G3 Genes | Genomes | Genetics

GENOME REPORT

Genome Assemblies of Two Rare Opportunistic Yeast Pathogens: *Diutina rugosa* (syn. *Candida rugosa*) and *Trichomonascus ciferrii* (syn. *Candida ciferrii*)

Verónica Mixão,*,†,‡ Ester Saus,*,†,‡ Antonio Perez Hansen,§ Cornelia Lass-Florl,§ and Toni Gabaldón*,†,‡,,††,1**

*Centre for Genomic Regulation, The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain, †Life Sciences Department. Barcelona Supercomputing Center (BSC). Jordi Girona, 29. 08034 Barcelona, Spain, ‡Mechanisms of Disease Department. Institute for Research in Biomedicine (IRB). Barcelona, Spain, §Division of Hygiene and Medical Microbiology, Innsbruck Medical University, Austria, **Universitat Pompeu Fabra (UPF), Barcelona, Spain, and ††ICREA, Pg. Lluis Companys 23, Barcelona 08010, Spain

ORCID IDs: 0000-0001-6669-0161 (V.M.); 0000-0003-0019-1735 (T.G.)

ABSTRACT Infections caused by opportunistic yeast pathogens have increased over the last years. These infections can be originated by a large number of diverse yeast species of varying incidence, and with distinct clinically relevant phenotypic traits, such as different susceptibility profiles to antifungal drugs, which challenge diagnosis and treatment. *Diutina rugosa* (syn. *Candida rugosa*) and *Trichomonascus ciferrii* (syn. *Candida ciferrii*) are two opportunistic rare yeast pathogens, which low incidence (< 1%) limits available clinical experience. Furthermore, these yeasts have elevated Minimum Inhibitory Concentration (MIC) levels to at least one class of antifungal agents. This makes it more difficult to manage their infections, and thus they are associated with high rates of mortality and clinical failure. With the aim of improving our knowledge on these opportunistic pathogens, we assembled and annotated their genomes. A phylogenomics approach revealed that genes specifically duplicated in each of the two species are often involved in transmembrane transport activities. These genomes and the reconstructed complete catalog of gene phylogenies and homology relationships constitute useful resources for future studies on these pathogens.

KEYWORDS
Diutina rugosa
Trichomonascus ciferrii
genome assembly
yeast pathogen
Candida rugosa
Candida ciferrii

Candida species are the most common cause of hospital-acquired fungal infections, very often leading to patient's death (Pfaller and Diekema 2007; Lass-Flörl 2009; Brown *et al.* 2012; Gabaldón and Carreté 2016). Although *Candida albicans*, *Candida glabrata* and *Candida parapsilosis* are the species with highest prevalence (Pfaller and Diekema 2007; Jordà-Marcos *et al.* 2007), in the last years the incidence of "rare yeast" infections has increased (Lass-Flörl 2009; Pfaller *et al.* 2012; Bretagne *et al.* 2017). By "rare yeasts" we mean ascomycetous yeasts that have very low prevalence (< 1% of clinical *Candida* infections) and have high Minimum Inhibitory Concentrations (MICs) toward at least one class of antifungal drugs (Guitard *et al.* 2013; Jung *et al.* 2015; Bretagne *et al.* 2017).

Diutina rugosa (syn. *Candida rugosa* (Khunnamwong *et al.* 2015)) and *Trichomonascus ciferrii* (syn. *Candida ciferrii* (Kurtzman and Robnett 2007)) are two "rare yeasts" (Pfaller *et al.* 2010). *D. rugosa* has been reported as a causative agent of veterinary infections (Moretti *et al.* 2000; Crawshaw *et al.* 2005; Scaccabarozzi *et al.* 2011), and therefore might have impact in industry and economics. Furthermore, it has been identified as the etiological agent of several clinical infections, including a clinical outbreak in Brazil (Lopes Colombo *et al.* 2003; Pfaller *et al.* 2010). Thus, this species is considered an emerging fungal pathogen (Pfaller *et al.* 2006; Minces *et al.* 2009). Indeed, in a 10-year multi-center study a 10-fold increase in

doi: https://doi.org/10.1534/g3.119.400762
Manuscript received July 17, 2019; accepted for publication September 30, 2019; published Early Online October 1, 2019.

Supplemental material available at figshare: https://doi.org/10.25387/g3.8945048.
[1]Corresponding author: Barcelona Supercomputing Centre (BSC-CNS), Jordi Girona, 29, 08034 Barcelona, Spain. E-mail: toni.gabaldon.bcn@gmail.com

the number of *D. rugosa* clinical cases was reported (Pfaller *et al.* 2010). *T. ciferrii* has also been reported as an opportunistic pathogen in some sporadic cases of infections in immunocompromised patients (Gunsilius *et al.* 2001; Pfaller *et al.* 2010; Saha *et al.* 2013; Villanueva-Lozano *et al.* 2016; Upadhyay *et al.* 2018). Both species were recently shown to present high MICs to azoles and echinocandins (Pfaller *et al.* 2006; Pérez-Hansen *et al.* 2019).

Next Generation Sequencing (NGS) is a powerful tool to study the genomic background of pathogens, which might reveal many of their features. In the last years, more and more studies performing NGS analysis on yeast pathogens were published and showed the relevance of whole-genome sequence for the study of pathogenic genomic determinants (Butler *et al.* 2009; Pryszcz *et al.* 2015; Schröder *et al.* 2016; Ropars *et al.* 2018; Mixão *et al.* 2019). In this context, we decided to sequence the genome of both *D. rugosa* and *T. ciferrii*, which will be useful for future studies on these opportunistic pathogens.

MATERIALS AND METHODS

Library preparation and genome sequencing

We sequenced the type strains for *D. rugosa* (CBS613) and *T. ciferrii* (CBS4856). Genomic DNA extraction was performed using the MasterPure Yeast DNA Purification Kit (Epicentre, United States) following manufacturer's instructions and all reagents mentioned are from the kit if not specified otherwise. Briefly, cultures were grown in an orbital shaker overnight (200 rpm, 30°) in 15 ml of YPD medium (Yeast extract-Peptone-Dextrose medium: 10 g of yeast extract, 20 g of bacto peptone and 50 ml of dextrose 40% in 1 L of distilled water). 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° for 15 min with 300 µl of yeast cell lysis solution (containing 1 µl of RNAse A). After being on ice for 5 min, 150 µl 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°. The pellet was washed twice with 70% cold ethanol and, once the pellet was dried, the sample was resuspended in 100 µl 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 an Illumina 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 µg of gDNA was fragmented by nebulization using the Covaris S2 instrument (Covaris Inc.) 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, 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 in paired-end reads of 125bp on Illumina's HiSeq2500. 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 has been deposited in short read archive (SRA) under the BioProject Accession No. PRJNA531406.

De novo genome assembly and phylome reconstruction

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 size below 31 bp and for the presence of adapters with Trimmomatic v0.36 (Bolger *et al.* 2014). The K-mer Analysis Toolkit v2.4.1 (KAT; (Mapleson *et al.* 2017)) was used to get the GC content and k-mer frequency distribution and estimate the expected genome size. SOAPdenovo v2.04 (Luo *et al.* 2012) was used to perform genome assembly. 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 v2.4.1 (Mapleson *et al.* 2017). Species identification was confirmed by BLASTn (Zhang *et al.* 2000) of the respective ITS region (accession: NR_111249.1 for *D. rugosa* and NR_111160.1 for *T. ciferrii*), as recommended (Stavrou *et al.* 2018). Genome annotation was performed with Augustus v3.1 (Stanke and Morgenstern 2005), using *Meyerozyma guilliermondii* and *Saccharomyces cerevisiae* as model organisms for *D. rugosa* and *T. ciferrii*, respectively. The Ascomycota dataset in BUSCO v3 (Waterhouse *et al.* 2018) was used to assess completeness.

Phylome reconstruction - *i.e.*, the complete collection of phylogenies for every gene encoded in the genome - was performed using the PhylomeDB pipeline (Huerta-Cepas *et al.* 2014), as described in (Pryszcz *et al.* 2015), considering twenty-seven species (Supplementary file 1). This was done for both *D. rugosa* and *T. ciferrii*, using their respective predicted proteomes as seed. These phylomes and the corresponding orthology and paralogy relationships are available for browsing or download in PhylomeDB (Huerta-Cepas *et al.* 2014) with ID 932 and 842, respectively. 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 (UniProt Consortium 2019) (accessed on April 30th, 2019), in order to determine the possible function associated with these genes, as well as their GO terms. An enrichment analysis was done using FatiGO (Al-Shahrour *et al.* 2007). Species-tree reconstruction was based on the final concatenated alignment of 469 single genes, comprising 297,788 amino-acid positions, with RAxML v8.2.4 (Stamatakis 2014), using the PROTGAMMALG substitution model and performing rapid bootstrapping with 1000 replicates. A BLASTp (Zhang *et al.* 2000) of the species-specific genes against the non-redundant database maintained by NCBI (NCBI Resource Coordinators 2015) (accessed on September 13th, 2019), considering only hits with e-value < 0.001 and query coverage $> 50\%$, was performed to determine whether these genes have homologs in species which were not considered for phylome reconstruction.

Table 1 Metrics of *D. rugosa* and *T. ciferrii* nuclear genome assemblies, with indication of their respective genome size, N50, GC content, coverage, percentage of mapped reads, variants per kilo-base (kb) and heterozygous (heter) variants per kb

Species (strain)	Size (Mb)	N50	GC (%)	Coverage (reads/position)	Mapped reads (%)	SNPs/kb	Heter SNPs/kb
Diutina rugosa (CBS613)	13.4	193 138	49.56%	175.9	64.07%	0.09	0.07
Trichomonascus ciferrii (CBS4856)	20.5	69 012	47.46%	209.6	98.35%	0.12	0.09

Read mapping and variant calling

Read mapping for all strains (Table 1) was performed with BWA-MEM v0.7.15 (Li 2013). Picard v2.1.1 (http://broadinstitute.github.io/picard/) 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).

Samtools v0.1.18 (Li *et al.* 2009) and Picard v2.1.1 (http://broadinstitute.github.io/picard/) 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 and filter variants with the tools HaplotypeCaller and VariantFiltration, respectively, as described by (Mixão *et al.* 2019). 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 genomecov v2.25.0 (Quinlan and Hall 2010).

Mitochondrial genome assembly

NOVOPlasty v2.7.2 (Dierckxsens *et al.* 2017) with default parameters was used to assemble *D. rugosa* and *T. ciferrii* mitochondrial genomes, taking as seed input the respective *Cox2* gene (accession numbers: KT832772.1 and DQ443088.1, respectively). The final assemblies were complete, as the assembly program was able to circularize each of them. Mitochondrial genome annotation was performed with MITOS2 (Bernt *et al.* 2013). Read mapping to these mitochondrial assemblies was performed as mentioned before for the nuclear genome.

Data availability

Data generated by this project can be found under the BioProject PRJNA531406, including sequencing data, genome assemblies and respective annotation. Phylomes can be found in PhylomeDB, with the phylome IDs 842 and 932. A list of species used for phylome reconstruction, and the results of the enrichment analysis can be found in Supplementary files 1 and 2, respectively. Plots related to the *k*-mer analysis in the genome assemblies are in Supplementary figure 1. Supplemental material available at figshare: https://doi.org/10.25387/g3.8945048.

RESULTS AND DISCUSSION

Genome sequencing and assembly

In this study we sequenced the type strains of *D. rugosa* and *T. ciferrii*, using an Illumina-based, pair-end sequencing strategy (see Materials and Methods). GC content and 27-mer count analyses of the sequencing reads revealed only one peak for each strain (Supplementary figure 1), suggesting that the two sequenced strains are highly homozygous. Based on the same 27-mer counts, we estimated genome sizes of approximately 13 Mb and 19 Mb for *D. rugosa* and *T. ciferrii*, respectively (see Materials and Methods). We next performed a *de novo* genome assembly for each of these species (see Materials and Methods). The final nuclear genome assembly of *D. rugosa* comprised 13.4 Mb, with 49.56% GC content and a N50 of 193,138 bp (Table 1). This assembly was divided in 171 contigs, of which 88 were longer than 25 kb, representing 97.7% of the genome. Automated gene prediction resulted in 5,821 protein-coding genes (see Materials and Methods). Despite the fact that the genome size was close to our estimations and similar to the one reported for the closely-related species *Diutina catenulata* (13.1 Mb), the number of predicted proteins was substantially lower than the 7,128 proteins annotated in the close relative *D. catenulata* (O'Brien *et al.* 2018). Furthermore, only 64.07% of the reads could be mapped to *D. rugosa* nuclear genome assembly. These observations made us question the completeness of our assembly. However, KAT (Mapleson *et al.* 2017) reported that 98.96% of 27-mers was represented in the assembly (Supplementary figure 1), and BUSCO (Waterhouse *et al.* 2018) reported 97.7% completeness of *D. rugosa* predicted proteome. Finally, most of the reads that did not map to the nuclear genome were found to correspond to the mitochondrial genome (see section "Mitochondrial genome assembly"). Thus, we consider that *D. rugosa* genome annotation is not significantly underestimating its gene content. It remains to be investigated whether the large number of proteins reported for *D. catenulata* is an annotation artifact or a real biological difference. There are no other available genomes for this genus, and the close relatives *M. guilliermondii* and *Scheffersomyces stipitis* have 5,920 and 5,841 annotated proteins, respectively (Jeffries *et al.* 2007; Butler *et al.* 2009).

The nuclear genome assembly of *T. ciferrii* comprised 20.5 Mb, with 47.46% GC content and a N50 of 60,012 bp (Table 1). This assembly entailed 584 contigs, of which 132 were longer than 25 kb, representing 84.3% of the genome. Genome annotation predicted 6,913 proteins (see Materials and Methods). To the best of our knowledge, there is no other genome assembly of the *Trichomonascus* genus published so far, which would allow us to have a better assessment of the quality of our assembly. Even so, 27-mer frequency analysis showed that 99.83% of *T. ciferrii* 27-mers was represented in the assembly (Supplementary figure 1), and BUSCO (Waterhouse *et al.* 2018) estimated 93.4% proteome completeness, suggesting a good representation of the *T. ciferrii* genome in our assembly. Read mapping and variant calling confirmed that both *D. rugosa* and *T. ciferrii* are highly homozygous, having 0.07 and 0.09 heterozygous SNPs/kb, respectively (Table 1). It is worth mentioning that both genomes present homozygous SNPs (0.02 SNPs/kb in *D. rugosa* and 0.03 SNPs/kb in *T. ciferrii*), which is unexpected as the reads were mapped on the respective assembly. This situation can probably be a result of errors introduced during the sequencing process or data analysis (*i.e.*, read assembly, read mapping or variant calling).

Mitochondrial genome assembly

As mentioned before, only 64.07% of *D. rugosa* sequencing reads mapped to the respective genome assembly. Thus, we decided to

assemble its mitochondrial genome (see Materials and Methods), in order to see whether the remaining reads could come from it. A final 41.8 kb circular mitochondrial genome assembly was obtained, suggesting that the assembly is complete (Figure 1a). Read mapping confirmed that 34.5% of *D. rugosa* sequencing reads corresponded to the mitochondrial genome, which suggests a high mitochondrial content in this yeast. We have assembled *T. ciferrii* mitochondrial genome as well, obtaining a circular assembly with 29.2 kb (Figure 1b), where 2.2% of *T. ciferrii* sequencing reads mapped. While we annotated 14 protein-coding genes in *D. rugosa* mitochondria, in *T. ciferrii* we annotated 16. The major difference between the two species involved the *nad4L* (associated to complex I) and *rps3* genes, which were absent in *D. rugosa*.

Comparative genomics

In order to elucidate particular characteristics of *D. rugosa* and *T. ciferrii* we decided to follow a comparative genomics approach, and compared their nuclear genomes/proteomes with other species. We reconstructed the complete collection of gene evolutionary histories (*i.e.*, the phylome) (Gabaldón 2008) for each of these two species, in the context of twenty-six other species (see Material and Methods, Supplementary file 1). We identified 770 species-specific genes for *D. rugosa* and 1,217 for *T. ciferrii*, from which only 247 and 391, respectively, had homologs in species which were not considered for phylome reconstruction (see Materials and Methods). In both species, species-specific genes were not enriched in any particular function. Interestingly, genes specifically duplicated in each of the two species seemed to be enriched in transmembrane transport activities, as well as, oxidoreductase activity (detailed information can be found in Supplementary file 2). As can be observed in the species tree (Figure 2a), *D. rugosa* belongs to the CUG-Ser1 clade, while *T. ciferrii* is close to *Yarrowia lipolytica*. This shows that although very distantly related these two emergent pathogens present gene duplications affecting similar functions. Furthermore, in the case of *D. rugosa*, it is worth noting an enrichment in aspartic-type endopeptidase activity and ferrichrome transporter activity (Supplementary file 2), as both have been reported as important for pathogenic behavior, particularly in *Candida* species (Heymann *et al.* 2002; Naglik *et al.* 2003).

An earlier study on *D. catenulata* genome revealed an interesting break in the *MAT* locus of this species (O'Brien *et al.* 2018). By comparison with the *MAT* locus of *M. guilliermondii* (Reedy *et al.* 2009), these authors reported the absence of *PAP* gene close to *MAT alpha1* (Figure 2b), being the *PAP* gene instead in a different contig of *D. catenulata* genome assembly (O'Brien *et al.* 2018). Furthermore, they found that this gene was phylogenetically closer to *PAP a* than to *PAP alpha* (O'Brien *et al.* 2018). To assess whether this characteristic is shared within the *Diutina* genus, we here inspected the *MAT* locus of *D. rugosa*. Contrary to *D. catenulata* (O'Brien *et al.* 2018), *D. rugosa MAT* locus corresponded to the *MAT a* allele, where, similarly to *M. guilliermondii*, we could only find *MAT a2* (Figure 2b). Moreover, when comparing to *M. guilliermondii* (Reedy *et al.* 2009), we could not identify any particular rearrangement in this locus. Regarding the *MAT* locus of *T. ciferrii*, we observed that, in contrast to *Y. lipolytica* where both *MAT a* and *MAT alpha* were described (Butler *et al.* 2004), it only presents the *MAT alpha* allele (Figure 2b). It is worth to mention that although there is a protein-coding gene in the place of *MAT alpha2*, this protein does not present any homolog and therefore we were only able to identify *MAT alpha1* in *T. ciferrii* genome (Figure 2b).

Concluding remarks

We have here reported the genomes of two emergent yeast pathogens, which are phylogenetically very distantly related, namely *D. rugosa* and *T. ciferrii*. These two reference genomes provide an important resource for the assessment of relevant aspects of these yeasts, including the genetic bases of their clinically relevant traits, as virulence and drug resistance. In addition, the two phylomes, which include a full

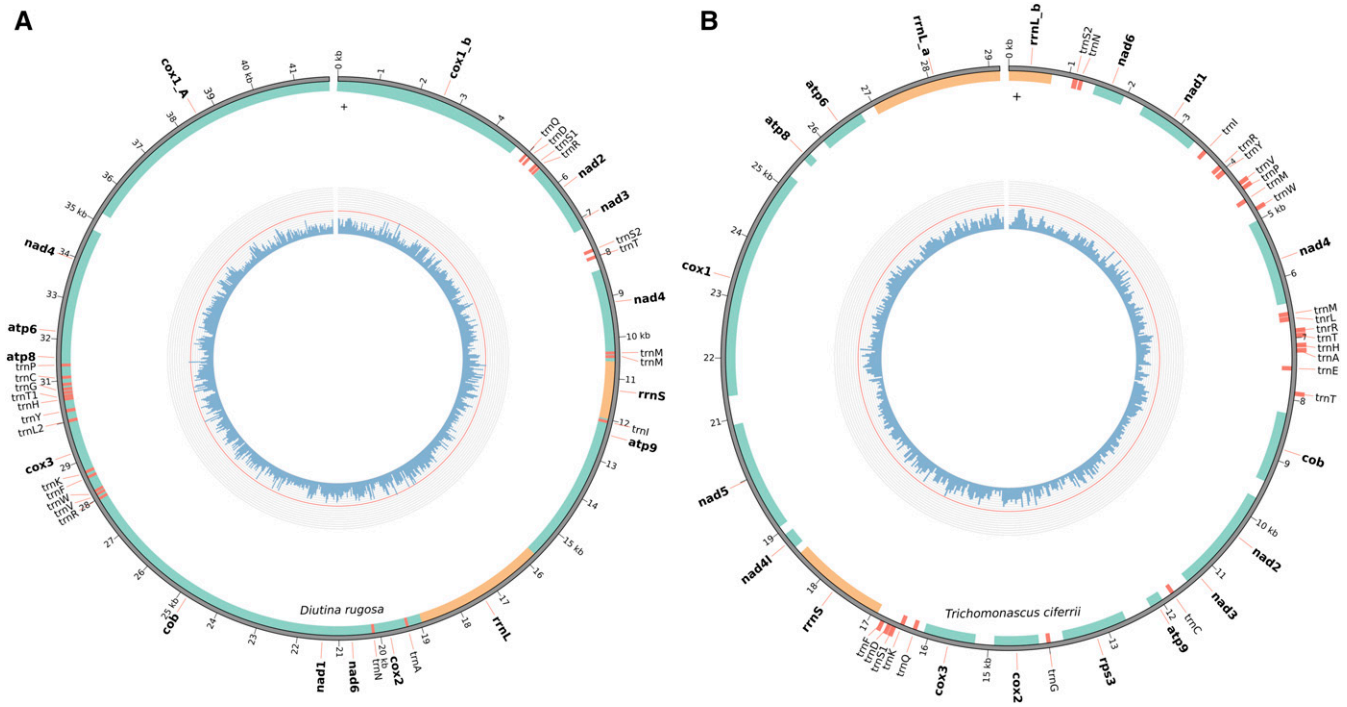

Figure 1 Mitochondrial genome representation of A) *D. rugosa* and B) *T. ciferrii*. Protein-coding genes are marked in green, tRNA genes are marked in red, and ribosomal genes are marked in orange. The blue histogram in the center represents the GC content variation.

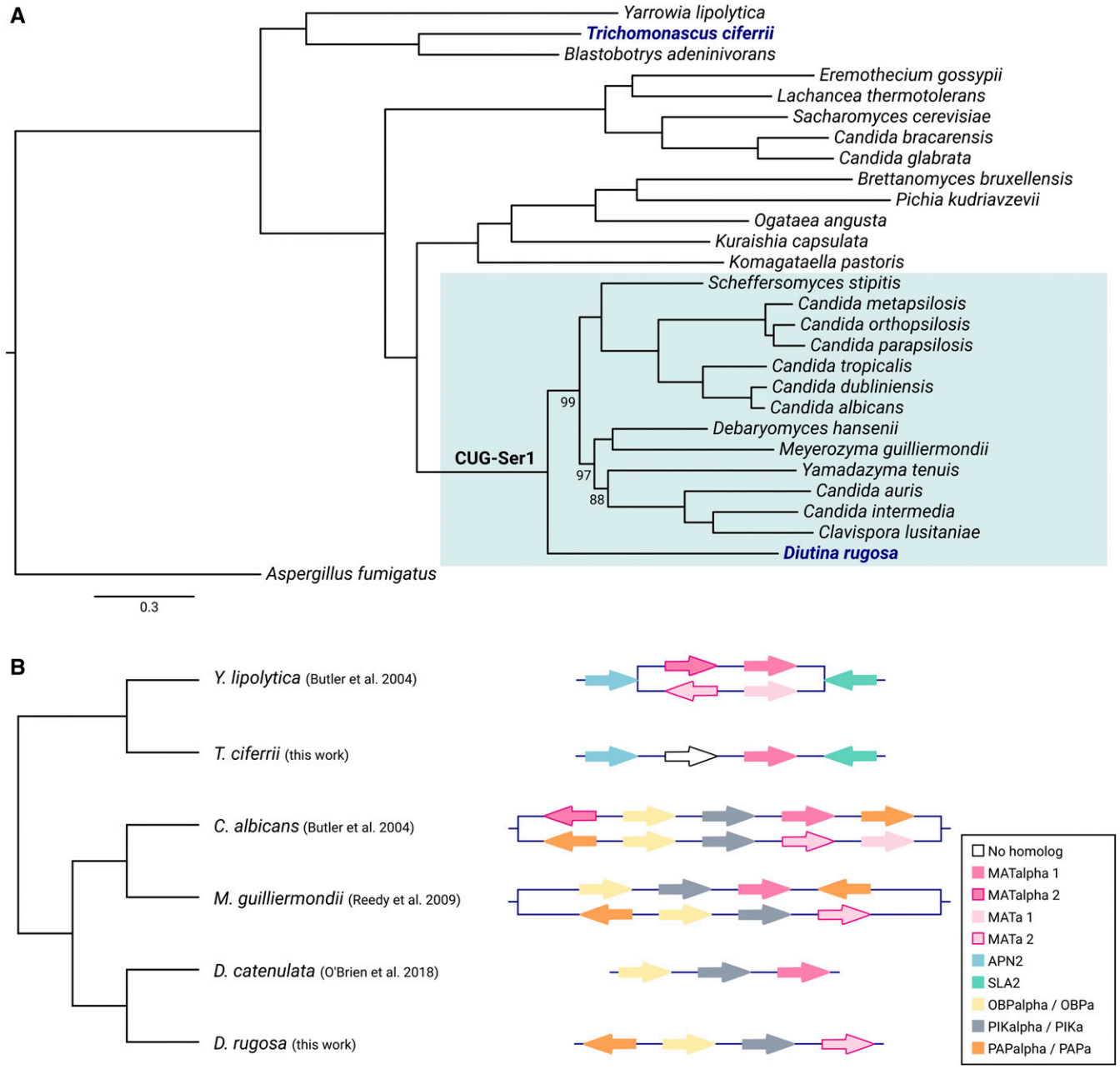

Figure 2 Comparative genomics of *D. rugosa* and *T. ciferrii* genomes. A) Maximum Likelihood phylogenetic tree of the concatenated alignment of 469 single genes, comprising 297,788 amino-acid positions. When the branch support is different from 100 the values are presented close to the respective branch. CUG-Ser1 clade is highlighted in blue. *D. rugosa* and *T. ciferrii* are marked in dark blue and bold. B) Schematic representation of the *MAT* locus of *D. rugosa* and *T. ciferrii* in comparison with closely-related species. The tree presents their phylogenetic relationship, but the branch length does not correspond to their phylogenetic distance. Each arrow represents a different gene with the color indicating the gene name.

repertoire of gene evolutionary histories and a catalog of orthologs and paralogs, can be used to trace the origin and evolution of genes of interest. Therefore, the data provided by this publication will certainly be of interest for the study of emergent yeast pathogens.

ACKNOWLEDGMENTS

This work received funding from the European Union's Horizon 2020 Research and Innovation Program 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-founded by European Regional Development Fund (ERDF); from the CERCA Program/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 No. ERC-2016-724173, and MSCA-747607. TG also receives support from an INB grant (PT17/0009/0023 – ISCIII-SGEFI/ERDF). CLF received funding from Christian Doppler Laboratory

for Fungal Infections: Avoid, find, and treat! The authors thank all Gabaldón lab members for helpful discussions on this study, specially Marina Marcet-Houben.

LITERATURE CITED

Al-Shahrour, F., P. Minguez, J. Tárraga, I. Medina, E. Alloza *et al.*, 2007 FatiGO: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. Nucleic Acids Res. 35: W91–W96. https://doi.org/10.1093/nar/gkm260

Bernt, M., A. Donath, F. Jühling, F. Externbrink, C. Florentz *et al.*, 2013 MITOS: improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69: 313–319. https://doi.org/10.1016/j.ympev.2012.08.023

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

Bretagne, S., C. Renaudat, M. Desnos-Ollivier, K. Sitbon, O. Lortholary *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. https://doi.org/10.1093/jac/dkx045

Brown, G. D., D. W. Denning, N. A. R. Gow, S. M. Levitz, M. G. Netea *et al.*, 2012 Hidden Killers: Human Fungal Infections. Science Translational Medicine 4: 165rv13–165rv13.

Butler, G., C. Kenny, A. Fagan, C. Kurischko, C. Gaillardin *et al.*, 2004 Evolution of the MAT locus and its Ho endonuclease in yeast species. Proc. Natl. Acad. Sci. USA 101: 1632–1637. https://doi.org/10.1073/pnas.0304170101

Butler, G., M. D. Rasmussen, M. F. Lin, M. A. S. Santos, S. Sakthikumar *et al.*, 2009 Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature 459: 657–662. https://doi.org/10.1038/nature08064

Crawshaw, W. M., N. R. MacDonald, and G. Duncan, 2005 Outbreak of *Candida rugosa* mastitis in a dairy herd after intramammary antibiotic treatment. Vet. Rec. 156: 812–813. https://doi.org/10.1136/vr.156.25.812

Dierckxsens, N., P. Mardulyn, and G. Smits, 2017 NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45: e18.

Gabaldón, T., 2008 Large-scale assignment of orthology: back to phylogenetics? Genome Biol. 9: 235. https://doi.org/10.1186/gb-2008-9-10-235

Gabaldón, T., and L. Carreté, 2016 The birth of a deadly yeast: tracing the evolutionary emergence of virulence traits in *Candida glabrata*. FEMS Yeast Res. 16: fov110. https://doi.org/10.1093/femsyr/fov110

Guitard, J., A. Angoulvant, V. Letscher-Bru, C. L'Ollivier, M. Cornet *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. https://doi.org/10.3109/13693786.2013.807444

Gunsilius, E., C. Lass-Flörl, C. M. Kähler, G. Gastl, and A. L. Petzer, 2001 *Candida ciferrii*, a new fluconazole-resistant yeast causing systemic mycosis in immunocompromised patients. Ann. Hematol. 80: 178–179. https://doi.org/10.1007/s002770000252

Gurevich, A., V. Saveliev, N. Vyahhi, and G. Tesler, 2013 QUAST: quality assessment tool for genome assemblies. Bioinformatics 29: 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Heymann, P., M. Gerads, M. Schaller, F. Dromer, G. Winkelmann *et al.*, 2002 The siderophore iron transporter of *Candida albicans* (Sit1p/Arn1p) mediates uptake of ferrichrome-type siderophores and is required for epithelial invasion. Infect. Immun. 70: 5246–5255. https://doi.org/10.1128/IAI.70.9.5246-5255.2002

Huerta-Cepas, J., S. Capella-Gutiérrez, L. P. Pryszcz, M. Marcet-Houben, and T. Gabaldón, 2014 PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome. Nucleic Acids Res. 42: D897–D902. https://doi.org/10.1093/nar/gkt1177

Jeffries, T. W., I. V. Grigoriev, J. Grimwood, J. M. Laplaza, A. Aerts *et al.*, 2007 Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*. Nat. Biotechnol. 25: 319–326. https://doi.org/10.1038/nbt1290

Jordà-Marcos, R., F. Alvarez-Lerma, M. Jurado, M. Palomar, J. Nolla-Salas *et al.*, 2007 Risk factors for candidaemia in critically ill patients: a prospective surveillance study. Mycoses 50: 302–310. https://doi.org/10.1111/j.1439-0507.2007.01366.x

Jung, D. S., D. Farmakiotis, Y. Jiang, J. J. Tarrand, and D. P. Kontoyiannis, 2015 Uncommon *Candida* Species Fungemia among Cancer Patients, Houston, Texas, USA. Emerg. Infect. Dis. 21: 1942–1950. https://doi.org/10.3201/eid2111.150404

Khunnamwong, P., N. Lertwattanasakul, S. Jindamorakot, S. Limtong, and M.-A. Lachance, 2015 Description of *Diutina* gen. nov., *Diutina siamensis* f.a. sp. nov., and reassignment of *Candida catenulata*, *Candida mesorugosa*, *Candida neorugosa*, *Candida pseudorugosa*, *Candida ranongensis*, *Candida rugosa* and *Candida scorzettiae* to the genus *Diutina*. Int. J. Syst. Evol. Microbiol. 65: 4701–4709. https://doi.org/10.1099/ijsem.0.000634

Kurtzman, C. P., and C. J. Robnett, 2007 Multigene phylogenetic analysis of the *Trichomonascus*, *Wickerhamiella* and *Zygoascus* yeast clades, and the proposal of *Sugiyamaella* gen. nov. and 14 new species combinations. FEMS Yeast Res. 7: 141–151. https://doi.org/10.1111/j.1567-1364.2006.00157.x

Lass-Flörl, C., 2009 The changing face of epidemiology of invasive fungal disease in Europe. Mycoses 52: 197–205. https://doi.org/10.1111/j.1439-0507.2009.01691.x

Li, H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303.3997v.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lopes Colombo, A., A. S. Azevedo Melo, R. F. Crespo Rosas, R. Salomão, M. Briones *et al.*, 2003 Outbreak of *Candida rugosa* candidemia: an emerging pathogen that may be refractory to amphotericin B therapy. Diagn. Microbiol. Infect. Dis. 46: 253–257. https://doi.org/10.1016/S0732-8893(03)00079-8

Luo, R., B. Liu, Y. Xie, Z. Li, W. Huang *et al.*, 2012 SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1: 18. Erratum: 4: s13742-015-0069-2. https://doi.org/10.1186/2047-217X-1-18

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

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297–1303. https://doi.org/10.1101/gr.107524.110

Minces, L. R., K. S. Ho, P. J. Veldkamp, and C. J. Clancy, 2009 *Candida rugosa*: a distinctive emerging cause of candidaemia. A case report and review of the literature. Scand. J. Infect. Dis. 41: 892–897. https://doi.org/10.3109/00365540903161531

Mixão, V., A. P. Hansen, E. Saus, T. Boekhout, C. Lass-Florl *et al.*, 2019 Whole-Genome Sequencing of the Opportunistic Yeast Pathogen *Candida inconspicua* Uncovers Its Hybrid Origin. Front. Genet. 10: 383. https://doi.org/10.3389/fgene.2019.00383

Moretti, A., D. Piergili Fioretti, L. Boncio, P. Pasquali, and E. Del Rossi, 2000 Isolation of *Candida rugosa* from Turkeys. Zentralbl. Veterinarmed. B. 47: 433–439.

Naglik, J. R., S. J. Challacombe, and B. Hube, 2003 *Candida albicans* secreted aspartyl proteinases in virulence and pathogenesis. Microbiol. Mol. Biol. Rev. 67: 400–428. https://doi.org/10.1128/MMBR.67.3.400-428.2003

NCBI Resource Coordinators, 2015 Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 43: D6–D17. https://doi.org/10.1093/nar/gku1130

O'Brien, C. E., C. G. P. McCarthy, A. E. Walshe, D. R. Shaw, D. A. Sumski *et al.*, 2018 Genome analysis of the yeast *Diutina catenulata*, a member of the Debaryomycetaceae/Metschnikowiaceae (CTG-Ser) clade. PLoS One 13: e0198957. https://doi.org/10.1371/journal.pone.0198957

Pérez-Hansen, A., C. Lass-Flörl, and M. Lackner, and Rare Yeast Study Group, 2019 Antifungal susceptibility profiles of rare ascomycetous yeasts. J. Antimicrob. Chemother. 74: 2649–2656. https://doi.org/10.1093/jac/dkz231

Pfaller, M. A., M. Castanheira, and R. N. Jones, 2012 Advances in Antifungal Susceptibility Testing of *Candida*, 2010–2012. Curr. Fungal Infect. Rep. 6: 141–153. https://doi.org/10.1007/s12281-012-0092-z

Pfaller, M. A., and D. J. Diekema, 2007 Epidemiology of Invasive Candidiasis: a Persistent Public Health Problem. Clin. Microbiol. Rev. 20: 133–163. https://doi.org/10.1128/CMR.00029-06

Pfaller, M. A., D. J. Diekema, A. L. Colombo, C. Kibbler, K. P. Ng *et al.*, 2006 *Candida rugosa*, an emerging fungal pathogen with resistance to azoles: geographic and temporal trends from the ARTEMIS DISK antifungal surveillance program. J. Clin. Microbiol. 44: 3578–3582. https://doi.org/10.1128/JCM.00863-06

Pfaller, M. A., D. J. Diekema, D. L. Gibbs, V. A. Newell, D. Ellis *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. https://doi.org/10.1128/JCM.02117-09

Pryszcz, L. P., and T. Gabaldón, 2016 Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Res. 44: e113. https://doi.org/10.1093/nar/gkw294

Pryszcz, L. P., T. Németh, E. Saus, E. Ksiezopolska, E. Hegedűsová *et al.*, 2015 The Genomic Aftermath of Hybridization in the Opportunistic Pathogen *Candida metapsilosis*. PLoS Genet. 11: e1005626. Erratum: 12: e1006202. https://doi.org/10.1371/journal.pgen.1005626

Quinlan, A. R., and I. M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842. https://doi.org/10.1093/bioinformatics/btq033

Reedy, J. L., A. N. Floyd, and J. Heitman, 2009 Mechanistic plasticity of sexual reproduction and meiosis in the *Candida* pathogenic species complex. Curr. Biol. 19: 891–899. https://doi.org/10.1016/j.cub.2009.04.058

Ropars, J., C. Maufrais, D. Diogo, M. Marcet-Houben, A. Perin *et al.*, 2018 Gene flow contributes to diversification of the major fungal pathogen Candida albicans. Nat. Commun. 9: 2253. https://doi.org/10.1038/s41467-018-04787-4

Saha, K., N. K. Sit, A. Maji, and D. Jash, 2013 Recovery of fluconazole sensitive *Candida ciferrii* in a diabetic chronic obstructive pulmonary disease patient presenting with pneumonia. Lung India 30: 338–340. https://doi.org/10.4103/0970-2113.120614

Scaccabarozzi, L., C. Locatelli, G. Pisoni, G. Manarolla, A. Casula *et al.*, 2011 Short communication: Epidemiology and genotyping of *Candida rugosa* strains responsible for persistent intramammary infections in dairy cows. J. Dairy Sci. 94: 4574–4577. https://doi.org/10.3168/jds.2011-4294

Schröder, M. S., K. Martinez de San Vicente, T. H. R. Prandini, S. Hammel, D. G. Higgins *et al.*, 2016 Multiple Origins of the Pathogenic Yeast *Candida orthopsilosis* by Separate Hybridizations between Two Parental Species. PLoS Genet. 12: e1006404. https://doi.org/10.1371/journal.pgen.1006404

Stamatakis, A., 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stanke, M., and B. Morgenstern, 2005 AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33: W465–W467. https://doi.org/10.1093/nar/gki458

Stavrou, A. A., V. Mixão, T. Boekhout, and T. Gabaldón, 2018 Misidentification of genome assemblies in public databases: The case of *Naumovozyma dairenensis* and proposal of a protocol to correct misidentifications. Yeast 35: 425–429. https://doi.org/10.1002/yea.3303

Thorvaldsdóttir, H., J. T. Robinson, and J. P. Mesirov, 2013 Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14: 178–192. https://doi.org/10.1093/bib/bbs017

UniProt Consortium, 2019 UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 47: D506–D515. https://doi.org/10.1093/nar/gky1049

Upadhyay, S., T. Wadhwa, and S. Sarma, 2018 A Series of Six Cases of *Candida Ciferrii* Infection in a Tertiary Care Centre of North India. Journal of Advances in Medicine and Medical Research 27: 1–5. https://doi.org/10.9734/JAMMR/2018/43742

Villanueva-Lozano, H., R. J. Treviño-Rangel, C. L. Hernández-Balboa, G. M. González, and M. F. Martínez-Reséndez, 2016 An unusual case of *Candida ciferrii* fungemia in an immunocompromised patient with Crohn's and Mycobacterium bovis disease. J. Infect. Dev. Ctries. 10: 1156–1158. https://doi.org/10.3855/jidc.8228

Waterhouse, R. M., M. Seppey, F. A. Simão, M. Manni, P. Ioannidis *et al.*, 2018 BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol. 35: 543–548. https://doi.org/10.1093/molbev/msx319

Zhang, Z., S. Schwartz, L. Wagner, and W. Miller, 2000 A greedy algorithm for aligning DNA sequences. J. Comput. Biol. 7: 203–214. https://doi.org/10.1089/10665270050081478

Communicating editor: A. Rokas