



# Whole-Genome Sequencing of the Opportunistic Yeast Pathogen *Candida inconspicua* Uncovers Its Hybrid Origin

Verónica Mixão<sup>1,2</sup>, Antonio Perez Hansen<sup>3</sup>, Ester Saus<sup>1,2</sup>, Teun Boekhout<sup>4,5</sup>, Cornelia Lass-Flörl<sup>3</sup> and Toni Gabaldón<sup>1,2,6\*</sup>

<sup>1</sup> Centre for Genomic Regulation, Barcelona Institute of Science and Technology, Barcelona, Spain, <sup>2</sup> Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain, <sup>3</sup> Division of Hygiene and Medical Microbiology, Innsbruck Medical University, Innsbruck, Austria, <sup>4</sup> Westerdijk Fungal Biodiversity Institute, Utrecht, Netherlands, <sup>5</sup> Institute of Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, Netherlands, <sup>6</sup> Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

## OPEN ACCESS

### Edited by:

Jean Marie François,  
UMR 5504, Laboratoire d'Ingénierie  
des Systèmes Biologiques et des  
Procédés (LISBP), France

### Reviewed by:

Christina A. Cuomo,  
Broad Institute, United States  
Santiago Castillo Ramírez,  
National Autonomous University  
of Mexico, Mexico

### \*Correspondence:

Toni Gabaldón  
toni.gabaldon.bcn@gmail.com

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Genetics

**Received:** 12 February 2019

**Accepted:** 09 April 2019

**Published:** 25 April 2019

### Citation:

Mixão V, Hansen AP, Saus E,  
Boekhout T, Lass-Flörl C and  
Gabaldón T (2019) Whole-Genome  
Sequencing of the Opportunistic  
Yeast Pathogen *Candida inconspicua*  
Uncovers Its Hybrid Origin.  
*Front. Genet.* 10:383.  
doi: 10.3389/fgene.2019.00383

Fungal infections such as those caused by *Candida* species are increasingly common complications in immunocompromised patients. The list of causative agents of candidiasis is growing and comprises a set of emerging species whose relative global incidence is rare but recurrent. This is the case of *Candida inconspicua*, which prevalence has increased 10-fold over the last years. To gain novel insights into the emergence of this opportunistic pathogen and its genetic diversity, we performed whole genome sequencing of the type strain (CBS180), and of 10 other clinical isolates. Our results revealed high levels of genetic heterozygosity structured in non-homogeneous patterns, which are indicative of a hybrid genome shaped by events of loss of heterozygosity (LOH). All analyzed strains were hybrids and could be clustered into two distinct clades. We found large variability across strains in terms of ploidy, patterns of LOH, and mitochondrial genome heterogeneity that suggest potential admixture between hybrids. Altogether, our results identify a new hybrid species with virulence potential toward humans and underscore the potential role of hybridization in the emergence of novel pathogenic lineages.

**Keywords:** *Candida inconspicua*, hybrid, yeast, pathogen, genome

## INTRODUCTION

Fungal infections are an increasingly common problem in hospital environments, very often leading to patient's death (Pfaller and Diekema, 2007; Lass-Flörl, 2009; Brown et al., 2012; Gabaldón and Carreté, 2016). Historically, *Candida* species have been the most common causative agents of hospital-acquired fungal infections (Pfaller and Diekema, 2007; Lass-Flörl, 2009; Brown et al., 2012; Gabaldón and Carreté, 2016). Patients at particular risk include those in the intensive care unit, those undergoing surgery, and patients with solid tumor or hematological malignancy (Lass-Flörl, 2009). Reported mortality rates for candidemia range from 28 to 59% in European surveys and depend on species, underlying disease conditions, and geographical location (Jordà-Marcos et al., 2007; Lass-Flörl, 2009). *Candida albicans* is the most common cause of candidemia, accounting for more than 50% of the cases, followed by *Candida glabrata* and *Candida parapsilosis*

(Holzheimer and Dralle, 2002; Jordà-Marcos et al., 2007; Pfaller and Diekema, 2007). However, the epidemiology of candidemia has shifted in recent years, with the incidence of rare species becoming increasingly important in the clinical setting (Lass-Flörl, 2009; Pfaller et al., 2012; Sardi et al., 2013; Gabaldón and Carreté, 2016; Gabaldón et al., 2016; Bretagne et al., 2017). For instance, *Candida auris* is an emerging multi-drug resistant pathogen responsible for many outbreaks all over the world in the last few years (Forsberg et al., 2019).

*Candida inconspicua* was firstly described as *Torulopsis inconspicua* in Lodder and Kreger-van Rij (1952) and later reclassified in *Candida* (Yarrow and Meyer, 1978). The species belongs to the *Pichia cactophila* clade, together with *Pichia kudriavzevii* [synonym *Candida krusei* (Douglass et al., 2018)], *Pichia norvegensis*, *P. cactophila*, and *Pichia pseudocactophila* (Kurtzman et al., 2011). *C. inconspicua* is genetically similar and phenotypically identical to *P. cactophila* and it has been suggested that they represent different sexual stages of the same species (Kurtzman et al., 2008, 2011; Guitard et al., 2015). *C. inconspicua* has also been misidentified as other members of the clade such as *P. norvegensis* (Majoros et al., 2003; Guitard et al., 2015). Many studies support that *C. inconspicua* can often be found in lactic products, including milk, cheese, or butter (Minervini et al., 2001; Suzzi et al., 2003; Krukowski et al., 2006; Callon et al., 2007; Kurtzman et al., 2011). In addition, it has recently been reported in traditional alcoholic beverages such as oil palm wine and a sorghum beer called *tchapalo* (Egue et al., 2018).

*Candida inconspicua* is also responsible for clinical infections, more prominently in European countries (Majoros et al., 2005; Pfaller et al., 2010; Guitard et al., 2013). A more than 10-fold increase in *C. inconspicua* infections between 1997–2000 and 2001–2004 (increase of 9 to 276 cases) followed by an apparent stabilization has been reported by a multi-center study (Pfaller et al., 2010). The majority of *C. inconspicua* infections are associated with osteomyelitis, oropharyngeal and esophageal candidiasis in HIV positive patients, as well as with candidemia in patients with hematological malignancies (Majoros et al., 2005). Frequently, *C. inconspicua* isolates derive from colonization of the digestive and respiratory tracts from unknown sources (Guitard et al., 2013). However, the above mentioned reported isolations make contaminated milk or other food products a possible source for the infecting strains. *C. inconspicua* was previously described as presenting a low susceptibility to fluconazole and other antifungal agents (Majoros et al., 2005; Pfaller et al., 2010; Guitard et al., 2013). For instance, Pfaller et al. (2010) reported that, depending on the site of isolation, the frequency of fluconazole resistant strains could range between 26.1% (skin and soft tissue) and 62.9% (genital tract), thus indicating a high phenotypic heterogeneity among *C. inconspicua* isolates.

To shed light on the genetic makeup and diversity of this emerging opportunistic pathogen we undertook the whole genome sequencing and assembly of the type strain and compared the genomic sequences of 10 other clinical isolates. We found that *C. inconspicua* has a highly heterozygous genome with patterns suggestive of a hybrid origin. We discuss this finding in comparison with two other medically important *Candida* hybrid lineages: *Candida metapsilosis* and *Candida*

*orthopsilosis* (Pryszcz et al., 2014, 2015; Schröder et al., 2016). Following *C. metapsilosis*, *C. inconspicua* is the second reported case of an opportunistic *Candida* human pathogen for which all clinical strains analyzed so far are hybrids, suggesting that hybridization may be at the root of its ability to infect humans (Mixão and Gabaldón, 2018).

## MATERIALS AND METHODS

### Library Preparation and Genomic DNA Sequencing

For this project we sequenced *C. inconspicua* type strain (CBS180), and 10 other clinical isolates. These isolates were sent by collaborating laboratories in the frame of an antifungal susceptibility test project. They were grown on Sabouraud's 2% dextrose agar at 30°C for 72 h and re-identified using direct extraction method by MALDI-TOF (MALDI-Biotyper, Bruker, Daltonics, Database version, United States). Genomic DNA extraction of the 11 *C. inconspicua* strains was performed using the MasterPure Yeast DNA Purification Kit (Epicentre, United States) following manufacturer's instructions. Briefly, *C. inconspicua* cultures were grown in an orbital shaker overnight (200 rpm, 30°C) in 15 ml of YPD medium. Cells were harvested using 4.5 ml of each culture by centrifugation at maximum speed for 2 min, and then they were lysed at 65°C for 15 min with 300  $\mu$ l of yeast cell lysis solution (containing 1  $\mu$ l of RNase A). After being on ice for 5 min, 150  $\mu$ l of MPC protein precipitation reagent were added into the samples, and they were centrifuged at 16.000 g for 10 min to pellet the cellular debris. The supernatant was transferred to a new tube, DNA was precipitated using 100% cold ethanol and centrifuging the samples at 16.000 g, 30 min, 4°C. The pellet was washed twice with 70% cold ethanol and, once the pellet was dried, the sample was resuspended in 100  $\mu$ l of TE. All gDNA samples were cleaned to remove the remaining RNA using the Genomic DNA Clean & Concentrator kit (Epicentre) according to manufacturer's instructions. Total DNA integrity and quantity of the samples were assessed by means of agarose gel, NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, United States) and Qubit dsDNA BR assay kit (Thermo Fisher Scientific).

Whole-genome sequencing was performed at the Genomics Unit from Centre for Genomic Regulation (CRG) with a HiSeq2500 machine. Libraries were prepared using the NEBNext Ultra DNA Library Prep kit for Illumina (New England BioLabs, United States) according to manufacturer's instructions. All reagents subsequently mentioned are from the NEBNext Ultra DNA Library Prep kit for Illumina if not specified otherwise. 1  $\mu$ g of gDNA was fragmented by nebulization in Covaris to a size of  $\sim$ 600 bp. After shearing, the ends of the DNA fragments were blunted with the End Prep Enzyme Mix, and then NEBNext Adaptors for Illumina were ligated using the Blunt/TA Ligase Master Mix. The adaptor-ligated DNA was cleaned-up using the MinElute PCR Purification kit (Qiagen, Germany) and a further size selection step was performed using an agarose gel. Size-selected DNA was then purified using the QIAGEN Gel Extraction Kit with MinElute columns (Qiagen) and library amplification was performed by PCR with the

NEBNext Q5 Hot Start 2X PCR Master Mix and index primers (12–15 cycles). A further purification step were done using AMPure XP Beads (Agencourt, United States). Final libraries were analyzed using Agilent DNA 1000 chip (Agilent) to estimate the quantity and check size distribution, and they were then quantified by qPCR using the KAPA Library Quantification Kit (KapaBiosystems, United States) prior to amplification with Illumina's cBot. Libraries were loaded and sequenced  $2 \times 125$  on Illumina's HiSeq 2500. Base calling was performed using Illumina pipeline software. In multiplexed libraries, we used 6 bp internal indexes (5' indexed sequences). De-convolution was performed using the CASAVA software (Illumina, United States). Sequence data of the genomes has been deposited in short read archive (SRA) under the BioProject Accession No. PRJNA517794.

## De novo Genome Assembly and Read Mapping

Raw sequencing data was inspected with FastQC v0.11.5<sup>1</sup>. 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). The K-mer Analysis Toolkit (KAT; Mapleson et al., 2017) was used to get the GC content and *k*-mer frequency of CBS180 reads and estimate the expected genome size. SPAdes v3.9 (Bankevich et al., 2012) was used to perform the genome assembly using this strain. Afterward, redundant contigs were removed with Redundans v0.13c (Pryszcz and Gabaldón, 2016) using default parameters, i.e., 51% minimum identity and at least 80% overlap. The quality of the assembly was inspected with Quast v4.5 (Gurevich et al., 2013) and KAT (Mapleson et al., 2017). Genome annotation was performed with Augustus v3.1 (Stanke and Morgenstern, 2005), using *C. albicans* as model organism. BUSCO v3 (Waterhouse et al., 2017) was used to assess the completeness predicted proteome considering the Ascomycota database. This genome assembly and annotation have been deposited at DDBJ/ENA/GenBank under the Accession No. SELW00000000. The version described in this paper is version SELW01000000.

Phylome reconstruction was performed using the PhylomeDB pipeline (Huerta-Cepas et al., 2014) as described by Pryszcz et al. (2015), using the predicted proteome as seed, and considering other twenty-one Saccharomycotina species (Table A in Supplementary file 1). A second phylome considering only proteins predicted in scaffolds > 10 kb was also reconstructed to confirm the obtained results. The presented results correspond to the phylome considering all proteins. *C. inconspicua* phylomes are available in PhylomeDB (Huerta-Cepas et al., 2014) with the ID 454 and 498 (this one only considering scaffolds > 10 kb). Gene gain and loss analysis in seed branch was performed based on the phylome results. A BLASTp (Zhang et al., 2000) was performed against the UniProt database, in order to determine the possible function associated to these genes, as well as their GO terms. An enrichment analysis was done using FatiGO (Al-Shahrour et al., 2004), and the results were summarized with REVIGO (Supek et al., 2011).

Read mapping for all strains (Table 1) was performed with BWA-MEM v0.7.15 (Li, 2013). Picard v2.1.1<sup>2</sup> was used to sort the resulting file by coordinate, as well as to mark duplicates, create the index file, and obtain mapping statistics. Mapping results were inspected with IGV version 2.0.30 (Thorvaldsdóttir et al., 2013). Mapping coverage was determined with SAMtools v0.1.18 (Li et al., 2009).

## Variant Calling and Ploidy Estimation

Samtools v0.1.18 (Li et al., 2009) and Picard v2.1.1<sup>2</sup> were used, respectively, to index the reference and create a dictionary to be used in subsequent variant calling steps. GATK v3.6 (McKenna et al., 2010) was used to call variants with the tool HaplotypeCaller set with `-genotyping_mode DISCOVERY -stand_emit_conf 10 -stand_call_conf 30 -ploidy 2 -nct 8`. The tool VariantFiltration of the same program was used to filter the vcf files with the following parameters: `-clusterSize 5 -clusterWindowSize 20 -genotypeFilterName "heterozygous" -genotypeFilterExpression "isHet = =1" -filterName "bad_quality" -filter "QD < 2.0 || MQ < 40 || FS > 60.0 || HaplotypeScore > 13.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0" -filterExpression "DP < = 20" -filterName "DepthofQuality."` In order to determine the number of SNPs/kb, a file containing only SNPs was generated with the SelectVariants tool. Moreover, for this calculation only positions in the reference with 20 or more reads were considered for the genome size, and these were determined with bedtools genomcov v2.25.0 (Quinlan and Hall, 2010).

To estimate the ploidy of each strain, nQuire histotest (Weiß et al., 2018) was used to test which distribution (diploid, triploid, or tetraploid) fits better to the variant frequency data. Given that for some of the strains the results were not clear, the allele frequency of each heterozygous variant was calculated by dividing the number of reads supporting the alternative haplotype by the total of reads mapping at that position. Allele frequency density was plot for each strain considering only scaffolds with more than 100 kb.

## Determination of the Different Hybridization Events and Parental Divergence

To determine heterozygous and loss of heterozygosity (LOH) blocks, the procedure applied and validated by Pryszcz et al. (2015) was used. Briefly, bedtools merge v2.25.0 (Quinlan and Hall, 2010) with a window of 100 bp was used to define heterozygous regions, and by opposite, LOH blocks would be all non-heterozygous regions in the genome. The minimum LOH and heterozygous block size was established at 100 bp. Due to the aneuploidies observed for *C. inconspicua*, contrary to what Pryszcz and colleagues performed, no coverage filter was applied.

As mentioned in section "Results," all *C. inconspicua* strains analyzed here were found to be hybrids. The current divergence between the parental genomes was calculated dividing the number of heterozygous positions by the total size of

<sup>1</sup><http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

<sup>2</sup><http://broadinstitute.github.io/picard/>



**TABLE 1** | List of *Candida inconspicua* strains used in this project, with indication of their respective clade, place of collection, specimen, number of heterozygous variants, level of loss of heterozygosity, and estimated overall ploidy.

Strain	Clade	Country	Specimen	Heterozygous variants/kb	Estimated LOH (> 100 bp)	nQuire most probable ploidy
14ANR23920	Clade 1	Germany	Blood	14.66	65.18%	Diploid
9_16	Clade 1	Belgium	Blood	14.57	65.43%	Diploid
CI1	Clade 1	Austria	Abdominal fluid	14.15	66.45%	Diploid
CBS180*	Clade 2	Netherlands	Sputum	14.34	65.11%	Diploid
110_10	Clade 2	Austria	Blood	18.18	55.98%	Triploid
1282	Clade 2	Romania	Swab	18.88	54.06%	Triploid
CNM_CL6867	Clade 2	Spain	Swab	19.35	53.01%	Triploid
IUM_96-0030	Clade 2	Italy	Swab	18.84	54.21%	Triploid
LL867	Clade 2	Spain	Blood	19.76	51.94%	Triploid
NRZ_BK_345	Clade 2	Germany	Blood	19.70	52.78%	Triploid
UCSC_1590	Clade 2	Italy	Blood	19.30	53.06%	Triploid

Non-type strains were collected in the last 10 years in the framework of antifungal susceptibility testing. \* Type strain.

heterozygous blocks. Another important step was to check whether all strains were originated from the same hybridization event. For this, pairwise comparisons between overlapping LOH blocks were performed using bedtools jaccard v2.25.0 (Quinlan and Hall, 2010), which allowed us to get the number of nucleotides in the intersection of the two strains over the number of nucleotides present in their union. Moreover, assuming that LOH blocks with exactly the same boundaries are not independent events, it was decided to repeat the pairwise comparisons, but this time instead of analyzing the number of nucleotides in LOH regions in both strains, it was decided to get the number of LOH blocks with exactly the same boundaries in both strains. In this case, in order to avoid false positives, short LOH blocks (<1 kb), as well as very short scaffolds (<10 kb) were not considered for the analysis.

## Phylogenetic Analysis

To obtain the phylogenetic relationship between the 11 strains, FastaAlternateReferenceMaker tool of GATK v3.6 (McKenna et al., 2010) was used to obtain the genome sequence of each strain substituting each position with a homozygous SNP by the respective allele. Furthermore, bedtools subtract and bedtools getfasta v2.25.0 (Quinlan and Hall, 2010) were used to remove from these new sequences all positions with a heterozygous SNP or an INDEL in at least one of the strains. In the end, as INDELS were not considered for the analysis, the sequences of the 11 strains presented exactly the same size, constituting a sequence alignment of 9,971,439 bp. A Maximum-likelihood tree representative of this alignment was obtained with RAxML v8.2.8 software (Stamatakis, 2014), using the GTRCAT model. The same approach was applied to obtain a phylogeny for MAT locus.

## Lineage Prediction and Detection of Recombination

To predict the number of lineages and clusters in our dataset, as well as to detect traces of admixture between the different strains, we used fastGEAR (Mostowj et al., 2017). For that, this

program was set to complete mode, and the same alignment used for the phylogenetic analysis was given as input. All the other parameters were set to default. To have a control of the two expected scenarios, we decided to do the same analysis for *C. metapsilosis* and *C. orthopsilosis*, as representatives of a unique hybridization event and multiple hybridization events, respectively (Pryszcz et al., 2015; Schröder et al., 2016). In both situations, all Illumina paired-end sequencing libraries available (BioProjects PRJEB4430, PRJEB1698 and PRJNA322245), as well as five extra libraries for *C. metapsilosis* and two other libraries for *C. orthopsilosis*, which will be soon publicly available under the BioProject PRJNA520893 (manuscript in preparation) were used for read mapping, post-processing analysis, variant calling and sequence alignment as mentioned before for *C. inconspicua*.

## Mitochondrial Genome Assembly

The mitochondrial genome assembly for *C. inconspicua* was performed using the filtered Illumina paired-end reads of CI1 strain. NOVOPlasty v2.7.2 (Dierckxsens et al., 2017) with default parameters was used to assemble this genome, taking as seed input *C. inconspicua* Cox2 gene (Accession No. EF599394.1). A final 31 kb contig was obtained. NUCmer algorithm of MUMmer v3.1 (Kurtz et al., 2004) was used to align this final assembly against *Pichia kluyveri* mitochondrial genome (Accession No. NC\_022158.1). MUMmerplot algorithm of MUMmer v3.1 (Kurtz et al., 2004) was used to visualize this alignment and see that our final 31 kb scaffold covers a big part of *P. kluyveri* mitochondrial genome (Supplementary Figure 1). Mitochondrial genome annotation was performed with MITOS2 (Bernt et al., 2013). Read mapping and variant calling of all strains against this final mitochondrial assembly was performed as mentioned before. The mitochondrial genome of each strain was reconstructed with FastaAlternateReferenceMaker tool of GATK v3.6 (McKenna et al., 2010), using IUPAC code to solve heterozygous positions. A NJ phylogenetic tree was generated with SplitsTree v4 (Huson and Bryant, 2006). To compare the topology of mitochondrial and nuclear trees, RAxML v8.2.8

software (Stamatakis, 2014) was used to compute per site log Likelihoods for each tree given each of the alignments. Consel v1.2 (Shimodaira and Hasegawa, 2001) was used to assess the confidence that a given tree could correspond to a given alignment.

### Antifungal Susceptibility Test

To test whether the different clades of *C. inconspicua* presented different susceptibilities to antifungal drugs, we performed antifungal susceptibility tests on all 11 strains using two different methods (Etest gradient strips and EUCAST). By EUCAST broth-microdilution, which is one of the main international reference methods, the following drugs were tested: Itraconazole (Sigma, Rowville, Australia), Posaconazole (Schering-Plough, Kenilworth, NJ, United States), Isavuconazole (Basilea, Basel, Switzerland), Fluconazole (Sigma), Voriconazole (Sigma), Anidulafungin (Pfizer, New York, NY, United States), Micafungin (Astellas, Munich, Germany), Caspofungin (Sigma), and Amphotericin B (Sigma) in Cellstar plates (Cellstar Cat-No. 655180, Greiner Bio-One, United States). Pre-cultures were grown on Sabouraud’s 2% dextrose agar at 30°C for 24 h for all method used. The RPMI used for the different media was provided by Sigma (RPMI-1640 Medium, R6504-50L). Broth-microdilution was performed according to EUCAST guidelines (Arendrup et al., 2017) with minor modifications. To ensure proper growth, the incubation time was prolonged to 48 h, and the optical density threshold of the plate reader reading was lowered to 0.1. Plates were evaluated at 48 h both

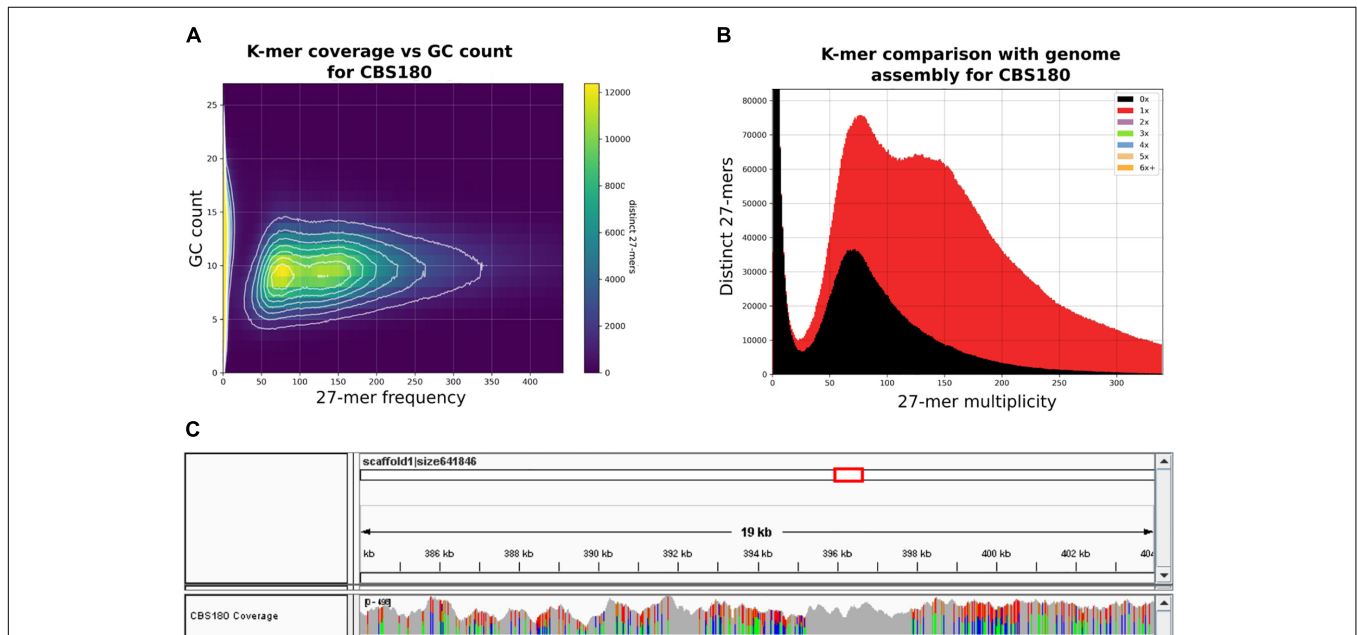
visually and by plate reader (Microplate Reader model 680, Bio-Rad, United States). *Candida parapsilosis* ATCC 22019 or *P. kudriavzevii* ATCC 6258 were used as quality control.

On the other hand, a commercial test was also used. Specifically, Etest strips with Itraconazole, Posaconazole, Isavuconazole, Fluconazole, Voriconazole, Anidulafungin, Micafungin, Caspofungin, and Amphotericin B (all bioMérieux SA, France, except Isavuconazole, which was provided by Liofilchem, Italy) were used as indicated by the respective manufacturers. Plates were incubated at 37°C and visually read after 48 h to coincide with the EUCAST conditions.

## RESULTS

### Evidence for the Hybrid Nature of *C. inconspicua*

To uncover the genomic features of *C. inconspicua* we sequenced the type strain CBS180 using an Illumina-based, pair-end sequencing strategy (see section “Materials and Methods”). GC content and 27-mer count analyses of the sequencing reads revealed two peaks with similar GC content but different coverage (Figures 1A,B). The first peak presented roughly half of the coverage of the second (Figure 1B), and therefore it could correspond to highly heterozygous regions of a diploid genome. We next performed a *de novo* genome assembly using SPAdes (Bankevich et al., 2012) and redundans (Pryszcz and Gabaldón, 2016), an assembly pipeline



**FIGURE 1 |** Heterozygosity patterns in *Candida inconspicua* type strain genome. **(A)** 27-mer frequency in CBS180 genomic reads and respective GC content. **(B)** 27-mer frequency in CBS180 genomic reads and indication of their presence (red) or absence (black) in the genome assembly. Although both 27-mer count peaks were represented in the genome assembly, roughly half of the reads corresponding to the first one was excluded. These reads probably correspond to redundant heterozygous contigs that were intentionally removed during the reduction step of the assembly process (see section “Materials and Methods”). **(C)** IGV coverage track of a 19 kb region of *C. inconspicua* scaffold 1. Colors represent polymorphic positions. Highly heterozygous regions are clearly interspaced by blocks of loss of heterozygosity (LOH).

tailored for highly heterozygous genomes (see section “Materials and Methods”). The final assembly comprised 10,353,411 bp divided in 744 contigs (76 longer than 50 kb, representing 80.16% of the genome) with 35.1% GC content and a N50 of 100,257 bp. Genome annotation predicted 5,079 proteins (see section “Materials and Methods”). The final genome size and the number of proteins that we have obtained are similar to what was previously described for the closely related species *P. kudriavzevii* (10.9 Mb and 4,949 predicted proteins; Cuomo et al., 2017), suggesting the completeness of the *C. inconspicua* genome. Indeed, 99.13% of CBS180 reads aligned to the assembly, and the missing 27-mers were in heterozygous regions, possibly corresponding to redundant contigs (Figure 1B). Furthermore, our predicted proteome has a 89% completeness as assessed by BUSCO (see section “Materials and Methods”). Phylome reconstruction (Gabaldón, 2008) in the context of twenty-one other Saccharomycotina species (see section “Materials and Methods” and Table A in Supplementary file 1) identified 501 species-specific genes. Furthermore, genes specifically duplicated in *C. inconspicua* seemed to be enriched in transmembrane transport and drug export functions, among others (Table B in Supplementary file 1).

Mapping of CBS180 reads against its own genome assembly followed by variant calling showed the presence of 14.36 variants/kb, of which most (14.34 variants/kb) corresponded to heterozygous positions (Table 1 and Supplementary file 2). Importantly, these variants were not homogeneously distributed throughout the genome, but rather formed stretches of highly heterozygous sequences separated by what appeared to be LOH, as it has been observed in previously analyzed genomes from hybrids of the *Candida* clade (Pryszcz et al., 2014, 2015; Schröder et al., 2016). These LOH blocks were flanked by heterozygous blocks with relatively constant levels of heterozygosity (36.4 heterozygous variants/kb). An illustrative example of such patterns is presented in Figure 1C. The high level of sequence divergence between the two haplotypes in the heterozygous blocks (3.64%) is much higher than the divergence observed between most distantly related strains of well-recognized yeast species (i.e., 1.1% for *Saccharomyces cerevisiae*) (Peter et al., 2018). Altogether, these analyses highly suggested that *C. inconspicua* type strain is a hybrid with a chimeric genome.

## Genome Heterogeneity Across *C. inconspicua* Strains

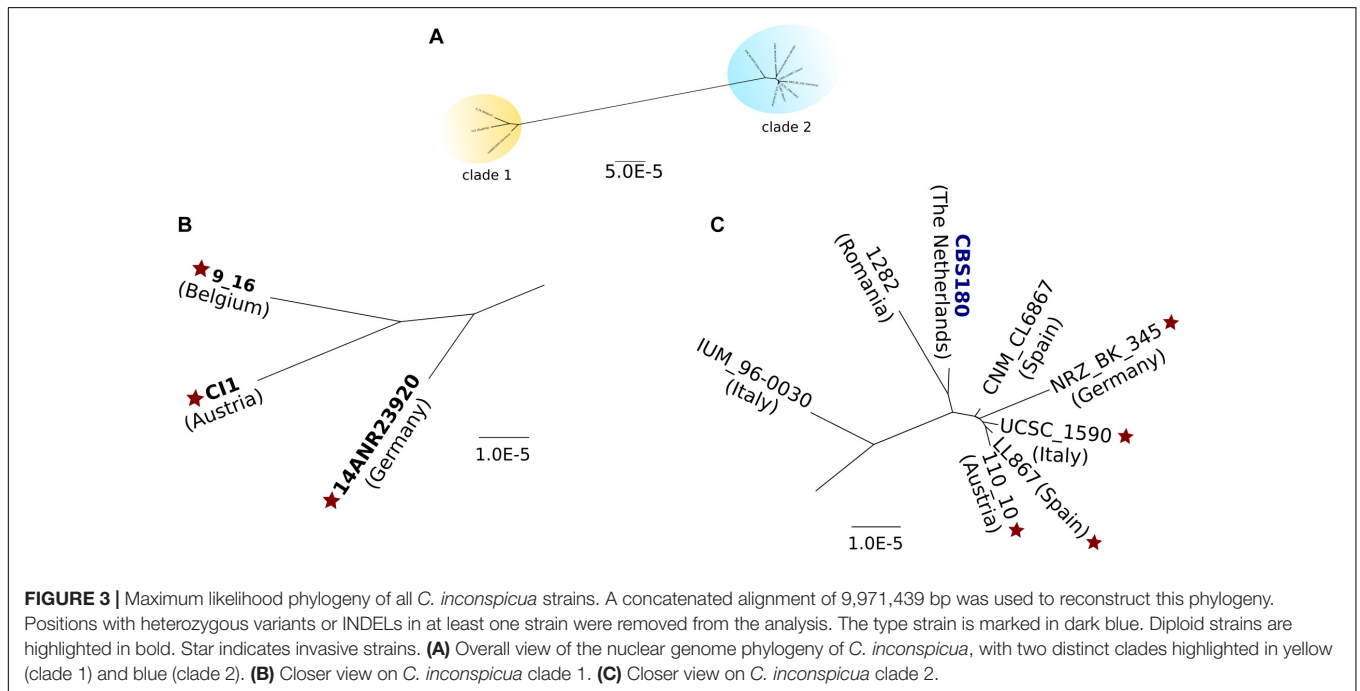
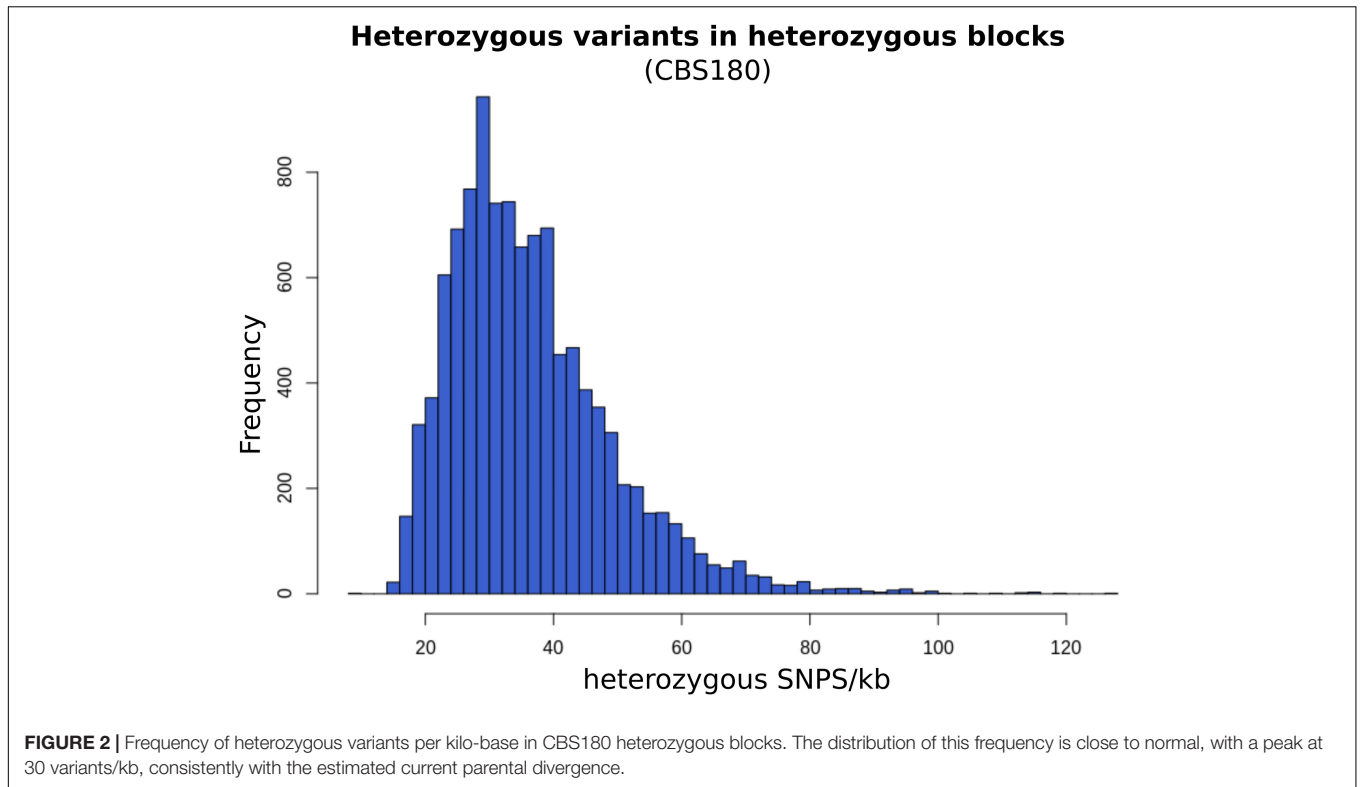
To gain a better insight into this species, 10 other clinical strains were sequenced, and their sequencing reads were mapped against the reference genome assembly described above (Table 1). All analyzed strains were highly heterozygous, with 14.15–19.76 heterozygous variants/kb, and all of them presented highly heterozygous genomic regions interspersed with LOH blocks (Table 1, Supplementary file 2 and Supplementary Figure 2). Thus, similar to the previously described hybrids in *C. metapsilosis* and *C. orthopsilosis* (Pryszcz et al., 2014, 2015; Schröder et al., 2016), *C. inconspicua* clinical isolates seemed to comprise a majority of hybrid strains, with all 11 strains analyzed so far being hybrids. The presence of both a and alpha

mating-types in the MAT locus suggests matting as a possible origin of the hybrids (Supplementary Figure 3).

Similar to previous studies (Pryszcz et al., 2014), the non-homogeneous distribution of heterozygous variants throughout the genome allowed us to define blocks of heterozygosity (see section “Materials and Methods”), which correspond to regions that retain genetic material from both hybridized lineages. On average, each strain presented 12,044 heterozygous blocks with an average size of 326 bp each, overall covering 38.15% of the genome, and comprising 82.41% of the heterozygous variants (Supplementary file 2). Based on the density of heterozygous variants within heterozygous blocks, we estimated that the current sequence divergence at the nucleotide level between the two parental lineages is approximately 3.72% (Figure 2).

We next used called SNPs to reconstruct phylogenetic relationships between the sequenced *C. inconspicua* strains using a maximum likelihood approach (see section “Materials and Methods”). The resulting strain phylogeny revealed the presence of at least two distinct clades (Figure 3), with the strains 14ANR23920, CI1, and 9\_16 forming one clade (clade 1) separated by a long branch from another clade comprising the remaining strains (clade 2). Within clade 2, IUM\_96-0030 appeared at a relatively long distance from the rest of the clade, as did, to a lesser extent, CBS180 and 1282 (Figure 3). The two clades were not related with the geographical distribution of the different strains, but invasive strains seemed to form two clusters, one in each of the clades (Figure 3 and Table 1). Furthermore, the levels of susceptibility of each strain to azoles, echinocandins, or amphotericin B (see section “Materials and Methods”) seemed to be unrelated with their position in the phylogeny (Supplementary file 3).

The existence of two separate hybrid clades pointed to two possible scenarios: (i) a unique hybridization event followed by ancestral divergence of the two clades; or (ii) two independent hybridization events between the same parental lineages, each originating one of the clades. To explore these scenarios, we first analyzed the MAT locus of the different strains (Supplementary Figure 3). While for mating-type alpha we could identify two major clades, which coincide with the nuclear genome phylogeny, for mating-type a, a third clade formed by CBS180 and IUM\_96-0030 was detected (Supplementary Figure 3). Thus, this analysis did not allow us to clearly support one or the other scenario. In a second attempt to explore these scenarios, we used a previously described approach based on the comparison of patterns of LOH blocks between strains (Pryszcz et al., 2015), where the presence of large LOH blocks with similar boundaries is indicative of a common origin. To do so, we defined LOH blocks for each *C. inconspicua* strain (see section “Materials and Methods”). On average, each strain presented 20,620 blocks with an average size of 292 bp, which covered 57.93% of the genome. Expectedly, pairwise comparisons of the overlap between LOH of each two strains revealed the same two clades as the phylogenetic analysis, but with CBS180, 1282 and IUM\_96-0030 not being clearly classified to any of them (Supplementary Figure 4). To investigate whether the two clades resulted from two different hybridization events we identified LOH blocks with exactly the same boundaries in each



pairwise comparison of the strains. Considering only blocks at least 1 kb-long present in scaffolds longer than 10 kb resulted in the same two clades (**Supplementary file 2**). However, all pairwise comparisons shared at least five large LOH blocks with exact boundaries. Indeed, disregarding any size limit in LOH blocks or scaffold size, all strains shared 633 blocks (3.07% of the

average number of blocks, **Supplementary Figure 5** as example). The relative low number of shared blocks did not allow us to clearly discard the possibility that more than one hybridization event have occurred. In fact, cluster and lineage prediction (see section “Materials and Methods”), using *C. metapsilosis* and *C. orthopsilosis* as control scenarios for one and more than one



hybridization events, respectively, clearly identified two different lineages in *C. inconspicua* (**Supplementary Figure 6**). Therefore, we considered that the most probable scenario is that all these strains were originated by two hybridization events, each one generating one of the clades.

## Aneuploidies and Mitochondrial Genome Heterogeneity Suggest Possible Admixture Among *C. inconspicua* Hybrids

We next estimated the ploidy level of each *C. inconspicua* strain using patterns of allele frequency (see section “Materials and Methods”). Our results pointed to the existence of two different groups of strains. Firstly, CBS180, and the three strains of clade 1, 14ANR23920, 9\_16, and CI1, were mainly diploid ( $r^2 > 0.9$ , **Table 1**). In contrast, the remaining seven strains presented a triploid model as the one with best support, although the results were inconclusive (**Supplementary file 4**). A closer inspection of allele frequency plots for individual scaffolds suggested the existence of a large number of aneuploidies, as shown by the coexistence of patterns compatible with tetraploid, triploid and diploid configurations, or a mixture thereof, in different scaffolds of a given strain (**Supplementary Figure 7**). The number of such aneuploidies was very reduced in the first group of strains with a largely diploid structure. All strains with large levels of aneuploidies belonged to the phylogenetic clade 2. In this regard, the type strain (CBS180) was the only diploid strain within a clade consisting mostly of aneuploid strains. This anomaly was perhaps related to the fact that the type strain has been conserved in isolated culture for a long time, likely promoting a fast diploidization, due to its frequent subculturing. Aneuploidies in the other strains may indicate intermediate levels of diploidization from an allotetraploid parental.

In another attempt to clarify the origin of these hybrid strains, we assembled a 31 kb region of *C. inconspicua* mitochondrial genome (see section “Materials and Methods”). This region was obtained from CI1 strain (clade 1), as with the type strain (CBS180) we could only get a highly fragmented mitochondrial assembly (see section “Materials and Methods”). The reads of all strains were mapped against this region. Overall, the patterns of variation in the mitochondrial genome were consistent with the nuclear genome phylogeny (**Figure 3** and **Supplementary Figures 8, 9**), with most polymorphisms likely resulting from accumulation of SNPs through time in the same mitochondrial genome background, rather than representing two different mitochondrial genomes each coming from a different parental species of the hybrids. Interestingly, mitochondrial genomes from all strains of clade 2 revealed some short deleted regions, from which we highlight a major 1.5 kb deletion in *Cox1* (**Supplementary Figure 8**), corresponding to a LAGLIDAG endonuclease domain, which is a mobile element.

As expected for a scenario of hybridization, the phylogenetic tree of the mitochondrial genome (**Supplementary Figure 9**) did not present the same topology as the one of the nuclear genome (AU test  $p\text{-val} < 0.033$ ). Even so, we could identify exactly the same clades and sub-clades in the two trees a

(**Supplementary Figure 9**). Importantly, the IUM\_96-0030 strain showed heterogeneous variation patterns consistent with the presence of mitochondrial sequences from the two clades (**Supplementary Figure 8**). For instance, besides the presence of heterozygous SNPs in the mitochondrial sequence, this strain presented some coverage in the above mentioned 1.5 kb deletion, indicating that this strain presented heteroplasmy (i.e., presence of two mitochondrial sequences within the same cell). This suggested that this clade 2 strain may have fused with a *C. inconspicua* strain belonging to clade 1. Indeed, fastGEAR (Mostowy et al., 2017) identified recent recombination in some nuclear regions of IUM\_96-0030 whose source was clade 1 (**Supplementary file 5**). The unbalanced representation of the two types of mitochondria in IUM\_96-0030 could be related to an unbalance in the inheritance of the mitochondrial genome, which was previously postulated to occur in yeast hybrids (Verspohl et al., 2018). This pointed to the occurrence of recent admixture between different *C. inconspicua* strains. Considering this, a plausible scenario is that the aneuploidies mentioned above and the mitochondrial genome heterogeneity described here are not unrelated phenomena, and both indicate that several strains of clade 2 result from recent fusions between *C. inconspicua* hybrids.

## DISCUSSION

*Candida inconspicua* is an opportunistic pathogenic yeast which is increasingly reported as a source of infection and often presents antifungal resistance (Sugita et al., 2004; Cendejas-Bueno et al., 2010; Guitard et al., 2013; Arendrup and Patterson, 2017). In this work, we have *de novo* assembled its type strain (CBS180) and sequenced 10 additional clinical isolates to obtain a better understanding of its recent evolution. Our results show compelling evidence for a hybrid nature of the *C. inconspicua* lineage, with all strains analyzed so far being hybrids between the same two parental lineages. Sequenced strains clearly clustered into two distinct clades. Although we consider that our results are not sufficiently conclusive, the low frequency of shared LOH blocks, and the identification of two different lineages by fastGEAR (Mostowy et al., 2017) make two independent hybridization events as the most plausible scenario.

All strains analyzed in this work were collected in Europe and the two clades do not show a particular enrichment in any geographical area. Although the inclusion of a larger number of strains may reveal geographical patterns in the future, this pattern is reminiscent of those of *C. metapsilosis*, where only a single hybrid lineage was found to have a global distribution, and *C. orthopsilosis*, where four different clades representing independent hybridization events have been identified but each of which has a widespread distribution (Pryszcz et al., 2015; Schröder et al., 2016).

Hybrid genomes present high levels of heterozygosity which may result in negative epistatic interactions and, consequently, reduced fitness (Mixão and Gabaldón, 2018). Such negative effects of hybridization can sometimes be compensated by emerging phenotypic properties that enable adaptation to a new niche, and therefore may offer a competitive advantage in certain



circumstances. In any case, hybridization is generally followed by different processes that lead to a gradual stabilization of these genomes, like whole genome duplication, LOH, or partial or complete chromosome loss (Mixão and Gabaldón, 2018). In *C. inconspicua*, we could observe different aneuploidies, which might represent intermediate ploidy stages to achieve the so-called genome stabilization. For instance, we could identify two groups of strains. A first group, mainly diploid, with less heterozygous variants and a higher level of LOH, and a second one with different ploidy levels, and less LOH. These two groups are almost completely coincident with the two phylogenetic clades, except for the mainly diploid type strain CBS180, which is intermingled in a clade of mostly aneuploid strains (clade 2, **Figure 3**). This clear diploid status of CBS180 is atypical of strains in clade 2. Furthermore, contrary to what is generally expected for hybrid genomes, where the improper chromosome pairing can cause problems during meiosis (Mixão and Gabaldón, 2018), CBS180 is able to enter meiosis and form ascospores (Guitard et al., 2015). This might be related to the apparent genomic stabilization that we have observed for this strain. Indeed, this strain is the type strain of *C. inconspicua* and therefore is being kept in collections for many years, completely isolated from all the other strains. We cannot be certain of whether this strain was diploid when it was collected or not, but we believe that the isolated environment and the recurrent subculturing and consequent bottleneck to which this strain is subject can have helped the genome stabilization and contributed to its different genomic patterns when compared to all the other strains.

The high prevalence of aneuploidies in some *C. inconspicua* strains is unexpected when compared to the evolutionary patterns observed in other hybrids (Pryszcz et al., 2014, 2015; Schröder et al., 2016). Additionally, the analysis of their mitochondrial genomes suggests the occurrence of crosses between different *C. inconspicua* strains and enabled us to distinguish the same three sub-clades of strains within clade 2 as the nuclear phylogeny (**Figure 3** and **Supplementary Figure 8**). For instance, a sub-clade formed exclusively by IUM\_96-0030, another one formed by CBS180 and 1282, and a third one with the remaining strains of clade 2. This suggests that clade 2 is formed by three sub-clades of hybrid strains, which are now in a process of diploidization after probably independent crossing events of two *C. inconspicua* strains. A possible scenario is that, in a process similar to the *C. albicans* parasexual cycle (Forche et al., 2008), two diploid hybrid strains form a tetraploid intermediate that would subsequently lose chromosomes until a diploid state is regained. Another possible explanation is that possibly originated ascospores can eventually cross with other *C. inconspicua* strains, working as a source of variability.

## CONCLUSION

*Candida inconspicua* is a lineage comprising opportunistic pathogens with a hybrid origin. Although the number of tested strains is low, the absence of homozygous parentals among clinical isolates suggests that the parental lineages are/might be

less able to cause infections when compared to hybrid strains. This adds to a growing list of hybrid yeast opportunist lineages and underscores the relevance of hybridization in the origin of new virulent lineages (Mixão and Gabaldón, 2018). The level of genetic variability among *C. inconspicua* hybrid strains is high, including distinct levels of aneuploidies and the presence of mitochondrial heterogeneity. This suggests that *C. inconspicua* hybrids are plastic and prone to adapt to new environments (Mixão and Gabaldón, 2018). Given the medical importance of this species, this should represent a special concern, as this high genomic plasticity may also correlate to a larger phenotypic diversity and a higher propensity to adapt to antifungal drugs and develop new resistances. Thus, more studies to identify new hybrid pathogens, as well as to try to understand how they shape their genomes, and how they can adapt to new environments should be performed. Indeed, it would be very interesting to analyze the genome of environmental strains, to understand if they are also hybrids, or if the hybridization event was the trigger that made this species become a pathogen, as it is suggested for *C. metapsilosis* (Pryszcz et al., 2015).

## DATA AVAILABILITY

Raw sequencing reads generated for this study can be found in SRA under the BioProject Accession No. PRJNA517794. Genome assembly and annotation are available under the same BioProject.

## AUTHOR CONTRIBUTIONS

AH, TB, and CL-F provided the strains and strain information. AH and CL-F performed susceptibility experiments. ES extracted DNA and prepared material for sequencing. VM performed all bioinformatics analyses. TG supervised the study. TG and VM wrote a first draft of the manuscript. All authors contributed to the final manuscript.

## FUNDING

This work received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No. H2020-MSCA-ITN-2014-642095. TG group also acknowledges support from the Spanish Ministry of Economy, Industry, and Competitiveness (MEIC) for the EMBL partnership, and grants 'Centro de Excelencia Severo Ochoa 2013–2017' SEV-2012-0208, and BFU2015-67107 co-funded 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 Agreement No. ERC-2016-724173. TG also receives support from a INB grant (PT17/0009/0023 – ISCIII-SGEFI/ERDF). CL-F received funding from Christian Doppler Laboratory for Fungal Infections: Avoid, find, and treat!

## ACKNOWLEDGMENTS

The authors thank all Gabaldón lab members for the helpful discussions and comments on this work, specially Marina Marcet-Houben for the help in the phylome analysis.

## REFERENCES

- Al-Shahrour, F., Diaz-Uriarte, R., and Dopazo, J. (2004). FatiGO: a web tool for finding significant associations of gene ontology terms with groups of genes. *Bioinformatics* 20, 578–580. doi: 10.1093/bioinformatics/btg455
- Arendrup, M. C., Meletiadi, J., Mouton, J. W., Lagrou, K., Hamal, P., Guinea, J., et al. (2017). *Method for the Determination of Broth Dilution Minimum Inhibitory Concentrations of Antifungal Agents For Yeasts. EUCAST E.DEF 7.3.1*. Available at: [http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST\\_files/AFST/Files/EUCAST\\_E\\_Def\\_7\\_3\\_1\\_Yeast\\_testing\\_definitive.pdf](http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/AFST/Files/EUCAST_E_Def_7_3_1_Yeast_testing_definitive.pdf) (accessed January 2017).
- Arendrup, M. C., and Patterson, T. F. (2017). Multidrug-resistant candida: epidemiology, molecular mechanisms, and treatment. *J. Infect. Dis.* 216(Suppl\_3), S445–S451. doi: 10.1093/infdis/jix131
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., et al. (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* 69, 313–319. doi: 10.1016/j.ympev.2012.08.023
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bretagne, S., Renaudat, C., Desnos-Ollivier, M., Sitbon, K., Lortholary, O., Dromer, F., et al. (2017). Predisposing factors and outcome of uncommon yeast species-related fungaemia based on an exhaustive surveillance programme (2002–14). *J. Antimicrob. Chemother.* 72, 1784–1793. doi: 10.1093/jac/dkx045
- Brown, G. D., Denning, D. W., Gow, N. A. R., Levitz, S. M., Netea, M. G., and White, T. C. (2012). Hidden killers: human fungal infections. *Sci. Transl. Med.* 4:165rv13. doi: 10.1126/scitranslmed.3004404
- Callon, C., Duthoit, F., Delbès, C., Ferrand, M., Le Frileux, Y., De Crémoux, R., et al. (2007). Stability of microbial communities in goat milk during a lactation year: molecular approaches. *Syst. Appl. Microbiol.* 30, 547–560. doi: 10.1016/j.syapm.2007.05.004
- Cendejas-Bueno, E., Gomez-Lopez, A., Mellado, E., Rodriguez-Tudela, J. L., and Cuenca-Estrella, M. (2010). Identification of pathogenic rare yeast species in clinical samples: comparison between phenotypic and molecular methods. *J. Clin. Microbiol.* 48, 1895–1899. doi: 10.1128/JCM.00336-10
- Cuomo, C. A., Shea, T., Yang, B., Rao, R., and Forche, A. (2017). Whole genome sequence of the heterozygous clinical isolate *Candida krusei* 81-B-5. G3 7, 2883–2889. doi: 10.1534/g3.117.043547
- Dierckxsens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45:e18. doi: 10.1093/nar/gkw955
- Douglass, A. P., Offei, B., Braun-Galleani, S., Coughlan, A. Y., Martos, A. A. R., Ortiz-Merino, R. A., et al. (2018). Population genomics shows no distinction between pathogenic *Candida krusei* and environmental *Pichia kudriavzevii*: one species, four names. *PLoS Pathog.* 14:e1007138. doi: 10.1371/journal.ppat.1007138
- Egue, L. A. N., Bouatenin, J.-P. K. M., N’guessan, F. K., and Koussémon-Camara, M. (2018). Virulence factors and determination of antifungal susceptibilities of *Candida* species isolated from palm wine and sorghum beer. *Microb. Pathog.* 124, 5–10. doi: 10.1016/j.micpath.2018.08.007
- Forche, A., Alby, K., Schaefer, D., Johnson, A. D., Berman, J., and Bennett, R. J. (2008). The parasexual cycle in *Candida albicans* provides an alternative pathway to meiosis for the formation of recombinant strains. *PLoS Biol.* 6:e110. doi: 10.1371/journal.pbio.0060110
- Forsberg, K., Woodworth, K., Walters, M., Berkow, E. L., Jackson, B., Chiller, T., et al. (2019). *Candida auris*: the recent emergence of a multidrug-resistant fungal pathogen. *Med. Mycol.* 57, 1–12. doi: 10.1093/mmy/myy054
- Gabaldón, T. (2008). Large-scale assignment of orthology: back to phylogenetics? *Genome Biol.* 9:235. doi: 10.1186/gb-2008-9-10-235
- Gabaldón, T., and Carreté, L. (2016). The birth of a deadly yeast: tracing the evolutionary emergence of virulence traits in *Candida glabrata*. *FEMS Yeast Res.* 16:fov110. doi: 10.1093/femsyr/fov110
- Gabaldón, T., Naranjo-Ortiz, M. A., and Marcet-Houben, M. (2016). Evolutionary genomics of yeast pathogens in the *Saccharomycotina*. *FEMS Yeast Res.* 16:fow064. doi: 10.1093/femsyr/fow064
- Guitard, J., Angoultant, A., Letscher-Bru, V., L’Ollivier, C., Cornet, M., Dalle, F., et al. (2013). Invasive infections due to *Candida norvegensis* and *Candida inconspicua*: report of 12 cases and review of the literature. *Med. Mycol.* 51, 795–799. doi: 10.1093/mmy/mnt086
- Guitard, J., Atanasova, R., Brossas, J. Y., Meyer, I., Gits, M., Marinach, C., et al. (2015). *Candida inconspicua* and *Candida norvegensis*: new insights into identification in relation to sexual reproduction and genome organization. *J. Clin. Microbiol.* 53, 1655–1661. doi: 10.1128/JCM.02913-14
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086
- Holzheimer, R. G., and Dralle, H. (2002). Management of mycoses in surgical patients - review of the literature. *Eur. J. Med. Res.* 7, 200–226.
- Huerta-Cepas, J., Capella-Gutiérrez, S., Pryszcz, L. P., Marcet-Houben, M., and Gabaldón, T. (2014). PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome. *Nucleic Acids Res.* 42, D897–D902. doi: 10.1093/nar/gkt1177
- Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030
- Jordà-Marcos, R., Álvarez-Lerma, F., Jurado, M., Palomar, M., Nolla-Salas, J., León, M. A., et al. (2007). Risk factors for candidaemia in critically ill patients: a prospective surveillance study. *Mycoses* 50, 302–310. doi: 10.1111/j.1439-0507.2007.01366.x
- Krukowski, H., Lisowski, A., Różański, P., and Skórka, A. (2006). Yeasts and algae isolated from cows with mastitis in the south-eastern part of Poland. *Pol. J. Vet. Sci.* 9, 181–184.
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. doi: 10.1186/gb-2004-5-2-r12
- Kurtzman, C., Fell, J. W., and Boekhout, T. (2011). *The Yeasts: A Taxonomic Study*. Amsterdam: Elsevier.
- Kurtzman, C. P., Robnett, C. J., and 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 Res.* 8, 939–954. doi: 10.1111/j.1567-1364.2008.00419.x
- Lass-Flörl, C. (2009). The changing face of epidemiology of invasive fungal disease in Europe. *Mycoses* 52, 197–205. doi: 10.1111/j.1439-0507.2009.01691.x
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 1303.3997:1–3.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Lodder, J., and Kreger-van Rij, N. J. W. (1952). *The Yeasts. A Taxonomic Study*, 1st Edn. Amsterdam: North-Holland Publishing Company, 671.
- Majoros, L., Kardos, G., Belák, A., Maráz, A., Asztalos, L., Csánky, E., et al. (2003). Restriction enzyme analysis of ribosomal DNA shows that *Candida inconspicua* clinical isolates can be misidentified as *Candida norvegensis* with traditional

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00383/full#supplementary-material>

- diagnostic procedures. *J. Clin. Microbiol.* 41, 5250–5253. doi: 10.1128/jcm.41.11.5250-5253.2003
- Majoros, L., Kardos, G., Szabó, B., Kovács, M., and Maráz, A. (2005). Fluconazole susceptibility testing of *Candida inconspicua* clinical isolates: comparison of four methods. *J. Antimicrob. Chemother.* 55, 275–276. doi: 10.1093/jac/dkh539
- Mapleson, D., Accinelli, G. G., Kettleborough, G., Wright, J., and Clavijo, B. J. (2017). KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* 33, 574–576. doi: 10.1093/bioinformatics/btw663
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a mapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Minervini, F., Montagna, M. T., Spilotros, G., Monaci, L., Santacroce, M. P., and Visconti, A. (2001). Survey on mycoflora of cow and buffalo dairy products from Southern Italy. *Int. J. Food Microbiol.* 69, 141–146. doi: 10.1016/s0168-1605(01)00583-9
- Mixão, V., and Gabaldón, T. (2018). Hybridization and emergence of virulence in opportunistic human yeast pathogens. *Yeast* 35, 5–20. doi: 10.1002/yea.3242
- Mostowy, R., Croucher, N. J., Andam, C. P., Corander, J., Hanage, W. P., and Marttinen, P. (2017). Efficient inference of recent and ancestral recombination within bacterial populations. *Mol. Biol. Evol.* 34, 1167–1182. doi: 10.1093/molbev/msx066
- Peter, J., De Chiara, M., Friedrich, A., Yue, J. X., Pflieger, D., and Bergström, A. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556, 339–344. doi: 10.1038/s41586-018-0030-5
- Pfaller, M., Neofytos, D., Diekema, D., Azie, N., Meier-Kriesche, H.-U., Quan, S.-P., et al. (2012). Epidemiology and outcomes of candidemia in 3648 patients: data from the prospective antifungal therapy (PATH Alliance®) registry, 2004–2008. *Diagn. Microbiol. Infect. Dis.* 74, 323–331. doi: 10.1016/j.diagmicrobio.2012.10.003
- Pfaller, M. A., and Diekema, D. J. (2007). Epidemiology of invasive candidiasis: a persistent public health problem. *Clin. Microbiol. Rev.* 20, 133–163. doi: 10.1128/cmr.00029-06
- Pfaller, M. A., Diekema, D. J., Gibbs, D. L., Newell, V. A., Ellis, D., Tullio, V., 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. *J. Clin. Microbiol.* 48, 1366–1377. doi: 10.1128/JCM.02117-09
- Pryszcz, L. P., and Gabaldón, T. (2016). Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* 44:e113. doi: 10.1093/nar/gkw294
- Pryszcz, L. P., Németh, T., Gácsér, A., and Gabaldón, T. (2014). Genome comparison of *Candida orthopsilosis* clinical strains reveals the existence of hybrids between two distinct subspecies. *Genome Biol. Evol.* 6, 1069–1078. doi: 10.1093/gbe/evu082
- Pryszcz, L. P., Németh, T., Saus, E., Ksiezopolska, E., Hegedúsová, E., Nosek, J., et al. (2015). The genomic aftermath of hybridization in the opportunistic pathogen *Candida metapsilosis*. *PLoS Genet.* 11:e1005626. doi: 10.1371/journal.pgen.1005626
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Sardi, J. C. O., Scorzoni, L., Bernardi, T., Fusco-Almeida, A. M., and Mendes Giannini, M. J. S. (2013). *Candida* species: current epidemiology, pathogenicity, biofilm formation, natural antifungal products and new therapeutic options. *J. Med. Microbiol.* 62(Pt 1), 10–24. doi: 10.1099/jmm.0.045054-0
- Schröder, M. S., Martínez de San Vicente, K., Prandini, T. H. R., Hammel, S., Higgins, D. G., Bagagli, E., et al. (2016). Multiple origins of the pathogenic yeast *Candida orthopsilosis* by separate hybridizations between two parental species. *PLoS Genet.* 12:e1006404. doi: 10.1371/journal.pgen.1006404
- Shimodaira, H., and Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17, 1246–1247. doi: 10.1093/bioinformatics/17.12.1246
- Stamatakis, A. (2014). RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33, W465–W467. doi: 10.1093/nar/gki458
- Sugita, T., Takeo, K., Ohkusu, M., Virtudazo, E., Takashima, M., Asako, E., et al. (2004). Fluconazole-resistant pathogens *Candida inconspicua* and *C. norvegensis*: DNA sequence diversity of the rRNA intergenic spacer region, antifungal drug susceptibility, and extracellular enzyme production. *Microbiol. Immunol.* 48, 761–766. doi: 10.1111/j.1348-0421.2004.tb03602.x
- Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6:e21800. doi: 10.1371/journal.pone.0021800
- Suzzi, G., Schirone, M., Martuscelli, M., Gatti, M., Fornasari, M. E., and Neviani, E. (2003). Yeasts associated with Manteca. *FEMS Yeast Res.* 3, 159–166. doi: 10.1016/S1567-1356(02)00183-6
- Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192. doi: 10.1093/bib/bbs017
- Verspohl, A., Pignedoli, S., and Giudici, P. (2018). The inheritance of mitochondrial DNA in interspecific *Saccharomyces* hybrids and their properties in winemaking. *Yeast* 35, 173–187. doi: 10.1002/yea.3288
- Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Kliuchnikov, G., et al. (2017). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543–548. doi: 10.1093/molbev/msx319
- Weiß, C. L., Pais, M., Cano, L. M., Kamoun, S., and Burbano, H. A. (2018). nQuire: a statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinformatics* 19:289. doi: 10.1186/s12859-018-2128-z
- Yarrow, D., and Meyer, S. A. (1978). Proposal for amendment of the diagnosis of the genus *Candida* Berkhout nom. cons. *Int. J. Syst. Bacteriol.* 28, 611–615. doi: 10.1099/00207713-28-4-611
- Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* 7, 203–214. doi: 10.1089/10665270050081478

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Mixão, Hansen, Saus, Boekhout, Lass-Flörl and Gabaldón. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.