OXFORD GENETICS

GENETICS, 2021, 217(2), iyaa004
DOI: 10.1093/genetics/iyaa004
Advance Access Publication Date: 20 November 2020
Fungal Genetics and Genomics

Extreme diversification driven by parallel events of massive loss of heterozygosity in the hybrid lineage of Candida albicans

Verónica Mixão,[1,2] Ester Saus,[1,2] Teun Boekhout,[3,4] and Toni Gabaldón[1,2,5,*]

[1]Life Sciences Department, Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain
[2]Mechanisms of Disease Department, Institute for Research in Biomedicine (IRB), The Barcelona Institute of Science and Technology, 08028 Barcelona, Spain
[3]Westerdijk Fungal Biodiversity Institute, 3584 CT Utrecht, The Netherlands
[4]Institute of Biodiversity and Ecosystem Dynamics, University of Amsterdam, 1090 GE Amsterdam, The Netherlands
[5]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

*Corresponding author: Life Sciences Department, Barcelona Supercomputing Centre (BSC-CNS), Jordi Girona, 29, 08034 Barcelona, Spain. toni.gabaldon.bcn@gmail.com

Abstract

Candida albicans is the most commonly reported species causing candidiasis. The taxonomic classification of C. albicans and related lineages is controversial, with Candida africana (syn. C. albicans var. africana) and Candida stellatoidea (syn. C. albicans var. stellatoidea) being considered different species or C. albicans varieties depending on the authors. Moreover, recent genomic analyses have suggested a shared hybrid origin of C. albicans and C. africana, but the potential parental lineages remain unidentified. Although the genomes of C. albicans and C. africana have been extensively studied, the genome of C. stellatoidea has not been sequenced so far. In order to get a better understanding of the evolution of the C. albicans clade, and to assess whether C. stellatoidea could represent one of the unknown C. albicans parental lineages, we sequenced C. stellatoidea type strain (CBS 1905). This genome was compared to that of C. albicans and of the closely related lineage C. africana. Our results show that, similarly to C. africana, C. stellatoidea descends from the same hybrid ancestor as other C. albicans strains and that it has undergone a parallel massive loss of heterozygosity.

Keywords: Candida stellatoidea; Candida albicans; yeast; comparative genomics

Introduction

Candida albicans is one of the most important fungal pathogens, being the *Candida* species with the highest impact on human health (Turner and Butler 2014). This species comprises a genetically diverse group of isolates, among which we can find *Candida africana* (syn. *C. albicans* clade 13 or *C. albicans* var. *africana*) and *Candida stellatoidea* (syn. *C. albicans* var. *stellatoidea*) (Meyer 1979; Ropars *et al.* 2018). These two lineages are genetically similar to *C. albicans*, but phenotypically different, which has made their taxonomic classification a controversy (Meyer 1979; Kwon-Chung *et al.* 1988; Jacobsen *et al.* 2008). Evidence suggests that *C. albicans* is mostly asexual with clonal reproduction and lacks a normal sexual cycle, possessing instead an alternative parasexual cycle through which different strains can mate (Forche *et al.* 2008; Alby and Bennett 2010). Therefore, it is difficult to properly establish species boundaries. With the advent of next-generation sequencing technologies, the genomes of multiple isolates of *C. albicans* and *C. africana* became available, allowing a better understanding of their evolutionary relationship (Ropars *et al.* 2018; Bensasson *et al.* 2019; Mixão and Gabaldón 2020). Indeed, recent comparative genomics analyses have revealed that *C. albicans* has a hybrid ancestor with unknown parental lineages and that *C. africana* is a descendent of the same hybridization event, which possibly underwent a massive loss of heterozygosity (LOH) (Mixão and Gabaldón 2020). Nevertheless, due to the absence of genome sequencing data for *C. stellatoidea*, the role of this lineage in this evolutionary puzzle remains unclear.

Candida stellatoidea was firstly described as *Monilia stellatoidea* in 1938 (Jones and Martin 1938) and reclassified 1 year later within the genus *Candida* (Langeron and Guerra 1939). The major phenotypic differences between this lineage and *C. albicans* are that *C. stellatoidea* cannot use sucrose as a carbon source and presents a lower growth rate (Kwon-Chung *et al.* 1988). Moreover, it is less pathogenic in animal models than in *C. albicans* (Kwon-Chung et al. 1988). In 1979, the analysis of some genetic markers revealed a high similarity between these two lineages and therefore they were proposed to be ranked as the same species (Meyer 1979). Further analyses on *C. stellatoidea* suggested that it should be divided into two types (type I —comprising the type strain CBS 1905— and type II), which differ in their electrophoretic karyotypes (Kwon-Chung *et al.* 1988). While type I corresponds to *C. stellatoidea sensu stricto*, type II corresponds to a sucrose-negative *C. albicans* which is more virulent than strains of type I (Kwon-Chung *et al.* 1988). Therefore, depending on the

Received: September 21, 2020. **Accepted:** November 3, 2020

type of *C. stellatoidea* used in previous works, the level of similarity with *C. albicans* could differ. Nevertheless, as chromosomal rearrangements could recover the phenotypes of both *C. stellatoidea* types, the idea that these lineages correspond to the same species was reinforced (Kwon-Chung *et al.* 1988; Wickes *et al.* 1991). Despite all these apparent similarities, more recent studies have shown that *C. stellatoidea* type I differs substantially from *C. albicans* at the DNA level, which contributed once again to the uncertainty in the respective taxonomic classification (Pujol *et al.* 1997; Biswas *et al.* 2001; Jacobsen *et al.* 2008).

Although a high number of isolates of the *C. albicans* clade were sequenced, including *C. africana* (Ropars *et al.* 2018; Bensasson *et al.* 2019), to the best of our knowledge none of them corresponds to *C. stellatoidea*. To gain a better understanding of the evolutionary relationship between these lineages, and assess the role of *C. stellatoidea* in the hybridization event at the basis of *C. albicans*, we sequenced the genome of the *C. stellatoidea* type strain (CBS 1905) and performed a comparative genomics analysis with *C. albicans* and *C. africana*.

Materials and methods

Library preparation and genome sequencing

Genomic DNA extraction was performed using the MasterPure Yeast DNA Purification Kit (Epicentre, USA) following the manufacturer's instructions. Briefly, CBS 1905 was 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 µl of yeast cell lysis solution (containing 1 µl of RNAse A). After being on the ice for 5 min, 150 µl of MPC protein precipitation reagent was 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 µl of TE. The samples were cleaned to remove the remaining RNA using the Genomic DNA Clean & Concentrator Kit (Epicentre) according to the manufacturer's instructions. Total DNA integrity and quantity of the sample were assessed by means of agarose gel, NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, USA) and Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific).

Whole-genome sequencing was performed at the Genomics Unit from the 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, USA) according to the manufacturer's instructions. All reagents subsequently mentioned are from the NEBNext Ultra DNA Library Prep Kit for Illumina if not specified otherwise. One microgram of gDNA was fragmented by ultrasonic acoustic energy in Covaris to a size of ~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 was done using AMPure XP Beads (Agentcourt, USA). The final library was analyzed using Agilent DNA 1000 chip (Agilent) to estimate the quantity and check size distribution, and it was then quantified by quantitative PCR (qPCR) using the KAPA Library Quantification Kit (KapaBiosystems, USA) prior to amplification with Illumina's cBot. The library was loaded and sequenced 2 × 125 on Illumina's HiSeq2500. Base calling was performed using Illumina pipeline software. De-convolution was performed using the CASAVA software (Illumina, USA).

Read mapping and variant calling

Raw sequencing data were inspected with FastQC v0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Paired-end reads were filtered for quality below 10- or 4-bp sliding windows with average quality per base of 15 and for the presence of adapters with Trimmomatic v0.36 (Bolger *et al.* 2014). A minimum read size was set to 31 bp. The K-mer Analysis Toolkit v2.4.1 (KAT; Mapleson *et al.* 2017) was used to get the *k*-mers frequency of *C. stellatoidea* sequencing libraries using default parameters ($k = 27$). This program was also used to inspect the representation of each *k*-mer in *C. albicans* SC5314 haplotype A (Muzzey *et al.* 2013).

The alignment of *C. stellatoidea* reads to *C. albicans* SC5314 haplotype A (Muzzey *et al.* 2013) and posterior 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 to align the reads, and GATK v4.0.2.1 (McKenna *et al.* 2010) and SAMtools v1.9 (Li *et al.* 2009) were used for sorting, marking duplicates, creating the index file, and obtaining the mapping statistics. GATK v4.0.2.1 (McKenna *et al.* 2010) was also used to call variants with HaplotypeCaller, requiring a minimum depth of coverage of 20 reads. The mapping results were visually inspected with IGV version 2.4.14 (Thorvaldsdóttir *et al.* 2013). In order to determine the number of SNPs/kb, only positions with 20 or more reads were considered for the genome size, and these were determined with bedtools genomecov v2.25.0 (Quinlan and Hall 2010).

Heterozygous and homozygous blocks definition

To define heterozygous and LOH blocks, and considering that the use of a fixed window size cannot properly identify region boundaries and may average their levels of heterozygosity (Pryszcz *et al.* 2014; Mixão and Gabaldón 2020), we applied a procedure previously used to study the *C. albicans* clade and other *Candida* hybrids (Pryszcz *et al.* 2014; Mixão *et al.* 2019; Mixão and Gabaldón 2020), and which is an adaptation of the procedure validated by Pryszcz *et al.* (2015). Briefly, bedtools merge v2.25.0 (Quinlan and 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. The minimum LOH and heterozygous blocks size was established at 100 bp. The current sequence divergence between the haplotypes in heterozygous regions was calculated by dividing the number of heterozygous positions by the total size of heterozygous blocks.

Comparison of SNPs between *C. albicans* and *C. stellatoidea*

To assess whether *C. stellatoidea* has the same hybrid ancestor as *C. albicans*, we used the same three groups of *C. albicans* strains as used in a previous work (group 1 = CEC4492 [clade 11], SC5314 [clade 1], CEC4497 [clade 4], and CEC5026 [clade 9]; group 2 = CEC1426 [clade 11], CEC3544 [clade 1], CEC3716 [clade 4], and CEC3533 [clade 9]; and group 3 = CEC3601 [clade 11], CEC3560 [clade 1], CEC2021 [clade 2], and CEC3557 [clade 3]) (Ropars *et al.*

2018; Mixão and Gabaldón 2020). Moreover, we followed the approach previously described for the comparison of *C. albicans* and *C. africana* (Mixão and Gabaldón 2020). Briefly, we considered the ancestral heterozygous positions (*i.e.* those present in the hybrid ancestor) previously determined for *C. albicans* (Mixão and Gabaldón 2020) and compared their alleles to the allele present in the same position in *C. stellatoidea*. For each of these positions, we counted how many ancestral alleles were shared (the same genotype was found in *C. albicans* and *C. stellatoidea*), or corresponded to one of the haplotypes (haplotype A or B), or were undefined (none of the previous options was observed). If *C. albicans* and *C. stellatoidea* shared a hybrid origin, we would expect a high proportion of positions to be shared, for at least one allele, in the two lineages.

Phylogenetic analysis

For the phylogenetic analysis of *C. albicans, C. africana*, and *C. stellatoidea*, we considered the same sample of *C. albicans* (61) and *C. africana* (8) strains that was used in the past (Mixão and Gabaldón 2020) and added the type strain of *C. stellatoidea*. A concatenated alignment of all the positions which are homozygous in all the strains was generated by removing heterozygous positions and indels from the reference genome using bedtools subtract (Quinlan and Hall 2010). A maximum-likelihood tree representative of the final alignment was obtained with RAxML v8.2.8 (Stamatakis 2014), using the GTRCAT model.

In silico reconstruction of a hybrid between *C. stellatoidea* and *C. africana*

To discard the possibility that *C. stellatoidea* and *C. africana* are the parental lineages of the hybrid ancestor of *C. albicans*, and the detected mosaic of haplotypes is an artifact generated by chromosomal recombination, we computationally created synthetic hybrid between *C. africana* CEC4103 and *C. stellatoidea* CBS 1905. For this, information on homozygous SNPs was used to replace the nucleotide present in *C. albicans* SC5314 haplotype A for each of the two strains creating two new genomes. Then, genomic reads were simulated for these new genomes using wgsim (https://github.com/lh3/wgsim). Posterior read mapping and variant calling on *C. albicans* haplotype A (Muzzey *et al.* 2013) were performed as described in the previous sections. A final comparison between the heterozygous variants of this *in silico* hybrid and *C. albicans* strains was used to determine how many *C. albicans* heterozygous positions could be recovered in the hybrid. As heterozygous positions of the parental lineages could not be used to infer the final genotype of a putative cross, heterozygous positions of CEC4103 and SC5314 were excluded from the analysis. If *C. stellatoidea* and *C. africana* were the parental lineages of *C. albicans*, we would expect the majority of *C. albicans* ancestral heterozygous positions to be present in the *in silico* created hybrid genome.

Data availability

Sequencing data generated by this project can be found under the BioProject PRJNA650282 in NCBI database. The remaining sequencing libraries used in this work were generated by Ropars *et al.* (2018) and can be found under the BioProject PRJNA432884.

Supplementary material is available at figshare DOI: https://doi.org/10.25386/genetics.13207547.

Results and discussion

The genomic patterns of *C. stellatoidea* reveal its hybrid nature

To get a better understanding of *C. albicans* clade evolution, we sequenced the genome of the type strain of *C. stellatoidea* (CBS 1905). Given the expected genetic similarity between this strain and *C. albicans* haplotypes (Meyer 1979; Jacobsen *et al.* 2008), the analysis of the respective sequencing data was performed through *k*-mer frequency analysis, read mapping and posterior variant calling on *C. albicans* SC5314 haplotype A reference genome (Muzzey *et al.* 2013). The first inspection of *C. stellatoidea* genome based on *k*-mer frequency corroborated the genetic similarity with *C. albicans*. This analysis revealed the presence of two peaks of coverage, showing that the respective levels of heterozygosity are sufficiently high to detect the first peak (Supplementary Figure S1). These results are compatible with the scenario previously described for *C. albicans* and *C. africana* (Mixão and Gabaldón 2020). Indeed, *C. stellatoidea* has 10.6 variants/kb when compared to *C. albicans* reference genome haplotype A, from which 2.9 variants/kb correspond to heterozygous positions (Table 1). As was previously observed for *C. albicans* and *C. africana*, the heterozygous variants of *C. stellatoidea* are not spread around the genome, but rather form stretches of blocks of heterozygosity separated by what appears to be blocks of LOH (Supplementary Figure S2). Interestingly, similar to what was observed for *C. africana* (Mixão and Gabaldón 2020), LOH blocks in *C. stellatoidea* cover ~95% of the genome, and the haplotype divergence in heterozygous regions is 3.55% with a single density peak (Table 1 and Supplementary Figure S3). As discussed in previous works, the accumulation of mutations in asexual organisms could lead to the occurrence of divergent homolog chromosomes in the same cell (Halkett *et al.* 2005). Nevertheless, as in *C. albicans* LOH is expected to simultaneously occur (Ene *et al.* 2018), such a scenario is expected to result in heterozygous blocks with different haplotype divergence (LOH would erase mutations with time) (Mixão and Gabaldón 2020). This expectation contrasts with our observation of a single density peak. Instead, the observed genomic patterns are similar to what was previously observed in the increasing number of *Candida* hybrids (Pryszcz *et al.* 2015; Schröder *et al.* 2016; Mixão *et al.* 2019; Mixão and Gabaldón 2020), suggesting that *C. stellatoidea* also has a hybrid nature but, as *C. africana* (Ropars et al. 2018; Mixão and Gabaldón 2020), has undergone a massive LOH.

Candida stellatoidea shares a hybrid ancestor with *C. albicans* and *C. africana*

To assess whether *C. stellatoidea* descends from the previously described hybrid ancestor of *C. albicans* and *C. africana* (Mixão and Gabaldón 2020), we compared the genomic variability of these three lineages. To this end, we took advantage of the previously applied methodology for comparison of *C. albicans* and *C. africana*, in which ancestral heterozygous positions of *C. albicans* (*i.e.* those present before the current clades diverged) were identified in the overlap of heterozygous blocks of *C. albicans* strains (701 heterozygous blocks) and the presence/absence of the respective alleles in the other lineage was quantified (see *Materials and methods*) (Mixão and Gabaldón 2020). The comparison of *C. albicans* with *C. stellatoidea* revealed that, similarly to what was determined in the past for *C. africana* (Mixão and Gabaldón 2020), 99% of the 2423 identified *C. albicans* ancestral heterozygous positions have at least one of the alleles present in *C. stellatoidea*. Furthermore, the allele present in this lineage corresponds 54.5% of the times

Table 1 Summary of the genomic patterns of *C. stellatoidea* type strain and the average numbers previously reported for 61 *C. albicans* and eight *C. africana* strains (Mixão and Gabaldón 2020), with indication of lineage name, percentage of mapped reads, all variants per kilobase (kb), heterozygous variants per kb, LOH level (percentage of the genome classified as LOH), and estimated haplotype divergence

Lineage	Mapped reads (%)	Variants/kb	Heterozygous variants/kb	LOH level (>100 bp) (%)	Haplotype divergence (%)
C. stellatoidea	98.84	10.64	2.90	94.57	3.55
C. albicans	98.88	8.60	6.70	85.74	3.52
C. africana	99.03	10.51	2.64	95.16	3.71

to haplotype A and 45.5% of the times to haplotype B. These results indicate that, similarly to *C. africana*, *C. stellatoidea* homozygous regions are a mosaic of *C. albicans* hybrid ancestor haplotypes, and therefore *C. stellatoidea* descends from the same hybridization event as *C. albicans* and *C. africana*. Of note, depending on the *C. albicans* strain used for comparison, between 183 and 194 heterozygous blocks had both haplotypes represented in *C. stellatoidea*, which is compatible with the occurrence of recent chromosome recombination in these lineages (Wang *et al.* 2018).

Considering the observed levels of similarity between the genomic patterns of *C. stellatoidea* and *C. africana*, we hypothesized that they could correspond to the same lineage. To test this hypothesis, we assessed the evolutionary relationship between *C. albicans*, *C. africana*, and *C. stellatoidea*. To this end, we reconstructed a maximum likelihood phylogeny of their homozygous positions including 61 *C. albicans* strains, eight *C. africana*, and the type strain of *C. stellatoidea* (see *Materials and methods*). As Figure 1 shows, these three lineages are genetically different, forming three different clades, with *C. africana* and *C. stellatoidea* being equally distant to *C. albicans*. These are remarkable findings, as, together with the absence of a clear overlap in the heterozygosity patterns of the two lineages (Supplementary Figure S2), they support the occurrence of two independent massive losses of heterozygosity after a single hybridization event. These results contribute to the understanding of the different phenotypes observed in these three lineages. The phenotypic relevance of massive LOH in *C. stellatoidea* and *C. africana* is not explored in this study. However, it is important to note that, from these three lineages, *C. albicans* is the most virulent one (Kwon-Chung *et al.* 1988, 1990; Borman *et al.* 2013), and also the most heterozygous one. This observation is in accordance with previous theories that suggest that high levels of heterozygosity may provide a higher plasticity to survive in hostile environments as the human body (Mixão and Gabaldón 2018). Future studies should address this hypothesis and try to assess whether indeed heterozygosity is relevant for the pathogenicity of some yeasts.

Candida stellatoidea and *C. africana* are not *C. albicans* parental lineages

As mentioned before, the observation of a mosaic of *C. albicans* haplotypes in *C. africana* and *C. stellatoidea* homozygous regions indicates that they were initially heterozygous, as expected for a hybrid, and then lost their heterozygosity. Nevertheless, the existence of chromosomal recombination in *C. albicans* (Ropars *et al.* 2018; Wang *et al.* 2018) may have influenced the results, making both *C. stellatoidea* and *C. africana* look like a mosaic, while indeed they were the two parental lineages, and their heterozygous blocks were acquired through posterior recombination. This last scenario is supported by the equal distance of the two lineages to *C. albicans* (Figure 1). If this was the case, we would expect the ancestral heterozygous positions of *C. albicans* to have each of the alleles in *C. stellatoidea* and *C. africana* homozygous regions. Thus, we compared these genomic regions in the three lineages. Our analysis revealed that between 60% and 70% of the ancestral heterozygous positions in *C. albicans* have the same allele represented in *C. stellatoidea* and *C. africana*. Moreover, an *in silico* reconstruction of a hybrid of *C. stellatoidea* and *C. africana* could only recover between 21% and 22% of the ancestral heterozygous variants of each *C. albicans* strain. These results clearly show that *C. stellatoidea* and *C. africana* are not the parental lineages of *C. albicans*, solving any of the previous doubts. Altogether, these results point to a scenario in which all the three lineages share the same hybrid ancestor, and both *C. africana* and *C. stellatoidea* have massively and independently lost their heterozygosity. The eventual identification of the parental lineages involved in this event would undoubtedly provide additional support for this hypothesis.

Concluding remarks

Candida albicans is a yeast with medical importance and is used as a model for the study of the reproductive system in the CUG-Ser clade (Alby and Bennett 2010; Turner and Butler 2014). Recent analyses of multiple isolates of this species revealed genomic patterns which are shared between different clades, and an ancient hybridization event was hypothesized as the source of such patterns (Mixão and Gabaldón 2020). Although alternative scenarios, which were discussed by the same authors, could partially explain these observations, the increasing number of unrevealed *Candida* hybrids and the similarity of their genomic features to those observed in *C. albicans* additionally support the ancient hybridization event scenario (Pryszcz *et al.* 2015; Schröder *et al.* 2016; Mixão *et al.* 2019). Considering this, our results show that *C. stellatoidea* results from the same hybridization event as *C. albicans* and *C. africana* (Mixão and Gabaldón 2020), but currently they represent three distinct lineages (Figure 1). These findings represent an important step toward understanding the evolution of the *C. albicans* clade. Importantly, it is now clear that *C. albicans*, *C. africana*, and *C. stellatoidea* represent genetically divergent clades but share a hybrid ancestor.

Despite these new findings, the taxonomic classification of the clade is still unclear due to the difficulties in applying the biological species concept in these yeasts. The fact that *C. albicans* is mostly assexual with no evidence for the occurrence of meiosis (Forche *et al.* 2008; Butler *et al.* 2009) challenges the identification of species boundaries in the clade. In this context, next-generation sequencing was crucial to reveal a non-vertical evolution event in this clade. The evidence for the occurrence of hybridization in *C. albicans* (Mixão and Gabaldón 2020) complicates even further the delimitation of species boundaries and raises the question of the reproductive mechanisms underlying such an event. It is important to note that this problem is extended to other *Candida* species, where the advances of next-generation sequencing technologies are also uncovering several hybrid lineages (Schröder *et al.* 2016; Mixão *et al.* 2019). Therefore, there is an urgent need for proper guidelines that should be followed

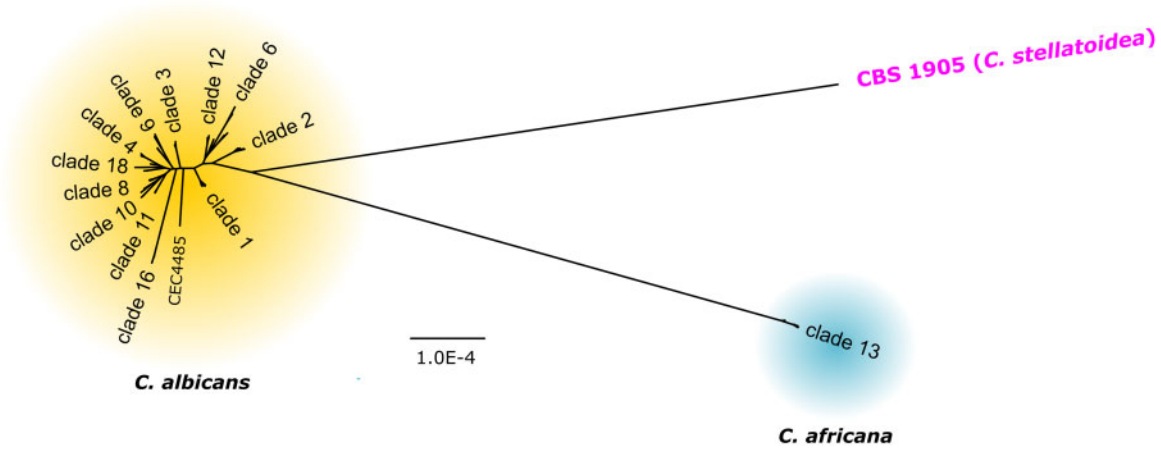

Figure 1. Maximum likelihood phylogeny of the 13,576,401-bp alignment considering the homozygous positions of 61 *C. albicans* strains, eight *C. africana* strains, and the type strain of *C. stellatoidea*. For a clear visualization, the code of the clade is shown close to the respective branch, instead of the respective names of the strains.

to the application of the species concept in these organisms. This clarification may have important repercussions in epidemiological studies, as different lineages have different phenotypes but may still be identified as the same species. For instance, in the particular case of *C. albicans* clade, it would be relevant to distinguish infections caused by each of the lineages and assess the correlation between the levels of heterozygosity and virulence. This could help us to understand not only the evolution of this clade but also the factors influencing the evolution of pathogens in general.

Acknowledgments

The authors thank all Gabaldón lab members, especially Marina Marcet-Houben, for helpful discussions.

Funding

This work was funded by the European Union's Horizon 2020 Research and Innovation Program under the Marie Sklodowska-Curie Grant Agreement No. H2020-MSCA-ITN-2014-642095. T.G. 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 Program under the grant agreements ERC-2016-724173 and MSCA-747607. T.G. also receives support from an INB Grant (PT17/0009/0023—ISCIII-SGEFI/ERDF).

Conflicts of interest: None declared.

Literature cited

Alby K, Bennett RJ. 2010. Sexual reproduction in the Candida clade: cryptic cycles, diverse mechanisms, and alternative functions. Cell Mol Life Sci. 67:3275–3285.

Bensasson D, Dicks J, Ludwig JM, Bond CJ, Elliston A, *et al.* 2019. Diverse lineages of *Candida albicans* live on Old Oaks. Genetics. 211:277–288.

Biswas SK, Yokoyama K, Wang L, Nishimura K, Miyaji M. 2001. Typing of *Candida albicans* isolates by sequence analysis of the cytochrome b gene and differentiation from *Candida stellatoidea*. J Clin Microbiol. 39:1600–1603.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120.

Borman AM, Szekely A, Linton CJ, Palmer MD, Brown P, *et al.* 2013. Epidemiology, antifungal susceptibility, and pathogenicity of *Candida africana* isolates from the United Kingdom. J Clin Microbiol. 51:967–972.

Butler G, Rasmussen MD, Lin MF, Santos MAS, Sakthikumar S, *et al.* 2009. Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature. 459:657–662.

Ene IV, Farrer RA, Hirakawa MP, Agwamba K, Cuomo CA, *et al.* 2018. Global analysis of mutations driving microevolution of a heterozygous diploid fungal pathogen. Proc Natl Acad Sci U S A. 115:E8688–E8697.

Forche A, Alby K, Schaefer D, Johnson AD, Berman J, *et al.* 2008. The parasexual cycle in *Candida albicans* provides an alternative pathway to meiosis for the formation of recombinant strains. PLoS Biol. 6:e110.

Halkett F, Simon J-C, Balloux F. 2005. Tackling the population genetics of clonal and partially clonal organisms. Trends Ecol Evol. 20:194–201.

Jacobsen MD, Boekhout T, Odds FC. 2008. Multilocus sequence typing confirms synonymy but highlights differences between *Candida albicans* and *Candida stellatoidea*. FEMS Yeast Res. 8:764–770.

Jones CP, Martin DS. 1938. Identification of yeastlike organisms isolated from the vaginal tracts of pregnant and nonpregnant women. Am J Obstet Gynecol. 35:98–106.

Kwon-Chung KJ, Wickes BL, Merz WG. 1988. Association of electrophoretic karyotype of *Candida stellatoidea* with virulence for mice. Infect Immun. 56:1814–1819.

Kwon-Chung KJ, Wickes BL, Salkin IF, Kotz HL, Sobel JD. 1990. Is *Candida stellatoidea* disappearing from the vaginal mucosa? J Clin Microbiol. 28:600–601.

Langeron M, Guerra P. 1939. Remarques sur le *Candida stellatoidea* (Jones et Martin 1938). Ann Parasitol Hum Comp. 17:257–260.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303.3997v, https://arxiv.org/pdf/1303.3997.pdf.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, *et al.* 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25:2078–2079.

Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ. 2017. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics. 33:574–576.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, *et al.* 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297–1303.

Meyer SA. 1979. DNA relatedness between physiologically similar strains and species of yeasts of medical and industrial importance. In: Garrattini, S, Paglialunga, S, Scrimshaw NS, editors. Single-Cell Protein Safety for Animal and Human Feeding. New York: Pergamon Press. p. 13–19.

Mixão V, Gabaldón T. 2018. Hybridization and emergence of virulence in opportunistic human yeast pathogens. Yeast. 35:5–20.

Mixão V, Gabaldón T. 2020. Genomic evidence for a hybrid origin of the yeast opportunistic pathogen *Candida albicans*. BMC Biol. 18:48.

Mixão V, Hansen AP, Saus E, Boekhout T, Lass-Florl C, *et al.* 2019. Whole-genome sequencing of the opportunistic yeast pathogen *Candida inconspicua* uncovers its hybrid origin. Front Genet. 10: 383.

Muzzey D, Schwartz K, Weissman JS, Sherlock G. 2013. Assembly of a phased diploid *Candida albicans* genome facilitates allele-specific measurements and provides a simple model for repeat and indel structure. Genome Biol. 14:R97.

Pegueroles C, Mixão V, Carreté L, Molina M, Gabaldón T. 2020. HaploTypo: a variant-calling pipeline for phased genomes. Bioinformatics. 36:2569–2571.

Pryszcz LP, Németh T, Gácser A, 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.

Pryszcz LP, Németh T, Saus E, Ksiezopolska E, Hegedűsová E, *et al.* 2015. The genomic aftermath of hybridization in the opportunistic pathogen *Candida metapsilosis*. PLoS Genet. 11:e1005626.

Pujol C, Renaud F, Mallié M, de Meeûs T, Bastide JM. 1997. Atypical strains of *Candida albicans* recovered from AIDS patients. Med Mycol. 35:115–121.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26:841–842.

Ropars J, Maufrais C, Diogo D, Marcet-Houben M, Perin A, *et al.* 2018. Gene flow contributes to diversification of the major fungal pathogen *Candida albicans*. Nat Commun. 9:2253.

Schröder MS, Martinez de San Vicente K, Prandini THR, Hammel S, Higgins DG, *et al.* 2016. Multiple origins of the pathogenic yeast *Candida orthopsilosis* by separate hybridizations between two parental species. PLoS Genet. 12:e1006404.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30: 1312–1313.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 14:178–192.

Turner SA, Butler G. 2014. The *Candida* pathogenic species complex. Cold Spring Harb Perspect Med. 4:a019778.

Wang JM, Bennett RJ, Anderson MZ. 2018. The genome of the human pathogen is shaped by mutation and cryptic sexual recombination. mBio. 9:e01205-18.

Wickes BL, Golin JE, Kwon-Chung KJ. 1991. Chromosomal rearrangement in *Candida stellatoidea* results in a positive effect on phenotype. Infect Immun. 59:1762–1771.

Communicating editor: A. Mitchell