NAR Breakthrough Article

Generating genomic platforms to study Candida albicans pathogenesis

Mélanie Legrand¹, Sophie Bachellier-Bassi¹,†, Keunsook K. Lee², Yogesh Chaudhari², Hélène Tournu³,⁴, Laurence Arbogast¹, Hélène Boyer¹, Murielle Chauvel¹, Vitor Cabral¹,⁵, Corinne Maufrais¹,⁶, Audrey Nesseir¹,⁵, Irena Maslanka², Emmanuelle Permal¹, Tristan Rossignol¹, Louise A. Walker², Ute Zeidler¹, Sadri Znaidi¹, Floris Schoeters³,⁴, Charlotte Majgier⁷, Renaud A. Julien⁷, Laurence Ma⁸, Magali Tichit⁸, Christiane Bouchier⁸, Patrick Van Dijck³,⁴, Carol A. Munro²,* and Christophe d’Enfert¹,*

¹Fungal Biology and Pathogenicity Unit, Department of Mycology, Institut Pasteur, INRA, Paris 75015, France, 
²Medical Research Council Centre for Medical Mycology at the University of Aberdeen, Institute of Medical Sciences, Foresterhill, Aberdeen AB25 2ZD, UK, 
³VIB-KU Leuven Center for Microbiology, Leuven 3001, Belgium, 
⁴Laboratory of Molecular Cell Biology, KU Leuven, Leuven 3001, Belgium, 
⁵Univ. Paris Diderot, Sorbonne Paris Cité, Cellule Pasteur, Paris 75015, France, 
⁶Institut Pasteur-Bioinformatics and Biostatistics Hub-C3BI, USR 3756 IP CNRS-Paris 75015, France, 
⁷Modul-Bio, Parc Scientifique Luminy Biotech II, Marseille 13009, France and 
⁸Institut Pasteur-Biomics Pole-CITECH-Paris 75015, France

Received February 05, 2018; Revised June 19, 2018; Editorial Decision June 20, 2018; Accepted June 25, 2018

ABSTRACT

The advent of the genomic era has made elucidating gene function on a 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 toward 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 5099 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://candidaorfome.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 en-

¹To whom correspondence should be addressed. Tel: +33 0 140613257; Fax: +33 0 140613456; Email: denfert@pasteur.fr
Correspondence may also be addressed to Carol Munro. Tel: +44 0 1224 437 485; Fax: +44 0 1224 437 506; Email: c.a.munro@abdn.ac.uk
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Present addresses:
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 Grégoir Mendel, Vincennes, 94300, France; Emmanuelle Permal, Toulouse 31000, France.
Tristan Rossignol, MICALIS, INRA, Jouy-en-Josas 78352, France.
Ute Zeidler, Sandorz, 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 Modeles Animaux, Institut Pasteur, Paris 75015, France.

Disclaimer: The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

© The Author(s) 2018. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
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 1670 out of the 6198 predicted protein-coding genes have been characterized as of 31 January 31, 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 approaches, 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 (6–10) or overexpression (11–13) mutant collections. Genome-wide ORF libraries or ORFeomes represent useful resources for the implementation of approaches used to elucidate gene function (14). Besides, the development of collections of overexpression mutants (12), 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 (15).

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, Caenorhabditis elegans, Schizosaccharomyces pombe, 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 (16–29). 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 (30). Most ORFeome libraries are created using the highly versatile Gateway™ cloning approach (31). The Gateway™ technology employs the recombination system of bacteriophage lambda with two sets of reactions: the BP reaction, catalyzed by Gateway™ BP Clonase™, facilitates recombination between attB and attP sequences, and the LR reaction, catalyzed by Gateway™ LR Clonase™, promotes recombination between attL and attR sequences (31).

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 (32). Here, we describe the generation and validation of the latest C. albicans ORFeome collection, which consists of 5099 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 gene 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 (33) that will allow systematic protein–protein interaction screening in C. albicans.

MATERIALS AND METHODS

Strains and growth conditions

Candida albicans strains used in this study are listed in Table 1. Candida 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.

Escherichia 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 tetracyclin. Solid media were obtained by adding 2% agar.

DH5α E. coli cells competent for transformation with DNA were prepared according to Hanahan et al. (34).

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'-GGGGACAAGTTTGTACAAAACAAGCAGGCTTG-3' to the 5' end of the first 30 nt of each ORF. Gene-specific reverse primers were designed by adding the sequence 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTC3' to the 5' end of the last 30 nt of each ORF excluding the Stop codon. Polymerase chain reaction
SC2H3
MTLa::SAT1-FLP
ura3 Δ::Imm434 IRO1/iro1 Δ::Imm434 his1 Δ::hisG ARG4
PGK1: (PADH1-cartTA,SAT1,PTET-WOR1)

C. albicans strains used in this study

| Strain          | Source       | Relevant Genotype | Reference |
|-----------------|--------------|-------------------|-----------|
| SC5314          | Wild-type    | ura3 Δ::Imm434/ura3 Δ::Imm434 his1 Δ::hisG/iro1 Δ::Imm434/iro1 Δ::Imm434 | (47)      |
| BWP17           |              | arg4 Δ::hisG/arg4 Δ::hisG/iro1 Δ::Imm434/iro1 Δ::Imm434 | (40)      |
| SN152           |              | arg4 Δ/arg4 Δ leu2 Δ/leu2 Δ his1 Δ/leu2 Δ his1 Δ URA3/ura3 Δ::Imm434 | (55)      |
| CEC161          | BWP17        | IRO1/iro1 Δ::Imm434/ura3 Δ::Imm434 his1 Δ::hisG/His1 arg4 Δ::hisG/ARG4 | (38)      |
| CEC369          | CEC161       | ura3 Δ::Imm434/ura3 Δ::Imm434 arg4 Δ::hisG/ARG4 his1 Δ::hisG/His1 | (38)      |
| CEC2907         | CEC161       | iro1 Δ::Imm434/iro1 Δ::Imm434/His1 arg4 Δ::hisG/ARG4 | (12)      |
| SC2H3           | SN152        | URA3/ura3 Δ::Imm434 IRO1/iro1 Δ::Imm434 5xLexAOp-ADH1b/His1 | (33)      |
| (LMBP10468)*    |              | 5xLexAOp-ADH1b/lacZ |           |
| SC2H3a          | SC2H3        | MTLa/MTLα arg4 Δ/arg4 Δ leu2 Δ/leu2 Δ his1 Δ/His1 | This study |
| SC2H3α          | SC2H3        | MTLα/MTLα: SAT1-FLP | This study |
| SC2H3α-pWOR1    | SC2H3a       | MTLα/MTLα: SAT1-FLP/MTLα arg4 Δ/arg4 Δ leu2 Δ/leu2 Δ his1 Δ/His1 | This study |
| (LMBP10470)*    |              | URA3/ura3 Δ::Imm434 IRO1/iro1 Δ::Imm434 5xLexAOp-ADH1b/His1 |           |
| SC2H3α-pWOR1    | SC2H3a       | MTLα/MTLα FRT arg4 Δ/arg4 Δ leu2 Δ/leu2 Δ his1 Δ/His1 | This study |
| (LMBP10469)*    |              | URA3/ura3 Δ::Imm434 IRO1/iro1 Δ::Imm434 5xLexAOp-ADH1b/His1 |           |

*Strain number in the BCCM collection

Microplates. After 36 h-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 Supplementary Table S2 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. Plasmid DNAs were fragmented using a Bioruptor sonicator. Sequencing libraries were prepared according to the KAPA LTP Library Preparation Kit protocol (KAPA Biosystems, catalogue number KK8232). The libraries were not made in duplicates and therefore pools (1–5) were sequenced only once. 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. The fastq files have been deposited into NCBI-SRA and are available under the BioProject PRJNA472959. In Supplementary Table S1, the sequencing pool (pools 1–5) is given for each cloned ORF. Because of a server crash, the dataset for pool2 had been lost, but the reads could be recovered from CLC format, and the headers reconstructed based on the sequenc-
ing information provided by the Institut Pasteur sequencing platform.

**CandidaOrfDB database**

CandidaOrfDB was created to integrate the data of the *C. albicans* ORFeome project and is available at [http://candidaorfeome.eu](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 Needleman-Wunsch 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](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/](https://www6.inra.fr/cirm_eng/)).

**Destination plasmids**

All destination plasmids constructed in this study were derived from CIp-P<sub>TET</sub>-GTW (12) and 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 Supplementary Table S2.

**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<sub>TET</sub>-GTW, yielding CIp-P<sub>TET</sub>-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<sub>TET</sub> from CIp-P<sub>TET</sub>-GTW with Aec65I 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 (36) 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 SalI 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 (12) 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 Sall–HindIII fragment from pCA-Dest1100 into the same sites of pUC18. For cloning of the 3xHA epitope, pCAMPY-3xHA (37) 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 Sall 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 (36) 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 Sall and NsiI and the GFP-bearing fragment ligated into the same sites of pCA-Dest1100, yielding pCA-Dest1110. For cloning the TAP-tag, the TAP-tag coding region was amplified from CIp10-P<sub>PKCI</sub>-GTW-TAPtag (12) 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 Sall and NsiI, and the TAP-containing fragment ligated in the same sites of pCA-Dest1100, yielding pCA-Dest1103.

**Promoters.** We replaced P<sub>TET</sub> with either P<sub>PKCI</sub> (obtained from CIp10-P<sub>PKCI</sub>-GTW-TAPtag, (12)), yielding the pCA-Dest12xx series; P<sub>TDH3</sub> (cut from pKS-P<sub>TDH3</sub>; P<sub>TDH3</sub> 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<sub>ACT1</sub>, cut from CIp10:P<sub>ACT1</sub>gLuc59 (38), yielding the pCA-Dest14xx series.

**Selection markers.** In all plasmids except the P<sub>TET</sub> series, the auxotrophic marker URA3 was replaced with the NAT1 marker conferring nourseothricin resistance. Briefly, a XbaI– DraIII fragment or Spel–Nael 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<sub>TET</sub> promoter with oligonucleotides UZ50 and UZ51, and using pFA6-SAT1 (39) 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](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 (35). Briefly, the UME6 ORF was transferred from the entry clone into each one of 15 expression vectors using the Gateway™ LR Clonase™ II Enzyme Mix (Invitrogen). After *E. coli* transformation, the plasmids were verified by EcoRV digestion. The NATI- and URA3-bearing expression plasmids were digested by Stul and transformed into *C. albicans* strain CEC369 (38) (that derives from CEC161 (38) = BWP17 (40) + HIS+ARG ) or CEC2907 (12), respectively, according to Walther and Wendland (41). Transformants were selected for prototrophy or nourseothricin resistance, respectively, and verified by PCR using primer ClpUL with (i) primer CIpUR for the *URA3* bearing plasmids and (ii) primer CgSAT1-rev for the *SATI*-bearing plasmids, that yield 1 kb and 1.6 kb products, respectively, if integration of the OE plasmid has occurred at the *RPS1* locus.

### Table 2. Destination vectors for use with the *Candida albicans* ORFeome. Plasmid names are based on the combination of modules

| Plasmid name* | Accession number | Selection marker | Promoter | N-Tag | C-Tag |
|--------------|------------------|-----------------|----------|-------|-------|
| pCA-DEST1100 | MG188268         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1101 | MG188269         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1103 | MG188271         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1110 | MG188272         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1100 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |
| pCA-DEST1130 | MG188287         | URA3            | P*TET    | HA3   |       |

* In bold: vectors validated with UME6

**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 (42). 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} and P_{TDH3} strains, a 30 ml culture in YPD or YPD+Atc3 was inoculated at OD600 = 0.2 or 2 × 10^6 cells/ml with a freshly grown colony and incubated at 30°C with shaking until OD600 reached ~1.

For the 3HA-tagged strains, 750 μl of cultures were collected by centrifugation, resuspended in 25 μl of water and diluted in 25 μl of 2 × Laemmli sample buffer. After 10 min at 100°C, proteins were stored at −80°C or directly submitted to electrophoresis.

For the TAP-tagged strains, 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 pH 9.5). 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-peroxidase 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 MTLα or MTLα locus using the SAT1 flipper cassette (39) in the two-hybrid strain background SC2H3 (33). To delete the MTLα locus, flanking regions were amplified with primers MTLα-5′F and MTLα-5′R, and with MTLα-3′F and MTLα-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 (39), at SacI/NotI and XhoI/KpnI sites respectively. SacI/KpnI deletion constructs were transformed in SC2H3 (33) and transformants were selected on YPD medium containing 200 μg/ml of nourseothricin. Positive SC2H3α 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 WOR fragment was amplified with primers WOR1F and WOR1R and cloned into pNIM1 vector (42) at SalI and BglII sites. SC2H3α and SC2H3α strains were transformed with pNIM1-WOR1 with selection on nourseothricin-containing medium. The resulting strains, SC2H3α-pWOR1 and SC2H3α-pWOR1, were grown in the presence of doxycycline (50 μg/ml) to induce the opaque state (43). 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 (33) were converted into 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 × 10^6 cells each in Spider medium (44) and incubated at 23°C for 24 h. Resulting tetraploid cells were selected on SC-LEU-ARG medium for 48 h incubation at 30°C, and transferred to SC-HIS-MET medium for protein–protein interaction detection. Chromosome loss was induced on sporulation medium at 37°C for 10 days as described previously (45).

RESULTS

Establishment of a sequence-validated Candida albicans ORFeome

Our objective was to develop a C. albicans ORFeome encompassing as many of the 6205 ORFs predicted in Assembly 21 of the C. albicans genome as possible (46). To this aim, forward and reverse oligonucleotides with, respectively, attB and attP sequences at their 5′ ends were synthesized for each of the 6205 ORFs (Supplementary Table S1), used in independent PCR reactions with genomic DNA of C. albicans strain SC5314 (47) and the resulting PCR products were cloned in pDONR207 using Gateway™ recombinational cloning (31). 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 6205 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 6179 remaining ORFs, 5099 (83%) 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; (48)). We defined 4 groups: (i) 4061 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) 108 sequences that do not carry any SNP but carry DIPs (No SNP, DIPs); (iii) 843 sequences that carry SNPs but not DIP (SNPs, no DIP); and (iv) 87 sequences that carry both SNPs and DIPs (SNPs and DIPs) (Figure 1A and Supplementary Table S1). DIPs, multiples of 3, vary from 3 to 33 nt in length. ORFs with DIPs non-multiples of three were also retained when present within introns.

Overall, 4061/6179 variation-free C. albicans ORFs (65.7%) are now available to the community. Among the 930 SNP-containing ORFs, 460 contain 1 SNP, 131 contain 2 SNPs, 78 contain 3 SNPs, 74 contain 4 SNPs and 187...
Figure 1. Statistics on the Candida albicans ORFeome. (A) Percentage of successfully cloned ORFs that are identical to the reference sequence (No SNPs, no DIPs) or that contain SNPs and/or DIPs. Only SNPs that do not introduce a STOP codon and DIPs multiple of 3 bp have been accepted. (B) Haplotypes distribution. Each cloned ORF has been assigned to HapA or HapB when there are differences between the two alleles of the ORF, and HapA/B when the two alleles of the ORF are identical (as defined in Assembly22). (C) Success rate on each chromosome. The graphs represent the number of reference ORFs and the number of successfully cloned ORFs on each chromosome, as well as the overall percentage of success for each chromosome. (D) Percentage of success based on ORF size.

ORFs contain 5 or more SNPs (up to 72 SNPs). For 36% of the validated ORFs, there was no difference between the two alleles in the reference sequences (HapA/B; Figure 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; Figure 1B) and haplotype B ORFs were cloned in 31% of cases (HapB; Figure 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%) (Figure 1C). Although ORFs up to 1.5 kb could be cloned with a success rate around 87% and ORFs between 1.5 and 4 kb could be cloned with a success rate around 77%, we observed a decreased success rate for the longer ORFs (Figure 1D). GC content did not seem to be the cause for either lack of PCR amplification or cloning failure as we did not see a statistically significant difference in GC content when comparing successfully cloned and failed-ORFs (data not shown).

Analysis of the 6875211 nucleotides of the 5099 sequence-validated ORFs revealed 3029 SNPs. The ratio of non-synonymous to synonymous changes due to SNPs was around 1. These nucleotide substitutions could be either real SNPs between our strain and the reference strain, or due to poor quality of some reference sequences, or mutations in the primers or PCR-induced mutations. Overall, the resulting maximum error rate amounted to $4.4 \times 10^{-4}$ i.e. 1 SNP every 2270 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 (Supplementary Tables S3 and 4). In this study, we also detected recombinant haplotypes for 116 of the cloned ORFs (Supplementary Table S5) that displayed characteristics of both reference haplotypes.

Taken together, our study provides an extensive, 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://candidaorfome.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 (Figure 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 (Figure 2). The reference sequence data have been extracted from CGD (48). The database can be queried through the ORF number (orf19.xxx or 19.xxx; (5)), the Assembly 22 name (Ci_XXXXXX 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 to Gateway™-adapted destination vectors. While destination vectors for ORF expression in hosts such as *E. coli* or *S. cerevisiae* are available (49,50), only a few destination vectors for ORF expression in *C. albicans* have been reported (12). 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 (12) 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 Figure 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, the promoter P*TDH3* for ORF expression in *C. albicans* genome at the *RPS1* locus, (iii) filamentation phenotype and/or (iv) detection of the 3xHA- or TAP-tagged Um6 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 five 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 previously described (12,51,52), overexpression of Um6 led to hyperfilamentation. This phenotype was observed for all strains, regardless of the tag used (Figure 4A and B). Nevertheless, the C-terminal TAP-tag appeared to hinder Um6 function. Indeed, protein tags can potentially alter protein localization and/or protein function. One strength of our system is that it provides a choice of tagging, that could minimize the effect of the tag on a specific protein function. 3xHA- or TAP-tagged Um6 was detected in the relevant strains and under the appropriate growth conditions (Figure 5A and B). We noticed extra bands migrating faster than the 3HA-tagged Um6 protein, which are likely to result from protein degradation. The 3HA-tagged Um6 protein was extracted by suspending the cells in 2× Laemml sample buffer and boiling at 100°C for 10 min. This rapid protocol might be responsible for the observed protein degradation.

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 (33). Hence, there was a need to engineer the system for high throughput screening based on rapid and convenient recombinant 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 pleomorphic 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
Figure 2. Snapshot of the CandidaOrfDB interface. An example of an ORF page is shown. In the ‘ORF details’ box, the different ID names, the length and the chromosome location of the ORF of interest are displayed. The haplotype assigned to the cloned ORF is noted (A, B or A/B if there is no allelic differences or equal number of differences against both haplotypes). The coordinates of introns are indicated when present. The summary results of the sequence analysis against the reference sequence are presented. The ‘SNP(s)’ box shows a table that lists the sequence differences between the cloned ORF and the reference sequences (Haplotypes A and B from Assembly22, and Assembly21 sequences). The ‘Nucleotide and Protein sequences’ boxes display the sequences with a color code for synonymous and non-synonymous SNPs. All sequences can be downloaded. The ‘Resources’ box displays links towards information that is relevant to each resource, i.e. oligonucleotide sequences for the BP clones, barcode sequence for the overexpression plasmids and the \( C.\) \( albicans\) overexpression strains, position of the clone in the different plates of the collection, plasmid sequences. The ‘Restrictions on the cloned sequence’ box indicates the existence of restriction sites, as well as the size of the expected fragments for enzymes that are used in regards to subsequent applications of these donor plasmids.

state by chromosome loss (reviewed in (53)). 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 \( a1/\alpha2\) complex, encoded at the mating type like (\( MTL\)) loci, acts as a repressor of opaque switching (54). In that context, the two-hybrid strain SC2H3 originating from the diploid \( a/\alpha\) SN152 strain (55) was deleted for the \( MTLa\) or \( MTLa\) locus to construct SC2H3a 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 (43,56).

A proof of principle of the whole procedure for protein–protein detection is shown in Figure 6. The MAP 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 SC2H3a a SC2H3a strains. 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 \( HIS1\) marker, allowing only the mating-derived strains to grow on histidine-free medium, whereas the original diploid strains could not (Figure 6B). It is possible that some strains reverted to the diploid state, but we did not perform FACS analysis to assess this as it is not relevant for the protein–protein interaction analysis. All transformants were able to grow in the
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_ (12,32,57). 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 5099 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 (16,17,19,22,24,30). Nevertheless, 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. 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 (58), 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, eukaryotic hosts other 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 (59). Indeed, in _C. albicans_, CUG is decoded as a serine instead of a leucine the majority of the...
Figure 4. Ume6-driven filamentation validates the primary set of 15 destination vectors. Inducible (A) and constitutive (B) overexpression of UME6 triggers filamentation. Isolates were grown in rich medium in presence or absence of ATc3 for 2–4 h at 30 °C. Cells were observed with a Leica DM RXA microscope (Leica Microsystems) with an x40 oil-immersion objective.

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% (27,30). Similar rates were observed in this work. Unlike other Gateway™ recombinational cloning projects (16,30), 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 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 two alleles of a heterozygous gene (60–62), 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 1081 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 (12,32,57). In this respect, the C. albicans ORFeome project is currently completing its second and third phases, which consist of transferring the 5099 cloned ORFs into barcoded destination vectors and generating a collection of C. albicans barcoded overexpression mutants, each mutant carrying one of the 5099 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
Figure 6. Proof of principle of two-hybrid based PPI detection via mating in Candida albicans. (A) Schematic representation of the concept. Diploid opaque MTLa bait-expressing cells were mixed with opaque MTLa prey-expressing cells to obtain tetraploids, as selected on leucine and arginine-free medium. Detection of protein-protein interaction was observed on medium lacking histidine as an indicator of expression of the two-hybrid readout marker. (B) Proof of principle using Hst7 as a bait and Cek1 as a prey, previously shown to interact (31). Cells of each type were spotted in a dilution series and growth was monitored on SC-leu-arg, which allowed growth of the tetraploid and diploid offspring and on SC-met-his, which allowed detection of PPI, only in those cells that are expressing both bait and prey proteins. As negative controls, tetraploids derived from the crossing of bait-expressing strains with empty prey vector transformed strains, and from the crossing of prey-expressing strains with strains expressing the empty bait vector grew on SC-leu-arg but not on SC-met-his.

valuable resource for the scientific community that should greatly facilitate future functional studies in C. albicans.

DATA AVAILABILITY
Plasmid sequences have been submitted to GenBank. Accession numbers are given in Table 2.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS
We are grateful to other members of our groups for their support throughout the development of the C. albicans ORFeome. We want to thank Cosmin Saveanu for his expertise in western-blotting and Thierry Ancelle for his expertise in statistical analysis.

FUNDING
Wellcome Trust [088858/Z/09/Z to C.A.M., C.D.]; European Commission (FINSysB) [PITN-GA-2008-214004 to C.D.]; Agence Nationale de la Recherche [KANJI, ANR-08-MIE-033-01 to C.D.; CANDICOL, ANR-10-01 to C.D.]; Belgian Science Policy Office, Interuniversity Attraction Poles Programme [IAP P7/28 to P.V.D.]; Medical Research Council New Investigator Award [G0400284 to C.A.M.]; MRC Centre for Medical Mycology and the University of Aberdeen [MR/M026663/1]; FINSysB Consortium PhD Fellowship [PITN-GA-2008-214004 to V.C.]; DIM-MallInf Région Ile-de-France PhD Fellowship (to A.N.); C. albicans ORFeome Project Postdoctoral Fellowship [WT088858MA to E.P.]; NPARI Consortium Postdoctoral Fellowship [LSHE-CT-2006-037692 to T.R.]; Programme Fungi from Institut Pasteur Maladies Infectieuses Postdoctoral Fellowship (to U.Z.); FINSysB Post-doctoral Fellowship [PITN-GA-
REFERENCES

1. Jones,T., Federspiel,N.A., Chibana,H., Dungan,J., Kalman,S., Magee,B.B., Newport,G., Thorstensen,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. Muzey,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 Biol., 14, R97.

3. d‘Enfert,C., Goyard,S., Rodriguez-Arnavelle,S., Frangeul,L., Jones,L., Tekaia,F., Bader,O., Albrecht,A., Castillio,L., Dominguez,E. et al. (2005) CandidaDB: a genome database for Candida albicans pathogenomics. Nucleic Acids Res., 33(Database Issue), D353–D357.

4. Braun,B.R., van Het Hoog,M., d‘Enfert,C., Morschhauser,J. (2013) Analysis of a fungus-specific ORFeome version 1.1: Yeast ORFeome: a resource for comparative molecular microbiology. BMC Genomics, 11, 470.

5. Nobile,C.J., Bruno,V.M., Richard,M.L., Davis,D.A. and Mitchell,A.P. (2003) Genetic control of chlamydospore formation in Candida albicans. Microbiology, 149, 3629–3637.

6. Roemer,T., Jiang,B., Davison,J., Ketela,T., Veillette,K., Breton,A., Roemer,T., Jiang,B., Damunken,A., Ketela,T., Veillette,K., Breton,A., et al. (2004) High-throughput generation of ORFeome: a resource that enables functional genomics. Proc. Natl. Acad. Sci. U.S.A., 101, 1318–1329.

7. Stynen,B., Van Dijck,P. and Tournu,H. (2010) A CUG codon adapted two-hybrid system for the pathogenic fungus Candida albicans ability to form single- and multi-strain biofilms. BMC Genomics, 13, 632.

8. Hanahan,D. (1983) Studies on transformation of human genes. Methods Enzymol., 100, 575–592.

9. Cevallos,E., Chauvel,M., Firon,A., Legrand,M., Nesseir,A., Cabral,V., Znadi,S., Goyard,S., Bachellier-Bassi,S., Firon,A., Legrand,M., Diogo,D., Nauleau,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.

10. Armstrong,C.M., Li,S., Jacotot,L., Bertin,N., Janky,R. et al. (2004) High-throughput generation of ORFeome using the Gateway system. Methods Mol. Biol., 220, 91–111.

11. 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.

12. Takahashi,Y. and Yoshida,M. (2009) Systematic cloning of an ORFeome using the Gateway system. Methods Mol. Biol., 577, 11–24.

13. 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.

14. Goyal,A., Zhao,J., Wise,K.J., Lopez-Hoyo,N., Jiang,L., Piccirillo,S., Yu,H. et al. (2004) Modular gene over-expression strategies for Candida albicans. BMC Genomics, 5, 349.

15. Nobile,C.J., Bruno,V.M., Richard,M.L., Davis,D.A. and Mitchell,A.P. (2003) Genetic control of chlamydospore formation in Candida albicans. Microbiology, 149, 3629–3637.

16. 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.

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

18. Hanahan,D. (1983) Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol., 166, 557–580.

19. Cabral,V., Znadi,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 Pathog., 10, e1004542.

20. Hanahan,D. (1983) Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol., 166, 557–580.

21. Wang,J., Yang,Y., Hu,L., Liu,X.F., Dong,C.X., Yang,J.Y., Hu,L.D., Liu,Y., Deng,C.X. et al. (2004) Genome-wide ORFeome cloning and analysis of Arabidopsis transcription factor genes. Plant Physiol., 135, 773–782.

22. Nobile,C.J., Bruno,V.M., Richard,M.L., Davis,D.A. and Mitchell,A.P. (2003) Genetic control of chlamydospore formation in Candida albicans. Microbiology, 149, 3629–3637.

23. Lamesch,P., Li,N., Milstein,S., Fan,C., Hao,T., Szabo,G., Hu,Z., et al. (2012) Modular gene over-expression strategies for Candida albicans. BMC Genomics, 5, 349.

24. Cabral,V., Chauvel,M., Firon,A., Legrand,M., Nesseir,A., Bachellier-Bassi,S., Chaudhari,Y., Munro,C.A. and D‘Enfert,C. (2012) ORFeome: a resource of comparative molecular microbiology. BMC Genomics, 11, 138.
36. 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.
37. 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. Eukaryot. Cell, 6, 2122–2138.
38. Enjalbert, B., Rachini, A., Vediappa, G., Pietrella, D., Spaccapelo, R., Vecchiarelli, A., Brown, A.J. and d’Enfert, C. (2009) A multifunctional, synthetic Gaussea princeps luciferase reporter for live imaging of Candida albicans infections. Infect. Immun., 77, 4847–4858.
39. 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.
40. Wilson, R.B., Davis, D. and Mitchell, A.P. (1999) Rapid hypothesis testing with Candida albicans through gene disruption with short homology regions. J. Bacteriol., 181, 1868–1874.
41. Walther, A. and Wendland, J. (2008) PCR-based gene targeting in Candida albicans. Nat. Protoc., 3, 1414–1421.
42. Park, Y.N. and Morschhauser, J. (2005) Tetracycline-inducible gene expression and gene deletion in Candida albicans. Eukaryot. Cell, 4, 1328–1342.
43. 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.
44. 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.
45. 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.
46. 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 Biol., 8, R52.
47. Fonzi, W.A. and Irwin, M.Y. (1993) Isogenic strain construction and gene mapping in Candida albicans. Genetics, 134, 717–728.
48. Skrzypek, M.S., Binkley, J., Binkley, G., Miyasato, S.R., Simison, M. and Sherlock, G. (2017) Candida Genome Database. http://www.candidagenome.org/
49. 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.
50. Halperz, 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.
51. Carlisle, P.L., Banerjee, M., Lazzell, A., Monteaudo, 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.
52. 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 Res., 9, 126–142.
53. Bennett, R.J. (2015) The parasexual lifestyle of Candida albicans. Curr. Opin. Microbiol., 28, 10–17.
54. 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.
55. 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. Eukaryot. Cell, 4, 298–309.
56. Zordan, R.E., Galgocyzy, 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.
57. Loli-Krippleber, R., Feri, A., Nguyen, M., Maufrais, C., Yansouni, J., d’Enfert, C. and Legrand, M. (2015) A FACScan-optimized screen identifies regulators of genome stability in Candida albicans. Eukaryot. Cell, 14, 311–322.
58. 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.
59. 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.
60. Coste, A., Turner, V., Ischer, F., Morschhauser, J., Porche, 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.
61. 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.
62. Muzzey, D., Sherlock, G. and Weissman, J.S. (2014) Extensive and coordinated control of allele-specific expression by both transcription and translation in Candida albicans. Genome Res., 24, 963–973.