Generating genomic platforms to study *Candida albicans* pathogenesis

Mélanie Legrand[1][§], Sophie Bachellier-Bassi[1][§], Keunsook K. Lee[2], Yogesh Chaudhari[2],

Hélène Tournu[3,4], Laurence Arbogast[1], Hélène Boyer[1], Murielle Chauvel[1], Vitor Cabral[1,5],

Corinne Maufrais[1,6], Audrey Nesseir[1,5], Irena Maslanka[2], Emmanuelle Permal[1], Tristan

Rossignol[1], Louise A. Walker[2], Ute Zeidler[1], Sadri Znaidi[1], Floris Schoeters[3,4], Charlotte

Majgier[7], Renaud A. Julien[7], Laurence Ma[8], Magali Tichit[8], Christiane Bouchier[8], Patrick

Van Dijck[3,4], Carol A. Munro[2,*], and Christophe d'Enfert[1,*]

[1] Fungal Biology and Pathogenicity Unit, Department of Mycology, Institut Pasteur, INRA,

Paris, 75015, France

[2] MRC Centre for Medical Mycology, Institute of Medical Sciences, University of Aberdeen,

Foresterhill, Aberdeen, UK

[3] VIB-KU Leuven Center for Microbiology, Leuven, Belgium

[4] Laboratory of Molecular Cell Biology, KU Leuven, Leuven, Belgium

[5] Univ. Paris Diderot, Sorbonne Paris Cité, Cellule Pasteur, Paris, 75015, France

[6] Institut Pasteur - Bioinformatics and Biostatistics Hub - C3BI, USR 3756 IP CNRS - Paris,

France

[7] Modul-Bio; Parc Scientifique Luminy Biotech II ; Marseille, France.

[8] Institut Pasteur - Biomics Pole - CITECH - Paris, France

[§] The authors wish it to be known that, in their opinion, the first two authors should be

regarded as joint First Authors

Corresponding authors: Christophe d'Enfert and Carol Munro

* To whom correspondence should be addressed. Tel: +33 (0) 140613257; Fax: +33 (0) 140613456; Email: christophe.denfert@pasteur.fr. Correspondence may also be addressed to Carol Munro, Tel: +44 (0) 1224-437485; Fax: +44 (0) 1224-437506; Email: c.a.munro@abdn.ac.uk.

Present Address: Yogesh Chaudhari, Biosciences Department, University of Exeter, Exeter, EX4 4QD, UK; Hélène Tournu, Department of Clinical Pharmacy, University of Tennessee, Memphis, TN, 38163, USA; Laurence Arbogast, Brain plasticity in response to the environment Group, Institut Pasteur, Paris, 75015, France; Hélène Boyer, Laboratoire d'Immunologie, Hôpital Henri Mondor, Créteil, 94000, France; Vitor Cabral, Bacterial Signalling Group, Instituto Gulbenkian de Ciência, Oeiras, 2780-156, Portugal; Audrey Nesseir, Lycée Grégor Mendel, Vincennes, 94300, France; Emmanuelle Permal, Toulouse, 31000, France; Tristan Rossignol, MICALIS, INRA, Jouy-enJosas, 78352, France; Ute Zeidler, Sandoz, Unterach, 4866, Austria; Sadri Znaidi, Laboratory of Molecular Microbiology, Vaccinology & Biotechnology Development, Institut Pasteur de Tunis, Tunis-Belvédère, 1002, Tunisia; Magali Tichit, Laboratoire Histopathologie Humaine et Modèles Animaux, Institut Pasteur, Paris, 75015, France.

ABSTRACT

The advent of the genomic era has made elucidating gene function at large scale a pressing challenge. ORFeome collections, whereby almost all ORFs of a given species are cloned and can be subsequently leveraged in multiple functional genomic approaches, represent valuable resources towards this endeavor. Here we provide novel, genome-scale tools for the study of *Candida albicans*, a commensal yeast that is also responsible for frequent superficial and disseminated infections in humans. We have generated an ORFeome collection composed of 5,102 ORFs cloned in a Gateway™ donor vector, representing 83% of the currently annotated coding sequences of *C. albicans*. Sequencing data of the cloned ORFs are available in the CandidaOrfDB database at http://candidaorfeome.eu. We also engineered 49 expression vectors with a choice of promoters, tags, and selection markers and demonstrated their applicability to the study of target ORFs transferred from the *C. albicans* ORFeome. In addition, the use of the ORFeome in the detection of protein-protein interaction was demonstrated. Mating-compatible strains as well as Gateway™-compatible two-hybrid vectors were engineered, validated and used in a proof of concept experiment. These unique and valuable resources should greatly facilitate future functional studies in *C. albicans* and the elucidation of mechanisms that underlie its pathogenicity.

INTRODUCTION

Over the last decade, there has been an exponential growth in the quantity of available genome sequence data due to the very rapid progress in sequencing technology. In 2004, the genome sequence of the human fungal pathogen *Candida albicans* was released as Assembly 19 (1). With the challenge of working with a heterozygous diploid organism, new computational methods had to be developed and resulted in the release in 2013 of Assembly 22, an assembly of a completely phased diploid genome sequence for the standard *C. albicans* reference strain SC5314 (2). This opened new perspectives to understand the genetic basis and functional mechanisms that underlie pathogenesis and evolution in this organism.

Although *C. albicans* gene sequences have been available to the community for more than a decade (3,4), only 1,670 out of the 6,198 predicted protein-coding genes have been characterized as of January 31st, 2018 according to the Candida Genome Database (5). Nowadays, the growing availability of whole-genome datasets has encouraged a shift towards the development of functional genomics and systems biology, enabling analysis of high-throughput whole-genome assays to better understand biological networks. In this context, major efforts have been made to generate large-scale deletion (6-9) or overexpression (10-12) mutant collections. Genome-wide ORF libraries or ORFeomes represent useful resources for the implementation of approaches used to elucidate gene function (13). Besides the development of collections of overexpression mutants (11), ORFeomes facilitate approaches aimed at evaluating protein subcellular localization and identifying protein-protein interactions (yeast two-hybrid) both at steady state and in response to environmental stimuli (14).

Large-scale cloning projects, with the goal of cloning all predicted ORFs into flexible recombinational vectors, have been described for several model organisms, including

Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, *Escherichia coli* K-12, *Arabidopsis thaliana*, *Xenopus laevis* and *Drosophila melanogaster*, as well as infectious microorganisms such as *Brucella melitensis, Plasmodium falciparum*, *Helicobacter pylori*, *Chlamydia pneumonia*, *Staphylococcus aureus* and viruses (15-28). Human ORFeomes have also been generated and made publicly available, with the latest version containing sequence-confirmed ORFs for more than 11,000 human genes (29). Most ORFeome libraries are created using the highly versatile Gateway™ cloning approach (30). The Gateway™ technology employs the recombination system of bacteriophage lambda with two sets of reactions: the BP reaction, catalysed by Gateway™ BP Clonase™, facilitates recombination between *attB* and *attP* sequences, and the LR reaction, catalysed by Gateway™ LR Clonase™, promotes recombination between *attL* and *attR* sequences (30). Previously, we have reported the development of a first generation *C. albicans* ORFeome, in which 644 full-length ORFs, encoding signaling proteins (transcription factors and kinases), as well as cell wall proteins and proteins involved in DNA processes (repair, replication and recombination), were cloned in the Gateway™ vector pDONR207 (31). Here we describe the generation and validation of the latest *Candida albicans* ORFeome collection, which consists of 5,102 ORFs in pDONR207, allowing the transfer of the cloned genes into a variety of *C. albicans* Gateway™-compatible expression vectors. To this end, we also provide a panel of 15 expression vectors that differ in the combination of promoter/selection marker they are carrying, and give different options regarding the tagging of the cloned gene product. These 15 expression vectors have been validated using the filament-specific transcription factor *UME6*. In addition, we also provide a secondary collection of 34 expression vectors with additional options regarding the choice of the promoter and the tagging of the cloned gene product. Finally, among the numerous applications of ORFeome collections, we present a

powerful association of the *C. albicans* ORFeome and a *C. albicans*-adapted two-hybrid system (32) that will allow systematic protein-protein interaction screening in *C. albicans*.

MATERIAL AND METHODS

Strains and growth conditions

C. albicans strains used in this study are listed in **Table 1**. *C. albicans* strains were routinely cultured at 30°C in YPD medium (1% yeast extract, 2% peptone, 2% dextrose), synthetic dextrose (SD) medium (0.67% yeast nitrogen base, 2% dextrose) or nourseothricin-containing YPD medium (YPD + 200 µg/ml Nourseothricin). Solid media were obtained by adding 2% agar.

E. coli strains were routinely cultured at 30°C or 37°C in LB or 2YT supplemented with 10 µg/ml gentamicin, 50 µg/ml kanamycin or 50 µg/ml ticarcillin. Solid media were obtained by adding 2% agar.

DH5α *E. coli* cells competent for transformation with DNA were prepared according to Hanahan *et al*. (33).

The ORFeome collection is being distributed by the Centre International de Ressources Microbiennes (CIRM). The 49 expression vectors are available from Addgene, while two-hybrid strains and vectors are now accessible from the Belgian Co-ordinated Collections of Micro-organisms (BCCM).

Generation of an entry clone collection

Primer design. The DNA sequence of *C. albicans* was obtained from the Candida Genome Database Release 21 (www.candidagenome.org), before Assembly 22 was released. Gene-specific forward primers were designed by adding the sequence 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTG-3' to the 5' end of the first 30

nucleotides of each ORF. Gene-specific reverse primers were designed by adding the sequence 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTC-3' to the 5' end of the last 30 nucleotides of each ORF excluding the Stop codon. PCR fragments amplified from these primers are compatible for C-terminal tagging with destination vectors containing the Gateway™ cassette A. A total of 6,205 primer pairs were obtained from Invitrogen in a 96-well format and are listed in **Supplemental Table S1**.

Gateway™ cloning of the C. albicans *ORFeome*. The detailed method for the cloning of *C. albicans* ORFs in the pDONR207 vector has been described (34). Briefly, ORFs, ranging from 90 bp to 5,295 bp, were amplified from genomic DNA of *C. albicans* strain SC5314 in 96-well plates using the Thermo Scientific Phusion High-Fidelity DNA Polymerase and 30 cycles of amplification, with elongation time varying from 1 to 3 min according to the ORF size. After ethanol precipitation, Gateway™-compatible amplified ORFs were recombined into pDONR207 (Invitrogen) using the Gateway™ BP Clonase™ II Enzyme Mix (Invitrogen). Reaction mixes containing pDONR207, the PCR products and the BP Clonase™ were incubated overnight in 96-well plates, at room temperature. After adding proteinase K (Invitrogen) and incubating 10 min at 37°C, the BP reactions were directly used for bacterial transformation. 45-50 μl of chemically competent *E. coli* DH5α were added to the BP reactions and incubated for 30 min at 4°C. After heat-shock at 42°C for 35 s, 150 μl of SOC medium (20 ml YPD + 2 ml LB + 1 ml Hepes) was added to the transformation reactions and the samples were covered with breathing films and incubated for 1.5 h at 37°C with shaking. Then 100 μl of each sample were plated onto LB agar containing 10 μg/ml Gentamicin and incubated overnight at 37°C. The remainder of the transformation reactions were stored at -80°C in 30% glycerol. A single colony from each transformation reaction was inoculated in 400 μl 2YT + 10 μg/ml Gentamicin in 96 deep-well microplates. After 36h-growth at 37°C, the cultures were used for plasmid extraction and for -80°C storage.

Validation of entry clones by DNA sequencing and bioinformatics analysis. Sanger sequencing of the 5'-end of each BP clone was performed with the 207VER-F oligonucleotide (see **Supplemental Table 2** for oligonucleotide sequences except those used to amplify ORFs) to confirm that the expected ORF had been cloned. The 3'-ends were also sequenced with the 207VER-R oligonucleotide to ensure that the oligonucleotides did not carry deletions. All inserts validated by Sanger sequencing were systematically subjected to full-length sequence analysis using Illumina technology as described in Chauvel *et al*. (11). If nonsense or frameshift mutations were observed in a cloned ORF, another colony was checked, or the ORF reamplified and cloned again. If these attempts were unsuccessful, cloning of the ORF was abandoned. Clones containing contiguous deletions or insertions of multiples of 3 bp were accepted. Missense mutations and those located within introns were accepted. It should be noted that further analysis of each mutation-containing ORF is required to determine whether or not mutations affect the function of the ORF.

CandidaOrfDB database

CandidaOrfDB was created to integrate the data of the *C. albicans* ORFeome project and is available at http://candidaorfeome.eu. The database is a relational database using the SQL database server Oracle with a web interface developed using J2EE. The alignment algorithm uses *biojava3-alignment* Java library, implemented with a NeedlemanWunsch aligner, configured with a gap open penalty of 5 and a gap extension penalty of 2. The substitution matrix used is named "nuc-4_4" and was created by Todd Lowe (https://github.com/sbliven/biojava/blob/master/biojava3-alignment/src/main/resources/nuc-4_4.txt). The alignment algorithm was specifically customized for CandidaOrfDB to work on segments of 500 codons, which provides an optimized alignment performance for *C. albicans* average sequence length.

Clone access

The *C. albicans* ORFeome clones are available at the Centre International de Ressources Microbiennes (CIRM - https://www6.inra.fr/cirm_eng/).

Destination plasmids

All destination plasmids constructed in this study were derived from CIp-P$_{TET}$-GTW (11) and were confirmed by Sanger sequencing. Plasmid sequences have been submitted to GenBank. Accession numbers are given in **Table 2**. All oligonucleotides used for PCR and/or sequencing are listed in **Supplemental Table 2**.

SP cloning. First, a spacer sequence (SP) was amplified from the *E. coli kanR* gene borne on pCR-topo-blunt plasmid (Invitrogen) with primers SP3 and SP4. The PCR product was cloned in the SacII site of CIp-P$_{TET}$-GTW, yielding CIp-P$_{TET}$-GTW-SP, also referred to as pCA-Dest1100. This 390 bp-long sequence is flanked by the SP1 and SP2 Illumina paired-end sequencing primers 1 and 2 and can be used to insert molecular barcodes if expression plasmids are to be used in signature-tagged mutagenesis approaches.

Epitope tags. We then inserted in frame tags either upstream of attR1 or downstream of attR2 to allow N-terminal or C-terminal protein tagging, respectively. For the cloning of N-terminal tags, we used pUC-attR1-CmR, containing a PciI-BglII fragment from CIp-GTW cloned into the PciI and BamHI sites of pUC18. CIp-GTW was obtained after deleting P$_{TET}$ from CIp-P$_{TET}$-GTW with Acc65I and HpaI. For cloning of the 3xHA coding sequence, two oligonucleotides (oligo1_PciHABsrG and oligo2_PciHABsrG) containing three HA epitopes were hybridized, gel purified, and cloned into pUC-attR1-CmR cut with PciI and BsrGI. The resulting plasmid was then used as a PCR template with primers GTW02 and GTW03. The resulting PCR product was cut with BspEI and HpaI, and cloned into the same sites of pCA-

Dest1100 to yield pCA-Dest1110. For cloning of the GFP tag, the GFP gene was amplified from pFA-GFP-URA3 (35) with primers GTW04 and GTW05. The PCR product was cut with PciI and EcoRV and ligated into the same sites of pUC-attR1-CmR. The resulting plasmid was then used as a PCR template with primers GTW07 and GTW03. The resulting PCR product was cut with ScaI and BspEI and ligated into pCA-Dest1100 cut with HpaI and BspEI, yielding pCA-Dest1120. For cloning of the TAP-tag, the TAP-tag coding region was PCR amplified from pFA-TAP-URA3 (11) with primers GTW13 and GTW14. The resulting PCR product was then cut with PciI and BsrGI, and ligated into the same sites of pUC-attR1-CmR. The resulting plasmid was then digested with EcoRV and BspEI and the smallest restriction fragment was ligated into pCA-Dest1100 cut with HpaI and BspEI to yield pCA-Dest1130. For the cloning of C-terminal tags, we generated pUC-attR2 by inserting a SalI-HindIII fragment from pCA-Dest1100 into the same sites of pUC18. For cloning of the 3xHA epitope, pCaMPY-3xHA (36) was used as a PCR template with primers GTW20 and GTW21, and the resulting PCR product cut with BsrGI and NsiI was ligated in the same sites of pUC-attR2. The resulting plasmid was cut with SalI and NsiI and the fragment cloned in the same sites of pCA-Dest1100, yielding pCA-Dest1101. For cloning the GFP tag, the GFP gene was amplified from pFA-GFP-URA3 (35) with primers GTW11 and GTW12. The resulting PCR product was cut with NsiI and EcoRV and ligated into the same sites of pUC-attR2. The resulting plasmid was then cut with SalI and NsiI and the GFP-bearing fragment ligated into the same sites of pCA-Dest1100, yielding pCA-Dest1102. For cloning of the TAP-tag, the TAP-tag coding region was amplified from CIp10-P$_{PCK1}$-GTW-TAPtag (11) with primers GTW15 and GTW16. The PCR product was cut with BsrGI and NsiI and cloned in the same sites of pUC-attR2. The resulting plasmid was digested with SalI and NsiI, and the TAP-containing fragment ligated in the same sites of pCA-Dest1100, yielding pCA-Dest1103.

Promoters. We replaced P$_{TET}$ with either P$_{PCK1}$ (obtained from CIp10-P$_{PCK1}$-GTW-TAPtag, (11)), yielding the pCA-Dest12xx series; P$_{TDH3}$ (cut from pKS-P$_{TDH3}$; P$_{TDH3}$ was PCR amplified from SC5314 genomic DNA with oligos SZ11 and SZ12, and the fragment inserted into pBluescript-KS (+) cut with XhoI and EcoRV) yielding the pCA-Dest13xx series; or P$_{ACT1}$, cut from CIp10::P$_{ACT1}$-gLUC59 (37), yielding the pCA-Dest14xx series.

Selection markers. In all plasmids except the P$_{TET}$ series, the auxotrophic marker *URA3* was replaced with the *NAT1* marker conferring nourseothricin resistance. Briefly, a XbaI-DraIII fragment or SpeI-NaeI fragment was cut from pUC-NAT1 and inserted into the similarly cut vectors. pUC-NAT1 was built by amplifying *NAT1* under the control of P$_{TEF1}$ promoter with oligonucleotides UZ50 and UZ51, and using pFA6-SAT1 (38) as a template, and cloning the AatII-cut fragment into the same site of pUC18.

The 49 expression vectors are available from Addgene (https://www.addgene.org).

Validation of 15 expression plasmids

Construction of C. albicans UME6-*overexpression strains*. Detailed methods for the transfer of *C. albicans* ORFs from pDONR207 into the expression plasmids as well as the integration of the resulting expression plasmids at the *RPS1* locus have been described (34). Briefly, the *UME6* ORF was transferred from the entry clone into each one of 15 expression vectors by using the Gateway™ LR Clonase™ II Enzyme Mix (Invitrogen). After *E. coli* transformation, the plasmids were verified by EcoRV digest. The *URA3*- and *NAT1*-bearing expression plasmids were digested by StuI and transformed into *C. albicans* strain CEC161 or CEC369, respectively, according to Walther and Wendland (39). Transformants were selected for prototrophy or nourseothricin resistance, respectively, and verified by PCR using primer CIpUL with (i) primer CIpUR for the *URA3*-bearing plasmids and (ii) primer

CgSAT1-rev for the *SAT1*-bearing plasmids, that yield 1 kb and 1.6 kb products, respectively, if integration of the OE plasmid has occurred at the *RPS1* locus.

Induction of the Tet-On system. Overexpression from P_{TET} was achieved by the addition of anhydrotetracycline (ATc, 3 μg/ml; Fisher Bioblock Scientific) in YPD at 30°C (40). Overexpression experiments were carried out in the dark, as ATc is light sensitive.

Microscope analysis for filamentation. Cells were observed with a Leica DM RXA microscope (Leica Microsystems) with an x40 oil-immersion objective.

Analysis of TAP- and 3HA-tagged proteins by Western blots.

For the P_{TET}, P_{TDH3} and P_{ACT1} strains, a 30 ml culture in YPD or YPD+ATc3 was inoculated at OD600 = 0.2 with a freshly grown colony and incubated at 30°C with shaking until OD600 reaches ~1. For the P_{PCK1} strain, cells were scraped off the agar of Petri dish cultures and resuspended in 1 ml dH$_2$O after centrifugation. The cell suspension was used to inoculate a 40 ml SD culture at OD600 = 0.2.

20 ODs of exponentially growing cells were collected by centrifugation and resuspended in lysis buffer (9M Urea, 1.5% w/v dithiothreitol, 2-4% w/v CHAPS, and 1.5 M Tris pH9.5). Homogenization of the cells was achieved using a FastPrep bead beater. After homogenization, the lysed cells were centrifuged at 13,000 rpm for 10 min at room temperature, and the supernatant containing the solubilized proteins was used directly or stored at -80 °C.

Proteins were separated on an Invitrogen 8% NuPage gel, transferred onto nitrocellulose. TAP- and 3HA-tagged proteins were detected using peroxidase-coupled anti-peroxidase and anti-HA-peroxydase antibodies (Sigma and Roche, respectively) and an ECL kit (GE Healthcare).

Construction of opaque mating-compatible two-hybrid strains and Gateway™-compatible two-hybrid vectors

Mating-compatible strains were generated by deletion of MTL**a** or MTLα locus using the *SAT1* flipper cassette (38) in the two-hybrid strain background SC2H3 (32). To delete the MTL**a** locus, flanking regions were amplified with primers MTLa-5'F and MTLa-5'R, and with MTLa-3'F and MTLa-3'R. To delete the MTLα locus, homologous regions were amplified with primers MTLα-5'F and MTLα-5'R, and with MTLα-3'F and MTLα-3'R. Both fragments for each locus were cloned into pSAT1 (38), at SacI/NotI and XhoI/KpnI sites respectively. SacI/KpnI deletion constructs were transformed in SC2H3 (32) and transformants were selected on YPD medium containing 200 μg/ml of nourseothricin. Positive SC2H3a and SC2H3α strains were subsequently grown in maltose medium for induction of the recombinase and loss of the *SAT1* flipper cassette.

For induction of opaque switching, a *WOR1* fragment was amplified with primers WOR1F and WOR1R and cloned into pNIM1 vector (40) at SalI and BglII sites. SC2H3a and SC2H3α strains were transformed with pNIM1-WOR1 with selection on nourseothricin-containing medium. The resulting strains, SC2H3a-pWOR1 and SC2H3α−pWOR1, were grown in presence of doxycycline (50 μg/ml) to induce the opaque state (41). Opaque mating-compatible strains were finally transformed with prey or bait plasmid and selected on SC-ARG or SC-LEU respectively.

Bait and prey two-hybrid vectors, pC2HB and pC2HP respectively (32) were converted into Gateway™ destination vectors using the Gateway™ Vector Conversion System to generate pC2HB-GC and pC2HP-GC respectively. The genes encoding the bait protein Hst7 (Orf19.469) and the prey Cek1 (Orf19.2886) were transferred by LR reactions from the donor collection of vectors into pC2HB-GC and pC2HP-GC destination vectors respectively.

Mating approach and protein-protein interaction detection

Opaque cells of SC2H3a and SC2H3α, expressing VP16-Cek1 (Arg+) and lexA-Hst7 (Leu+) respectively, were crossed as follows. Opaque cells of both types were mixed at 1 x 10⁶ cells each in Spider medium (42) and incubated at 23°C for 24h. Resulting tetraploid cells were selected on SC-LEU-ARG medium for 48h incubation at 30°C, and transferred to SC-HIS-MET medium for protein-protein interaction detection. Chromosome loss was induced on pre-sporulation medium at 37°C for 10 days as described previously (43).

RESULTS

Establishment of a sequence-validated *Candida albicans* ORFeome

Our objective was to develop a *C. albicans* ORFeome encompassing as many of the 6,205 ORFs predicted in Assembly 21 of the *C. albicans* genome as possible (44). To this aim, forward and reverse oligonucleotides with, respectively, *attB* and *attP* sequences at their 5' ends were synthesized for each of the 6,205 ORFs (**Supplemental Table S1**), used in independent PCR reactions with genomic DNA of *C. albicans* strain SC5314 (45) and the resulting PCR products were cloned in pDONR207 using Gateway™ recombinational cloning (30). The resulting plasmids were then subjected to Sanger sequencing at the 5' and 3' ends of the cloned ORFs and to Illumina sequencing throughout the cloned ORFs.

Among the 6,205 ORF sequences (from Assembly 21) used for oligonucleotide design, 6 are no longer present in Assembly 22 while 20 are annotated as mitochondrial genes in Assembly 22. Among the 6,179 remaining ORFs, 5,102 (82.6%) were successfully cloned and full-length sequence-validated. Sequences were validated by comparing the entire sequence of each cloned ORF against the reference sequences for haplotype A and B of the *C. albicans* SC5314 genome available at the Candida Genome Database (CGD; version_A22-s07-m01-r18; (46)). We defined 4 groups: (i) 4,209 sequences that are 100% identical with a reference

sequence (**No SNP, no DIP**; where SNP stands for Single Nucleotide Polymorphism and DIP stands for Deletion/Insertion Polymorphism); (ii) 116 sequences that do not carry any SNP but carry DIPs (**No SNP, DIPs**); (iii) 682 sequences that carry SNPs but no DIP (**SNPs, no DIP**); and (iv) 95 sequences that carry both SNPs and DIPs (**SNPs and DIPs**) **(Fig. 1A; Supplemental table S1)**.

Overall, 4,209/6,179 variation-free *C. albicans* ORFs (68.1%) are now available to the community. For 35.9% of the validated ORFs, there was no difference between the two alleles in the reference sequences (**HapA/B; Fig. 1B**). When nucleotide differences were present in the two alleles of the reference ORF sequence, we did not notice any bias in the haplotype that was cloned: haplotype A ORFs were cloned in 33% of cases (**HapA; Fig. 1B**) and haplotype B ORFs were cloned in 31% of cases (**HapB; Fig. 1B**). On chromosomes R, 1, 2, 3, 4, 5 and 7, ORF cloning was uniformly successful, with a cloning success rate ranging from ~82% to ~86%. Chromosome 6 ORFs were slightly underrepresented (76.9%) **(Fig. 1C)**. Although ORFs up to 1.5kb could be cloned with a success rate around 87% and ORFs between 1.5 and 4kb could be cloned with a success rate around 77%, we observed a decreased success rate for the longer ORFs **(Fig. 1D)**.

The analysis of the 6,880,620 nucleotides of the 5,102 sequence-validated ORFs revealed 2077 SNPs. These nucleotide substitutions could be real SNPs between our strain and the reference strain, or due to poor quality of some reference sequences, or mutations in the primers, or misincorporation by the DNA polymerase. Overall, the resulting maximum error rate amounts to 3×10^{-4} or 1 SNP every 3,312 cloned nucleotides. Notably, the *C. albicans* SC5314 genome sequence available at CGD (version_A22-s07-m01-r18) contained 272 ORFs with sequence ambiguities for at least one of the alleles. These ambiguities included stretches of 'Ns' or IUPAC code nucleotides (Y, R, S, M, K, W). The ORFeome-generated sequencing data resolved sequence ambiguities for 110 of these ORFs (**Supplemental Tables**

3 and 4). In this study, we also detected recombinant haplotypes for 116 of the cloned ORFs (**Supplemental Table 5**) that display characteristics of both reference haplotypes.

Taken together, our study provides an almost comprehensive, extremely high quality *C. albicans* ORFeome.

CandidaOrfDB: a database for the *C. albicans* ORFeome

CandidaOrfDB was created to integrate the data of the *C. albicans* ORFeome project and is available at http://candidaorfeome.eu. CandidaOrfDB enables the scientific community to search for availability and quality of the clones. All data pertinent to the cloning process are stored in the database. This information includes the name and sequence of the primers used to amplify the ORFs from *C. albicans* SC5314 genomic DNA, the coordinates of the primers in their 96-well storage plates, and any sequencing information available on the clones. SNPs and DIPs, their location in the ORF and their influence on the corresponding amino acid sequences are shown (**Fig. 2**). Nucleotide sequence of the cloned ORF and the corresponding amino acid sequence are displayed with highlights of SNPs or DIPs relative to the closest haplotype sequence (**Fig. 2**). The reference sequence data have been extracted from CGD (46). The database can be queried through the ORF number (orf19.xxxx or 19.xxxx; (5)), the Assembly 22 name (Ci_XXXXXW or Ci_XXXXXC; (5)) or the gene name.

A collection of destination vectors for exploiting the *C. albicans* ORFeome

The *C. albicans* ORFeome described above was developed in pDONR207 as this allows subsequent transfer of the cloned ORFs in Gateway™-adapted destination vectors. While destination vectors for ORF expression in hosts such as *E. coli* or *S. cerevisiae* are available (47,48), only a few destination vectors for ORF expression in *C. albicans* have been reported (11). Therefore, we set out to establish a collection of Gateway™-adapted destination vectors

for constitutive or conditional expression of untagged or tagged ORFs in *C. albicans*. Our primary collection of 15 *C. albicans* destination vectors derived from CIp10S (11) provides a choice of two promoters, either the constitutive promoter P_{TDH3} or the inducible promoter P_{TET}, and the option for N- or C-terminal fusion to various epitope tags (3xHA or TAP) (**Table 2** and **Fig. 3**). Integration of the destination vectors and their derivatives at the *RPS1* locus in the *C. albicans* genome is promoted by StuI or I-SceI linearization. These CIp10-derived vectors carry either the auxotrophic marker *URA3* or the *NAT1* marker that confers resistance to nourseothricin and can be used with clinical isolates. However, P_{TET}-bearing plasmids cannot carry the *NAT1* marker since the transactivator needed for tetracycline-mediated induction of P_{TET} is borne on the *NAT1*-carrying pNIMX plasmid (11). All vectors harbor a so-called spacer sequence (SP) originating from the *E. coli* kanamycin resistance gene, for barcode insertion and subsequent Illumina-based barcode sequencing. A second set of 34 plasmids was also generated with the constitutive promoters P_{ACT1} or the inducible promoter P_{PCK1}, and the option for N- or C-terminal fusion to GFP (**Table 2**). All 49 destination vectors were designated with a standardized nomenclature, pCA-DESTijkl, whereby i stands for the transformation marker (1, *URA3*; 2, *NAT1*); j stands for the promoter (1, P_{TET}; 2, P_{PCK1}; 3, P_{TDH3}; 4, P_{ACT1}); k stands for N-terminal tagging (0, no tag; 1, 3xHA; 2, GFP; 3, TAP); and l stands for C-terminal tagging (0, no tag; 1, 3xHA; 2, GFP; 3, TAP) (**Table 2** and **Fig. 3**). All destination vectors are available from Addgene (https://www.addgene.org).

In order to validate that the primary set of 15 destination vectors could drive constitutive or conditional expression of untagged or tagged ORFs, the *UME6* ORF was transferred into each of them using LR Clonase™. *UME6* encodes a transcription factor whose overexpression has been shown to force filamentation in *C. albicans* (49,50). The validation criteria for the destination vectors were (i) success of the LR reaction, (ii) integration in the

C. albicans genome at the *RPS1* locus, (iii) filamentation phenotype and/or (iv) detection of the 3xHA- or TAP-tagged Ume6 protein by western blot. The LR reaction was successfully used to transfer *UME6* from the pDONR207::*UME6* plasmid to all 15 destination vectors. The resulting expression plasmids were successfully integrated at the *C. albicans RPS1* locus. The 5 strains containing P$_{TET}$ were induced overnight at 30°C by adding 3 μg/ml anhydrotetracycline (ATc3) to the culture medium, while the 10 strains containing P$_{TDH3}$ were grown overnight at 30°C in YPD. As described in Chauvel *et al.* (11), overexpression of Ume6 led to hyperfilamentation. This phenotype was observed for all strains, regardless the tags (**Fig. 4A and 4B**). 3xHA- or TAP-tagged Ume6 was detected in the relevant strains and under the appropriate growth conditions (**Fig. 5A and 5B**).

Taken together, these data present a collection of 49 available destination vectors and validate 15 of those for application with the *C. albicans* ORFeome.

Proof of concept of a two-hybrid matrix approach of protein-protein interaction detection via mating in *C. albicans*

The availability of the ORFeome collection is a prerequisite to large, systematic two-hybrid (2H) screenings in *C. albicans*. The Candida 2H (C2H) system was developed for targeted one to one protein-protein interaction detection (32). Hence, there was a need to engineer the system for high throughput screening based on rapid and convenient recombination cloning systems for the generation of prey/bait arrays of proteins, and to use a mating approach to circumvent the low efficiency of transformation of this fungus. The former was obtained by adapting the prey and bait recipient vectors for Gateway™-mediated transfer of the ORFeome library, yielding pC2HB-GC and pC2HP-GC, respectively. The latter involved the combined processes of white-opaque switching and the preparation of mating-compatible 2H strains. The pleiomorphic fungus *C. albicans* is characterized by a parasexual cycle, with the

possible mating of diploids, and generation of tetraploids, the absence of meiosis and the reversion to the diploid state by chromosome loss (reviewed in (51)). An essential pre-requisite to mating in *C. albicans* is the phenotypic switch from white to opaque, a stable but reversible switch regulated by environmental stimuli such as low temperatures, carbon dioxide and nutrients. The epigenetic switch is tightly linked to mating since the a1/α2 complex, encoded at the mating type like (*MTL*) loci, acts as a repressor of opaque switching (52). In that context, the two-hybrid strain SC2H3 originating from the diploid a/α SN152 strain was deleted for the *MTL*a or *MTL*α locus to construct SC2H3α and SC2H3a strains. The efficient switch to opaque in the hemizygote strains was achieved by the doxycycline-controlled expression of *WOR1*, a major regulator of the white to opaque switch (41,53).

A proof of principle of the whole procedure for protein-protein detection is shown in **Figure 6**. The MAP kinase kinase Hst7 was used as bait and linked to the DNA binding domain lexA, as part of the pC2HB-GC vector. Similarly, the Cek1 kinase was expressed as a prey protein, fused to the activation domain VP16, component of the pC2HP-GC vector. pC2HB-Hst7 and pC2HP-Cek1 were transformed into opaque SC2H3α and SC2H3a strains, respectively. Mating of the resulting transformants was achieved through selection for leucine and arginine prototrophy. Upon interaction of the two proteins, the reconstituted transcriptional module promoted the expression of the *HIS4* marker, allowing only the mating-derived strains to grow on a histidine-free medium, whereas the original diploid strains could not (**Fig. 6B**). Through a mechanism of chromosome loss, the crossed tetraploid strains return back to a diploid state. We showed that the plasmids were stably expressed in the diploid offspring strains as these strains conserved their ability to grow in absence of arginine and leucine, indicating the presence of the prey and bait vectors respectively. In addition, all these strains were able to grow in absence of histidine, indicative of the targeted protein-protein interaction.

DISCUSSION

A partial *C. albicans* Gateway™-adapted ORFeome (ORFeomeV1; 644 ORFs) and the corresponding *C. albicans* overexpression strains have already been successfully used to investigate morphogenesis, biofilm formation and genome dynamics in *C. albicans* (11,31,54). In this study, we have now generated the *C. albicans* ORFeomeV2 within the Gateway™ recombination cloning system, as well as a collection of destination vectors suitable for expression in *C. albicans*.

This new *C. albicans* ORFeome resource encompasses 5,102 ORFs (83% of the annotated ORFs) and will be made available to the community upon request. The resource is supported by the CandidaOrfDB database, providing information on the individual plasmids and their sequences. The *C. albicans* ORFeomeV2 is of unprecedented quality, as all ORFs in the collection are fully sequenced, which remains an exception in large-scale ORFeome projects where ORFs are rarely sequenced in their entirety and only the 5'- and 3'-ends of the cloned ORFs are sequenced to confirm identity and the absence of frameshift mutations in the primers (15,16,18,21,23,29). Given these high standards, the success rate of 83% that we achieved with the *C. albicans* ORFeomeV2 is quite remarkable.

In eukaryotes, ORFeomes are usually cloned from full-length cDNA libraries to account for the presence of introns and therefore splicing variants. However, only 6% of *C. albicans* genes have introns (55), rendering the generation of the *C. albicans* ORFeome more straightforward than for pluricellular eukaryotes. Indeed, ORFs were simply amplified from the genomic DNA of the reference strain SC5314 and therefore, 360 ORFs in the *C. albicans* ORFeomeV2 harbor introns, a matter that should be taken into consideration when expressing ORFs in a prokaryotic host and, possibly, other eukaryotic hosts than *C. albicans*.

A second aspect to be taken into consideration when using the ORFeome in hosts other than

C. albicans is the unusual codon usage in this species (56). Indeed, in *C. albicans*, CUG is decoded as a serine instead of a leucine the majority of the time. Hence, improper translation in hosts with a standard genetic code may distort protein structure and function.

Sanger sequencing of both 5' and 3' ends was performed to confirm ORF identity and exclude clones containing primer or recombination errors. Previous studies have reported mutation rates in primers of 3-10% (26,29). Similar rates were observed in this work. Unlike other Gateway™ recombinational cloning projects (15,29), we did not see major differences in cloning efficiency for ORFs up to 4 kb. A decrease in success rate was observed only for ORFs >4 kb. This size bias could be attributed to an increased difficulty to amplify the ORF by PCR, a reduced efficiency of the Gateway™ BP Clonase™ reactions with long PCR products, and the error rate of the PCR polymerase, which increases with longer products. In addition, we also corrected the sequences of 110 ORFs with sequence ambiguities (N-tracts and IUPAC code nucleotides) according to the Candida Genome Database and identified 116 ORFs that display a recombinant haplotype. Recombinant haplotypes could be explained by template switching during PCR cycles or by incorrectly phased SNPs in reference sequences. This could be addressed by performing long-read sequencing of the *C. albicans* SC5314 genome.

However, many challenges remain to be addressed. For heterozygous genes, only one allele is present in the ORFeome. Because functional differences have been reported between the 2 alleles of a heterozygous gene (57,58), it would be relevant to also clone and validate the second allele in these instances. In addition, oligonucleotides have been designed based on annotation of the haploid set of Assembly 19 of the *C. albicans* genome. As a consequence, genes of the *MTLα* locus have not been included in the ORFeome. Despite our efforts, the collection is still missing 1,081 clones. The next version of the *C. albicans* ORFeome could

extend gene coverage by adding the missing ORFs, allelic variants or strain-specific variations.

ORFeomes are essential to bridge the gap between genome annotation and systems biology and allow large-scale gene and protein characterization. The development of a collection of *C. albicans* constitutive or conditional overexpression strains is one of the applications of the *C. albicans* ORFeome that we have successfully explored (11,31,54). In this respect, the *C. albicans* ORFeome project is currently completing its second and third phases, which consist of transferring the 5,102 cloned ORFs into barcoded destination vectors and generating a collection of *C. albicans* barcoded overexpression mutants, each mutant carrying one of the 5,102 cloned ORFs under the control of the inducible P_{TET} promoter. CandidaOrfDB already provides information on the available overexpression plasmids and strains. In this study, we have paved the way for a second application of the *C. albicans* ORFeome through the development of tools for a 2H matrix approach of protein-protein interaction detection via mating in *C. albicans*. Our data show that mating of diploid *C. albicans* expressing a bait fused to the LexA DNA binding domain and a prey fused to the VP16 activation domain, respectively, allows protein-protein interactions to be tested. In *S. cerevisiae,* large-scale 2H screens can be performed in a matrix design whereby haploid strains expressing baits and preys are mated and protein-protein interactions are scored in the resulting diploids. Our toolkit now enables the implementation of such a matrix design for large-scale 2H screens in *C. albicans.* To this aim, a collection of 1500 prey clones has already been generated. As mentioned above, *C. albicans* codon usage is unusual, which limits the development of applications of the *C. albicans* ORFeome in species that use standard decoding such as *S. cerevisiae*. Our development of tools for large-scale 2H screens in *C. albicans* circumvents this limitation, opening the path for the characterization of the *C. albicans* interactome.

Defining the *C. albicans* interactome will undoubtedly impact our understanding of *C. albicans* pathogenesis in humans.

In summary, high-level gene coverage coupled with the versatility of the Gateway™ recombinational cloning, *C. albicans*-adapted functional genomics tools and full access to clones, make the *C. albicans* ORFeomeV2 a unique and valuable resource for the scientific community that should greatly facilitate future functional studies in *C. albicans*.
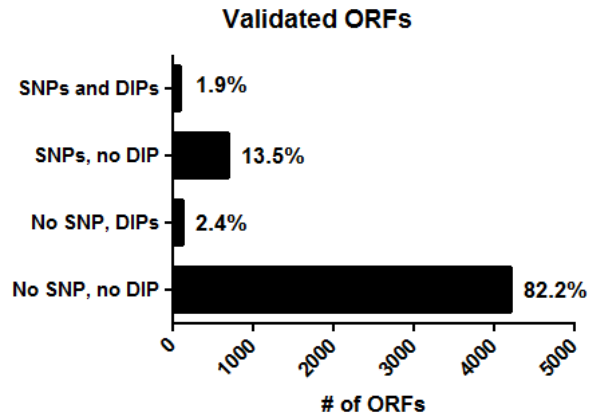
REFERENCES

Bibliography

1. Jones, T., Federspiel, N.A., Chibana, H., Dungan, J., Kalman, S., Magee, B.B., Newport, G., Thorstenson, Y.R., Agabian, N., Magee, P.T. *et al.* (2004) The diploid genome sequence of *Candida albicans*. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 7329-7334.

2. Muzzey, D., Schwartz, K., Weissman, J.S. and 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 biology*, **14**, R97.

3. d'Enfert, C., Goyard, S., Rodriguez-Arnaveilhe, S., Frangeul, L., Jones, L., Tekaia, F., Bader, O., Albrecht, A., Castillo, L., Dominguez, A. *et al.* (2005) CandidaDB : a genome database for *Candida albicans* pathogenomics. *Nucl. Acids Res.*, **33 Database Issue**, D353-357.

4. Braun, B.R., van Het Hoog, M., d'Enfert, C., Martchenko, M., Dungan, J., Kuo, A., Inglis, D.O., Uhl, M.A., Hogues, H., Berriman, M. *et al.* (2005) A Human-Curated Annotation of the *Candida albicans* Genome. *PLoS Genet.*, **1**, 36-57.

5. Skrzypek, M.S., Binkley, J., Binkley, G., Miyasato, S.R., Simison, M. and Sherlock, G. (2017) The Candida Genome Database (CGD): incorporation of Assembly 22, systematic identifiers and visualization of high throughput sequencing data. *Nucleic Acids Res*, **45**, D592-D596.

6. Roemer, T., Jiang, B., Davison, J., Ketela, T., Veillette, K., Breton, A., Tandia, F., Linteau, A., Sillaots, S., Marta, C. *et al.* (2003) Large-scale essential gene identification in *Candida albicans* and applications to antifungal drug discovery. *Mol. Microbiol.*, **50**, 167-181.

7. Xu, D., Jiang, B., Ketela, T., Lemieux, S., Veillette, K., Martel, N., Davison, J., Sillaots, S., Trosok, S., Bachewich, C. *et al.* (2007) Genome-wide fitness test and mechanism-of-action studies of inhibitory compounds in *Candida albicans*. *PLoS pathogens*, **3**, e92.

8. Homann, O.R., Dea, J., Noble, S.M. and Johnson, A.D. (2009) A phenotypic profile of the *Candida albicans* regulatory network. *PLoS genetics*, **5**, e1000783.

9. Noble, S.M., French, S., Kohn, L.A., Chen, V. and Johnson, A.D. (2010) Systematic screens of a *Candida albicans* homozygous deletion library decouple morphogenetic switching and pathogenicity. *Nat Genet*, **42**, 590-598.

10. Sahni, N., Yi, S., Daniels, K.J., Huang, G., Srikantha, T. and Soll, D.R. (2010) Tec1 mediates the pheromone response of the white phenotype of *Candida albicans*: insights into the evolution of new signal transduction pathways. *PLoS Biol*, **8**, e1000363.

11. Chauvel, M., Nesseir, A., Cabral, V., Znaidi, S., Goyard, S., Bachellier-Bassi, S., Firon, A., Legrand, M., Diogo, D., Naulleau, C. *et al.* (2012) A versatile overexpression strategy in the pathogenic yeast *Candida albicans*: identification of regulators of morphogenesis and fitness. *PloS one*, **7**, e45912.

12. Schillig, R. and Morschhauser, J. (2013) Analysis of a fungus-specific transcription factor family, the *Candida albicans* zinc cluster proteins, by artificial activation. *Mol. Microbiol.*, **89**, 1003-1017.

13. Rual, J.F., Hill, D.E. and Vidal, M. (2004) ORFeome projects: gateway between genomics and omics. *Curr. Opin. Chem. Biol.*, **8**, 20-25.

14. Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K. and Weissman, J.S. (2003) Global analysis of protein expression in yeast. *Nature*, **425**, 737-741.

15. Reboul, J., Vaglio, P., Rual, J.F., Lamesch, P., Martinez, M., Armstrong, C.M., Li, S., Jacotot, L., Bertin, N., Janky, R. *et al.* (2003) *C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nat. Genet.*, **34**, 35-41.

16. Gelperin, D.M., White, M.A., Wilkinson, M.L., Kon, Y., Kung, L.A., Wise, K.J., Lopez-Hoyo, N., Jiang, L., Piccirillo, S., Yu, H. *et al.* (2005) Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes Dev.*, **19**, 2816-2826.

17. Matsuyama, A. and Yoshida, M. (2009) Systematic cloning of an ORFeome using the Gateway system. *Methods Mol. Biol.*, **577**, 11-24.

18. Rajagopala, S.V., Yamamoto, N., Zweifel, A.E., Nakamichi, T., Huang, H.K., Mendez-Rios, J.D., Franca-Koh, J., Boorgula, M.P., Fujita, K., Suzuki, K. *et al.* (2010) The *Escherichia coli* K-12 ORFeome: a resource for comparative molecular microbiology. *BMC genomics*, **11**, 470.

19. Gong, W., Shen, Y.P., Ma, L.G., Pan, Y., Du, Y.L., Wang, D.H., Yang, J.Y., Hu, L.D., Liu, X.F., Dong, C.X. *et al.* (2004) Genome-wide ORFeome cloning and analysis of Arabidopsis transcription factor genes. *Plant Physiol.*, **135**, 773-782.

20. Underwood, B.A., Vanderhaeghen, R., Whitford, R., Town, C.D. and Hilson, P. (2006) Simultaneous high-throughput recombinational cloning of open reading frames in closed and open configurations. *Plant Biotechnol J*, **4**, 317-324.

21. Grant, I.M., Balcha, D., Hao, T., Shen, Y., Trivedi, P., Patrushev, I., Fortriede, J.D., Karpinka, J.B., Liu, L., Zorn, A.M. *et al.* (2015) The Xenopus ORFeome: A resource that enables functional genomics. *Dev. Biol.*, **408**, 345-357.

22. Bischof, J., Bjorklund, M., Furger, E., Schertel, C., Taipale, J. and Basler, K. (2013) A versatile platform for creating a comprehensive UAS-ORFeome library in Drosophila. *Development*, **140**, 2434-2442.

23. Dricot, A., Rual, J.F., Lamesch, P., Bertin, N., Dupuy, D., Hao, T., Lambert, C., Hallez, R., Delroisse, J.M., Vandenhaute, J. *et al.* (2004) Generation of the *Brucella melitensis* ORFeome version 1.1. *Genome Res.*, **14**, 2201-2206.

24. Aguiar, J.C., LaBaer, J., Blair, P.L., Shamailova, V.Y., Koundinya, M., Russell, J.A., Huang, F., Mar, W., Anthony, R.M., Witney, A. *et al.* (2004) High-throughput generation of *P. falciparum* functional molecules by recombinational cloning. *Genome Res.*, **14**, 2076-2082.

25. Hauser, R., Ceol, A., Rajagopala, S.V., Mosca, R., Siszler, G., Wermke, N., Sikorski, P., Schwarz, F., Schick, M., Wuchty, S. *et al.* (2014) A second-generation protein-protein interaction network of *Helicobacter pylori*. *Molecular & cellular proteomics : MCP*, **13**, 1318-1329.

26. Maier, C.J., Maier, R.H., Virok, D.P., Maass, M., Hintner, H., Bauer, J.W. and Onder, K. (2012) Construction of a highly flexible and comprehensive gene collection representing the ORFeome of the human pathogen *Chlamydia pneumoniae*. *BMC genomics*, **13**, 632.

27. Brandner, C.J., Maier, R.H., Henderson, D.S., Hintner, H., Bauer, J.W. and Onder, K. (2008) The ORFeome of *Staphylococcus aureus* v 1.1. *BMC genomics*, **9**, 321.

28. von Brunn, A., Teepe, C., Simpson, J.C., Pepperkok, R., Friedel, C.C., Zimmer, R., Roberts, R., Baric, R. and Haas, J. (2007) Analysis of intraviral protein-protein interactions of the SARS coronavirus ORFeome. *PloS one*, **2**, e459.

29. Lamesch, P., Li, N., Milstein, S., Fan, C., Hao, T., Szabo, G., Hu, Z., Venkatesan, K., Bethel, G., Martin, P. *et al.* (2007) hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. *Genomics*, **89**, 307-315.

30. Walhout, A.J., Temple, G.F., Brasch, M.A., Hartley, J.L., Lorson, M.A., van den Heuvel, S. and Vidal, M. (2000) GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol.*, **328**, 575-592.

31. Cabral, V., Znaidi, S., Walker, L.A., Martin-Yken, H., Dague, E., Legrand, M., Lee, K., Chauvel, M., Firon, A., Rossignol, T. *et al.* (2014) Targeted changes of the cell wall proteome influence *Candida albicans* ability to form single- and multi-strain biofilms. *PLoS pathogens*, **10**, e1004542.

32. Stynen, B., Van Dijck, P. and Tournu, H. (2010) A CUG codon adapted two-hybrid system for the pathogenic fungus *Candida albicans*. *Nucleic Acids Res*, **38**, e184.

33. Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.*, **166**, 557-580.
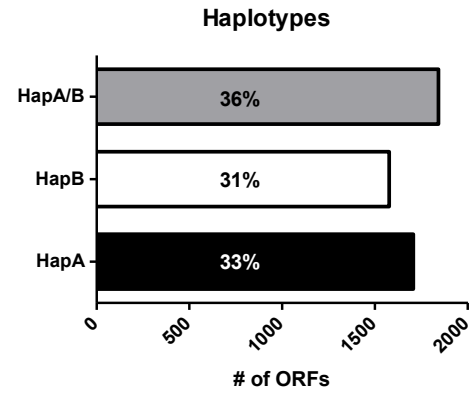
34.   Cabral, V., Chauvel, M., Firon, A., Legrand, M., Nesseir, A., Bachellier-Bassi, S., Chaudhari, Y., Munro, C.A. and d'Enfert, C. (2012) Modular Gene Over-expression Strategies for *Candida albicans*. *Methods Mol Biol*, **845**, 227-244.

35.   Gola, S., Martin, R., Walther, A., Dunkler, A. and Wendland, J. (2003) New modules for PCR-based gene targeting in *Candida albicans*: rapid and efficient gene targeting using 100 bp of flanking homology region. *Yeast*, **20**, 1339-1347.

36.   Liu, T.T., Znaidi, S., Barker, K.S., Xu, L., Homayouni, R., Saidane, S., Morschhauser, J., Nantel, A., Raymond, M. and Rogers, P.D. (2007) Genome-wide expression and location analyses of the *Candida albicans* Tac1p regulon. *Eukaryotic cell*, **6**, 2122-2138.

37.   Enjalbert, B., Rachini, A., Vediyappan, G., Pietrella, D., Spaccapelo, R., Vecchiarelli, A., Brown, A.J. and d'Enfert, C. (2009) A multifunctional, synthetic *Gaussia princeps* luciferase reporter for live imaging of *Candida albicans* infections. *Infect Immun*, **77**, 4847-4858.

38.   Reuss, O., Vik, A., Kolter, R. and Morschhauser, J. (2004) The *SAT1* flipper, an optimized tool for gene disruption in *Candida albicans*. *Gene*, **341**, 119-127.

39.   Walther, A. and Wendland, J. (2008) PCR-based gene targeting in *Candida albicans*. *Nat Protoc*, **3**, 1414-1421.

40.   Park, Y.N. and Morschhauser, J. (2005) Tetracycline-inducible gene expression and gene deletion in *Candida albicans*. *Eukaryotic cell*, **4**, 1328-1342.

41.   Huang, G., Wang, H., Chou, S., Nie, X., Chen, J. and Liu, H. (2006) Bistable expression of *WOR1*, a master regulator of white-opaque switching in *Candida albicans*. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 12813-12818.

42.   Liu, H., Kohler, J. and Fink, G.R. (1994) Suppression of hyphal formation in *Candida albicans* by mutation of a *STE12* homolog. *Science*, **266**, 1723-1726.

43.   Bennett, R.J. and Johnson, A.D. (2003) Completion of a parasexual cycle in Candida albicans by induced chromosome loss in tetraploid strains. *EMBO J.*, **22**, 2505-2515.

44.   van het Hoog, M., Rast, T.J., Martchenko, M., Grindle, S., Dignard, D., Hogues, H., Cuomo, C., Berriman, M., Scherer, S., Magee, B.B. *et al.* (2007) Assembly of the *Candida albicans* genome into sixteen supercontigs aligned on the eight chromosomes. *Genome biology*, **8**, R52.

45.   Fonzi, W.A. and Irwin, M.Y. (1993) Isogenic strain construction and gene mapping in *Candida albicans*. *Genetics*, **134**, 717-728.

46.   Skrzypek, M.S., Binkley, J., Binkley, G., Miyasato, S.R., Simison, M. and Sherlock, G. "Candida Genome Database" http://www.candidagenome.org/ (2015*)*.

47.   Alberti, S., Gitler, A.D. and Lindquist, S. (2007) A suite of Gateway cloning vectors for high-throughput genetic analysis in Saccharomyces cerevisiae. *Yeast*, **24**, 913-919.

48.   Hallez, R., Letesson, J.J., Vandenhaute, J. and De Bolle, X. (2007) Gateway-based destination vectors for functional analyses of bacterial ORFeomes: application to the Min system in Brucella abortus. *Appl. Environ. Microbiol.*, **73**, 1375-1379.

49.   Carlisle, P.L., Banerjee, M., Lazzell, A., Monteagudo, C., Lopez-Ribot, J.L. and Kadosh, D. (2009) Expression levels of a filament-specific transcriptional regulator are sufficient to determine *Candida albicans* morphology and virulence. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 599-604.

50.   Zeidler, U., Lettner, T., Lassnig, C., Muller, M., Lajko, R., Hintner, H., Breitenbach, M. and Bito, A. (2009) *UME6* is a crucial downstream target of other transcriptional regulators of true hyphal development in *Candida albicans*. *FEMS yeast research*, **9**, 126-142.

51.   Bennett, R.J. (2015) The parasexual lifestyle of *Candida albicans*. *Curr. Opin. Microbiol.*, **28**, 10-17.

52.   Miller, M.G. and Johnson, A.D. (2002) White-opaque switching in *Candida albicans* is controlled by mating-type locus homeodomain proteins and allows efficient mating. *Cell*, **110**, 293-302.

53. Zordan, R.E., Galgoczy, D.J. and Johnson, A.D. (2006) Epigenetic properties of white-opaque switching in *Candida albicans* are based on a self-sustaining transcriptional feedback loop. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 12807-12812.

54. Loll-Krippleber, R., Feri, A., Nguyen, M., Maufrais, C., Yansouni, J., d'Enfert, C. and Legrand, M. (2015) A FACS-optimized screen identifies regulators of genome stability in *Candida albicans*. *Eukaryotic cell*, **14**, 311-322.

55. Mitrovich, Q.M., Tuch, B.B., Guthrie, C. and Johnson, A.D. (2007) Computational and experimental approaches double the number of known introns in the pathogenic yeast *Candida albicans*. *Genome Res.*, **17**, 492-502.

56. Santos, M.A. and Tuite, M.F. (1995) The CUG codon is decoded in vivo as serine and not leucine in *Candida albicans*. *Nucleic Acids Res*, **23**, 1481-1486.

57. Coste, A., Turner, V., Ischer, F., Morschhauser, J., Forche, A., Selmecki, A., Berman, J., Bille, J. and Sanglard, D. (2006) A mutation in Tac1p, a transcription factor regulating CDR1 and CDR2, is coupled with loss of heterozygosity at chromosome 5 to mediate antifungal resistance in *Candida albicans*. *Genetics*, **172**, 2139-2156.

58. Holmes, A.R., Tsao, S., Ong, S.W., Lamping, E., Niimi, K., Monk, B.C., Niimi, M., Kaneko, A., Holland, B.R., Schmid, J. *et al.* (2006) Heterozygosity and functional allelic variation in the *Candida albicans* efflux pump genes CDR1 and CDR2. *Mol. Microbiol.*, **62**, 170-186.

59. Noble, S.M. and Johnson, A.D. (2005) Strains and strategies for large-scale gene deletion studies of the diploid human fungal pathogen *Candida albicans*. *Eukaryotic cell*, **4**, 298-309.
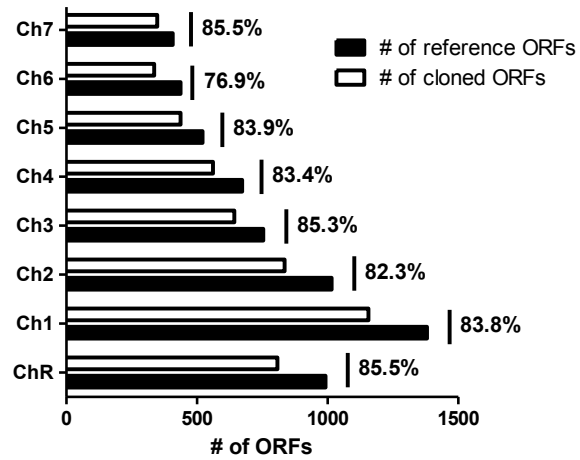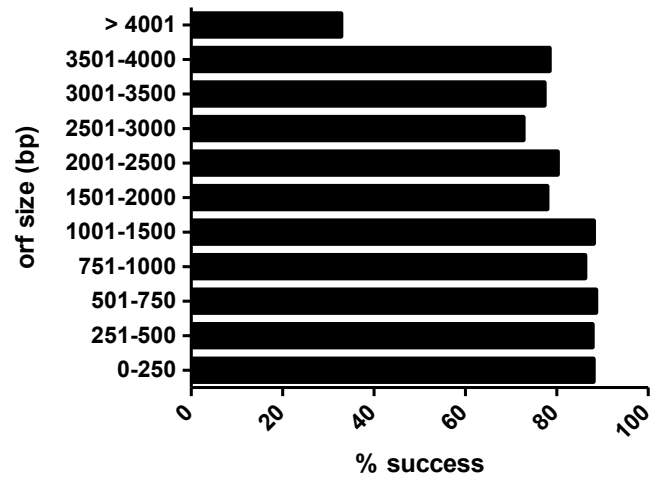
A.

Validated ORFs

SNPs and DIPs — 1.9%
SNPs, no DIP — 13.5%
No SNP, DIPs — 2.4%
No SNP, no DIP — 82.2%

# of ORFs

B.

Haplotypes

HapA/B — 36%
HapB — 31%
HapA — 33%

# of ORFs

C.

Summary of successfully cloned ORFs on each chromosome

Ch7 — 85.5%
Ch6 — 76.9%
Ch5 — 83.9%
Ch4 — 83.4%
Ch3 — 85.3%
Ch2 — 82.3%
Ch1 — 83.8%
ChR — 85.5%

■ # of reference ORFs
□ # of cloned ORFs

# of ORFs

D.

Success rate based on orf size

orf size (bp)

> 4001
3501-4000
3001-3500
2501-3000
2001-2500
1501-2000
1001-1500
751-1000
501-750
251-500
0-250

% success

THE CANDIDA ALBICANS
ORFEOME PROJECT

Search ORF | **Search**

ORF details

Name	ORF19.2249
A22 name	C2_06950C
Theoretical size	717 bp
Chromosome	2
Haplotype	B
Introns	
Comments	(BP16B03)3 SNPs, No DIP

Resources

Gateway Donor Plasmid	
CIp-PTET-GTW-derived plasmid - NOT AVAILABLE YET	
Barcoded dox-induced OE C.alb strain - NOT AVAILABLE YET	
Phenotypes	

SNP(s)

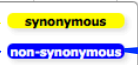|	Nucleotide					Amino acid			
Position	Cloned	A22_HapA	A22_HapB	A21	Position	Cloned	A22_HapA	A22_HapB	A21
119	G	A	A	A	40	S	N	N	N
234	A	G	A	A	78	L	L	L	L
448	G	A	G	G	150	D	N	D	D
492	A	C	A	A	164	I	I	I	I
537	T	C	C	C	179	I	I	I	I
629	G	A	A	A	210	R	K	K	K

Restrictions on cloned sequence

	Fragments	Position of cleavage sites	Fragment sizes (bp)
StuI	No cuts		
EcoRV	2	402	402, 312
BsrGI	2	481	481, 233

Nucleotide sequences

Cloned sequence
```
1   ATGTCGACTGCTGCCCCACAAGAAAATTCTCCTGAATTGA
41  AGTTATCGCCGAGTTCCAAACTTAATGGATTCGAATGGTG
81  GAGAAGAACACTTCTGTACCAAACTGGGTTAGGATTAACT
121 GCTGAAGAAAAAGCACAGTATGAATTGATTACAATAATA
161 GAAATCTTGCAGAGAAAATGCCAGCAATGCAATGAGTATCG
201 TAATTGGATGTTAGAATATTCTCCATCTGTTTTATTTATG
241 ATGGATCATATCAAGAAATTGACTAAAGATAAAGACACCAA
281 TACTAACCAAAGAAACATTACTTGTGATGTGCGTGATTT
321 CACAAAAGGTGGTGGATTTCATCCTGATGCAGGAATATTA
361 TTATGTGCCAACTGGATAAGAAATAACTGGCAATTGGAAG
401 ATATCCTAACACATGAGCTAGTTCATGTTTATGATTACTT
441 GAAGTTTGATCTCGATATGATGAACTTGAGCACACCATGCA
481 TGTACAGAAATAAGAGCTTTTATGCTTAGTGGTGGGTGTA
521 GAATCTGGCAAGAAATTAAGAAAACAGGATTGGGGAATTT
561 TGGGAAAAAATTCCAGGATTGTATCAAGACGGTGCTGTT
601 TTATCTGTCTCTAGTAATCCGATTTGTACATCACCTGAAG
641 AGGCAGAAAAAGTGGTCAGTACAGTTTGGAATTCATGTTT
681 TAATGACACAAGACCTTTCGAAAGAGTCTAT
```

A22_HapA sequence
```
1   ATGTCGACTGCTGCCCCACAAGAAAATTCTCCTGAATTGA
41  AGTTATCGCCGAGTTCCAAACTTAATGGATTCGAATGGTG
```

A22_HapB sequence
```
1   ATGTCGACTGCTGCCCCACAAGAAAATTCTCCTGAATTGA
41  AGTTATCGCCGAGTTCCAAACTTAATGGATTCGAATGGTG
```

A21 sequence
```
1   ATGTCGACTGCTGCCCCACAAGAAAATTCTCCTGAATTGA
41  AGTTATCGCCGAGTTCCAAACTTAATGGATTCGAATGGTG
```

synonymous | non-synonymous

Protein sequences

Cloned translated
```
1   MSTAAPQENSPELKLSPSSKLNGFEWWRRTLLYQTGLGLS
41  AEEKAQYEFDYNNRNLAEKCQQCNEYRNWMLEYSPSVLFM
81  MDHIKKLTKDKEPILTKENITCDVCDFTKGGGFHPDEGIL
121 LCANWIRNKWQLEDILTHELVHVYDYLKFDLDMNNLRHHA
161 CTEIRASMLSGECRIWQEIKKTGLGNFGKKFQDCIKRRAV
201 LSVSSNPICRSPEEAEKVVSTVWNSCFNDTRPFERVY
```

A22_HapA translated
```
1   MSTAAPQENSPELKLSPSSKLNGFEWWRRTLLYQTGLGLN
41  AEEKAQYEFDYNNRNLAEKCQQCNEYRNWMLEYSPSVLFM
```

A22_HapB translated
```
1   MSTAAPQENSPELKLSPSSKLNGFEWWRRTLLYQTGLGLN
41  AEEKAQYEFDYNNRNLAEKCQQCNEYRNWMLEYSPSVLFM
```

A21 translated
```
1   MSTAAPQENSPELKLSPSSKLNGFEWWRRTLLYQTGLGLN
41  AEEKAQYEFDYNNRNLAEKCQQCNEYRNWMLEYSPSVLFM
```

Barcodes insertion site

SP

PROMOTER

N-tag

Inducible
- P_{TET} – 1 – –
- P_{PCK1} – 2 – –

Constitutive
- P_{TDH3} – 3 – –
- P_{ACT1} – 4 – –

pCA-DESTijkl

GTW-cassette

C-tag

MARKER

URA3 1 – – –
SAT1 2 – – –

N-tag:
- NONE – – 0 –
- 3xHA – – 1 –
- GFP – – 2 –
- TAP – – 3 –

C-tag:
- NONE – – – 0
- 3xHA – – – 1
- GFP – – – 2
- TAP – – – 3

A.

Promoter: P_{TET}   P_{TDH3}

Marker: *URA3*   *URA3*   *SAT1*

Tag position: C-ter   N-ter   C-ter   N-ter

ATc3: − +   − +   −   −   −

→ 3HA-tagged Ume6 125kDa

B.

Promoter: P_{TET}   P_{TDH3}

Marker: *URA3*   *SAT1*   *URA3*

Tag position: C-ter   N-ter   N-ter   C-ter   N-ter   C-ter

ATc3: − +   − +   −   −   −   −

TAP-tagged Ume6 140kDa →

→ TAP-tagged Ume6 140kDa

A

MTLa/Δ

| lexA | bait | leu+ |

X

Δ/MTLα

| VP16 | prey | arg+ |

a/α tetraploid

| lexA | bait | leu+ |
| VP16 | prey | arg+ |

→ Histidine prototrophy
upon bait/prey interaction

B

SC-leu-arg
markers selection

SC-met-his
PPI selection

(1)
a-VP16*CEK1*
α-lexA*HST7*
a-VP16
α-lexA

(2)
α-lexA*HST7* **X** a-VP16*CEK1*
α-lexA*HST7* **X** a-VP16*CEK1*
α-lexA*HST7* **X** a-VP16
α-lexA **X** a-VP16*CEK1*

(3)
offspring from
α-lexA*HST7* **X** a-VP16*CEK1*

A.

	YPD	ATc3
CEC4164 P$_{TET}$-*UME6*		
CEC4167 P$_{TET}$-TAP-*UME6*		
CEC4170 P$_{TET}$-*UME6*-TAP		
CEC4165 P$_{TET}$-3HA-*UME6*		
CEC4168 P$_{TET}$-*UME6*-3HA		

B.

	N-Tag	C-Tag	
TAP	CEC4191	CEC4197	*URA3*
	CEC4192	CEC4198	*SAT1*
3HA	CEC4187	CEC4193	*URA3*
	CEC4533	CEC4194	*SAT1*