Title
Whole RNA-Sequencing and Transcriptome Assembly of Candida albicans and Candida africana under Chlamydospore-Inducing Conditions.

Permalink
https://escholarship.org/uc/item/3kz0f9zc

Journal
Genome biology and evolution, 9(7)

ISSN
1759-6653

Authors
Giosa, D
Felice, MR
Lawrence, Tj
et al.

Publication Date
2017-07-01

DOI
10.1093/gbe/evx143

Peer reviewed
Whole RNA-Sequencing and Transcriptome Assembly of *Candida albicans* and *Candida africana* under Chlamydospore-Inducing Conditions

Domenico Giosa¹, Maria Rosa Felice², Travis J. Lawrence³,⁴, Megha Gulati³, Fabio Scordino¹, Letterio Giuffré⁵, Carla Lo Passo², Enrico D’Alessandro⁵, Giuseppe Criseo², David H. Ardell³, Aaron D. Hernday³, Clarissa J. Nobile³, and Orazio Romeo¹,²,*

¹IRCCS Centro Neurolesi “Bonino-Pulejo,” Messina, Italy
²Department of Chemical, Biological, Pharmaceutical, and Environmental Sciences, University of Messina, Italy
³Department of Molecular and Cell Biology, University of California, Merced, CA
⁴Quantitative and System Biology Graduate Program, University of California, Merced, CA
⁵Department of Veterinary Sciences, Division of Animal Production, University of Messina, Italy

*Corresponding author: E-mail: oromeo@unime.it.

Accepted: July 25, 2017

Data deposition: The data sets supporting the conclusions of this article have been deposited in the Transcriptome Shotgun Assembly (TSA) database under accession numbers GEVV02000000 (CBS 11016 strain) and GEVW02000000 (GE1 strain), associated with BioProject IDs PRJNA327731 and PRJNA327736, respectively. The raw reads used for the transcriptome analysis have also been deposited in the NCBI Sequence Read Archive (SRA) under the following accession numbers SRR3747616 and SRR3745408.

Abstract

*Candida albicans* is the most common cause of life-threatening fungal infections in humans, especially in immunocompromised individuals. Crucial to its success as an opportunistic pathogen is the considerable dynamism of its genome, which readily undergoes genetic changes generating new phenotypes and shaping the evolution of new strains. *Candida africana* is an intriguing *C. albicans* biovariant strain that exhibits remarkable genetic and phenotypic differences when compared with standard *C. albicans* isolates. *Candida africana* is well-known for its low degree of virulence compared with *C. albicans* and for its inability to produce chlamydospores that *C. albicans*, characteristically, produces under certain environmental conditions. Chlamydospores are large, spherical structures, whose biological function is still unknown. For this reason, we have sequenced, assembled, and annotated the whole transcriptomes obtained from an efficient *C. albicans* chlamydospore-producing clinical strain (GE1), compared with the natural chlamydospore-negative *C. africana* clinical strain (CBS 11016). The transcriptomes of both *C. albicans* (GE1) and *C. africana* (CBS 11016) clinical strains, grown under chlamydospore-inducing conditions, were sequenced and assembled into 7,442 (GE1 strain) and 8,370 (CBS 11016 strain) high quality transcripts, respectively. The release of the first assembly of the *C. africana* transcriptome will allow future comparative studies to better understand the biology and evolution of this important human fungal pathogen.

Key words: *Candida africana*, *Candida albicans*, Chlamydospores, RNA-seq, transcriptome assembly, nTAR.

Introduction

*Candida albicans* is a diploid commensal fungus that asymptptomatically colonizes the gastrointestinal and genitourinary tracts, vagina and skin of healthy humans, and other warm-blooded animals (Neville et al. 2015). However, it is also a major opportunistic fungal pathogen capable of causing a variety of superficial infections in healthy individuals as well as more severe infections, typically in immunocompromised hosts (Papon et al. 2013). Genetically, this yeast is extraordinarily versatile and its population structure is exceptionally heterogeneous, correlating with its prevalence in different clinical samples and the notable pathogenic potential of the...
isolates (McManus and Coleman 2014). Nevertheless, a particular group of strains appears to be better adapted at causing vaginal infections than others and represents an interesting example of divergent evolution in C. albicans (Odds 2010). These strains were originally proposed to be a new Candida species, Candida africana (Tietz et al. 2001), however comparative genetic studies indicated that the differences between C. albicans and C. africana are insufficient to support its status as a new Candida species (Odds et al. 2007; Romeo and Criseo 2011).

Despite the high degree of genetic similarity shared between C. albicans and C. africana, there is considerable phenotypic diversity differentiating the two pathogens. Unlike C. albicans, C. africana is relatively limited in its abilities to assimilate distinct carbon sources, shows a low degree of virulence in disseminated infection models in mice, has a reduced growth rate, and exhibits a decreased ability to filament under standard filament-inducing conditions (Romeo et al. 2013; Pagniez et al. 2015; Felice et al. 2016). Another distinct feature of C. africana compared with C. albicans is its inability to form chlamydospores, large, spherical, thick-walled structures of unknown function that in the past were widely used as a taxonomic criterion to distinguish C. albicans from other Candida species.

Chlamydospore development in C. albicans is not well understood; we, therefore, decided to explore and compare the transcriptomes of C. albicans and C. africana in response to specific culture conditions that are known to induce chlamydospore formation in C. albicans.

In this study we report, for the first time, the whole-transcriptome profiles of C. albicans and C. africana biovariant strains grown under chlamydospore-inducing conditions.

## Materials and Methods

### Fungal Strains and Culture Conditions

In this study, we examined a typical C. albicans clinical strain (named GE1) and the reference clinical strain CBS 11016 also known as C. africana, a chlamydospore-negative biovariant of C. albicans (Romeo and Criseo 2009). Both Candida strains were isolated from vaginal samples of women suffering from vaginitis and were identified using conventional clinical laboratory and molecular methods (Romeo and Criseo 2008). The C. albicans GE1 strain shows a strong ability to produce chlamydospores when cultivated on plates containing corn meal agar (CMA) plus 1% tween 80 at 25°C for 48–72 h, whereas the C. africana CBS 11016 strain is unable to form chlamydospores under the same culture conditions (Romeo and Criseo 2009).

To stimulate chlamydospore production and induce the expression of genes associated with their formation, 15 ml of a standard overnight yeast suspension culture (containing ~10^7 cells/ml) was used to inoculate a 100 mm diameter Petri dish containing CMA plus 1% tween 80. The plates were incubated at 25°C and microscopically monitored until chlamydospores were produced by the positive GE1 strain, which occurred at 48 h, and cells were collected and subjected to total RNA extraction.

### RNA Extraction and Sequencing

A total of 1 × 10^8 yeast cells were used for RNA extraction using the Ribopure Yeast Purification kit (Thermofisher, Italy) following the manufacturer’s instructions. RNA quantification was carried out spectrophotometrically at 260 nm and 280 nm and RNA integrity was evaluated with an Agilent 2100 Bioanalyzer instrument using the RNA 6000 Nano kit (Agilent Technologies, Italy).

Before RNA-Seq, the expression levels of two genes highly expressed specifically during chlamydospore formation, CSP1 and CSP2, were examined by quantitative real time PCR as described previously (Palige et al. 2013). For cDNA synthesis, 2 μg of total RNA were digested with DNase I (Sigma-Aldrich, Italy) following manufacturer’s instructions and retrotranscribed by RevertAid First Strand cDNA Synthesis kit (Thermofisher, Italy) using oligo(dT) at 42°C for 1 h followed by a reverse transcriptase denaturation step at 70°C for 10 min.

The mRNA levels of CSP1 and CSP2 were quantified by StepOne Plus Real Time PCR system (Applied Biosystem, Italy) using SYBR Premix Ex Taq II (Takara, Clontech, Italy) and primers listed in table 1. Actin (ACT7) was used as housekeeping gene and the ∆ΔCt was calculated using the Livak and Schmittgen (2001) method. cDNA was also used to confirm the presence of two distinct “novel transcriptionally active regions” (nTARs) (table 2) detected during evaluation of the transcriptome assembly analysis. The sequences of these transcripts were retrieved from our assemblies, aligned with MEGA6 software (Tamura et al. 2013) and used to design specific oligonucleotide primers (table 1).

In vitro amplifications of the nTARs were performed separately using the DreamTaq PCR Master Mix (Thermofisher, Italy) plus 1 μl of cDNA and 0.5 mM of each primer pair (table 1). PCR reactions were carried out in a Bio-Rad T100 thermal cycler with preliminary denaturation at 95°C for 4 min followed by 35 cycles of denaturation at 94°C for 40 sec, annealing at 56°C for 40 sec, extension at 72°C for 45 sec and a final step of 10 min at 72°C. Amplicons were subjected to 2% agarose gel electrophoresis to confirm the presence of amplified fragments, expected size and the absence of nonspecific products. Subsequently, PCR products were purified using the QiAquick PCR Purification kit (Qiagen, Italy) and sequenced at the Eurofins Genomics (Ebersberg, Germany; www.eurofinsgenomics.eu) using the same primers used for PCR (table 1).

For whole RNA sequencing, ~4 μg (100 ng/μl) of high quality total RNA (OD_{260/280} ≥ 2.0; RIN value ≥ 9.6) from each
Candida strain, were sent to GATC Biotech (Konstanz, Germany) for poly-A filtering, fragmentation, random primed cDNA synthesis and library preparation for Illumina HiSeq single read sequencing (read length: 1 × 50 bp).

Transcriptome Assembly and Evaluation

Before assembling, raw reads were processed using the programs Scythe (version 0.994 BETA; https://github.com/vsbufalo/scythe; last accessed July 28, 2017) and Sickle (version 1.33; https://github.com/najoshi/sickle; last accessed July 28, 2017) to remove adapters and sequences with low Phred-scores (cutoff: ≥20).

For assembling whole transcriptomes, we used two different strategies: a de novo approach using BinPacker version 1.0 (Liu et al. 2016) and a reference based assembly using StringTie version 1.3.0 (Pertea et al. 2015) with different alignment programs: HISAT2 version 2.0.4 (Kim et al. 2015) and Subjunc version 1.22.3 (Liao et al. 2013). Both HISAT2 and Subjunc short read aligners produced similar results. The reference genome used was that of C. albicans SC5314, version A22-s05-m05-r03, retrieved from the Candida Genome Database (CGD) (www.candidagenome.org).

To evaluate the completeness of our assemblies, we used BUSCO version 1.22 (Simão et al. 2015), which quantitatively assesses the assembled transcriptomes by comparing them against a database (OrthoDB; www.orthodb.org) of universal single copy orthologs for fungi.

A full description of the command-line programs, including an R script used for Subjunc alignment in the reference based assembly, is provided in the supplementary files S1 and S2, Supplementary Material online, respectively.

Overall statistics of the two Candida transcriptomes assembled in this study were calculated using FAST version 1.06 (Lawrence et al. 2015) and SAMtools version 1.3.1 (Li et al. 2009).

Results

In this study, the C. albicans GE1 strain produced abundant chlamydospores after 48 h of incubation on CMA supplemented with 1% tween 80 whereas these structures were not observed in the C. africana CBS 11016 culture. To further demonstrate the lack of chlamydospore production in C. africana, we examined the induction of two specific genes, CSP1 and CSP2, which were reported to be highly upregulated during chlamydospore development in C. albicans (Palige et al. 2013). As expected the expression levels of these two genes were significantly elevated in the GE1 strain compared with CBS 11016 strain (data not shown).

Illumina whole transcriptome sequencing produced over 78 and 61 million raw reads for strains CBS 11016 and GE1, respectively (table 3). After quality filtering and trimming, ~0.5% (CBS 11016 strain) and 0.76% (GE1 strain) of the total reads were removed, leaving a large data set of high-quality reads for both de novo and reference guided assembling (average Q-score: ~38) (table 3). The number of complete, fragmented, and missing orthologs obtained by BUSCO (Simão et al. 2015) for all transcriptomes is shown in figure 1. Based on this result, it appears that the de novo
approach yielded a more complete transcriptome by a large margin.

To assess whether the transcripts were correctly assembled, an additional quality analysis was performed by mapping back the original reads to the assembled transcriptomes. The data showed that 95% of the reads mapped to both de novo assemblies produced by BinPacker while only 44.7% (C. africana) and 54% (C. albicans) of the reads mapped to the reference-based transcriptomes assembled by StringTie.

The de novo assembled transcriptomes (Genbank accession number: GEVW02000000 for GE1 and GEVC02000000 for CBS 11016) resulted in 7,442 contigs (≥200 bp; largest contig 10,652 bp; total consensus length of 12,270,608 bp) for the C. albicans GE1 strain and 8,370 contigs (≥200 bp; largest contig 10,102 bp; total consensus length of 10,507,121 bp) for the C. africana CBS 11016 strain, constituting ~43% and ~37%, respectively, of the entire C. albicans genome. Overall statistics of the two Candida transcriptomes assembled in this study are summarized in Table 3. A detailed list of the genes detected, including their annotations, is provided in the supplementary table S1, Supplementary Material online.

From our RNA-seq data we generated over 75 (C. africana) and 58 (C. albicans) million mappable reads (table 3) and identified two novel transcriptionally active regions (nTARs) that mapped outside of the annotated regions of the C. albicans reference genome used in this study (fig. 2). The presence of the two nTARs in our cDNA samples was further verified by standard Sanger sequencing, which indicated that both C. albicans and C. africana produce these transcripts.

In silico translation (http://web.expasy.org/translate; last accessed July 28, 2017) and BLAST analysis were unable to identify any homologous product currently available in the public databases and therefore our nTARs were classified as long noncoding RNAs (lncRNAs). Both lncRNAs, nTAR1 and nTAR2, map to two different unannotated regions of chromosome R (table 2 and fig. 2) between tD(GUC)7 (encoding a tRNA-Asp) and an uncharacterized ORF (CR_09460 C_A) and between SAP2 (encoding a major secreted aspartyl proteinase) and YHB5 (encoding a flavohemoglobin-related protein) (fig. 2).

### Discussions and Conclusions

The biological function of chlamydospores is unknown and the reasons why only C. albicans, and its closely related species

---

**Table 3**

Transcriptome Assembly Statistics for C. africana and C. albicans

<table>
<thead>
<tr>
<th></th>
<th>CBS 11016</th>
<th>GE1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total raw reads</td>
<td>78,360,457</td>
<td>61,354,627</td>
</tr>
<tr>
<td>Number of reads used for assembling</td>
<td>77,957,181</td>
<td>60,885,731</td>
</tr>
<tr>
<td>Average read length</td>
<td>50.77</td>
<td>50.74</td>
</tr>
<tr>
<td>Total mapped</td>
<td>75,736,643</td>
<td>58,898,936</td>
</tr>
<tr>
<td>Uniquely mapped</td>
<td>72,959,571</td>
<td>55,128,869</td>
</tr>
<tr>
<td>Multiply mapped</td>
<td>2,777,072</td>
<td>3,770,067</td>
</tr>
<tr>
<td>Unaligned</td>
<td>2,623,814</td>
<td>2,455,691</td>
</tr>
<tr>
<td>GC content (%)</td>
<td>34.3</td>
<td>34.4</td>
</tr>
<tr>
<td>Total assembled contigs</td>
<td>8,370</td>
<td>7,442</td>
</tr>
<tr>
<td>Total assembly length (bp)</td>
<td>10,507,121</td>
<td>12,270,608</td>
</tr>
<tr>
<td>Number of contigs ≥ 500 bp</td>
<td>5,582</td>
<td>6,418</td>
</tr>
<tr>
<td>Number of contigs ≥ 1000 bp</td>
<td>3,784</td>
<td>4,723</td>
</tr>
<tr>
<td>Longest contig (bp)</td>
<td>10,102</td>
<td>10,652</td>
</tr>
<tr>
<td>Contig N50</td>
<td>2,020</td>
<td>2,252</td>
</tr>
<tr>
<td>Number of contigs with ORF</td>
<td>7,051</td>
<td>7,048</td>
</tr>
</tbody>
</table>

---

![Fig. 1](image-url)  
**Fig. 1.** — Number of complete, fragmented, and missing orthologs obtained by BUSCO analysis (Simão et al. 2015) using both de novo and reference-guided assemblies of C. africana and C. albicans.
C. dubliniensis, have retained this phenotype throughout evolution remains a mystery (Palige et al. 2013; Navarathna et al. 2016).

Different hypotheses have been put forth to explain the presence of these characteristic structures in these two Candida species, but none have been definitely proven. It was proposed that chlamydospores could allow the survival of the fungus in harsh environmental conditions or, given that C. albicans and C. dubliniensis are restricted to humans, could facilitate persistence or survival of these species within the mammalian host (Staib and Morschhäuser 2007). However, rarely have chlamydospores been observed in infected tissues (Heineman et al. 1961; Ho and O’Day 1981; Chabasse et al. 1988; Cole et al. 1991) although a recent study documenting their formation in the kidneys of experimentally infected mice suggested a role in pathogenesis (Navarathna et al. 2016).

Evolutionarily, the assumption that chlamydospore production has persisted within C. albicans populations is not strictly true since a number of pathogenic strains, such as C. africana, have lost the ability to produce chlamydospores (Tietz et al. 2001; Romeo et al. 2013). These strains represent a group of phylogenetically distinct C. albicans strains (Odds 2010) and show many distinctive phenotypic traits, including profound changes in nutrient assimilation, reduced filamentation, and a low degree of virulence in insect and mammalian models (Borman et al. 2013; Romeo et al. 2013; Pagniez et al. 2015; Felice et al. 2016). This phenotypic diversity may be the result of a unique genetic background (Felice et al. 2016), however, the genomic sequence of C. africana has not yet been determined. For this reason, we decided to sequence the whole transcriptome of C. africana, compared with a standard C. albicans strain, in order to learn which genes are expressed under chlamydospore-inducing conditions.

In this study both transcriptomes were assembled using two different methods and our data suggested that a de novo assembly approach was advantageous, despite the availability of a closely related reference genome (Huang et al. 2016). The better performance of the de novo approach may be due, in part, to the nature of our sequencing data (consisting of 50 bp single reads), since most reference based transcriptome assemblers are optimized to take advantage of the additional information contained in paired-end data. However, it is also possible that genetic differences among strains could have contributed to the superior performance of
the de novo over reference-based transcriptome assemblies that we observed (Huang et al. 2016). In fact, genetically, the population structure of C. albicans is remarkably heterogeneous and C. africana represents the most evolutionary divergent lineage currently known (Odds et al. 2007; Odds 2010). The greater divergence of C. africana CBS 11016 from the reference C. albicans genome strain (SC5314) may explain why its reference-based assembly missed more orthologs (fig. 1) and mapped back fewer reads than the reference-based assembly using C. albicans GE1.

Another interesting finding of this study relates to the discovery of novel and uncharacterized transcriptionally active regions whose functions are at present unknown, but are believed to be of regulatory importance. One nTAR (nTAR2; table 2), although not obviously evident by RNA-seq read coverage (fig. 2), overlaps exactly the same nTAR (TFRW181) initially described by Sellam et al. (2010) in a C. albicans genome annotation paper using high-resolution tiling arrays, and later reported by Nobile et al. (2012) using RNA-seq to study the transcriptional circuitry controlling C. albicans biofilm formation. This nTAR, as well as the new nTAR1 reported here, was not found in any other sequenced Candida species (Candida glabrata CBS 138, Candida tropicalis MYA-3404, Candida parapsilosis CDC 317, Candida orthopsilosis Co 90-125, Candida lusitaniae ATCC 42720, Candida guillermondii ATCC 6260 and C. dubliniensis CD 36) accessible through the Candida genome database.

Overall the release of the first assembly of the C. africana transcriptome represents an important milestone for Candida research because it sets the framework for future genetic studies to determine the transcriptional network underlying chlamydospore development in C. albicans, and also provides a platform for future comparative studies between these two closely related strains, whose differences in phenotypic and genetic traits are remarkably vast.

**Supplementary Material**

Supplementary data are available at Genome Biology and Evolution online.

**Acknowledgments**

This study was funded in part by the University of Messina, Ordinary Research Program PRA2008/2009 (ORME097YB2) and by the National Institutes of Health (NIH) grant (R00AI100896 to C.J.N.). High performance computing (HPC) resources were supported by the National Science Foundation (NSF) (ACI-1429783).

**Literature Cited**


**Associate editor:** Maria Costantini