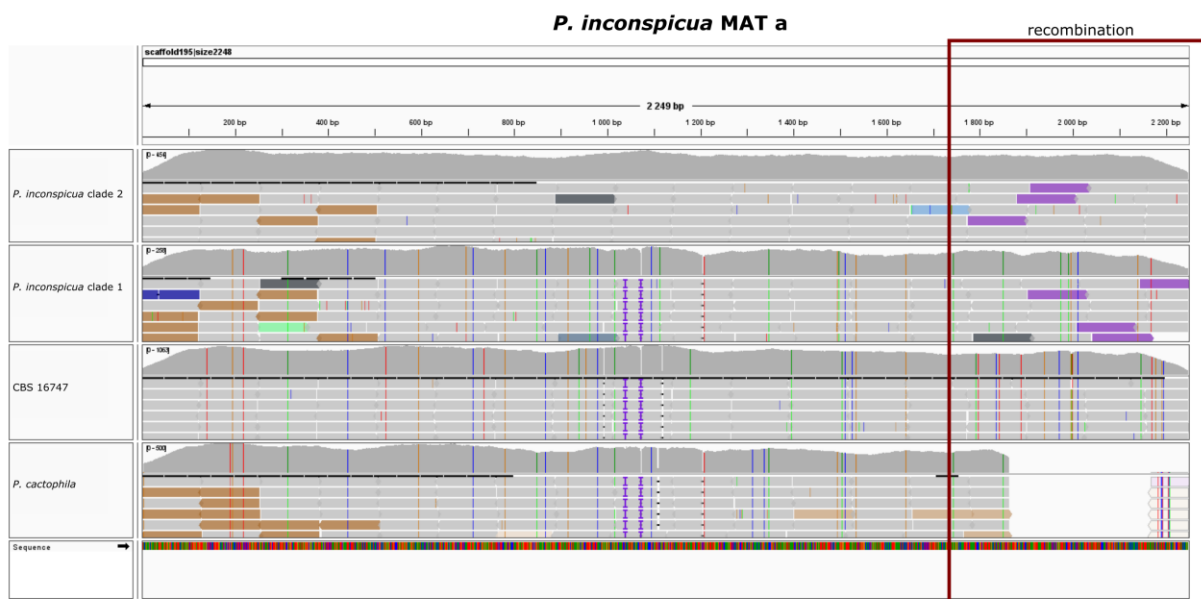
Density of the number of heterozygous SNPs per 100 bp in the heterozygous blocks



Supplementary Figure 2 – Density of the number of heterozygous SNPs per 100 bp in the heterozygous blocks determined for CBS 16747 (*P. alaskaensis* sp. nov.) and *P. cactophila* type strain when aligned to *P. inconspicua* type strain.

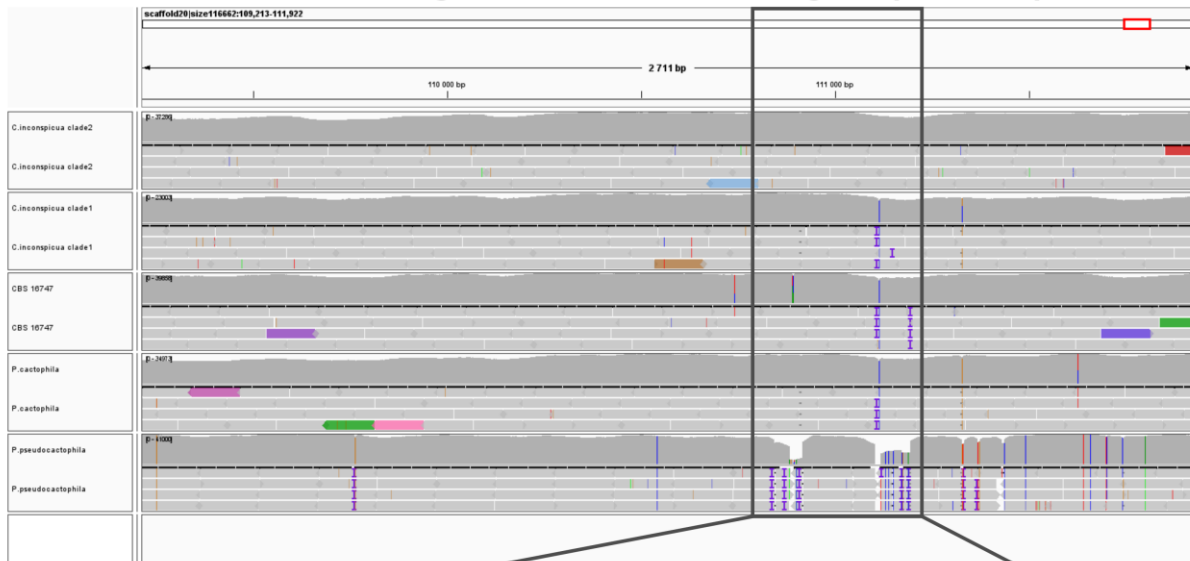


Supplementary Figure 3 – Phylogenetic tree reconstruction of the alignment of KOG1 (A), CLU1 (B), RFA1 (C) and VPS53 (D) phased haplotypes of *P. inconspicua* (CBS 180), CBS 16747 (*P. alaskaensis* sp. nov.) and *P. cactophila* type strain. For clearness reasons, CBS 180 is the only *P. inconspicua* strain indicated.

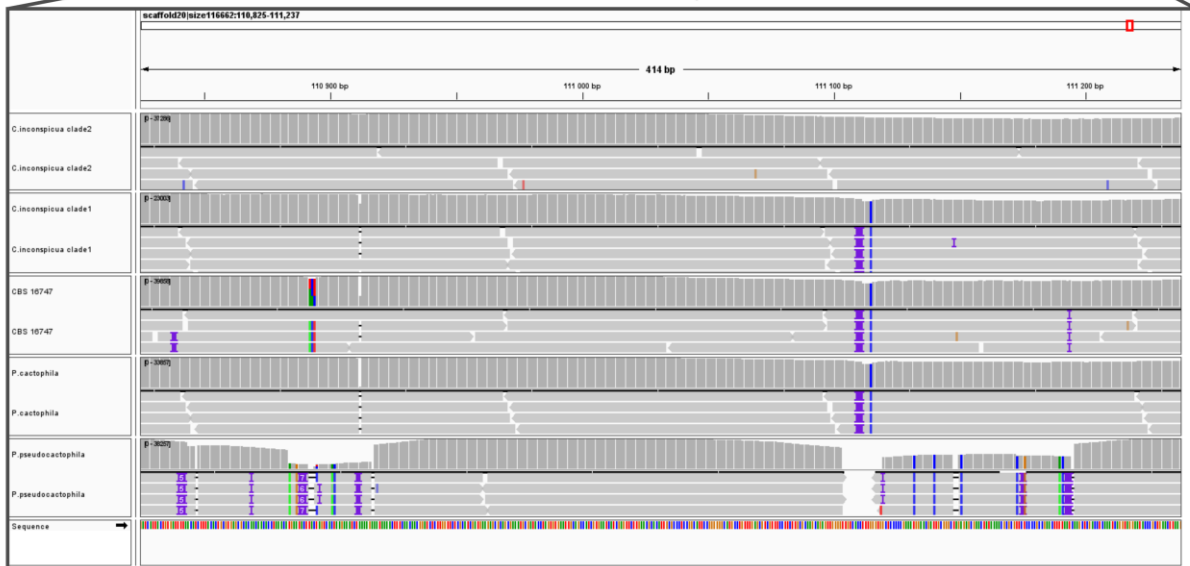
A**B**

Supplementary Figure 4 – IGV screenshot of the alignment of the library of *P. inconspicua* clade 2 (represented by the type strain CBS 180), *P. inconspicua* clade 1 (represented by 14ANR23920), CBS 16747 (*P. alaskaensis* sp. nov.) and *P. cactophila* on the MAT alpha (A) and MAT a (B) regions of *P. inconspicua* reference genome. In this figure, it can be seen the similarity of CBS 16747 MAT alpha to that of *P. inconspicua* clade 1 and their differences in MAT a. This figure also shows the differences between *P. cactophila* and *P. inconspicua* in the MAT alpha and their similarities in MAT a. The recombination event in the MAT a of *P. cactophila* is highlighted in red.

Alignment from 18S small subunit ribosomal RNA gene (partial sequence) until the 26S large subunit ribosomal RNA gene, partial sequence



Zoom-in ITS1 sequence



Supplementary Figure 5 – IGV screenshot of the alignment of the library of *P. inconspicua* clade 2 (represented by the type strain CBS 180), *P. inconspicua* clade 1 (represented by 14ANR23920), CBS 16747 (*P. alaskaensis* sp. nov.), *P. cactophila* and *P. pseudocactophila* on the region that goes from the 18S small subunit ribosomal RNA gene (partial sequence) until the 26S large subunit ribosomal RNA gene (partial sequence).

```

PICCA_alpha2_a2 ACAAACCTTATTGTTGTTTCACAAAAAATCTGCCACAAGAGTTCAGGATGTCCTTTATT
CANINC_alpha2  -----
CANINC_A2      ACAAACCTTATTGTTGTTTCACAAAAAATCTGCCACAAGAGTTCAGGATGTCCTTTATT

PICCA_alpha2_a2 TCTCAGGATCAATGACTTAAACTACATATTAACCTTTCATGTTTATGAGGATTCGTTTTTC
CANINC_alpha2  -----
CANINC_A2      TCTCAGGATCAA-----

PICCA_alpha2_a2 TGAATATAAGCTTTTCTAGGTAATCTCATAACGTAGGGCAGAAATAAAATTTGAAGCAA
CANINC_alpha2  -----
CANINC_A2      -----

PICCA_alpha2_a2 CTTTCGCGCAATAATAGTTTATCAACCTCAATTATTTCCCATCTTAATATTATTGCTGG
CANINC_alpha2  -----
CANINC_A2      -----

PICCA_alpha2_a2 CAGCGCGCAAGCTGCACTAGAAAGTTGATTATGAAACATACCCGAAAACAGGAACTGGAA
CANINC_alpha2  -----
CANINC_A2      -----

PICCA_alpha2_a2 TAATCTTTTGTATACCGTAAGCCCTCGTACTTTCTAGAGGAGTGTGATTAGAATACTGT
CANINC_alpha2  -----
CANINC_A2      -----

PICCA_alpha2_a2 GATATGCATTACCAGGAAGAGTAAGTTATAGGTTGTATCCTACGACAAGTTTTCAGTTAC
CANINC_alpha2  -----
CANINC_A2      -----

PICCA_alpha2_a2 TAACACTAACAAGATTGACAGAGATTACATTCATTTGAAGGATGTTTACGAGACTTAT
CANINC_alpha2  -----
CANINC_A2      -----

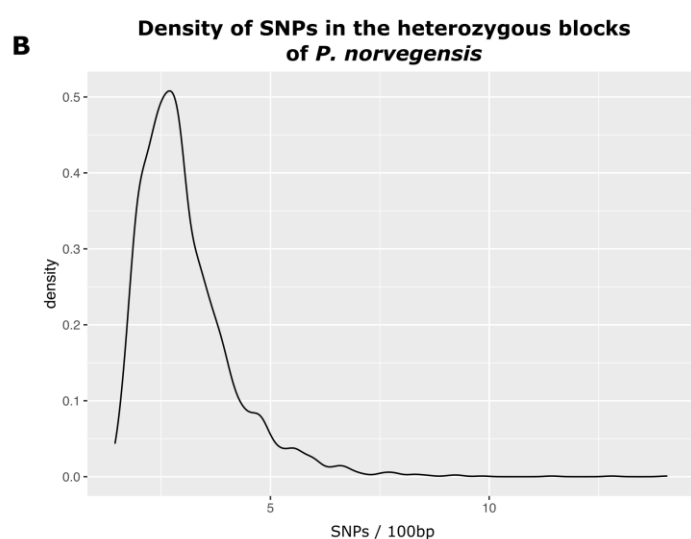
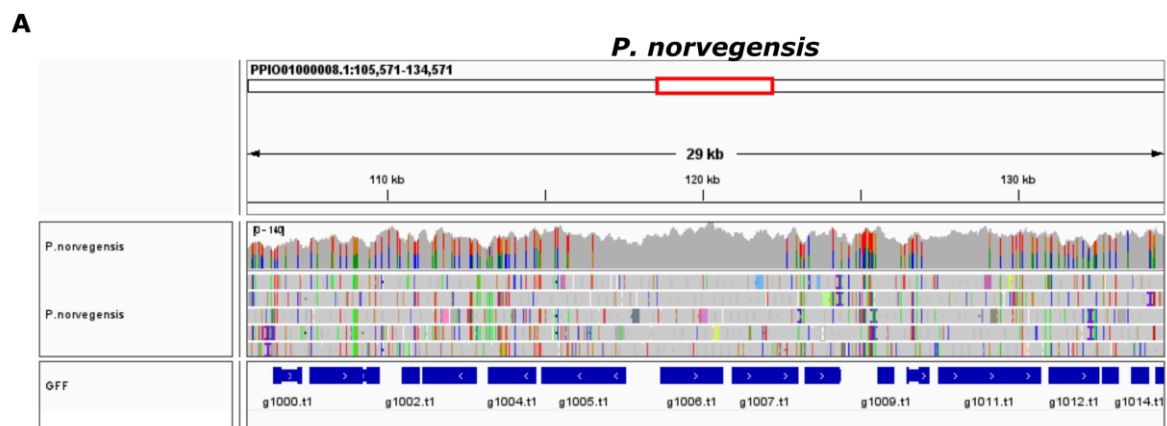
PICCA_alpha2_a2 CTAAAACCCCAATAGATGAAGAGTTCAAGACAGCACACTTGTGATGACATGTTAGTAT
CANINC_alpha2  -----ATGTTAGTAT
CANINC_A2      -----

PICCA_alpha2_a2 GTATCAAATTTGATTTTTTTTGACGAGTTCAACAATTAACCTTCACTGAAGAGTGATA
CANINC_alpha2  GTATCAAATTTGATTTTTTTTGACGAGTTCAACAATTAACCTTCACTGAAGAGTGATA
CANINC_A2      -----

PICCA_alpha2_a2 CTAACGTTGTTTAAATATAGAAGTTGAAGGAACTTTGAGTGGCACAATCGTCGACGAGCA
CANINC_alpha2  CTAACGTTGTTTAAATATAGAAGTTGAAGGAACTTTGAGTGGCACAATCGTCGACGAGCA
CANINC_A2      -----

```

Supplementary Figure 6 – Screenshot of a portion of the sequence alignment of *P. cactophila* MAT a that comprises the recombinant site between MAT a2 and MAT alpha2 (PICCA_alpha2_a2) and the MAT a2 (CANINC_a2) and alpha2 (CANINC_alpha2) sequences of *P. inconspicua*. The recombinant site is highlighted in pink.



Supplementary Figure 7 – Genomic patterns of *P. norvegensis*. A) IGV screenshot of *P. norvegensis* genome, showing the occurrence of heterozygous polymorphisms separated by what possibly are blocks of LOH. B) Density of the number of heterozygous SNPs per 100 bp in the heterozygous blocks determined for *P. norvegensis* when aligned to its own genome assembly.

Supplementary Table 1 Summary statistics of the genome assemblies of *P. cactophila* type strain, *P. pseudocactophila* type strain and *P. alaskaensis* sp. nov. (CBS 16747).

	<i>P. cactophila</i>	<i>P. pseudocactophila</i>	<i>P. alaskaensis</i>
Contigs	338	158	1165
Length (bp)	11,375,052	10,018,088	10,261,017
N50 (bp)	79,972	213,179	14,955
GC (%)	34.79	35.84	34.95
NCBI accession	GCA_039700745.1	GCA_039700735.1	GCA_040938015.1