Supporting Information

Padmanabhan *et al.* 10.1073/pnas.0809770105

SI Text

Sequence Analysis. First, we identified the orthologous ORFs in the *C. dubliniensis* genome by comparing the amino acid sequences of *CEN* adjacent ORFs of all chromosomes of *C. albicans*. We performed a BLAST search for the ORFs flanking the centromeres of *C. albicans* against the *C. dubliniensis* database (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/c_dubliniensis). Pairwise sequence alignment was performed using ClustalW (1) to analyze amino acid sequence homology between the ORFs of *C. albicans* and their homologs in *C. dubliniensis.* The synteny of the ORFs is maintained on both sides of the *CEN* regions in all chromosomes of *C. albicans* and *C. dubliniensis* except chromosome 6. Homologous ORFs in the *C. dubliniensis* genome were identified by a sequence analysis software Vector NTI (Invitrogen) where the minimum ORF length was set at 100 aa.

Homology Detection and Mutation Rate Measurement. We used Sigma (version 1.1.3) (2) and DIALIGN 2 (3) to align ORF-free centromeric and other intergenic sequences. Default parameters were used for both programs, but Sigma was given an auxiliary file of intergenic sequence from which to estimate a background model.

For protein-coding sequence, we ran WU-BLAST 2.0 (tblastn), querying each annotated coding region of *C. albicans* against the chromosome sequences of *C. dubliniensis.* Parameters used were "filter = seg matrix = blosum62 hspsepQmax = 1000 hspsepSmax = 2000." Hits with a summed P value of $\leq 1e$-30 were identified as potential orthologs. Criteria for ortholog assignment were sequence similarity and synteny (requiring at least two common syntenous immediate neighbors out of four). This led to 2,653 high-confidence predictions.

These orthologous genes were aligned (at the amino acid level) with T-Coffee (4). Then we considered the following seven amino acids when conserved, and coded by the indicated codons, in both species: GTn (valine), TCn (serine), CCn (proline), ACn (threonine), GCn (alanine), CGn (arginine), and GGn (glycine) (n, any nucleotide). Other synonymous codons, if any, were ignored. Leucine was ignored because of a variant codon, CTG, that codes for serine in these species. A naïve count of mutation rates in the third position yields 0.27. We improved on this by considering the genomewide bias for each codon, as follows: let the third-position conservation probability be q. Then if a third position nucleotide in *C. albicans* is b, in *C. dubliniensis* it stays b with probability q and mutates with probability $(1 - q)$. If it mutates, we assume that the probability of the new nucleotide is drawn from the known codon bias. For each amino acid A, we measure the individual mutation rates $P(b_2/b_1, A)$ for the third-position codon, changing from b_1 in *C. albicans* to b_2 in *C. dubliniensis* (the results are mathematically identical for evolution from a common ancestor), and solve for q; we then take the weighted average of q for all amino acids and all pairs of observed third-position nucleotides b_1 and b_2. This works out to $q = 0.58$, giving a mutation rate of 0.42. (Technically, this mutation rate is a slight overestimate, because we draw a mutated b_2 from a distribution that includes b_1; but it is a credible upper bound.)

Complementation Assay. To examine whether CdCse4p can complement CaCse4p function, we constructed a *C. albicans* strain where the first allele of Ca*CSE4* was disrupted using a *URA*-blaster cassette followed by recycling of the *URA3* marker, and the second allele was placed under control of the *PCK1* promoter (5). To disrupt the first Ca*CSE4* allele, a 4.9-kb *URA*-blaster-based Ca*CSE4* deletion cassette was released from pDC3 (6) as a SalI-SacI fragment and transformed BWP17, selecting for uridine prototrophy. The correct integrant (CAKS1b) was selected by Southern analysis. Thereafter, a Ura$^-$ strain, obtained by intrachromosomal recombination between *hisG* repeats resulting in the loss of *URA3* marker, was selected on medium containing 5-fluoroorotic acid (5-FOA). The correct revertant (CAKS2b) was identified by PCR analysis. To place the wild-type *CSE4* allele under regulation of the *PCK1* promoter in CAKS2b we linearized p*PCK1-CSE4* (6) by EcoRV and used it to transform strain CAKS2b, selecting transformants for uridine prototrophy. The desired integrant (CAKS3b) carrying the only full-length copy of *CSE4* under control of the *PCK1* promoter was identified by PCR analysis. CAKS3b can grow on succinate medium (where the *PCK1* promoter is induced) but is unable to grow on glucose medium (where *PCK1* promoter is repressed) (Fig. 2A). To test whether CdCse4p can complement CaCse4p function, we cloned both Cd*CSE4* and Ca*CSE4* genes in an *ARS2*/*HIS1* plasmid, pAB1 (7). A 2.14-kb fragment carrying Cd*CSE4* (CdChr3 coordinates 170,543–172,683) and a 2.13-kb fragment carrying Ca*CSE4* (CaChr3 coordinates 172,252–174,384) genes along with their respective promoters and terminators were amplified using FCdCSE4/RCdCSE4 and FCaCSE4/RCaCSE4 primer pairs (see Table S2 for primer sequences), respectively. These amplified Cd*CSE4* and Ca*CSE4* sequences were digested with SacI/HindIII and SacI/XbaI, respectively, and cloned into corresponding sites of pAB1 to get pAB1Cd*CSE4* and pAB1Ca*CSE4*. Subsequently CAKS3b was transformed with pAB1, pAB1Ca*CSE4*, or pAB1Cd*CSE4* and transformants were selected for histidine prototrophy on succinate medium followed by streaking on succinate and glucose-containing media.

Construction of CDM1 Carrying C-Terminally TAP-Tagged CdMIF2. Cd*MIF2* downstream sequence (from +1,634 to +2,198 with respect to the start codon of Cd*MIF2*) was PCR amplified with primer pair CdM3 (CGG GGT ACC GAT TGC AAG AAG TAC TAC ATA AGA GAG) and CdM4 (GCC CGA GCT CGC AGG TAA AAT TGT TCT TGA GGA GCC G), thereby introducing KpnI and SacI restriction sites (underlined). The resulting PCR-amplified fragment was digested with KpnI and SacI and cloned into corresponding sites of pUC19 to generate pCDM1. The TAP cassette along with the Ca*URA3* gene was released from plasmid pPK335 (8) as a BamHI-KpnI fragment and cloned into corresponding sites of pCDM1 to generate pCDM2. Subsequently the Cd*MIF2* ORF sequence from +1090 to +1548 was PCR amplified using primer pair CdM1 (ACG CGT CGA CCC CCC ACT GAT TAC GAT TAT GAA TCT GAT CC) and CdM2 (CAT GCC ATG GCC CAA TTC GTA TCG ATT TCT TCT GGT TTC) and cloned into pCDM2 as a NcoI-SalI fragment to get pCDM3. Finally, a 2-kb amplicon was PCR amplified by the primer pair CdM1 and CdM4 using pCDM3 as the template. This PCR fragment was used to transform the CdUM4B strain (9). The correct Ura$^+$ transformant (CDM1) was identified by PCR analysis.

Indirect Immunofluorescence. Asynchronously grown cells of Cd36 or CDM1 were fixed with 37% formaldehyde at room temperature for 1 h. Antibodies were diluted as follows: 1:30 for anti-∝-tubulin (YOL1/34) (Abcam), 1:500 for affinity-purified rabbit anti-Ca/CdCse4p (6) and rabbit anti-Protein A (Sigma), 1:500 for Alexa fluor 488 goat anti-rat IgG (Invitrogen), and 1:500 for Alexa fluor 568 goat anti-rabbit IgG (Invitrogen). The positions of nuclei of the cells were determined by staining with 4′,6-diamidino-2-phenylindole (DAPI) as described previously (6). Cells were examined at 100[mult] magnification on a confocal laser scanning microscope (LSM 510 META, Carl Zeiss). Using LSM 5 Image

Examiner, digital images were captured. Images were processed by Adobe PhotoShop software.

ChIP Assay. Chromatin immunoprecipitation (ChIP) by anti-CdCse4 antibodies followed by PCR analysis was done as described previously (10), which suggests that the predicted centromeric regions of all chromosomes of *C. dubliniensis* are enriched in centromeric-specific histone (CdCse4p) binding. Asynchronously grown culture of Cd36 was crosslinked with formaldehyde and sonicated to get chromatin fragments of an average size of 300–500 bp. The fragments were immunoprecipitated with anti-Ca/CdCse4p antibodies and checked by PCR. PCR was set up using 10 pmol of both forward and reverse primers (MWG Biotech and Ocimum Biosolutions), 5 μl of 10[mult] Taq buffer (Sigma), 5 μl of 2.5 mM dNTPs mix, 2 μl of DNA template, and 0.3 μl of Taq polymerase (Sigma) in a 50-μl reaction volume. PCR amplification was carried out using a PCR machine (Bio-Rad) with the following conditions: 1 min at 94 °C (denaturation), 30 s at 45 °C-55 °C (annealing temperature is variable with the primers used), and 1 min at 72 °C (extension). A final extension of 4 min was given at 72 °C. PCR with total DNA (1:10 dilution) and plus or minus antibody ChIP DNA fractions were performed using one-twenty-fifth of the template. The boundaries of the *CEN* regions on each chromosome of *C. dubliniensis* were mapped using semiquantitative ChIP-PCR in strain Cd36. Sequence-specific PCR primers were designed at ≈1-kb sequence intervals that span the putative *CEN* region of each chromosome of *C. dubliniensis* (Table S2). Cd*LEU2* PCR primers were used as an internal control in all PCR reactions. PCR amplification was performed and the PCR products were resolved on 1.5% agarose gels and band intensities were quantified using Quantity One 1-D Analysis Software (Bio-Rad). Enrichment values equal (+Ab) minus (−Ab) signals divided by the total DNA signal and were normalized to a value of 1 for *LEU2*. The PCR primers used in this study are listed in Table S2. Similarly, a ChIP assay to determine occupancy of TAP-tagged CdMif2p was performed using the strain CDM1 with anti-Protein A antibodies. All other conditions were identical to those described above for CdCse4p ChIP antibodies.

1. Thompson J-D, Higgins D-G, Gibson T-J (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
2. Siddharthan R (2006) Sigma: multiple alignment of weakly-conserved non-coding DNA sequence. *BMC Bioinformatics* 7:143.
3. Morgenstern B (1999) DIALIGN2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics* 15:211–218.
4. Notredame C, Higgins D, Heringa J (2000) T-Coffee: a novel method for multiple sequence alignments. *J Mol Biol* 302:205–217.
5. Leuker C-E, Sonneborn A, Delbruck S, Ernst J-F (1997) Sequence and promoter regulation of the *PCK1* gene encoding phosphoenolpyruvate carboxykinase of the fungal pathogen *Candida albicans*. *Gene* 192:235–240.
6. Sanyal K, Carbon J (2002) The CENP-A homolog CaCse4p in the pathogenic yeast *Candida albicans* is a centromere protein essential for chromosome transmission. *Proc Natl Acad Sci USA* 99:12969–12974.
7. Baum M, Sanyal K, Mishra P-K, Thaler N, Carbon J (2006) Formation of functional centromeric chromatin is specified epigenetically in *Candida albicans*. *Proc Natl Acad Sci USA* 103:14877–14882.
8. Corvey C, *et al.* (2005) Carbon source-dependent assembly of the Snf1p kinase complex in *Candida albicans*. *J Biol Chem* 280:25323–25330.
9. Staib P, Moran G-P, Sullivan D-J, Coleman D-C, Morschhauser J (2001) Isogenic strain construction and gene targeting in *Candida dubliniensis*. *J Bacteriol* 183:2859–2865.
10. Sanyal K, Baum M, Carbon J (2004) Centromeric DNA sequences in the pathogenic yeast *Candida albicans* are all different and unique. *Proc Natl Acad Sci USA* 101:11374–11379.
11. Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5:150–163.
12. Mishra P-K, Baum M, Carbon J (2007) Centromere size and position in *Candida albicans* are evolutionarily conserved independent of DNA sequence heterogeneity. *Mol Genet Genomics* 278:455–465.
13. Sullivan D-J, Westerneng T-J, Haynes K-A, Bennett D-E, Coleman D-C (1995) *Candida dubliniensis* sp. nov.: phenotypic and molecular characterization of a novel species associated with oral candidosis in HIV-infected individuals. *Microbiology* 141:1507–1521.
14. Wilson R-B, Davis D, Mitchell A-P (1999) Rapid hypothesis testing with *Candida albicans* through gene disruption with short homology regions. *J Bacteriol* 181:1868–1874.

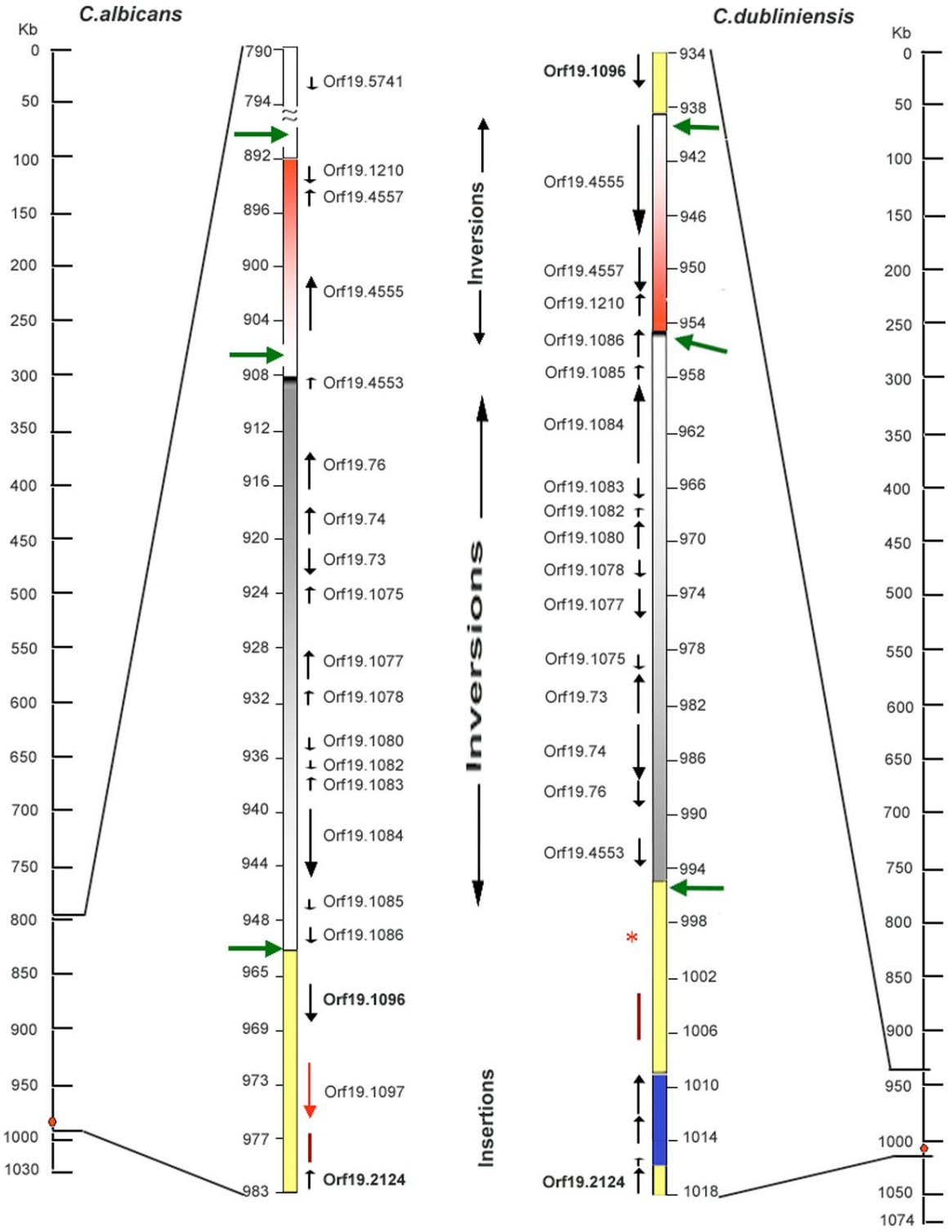

Fig. S1. Comparative analysis of the *CEN6* region of *C. albicans* and its orthologous region in *C. dubliniensis* showing genome rearrangement. Chromosomal maps of chromosome 6 of *C. albicans* and *C. dubliniensis* are shown, where the red dots represent the *CEN* regions. Black arrows along with the ORF numbers show the gene arrangement and the direction of transcription. Two paracentric inversions in *C. dubliniensis* are marked in shaded red and gray boxes. The direction of the shaded boxes (gradation of colors) represents the inversions that have occurred in *C. dubliniensis* when compared to *C. albicans*. The green arrows show the breakpoints where the inversions have occurred. The blue region in *C. dubliniensis* shows the region of insertions of ORFs from other chromosomes. The yellow regions are unaltered. The orange arrow shows Orf19.1097 in *C. albicans* and the orange star in the *C. dubliniensis* map shows that there is a premature termination codon in the Orf19.1097 homolog of *C. albicans* in *C. dubliniensis*. Brown bar, Cse4p-binding region.

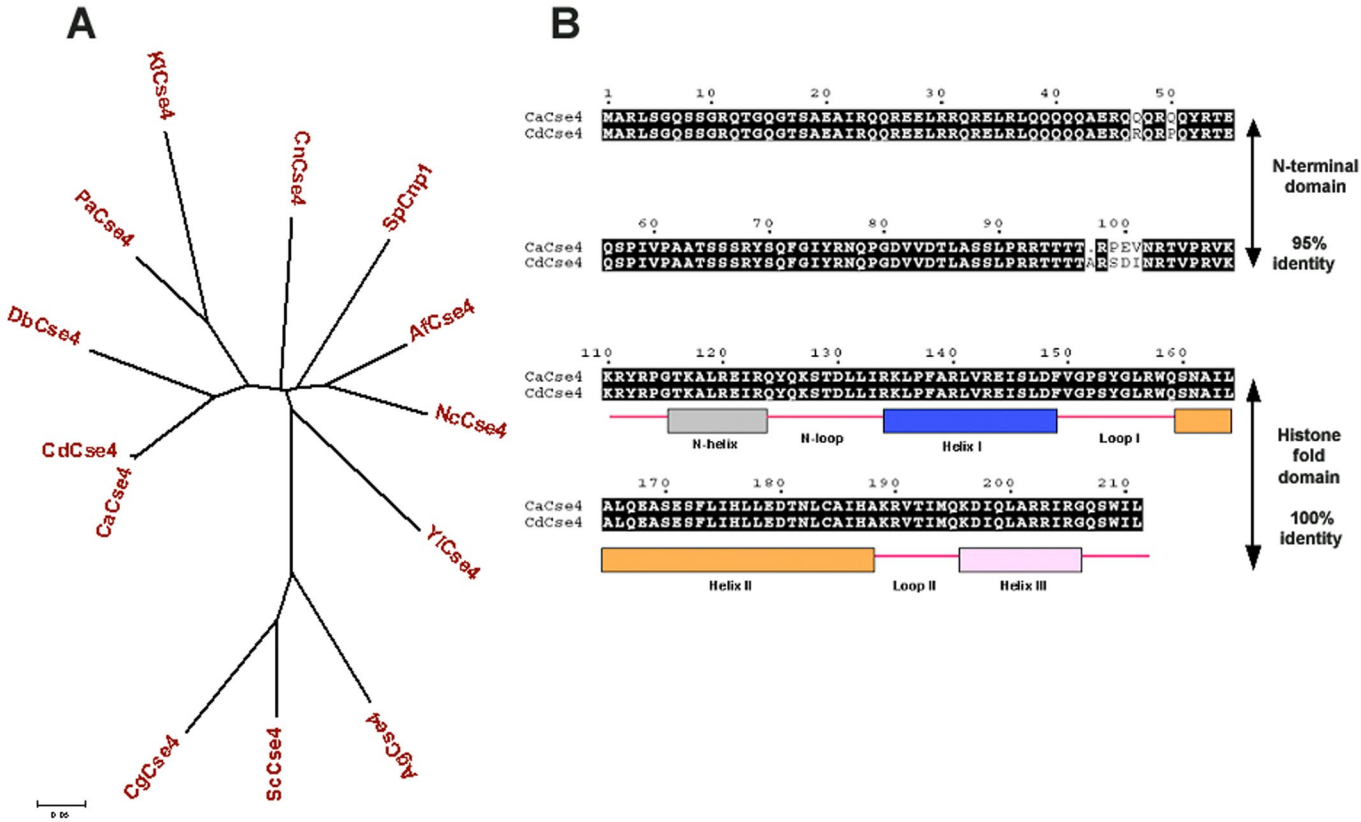

Fig. S2. The centromeric histone in *C. dubliniensis*, CdCse4p, belongs to the Cse4p/CENP-A family. (*A*) Phylogenetic tree of the Cse4 protein sequences in yeasts in the radiation format using the neighbor-joining method of Molecular Evolutionary Genetics Analysis version 3.1 (MEGA) software (11), showing that Cse4 proteins in *C. albicans* and *C. dubliniensis* are highly related. Ca, *Candida albicans*; Cd, *C. dubliniensis*; Db, *Debaryomyces hansenii*; Pa, *Pichia angusta*; Kl, *Kluyveromyces lactis*; Cn, *Cryptococcus neoformans*; Sp, *Schizosaccharomyces pombe*; Af, *Aspergillus fumigatus*; Nc, *Neurospora crassa*; Yl, *Yarrowia lipolytica*; Ag, *Ashbya gossypii*; Sc, *Saccharomyces cerevisiae*; Cg, *Candida glabrata.* (B) Pairwise comparison of Cse4p in *C. albicans* and *C. dubliniensis* showing homologies in the N-terminal region and the C-terminal histone-fold domain.

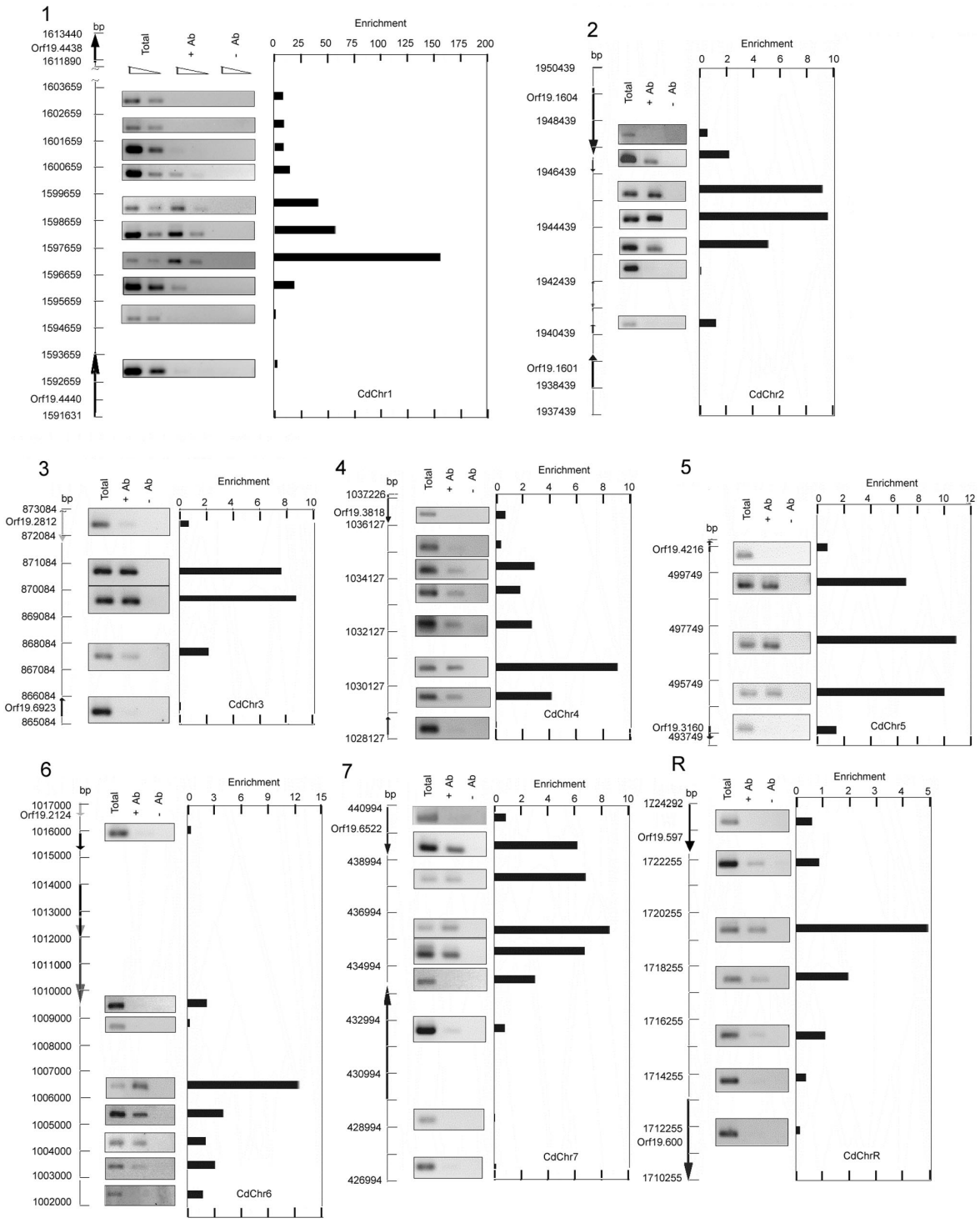

Fig. S3. Relative enrichment profiles of CdCse4p in various *C. dubliniensis* chromosomes. CdCse4p-associated chromosome regions were enriched by ChIP using anti-Ca/CdCse4p antibodies. Specific primers corresponding to putative centromere regions of *C. dubliniensis* were used to PCR amplify DNA fragments (150–300 bp) located at specific intervals as indicated (Table S2). PCR was performed on total, immunoprecipitated (+Ab), and beads-only control (−Ab) DNA fractions. Reverse images of ethidium bromide-stained PCR products resolved on 1.5% agarose gels are aligned with respect to their chromosomal map position of each *CEN* region. The coordinates of primer locations are based on the present version (May 16, 2007) of the *Candida dubliniensis* genome database. Enrichment values are calculated by determining the intensities of (+Ab) minus (−Ab) signals divided by the total DNA signals and are normalized to a value of 1 for the values obtained for a noncentromeric locus (Cd*LEU2*). The intensity of each band was determined by using Quantity One 1-D Analysis Software (Bio-Rad). The CdCse4p enrichment profiles on *C. dubliniensis* chromosomes at corresponding regions are indicated. Solid arrows and shaded arrows correspond to complete and incomplete ORFs, respectively, and indicate the direction of transcription.

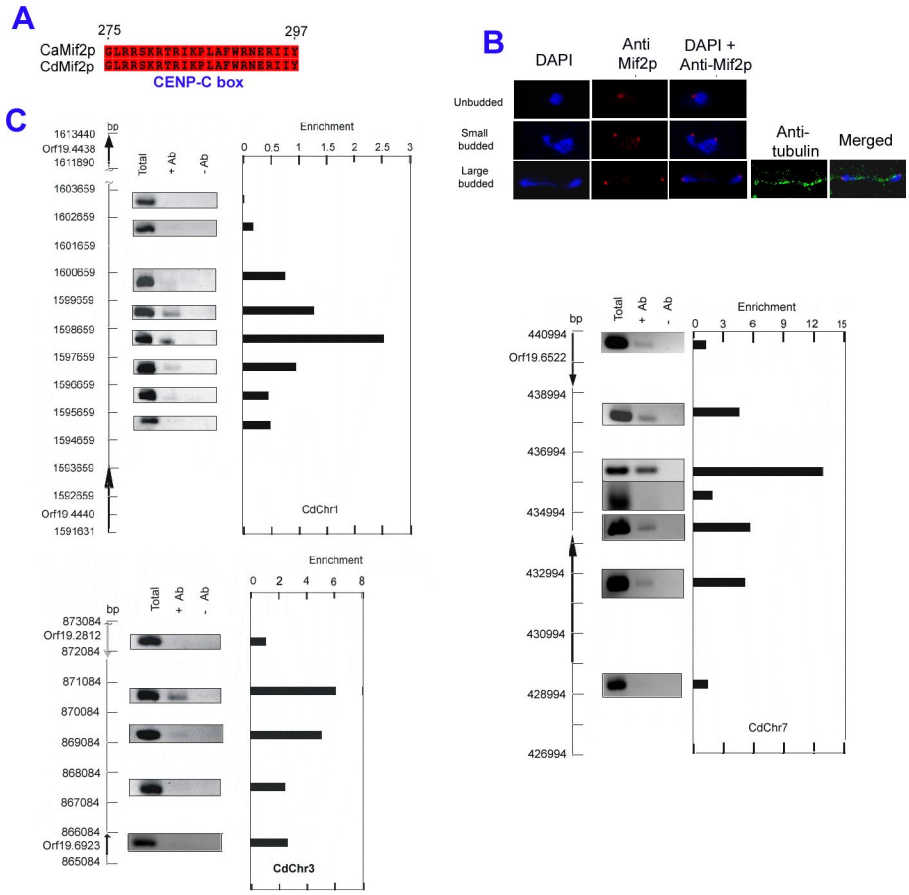

Fig. S4. The CENP-C homolog in *C. dubliniensis* (CdMif2p) is colocalized with CdCse4p. (A) Sequence alignment of CaMif2p and CdMif2p showing the conserved CENP-C block (red box). (B) Localization of CdMif2p at various stages of the cell cycle in *C. dubliniensis*. (C) ChIP enrichment profiles of CdMif2p on chromosomes 1, 3, and 7in the strain CDM1 by determining the intensities of (+Ab) minus (−Ab) signals divided by the total DNA signals and are normalized to a value of 1 for the values obtained for a noncentromeric locus (Cd*LEU2*) (see Table S2 for primers).

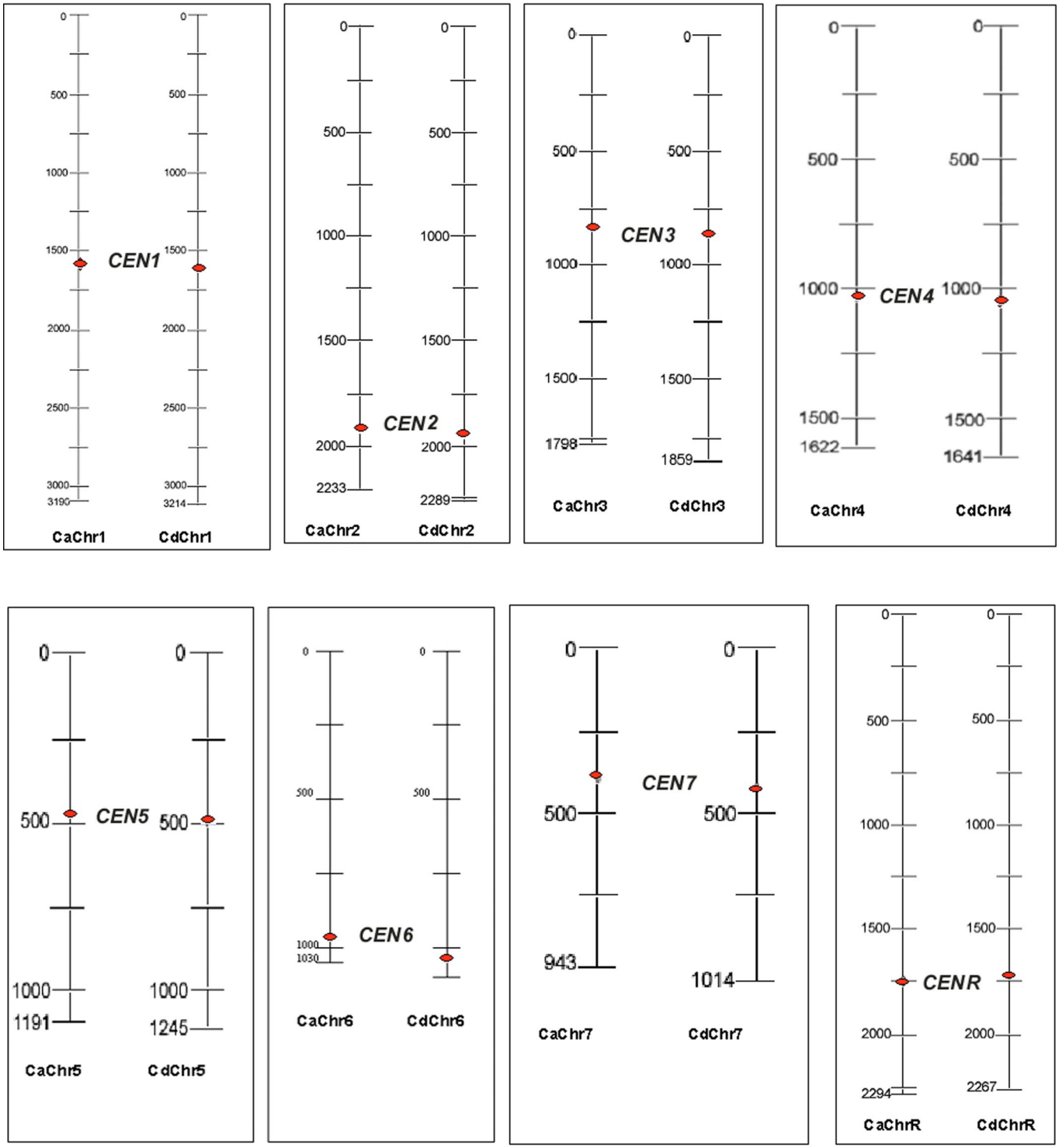

Fig. S5. Relative chromosomal positions of Cse4p-binding regions in *C. albicans* and *C. dubliniensis*. Red ovals, Cse4p-binding regions.

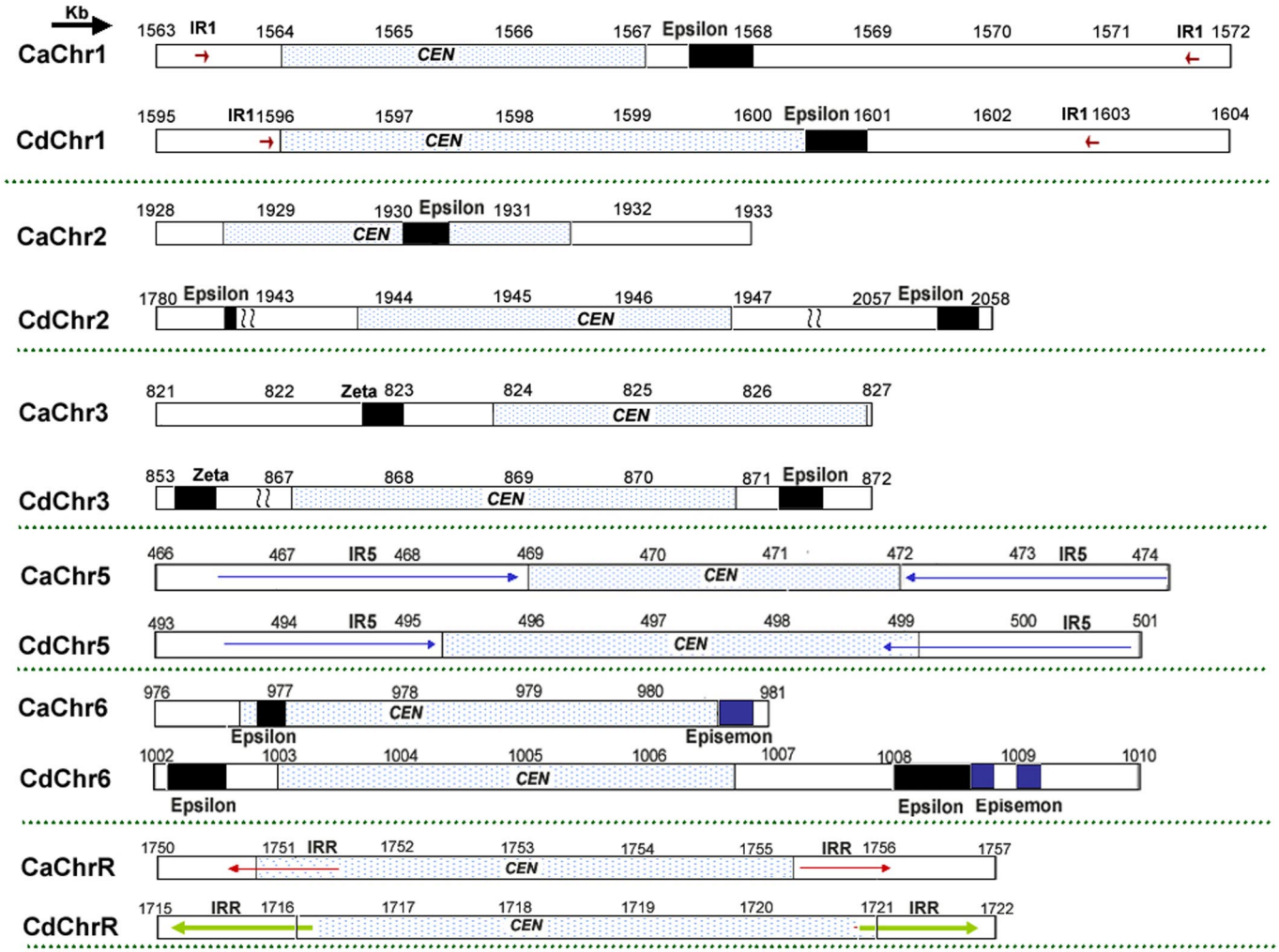

Fig. S6. Conserved blocks in the pericentric regions of various chromosomes of *C. dubliniensis* and *C. albicans*. The cyan stippled blocks represent the Cse4p-binding regions. DNA sequence stretches of various chromosomes having significant similarities (ClustalW scores >80) are shown by colored arrows as indicated. The numbers on the chromosomes represent their coordinates in their respective genome database. The direction of the arrows represents the orientation of repeats. A BLAST search was done to identify the repeats flanking the *CEN* region against the *C. dubliniensis* genome database with *C. albicans CEN* flanking repeats as the query sequences (12). The inverted repeats were observed in the chromosomes R, 1, and 5 of *C. albicans* and *C. dubliniensis* (see Table S4). The LTRs such as epsilon, zeta, episemon are also shown.

Table S1. Comparison of the amino acid sequence homology of the ORFs flanking the *CEN* regions in *C. albicans* and *C. dubliniensis*

			C. albicans		*C. dubliniensis*			
Chr no.	*C. albicans* ORF no.	*C. dubliniensis* ORF no.	Chromosomal coordinates	Amino acid length	Chromosomal coordinates	Amino acid length	Orientation	Amino acid homology (%)
1	4438	Cd36_06830	1580117–1581640	507	1611890–1613440	516	Direct	88
	4440	Cd36_06810	1559352–1561871	839	1591631–1594162	843	Direct	91
2	1601	Cd36_23540	1923194–1924363	389	1938439–1939608	389	Direct	99
	1604	Cd36_23560	1934775–1931570	916	1947203–1949623	806	Reverse	84
3	2812	Cd36_83930	828667–827105	503	871879–873366	495	Reverse	84
	6923	Cd36_83920	820347–821378	343	865253–866083	276	Direct	90
4	3818	Cd36_44310	1010148–1009312	278	1036396–1037226	276	Reverse	88
	3821	Cd36_44290	1000558–999371	395	1025948–1027126	392	Reverse	81
5	3160	Cd36_51930	467208–466702	168	493689–494072	127	Reverse	95
	4216	Cd36_51940	473741–474247	168	500592–500975	127	Direct	94
6	1096	Cd36_64780	965934–968573	879	934029–936683	884	Direct	84
	2124	Cd36_65100	982460–981390	353	1016599–1017672	357	Reverse	87
7	6522	Cd36_71800	431903–430173	586	439178–440899	573	Reverse	94
	6524	Cd36_71780	423631–422459	390	424821–425993	390	Reverse	99
R	597	Cd36_33630	1759087–1757405	560	1722610–1724292	560	Reverse	97
	600	Cd36_33620	1748818–1745649	1056	1710255–1713449	1064	Reverse	90

The coordinates of the *C. albicans* and *C. dubliniensis* chromosomes correspond to Assembly 20 of the *Candida albicans* Genome Database and the present version of the *Candida dubliniensis* Genome Database (May 16, 2007). As no true homolog of Orf19.3820 is present in chromosome 4 of *C. dubliniensis*, we considered Orf19.3818 (which is the adjacent ORF) for the sequence comparison. The new nomenclature for *C. dubliniensis* followed in the GeneDB database is also shown.

Table S2. Primers used

List of PCR primers used for ChIP assays

Primer	Sequence	Chromosomal locations
For Cd*CEN1*		
CdCEN1–1(F)	AAGCCCTTTGGATGTTGACTACGC	1593208–1593231
CdCEN1–2(R)	CCATCGACAGGGCCCATGTG	1593417–1593398
CdCEN1–3(F)	TATGATTATACCCCAATCCA	1595086–1595105
CdCEN1–4(R)	AGGATCAGTTACCAATGTTG	1595287–1595268
CdCEN1–3′(F)	CAACAATCAACAATTTCTGCTCCTCATG	1596131–1596158
CdCEN1–4′(R)	AAGTGGGTATCACCTTATTCGCAAATGA	1596368–1596341
CdCEN1–5(F)	CCTTTTTAAACGTGACACGCTCAAA	1597063–1597087
CdCEN1–6(R)	GGAAAAGTTGCGTGAGGAAATGGA	1597302–1597279
CdCEN1–5′(F)	CGGGTGCATCTAAGAAGGGTTTTA	1598062–1598085
CdCEN1–6′(R)	CAATATAACCTTGCACCCGTCAAATACG	1598347–1598320
CdCEN1–7(F)	GTTGCAGTGCATTGTACGAGGTAAGCTC	1599081–1599108
CdCEN1–8″ (R)	TGCAACTGATCCGAGACAACTTCAAAC	1599271–1599245
CdCEN1–7′(F)	GATCGCAAGCGAAGCACGAAATGAC	1600481–1600505
CdCEN1–8′(R)	CAATGTCTGTTCGACCACCATTCCC	1600721–1600697
CdCEN1–9(F)	AGAGCGAGCACCTGGTATTCCCAAG	1601290 – 1601314
CdCEN1–10(R)	CACCCAAAGCCCAGCTTAAATTCC	1601509–1601486
CdCEN1–9′(F)	TTTCAATTTAGCTGACTCCTTACCCTGG	1602167–1602194
CdCEN1–10′(R)	TTTTCGGTGATTTTGCCAAGAAGTTC	1602410 – 1602385
CdCEN1–11(F)	CAGCATTCATCCGGGTAAAGTGTTG	1603320–1603344
CdCEN1–12(R)	CAACGGATCCAAGGTCACCACATAG	1603543–1603519
Control (Noncentromeric locus in chromosome 7)		
CdLeu2–1(F)	AACTATCACAGTCTTGCCTGGTGA	119386–119409
CdLeu2–2(R)	ACAGCACCAGTGCCCCATTT	119618–119637
For Cd*CEN2*		
CdCEN2–1(F)	CGCGGTCCAAGAAGATAATC	1940515–1940534
CdCEN2–2(R)	CATCATGGGATGTAATTGCT	1940649–1940668
CdCEN2–3(F)	AGTGTAAGTCTTCGGGATAC	1942509–1942528
CdCEN2–4(R)	GTGAGCGAATAGAATAATTG	1942685–1942704
CdCEN2–5(F)	AGCTACATCTATTTTCAATGCACTC	1944606–1944630
CdCEN2–6(R)	AATTGCTCTGAAACAGCCAG	1944877–1944896
CdCEN2–7(F)	TATACCCCCGAATTAACAAGTGCGC	1943700–1943724
CdCEN2–8(R)	CAGTGCAGGTGCTTTCGTTTACCAG	1943847–1943871
CdCEN2–9(F)	CATCAGTTCAATTGATGGGGTTGTTCTG	1945542–1945569
CdCEN2–10(R)	AAACTGGCATAGCTTTTTGCATTATTGCC	1945736–1945764
CdCEN2–11(F)	ATTTCGAGAGGACTTGGTTCGTGC	1946646–1946669
CdCEN2–12(R)	CCGTACCCAAATAAAACTCCCAGC	1946844–1946867
CdCEN2–15(F)	TACAAAGCGGGTGATAAGGA	1947305–1947054
CdCEN2–16(R)	GGCGCAAAAGGAAATAGC	1947234–1947217
For Cd*CEN3*		
CdCEN3–1(F)	ACACTGTCTTGTCTTGTGTCTGAAGTCG	865133–865160
CdCEN3–2(R)	TTCTCTGTGTGTGGGCCCTCAGTAC	865293–865317
CdCEN3–3(F)	TCATCCATCATATCACAAATCCTACTG	867274–867300
CdCEN3–4(R)	GTTATTTTGAAAGTTGGGGAGAGGG	867456–867480
CdCEN3–5(F)	CCTACGACATGAACACATCAAACTACTC	869090–869117
CdCEN3–6(R)	TGCTTTTGTTGAAAACTTGCGAAAC	869243–869267
CdCEN3–7(F)	AGGCTAGTCGGTGGTTAACGGTTGTGTG	870638–870665
CdCEN3–8(R)	GACTCGGAATAAACACCATCGCCGATGC	870856–870883
CdCEN3–9(F)	GGTCCAATTAGAATCGGGTCGTTCCATG	872528–872555
CdCEN3–10(R)	CGTCATCCCTTCTATCTCTAACGTG	872683–872707
For Cd*CEN4*		
CdCEN4–1(F)	ATCATATCATGCAGCCCAACTCCG	1028245–1028268
CdCEN4–2(R)	CGGACGTAGTGAAACGATTGTTGG	1028410–1028433
CdCEN4–3(F)	ACAATTCCCAGTAAACCATTATAAAAG	1029835–1029861
CdCEN4–4(R)	CATTCATAATCTGATTTGTAGGCTC	1029965–1029989
CdCEN4–3′(F)	TGCTAAACGACCCCCTCAAAA	1030554–1030574
CdCEN4–4′(R)	GTACGACGATCATCAGCAACCAA	1030776–1030798
CdCEN4–5(F)	AATTAATTCGGATAGTTGGGGGAGACCG	1032446–1032473
CdCEN4–6(R)	ATTGAGCTGCTCACTTCACTGCCAC	1032619–1032643
CdCEN4–5′(F)	GCAGCGTTCTTGTGACCGTGAG	1033199–1033220
CdCEN4–6′(R)	TTGAATTGGACAGGGGCTTAGG	1033477–1033498
CdCEN4–7(F)	TGTGGTGGAGGGTCATCCATTTGTTGGTTG	1034406–1034435

Primer	Sequence	Chromosomal locations
CdCEN4–8(R)	GGCGACCCTCATGCACCCTACCAAATAAA	1034609–1034637
CdCEN4–7′(F)	AAGTACGGATGGTTGTTA	1035010 - 1035028
CdCEN4–8′(R)	TAGTCATTCTGCCATCTCTTAT	1035231–1035252
CdCEN4–9(F)	CCATGAACAAAAGGTTAGGTGGTGCTCC	1036158–1036185
CdCEN4–10(R)	GGGGAGTTGAATGGTGTGGTGTTAC	1036367–1036391
For Cd*CEN5*		
CdCEN5–7(F)	TCCAGCGTCAGACATTTTTCCAGT	494058–494081
CdCEN5–8(R)	TGCCCCGCGGTTGACAGT	494213–494230
CdCEN5–1(F)	TGGCCTCTCCCTTACAAAATTTGCCC	495324–495349
CdCEN5–2(R)	GGGAGATGAGGGGTGATTGAGGTAATAG	495504–495531
CdCEN5–3(F)	GCTCCAGTACCAACGAAAACGACTTC	496907–496932
CdCEN5–4(R)	GCATTTGAAAACTGCCAATGTAGTC	497035–497059
CdCEN5–5(F)	GCTGGGATAGTTTAGAGGCAGACTGTG	498944–498971
CdCEN5–6(R)	CCTCAATCACCCCTCATCTCCCTAC	499130–499155
CdCEN5–9(F)	AAGGGCAAGGAACAAGTCACAAGT	500673–500696
CdCEN5–10(R)	TATCAGCGCCGGTTTTAGCAC	500941–500961
For Cd*CEN6*		
CdCEN6–15(F)	GTGCCAACTTTCTCCTGAT	1002806–1002824
CdCEN6–16(R)	AGCGATTATTAAGTCTATGTGG	1002985–1002964
CdCEN6–13(F)	GAAGCAGCGACCCAACAGATAA	1003044–1003065
CdCEN6–14(R)	TTGAGCGAAATTGGGTAGAGTC	1003262–1003283
CdCEN6–5(F)	TGTCCATTCCCCAAACTTCATACGGACCAC	1004039–1004068
CdCEN6–6(R)	GAATGCTGGAAGGACTTGAGAAATG	1004175–1004199
CdCEN6–5′(F)	GAAACCAATAACAAGGAAAGAGTA	1005046–1005069
CdCEN6–6′(R)	CAATGGGAAAAAGAAATCAGTAG	1005313–1005335
CdCEN6–7(F)	GACGAGAGCATGTACTCAACTACGTGTC	1006472–1006499
CdCEN6–8(R)	GAATCTTGATTGAAATGCGAGGAAC	1006668–1006692
CdCEN6–9(F)	CATCCAATAACATTGATTTACTACTTTTAG	1008985–1009014
CdCEN6–10(R)	TTTTTTTTTCTCAAAGATTTAGCAG	1009115–1009139
CdCEN6–9′(F)	TGTACGATCAACCCAGAGTGC	1009504–1009524
CdCEN6–10′(R)	ACATGCCATTACCAACAACAGTC	1009749–1009771
CdCEN6–3(F)	TAGCTGTATTAAAAAATTCTGGCCGCATA	1015917–1015945
CdCEN6–4(R)	TCTGACAAAAAACCTCGTATGACCC	1016066–1016042
For Cd*CEN7*		
CdCEN7–1(F)	CTAGAGCTATGTTGTGACAGTCCACC	427615–427640
CdCEN7–2(R)	CTTCTGGAATTGAGCCAATCCCTAG	427777–427801
CdCEN7–3(F)	CTAGCTATTCAAGCATCCGTAGGCAGTC	429103–429130
CdCEN7–4(R)	CCCATACCCGGGTGGTGTAGTATAA	429228–429252
CdCEN7–5(F)	GTAGGCGCTACATATGAACTTCGTGC	436328–436354
CdCEN7–6(R)	AGATAATGTCTGAATGTCATTCGGG	436479–436504
CdCEN7–9′(F)	TCCAATGGGTGCTAAGATGAA	434047–434068
CdCEN7–10′(R)	TCCCGCCTGATTTTTGAA	434292–434310
CdCEN7–7(F)	TTATTTGATAGCCTAATTTCACCTGATG	438005–438031
CdCEN7–8(R)	ATTAACTGACTTTGAACCAGCAATG	438205–438230
CdCEN7–9(F)	AACGGTCACCTGATGAATAGAGTGGC	432732–432758
CdCEN7–10(R)	GACTGAAGCGTCCATACTTGGGATC	432956–432981
CdCEN7–11(F)	CCCAGAAGTATCCACTAGGGAACTTG	435240–435268
CdCEN7–12(R)	TTGTTCTGGTCAATGGTACAGCAAC	435365–435390
CdCEN7–13(F)	CACGCAACTAGAATGGCATGAATATATG	439500–439527
CdCEN7–14(R)	AGATCCGGTGTCTGTCTTATTGCTC	439630–439654
CdCEN7–15(F)	CCTGCGTTGTAATCATTTGTTGTC	440443–440466
CdCEN7–16(R)	TTACTCCGCCTTTGATCCCTATTT	440640–440617
For Cd*CENR*		
CdCENR-1(R)	ATTAAGGAGCTTCGTGAGGCTGTCG	1723671–1723647
CdCENR-2(F)	CATTTCCTTCAAAGGCACCGGGATG	1723429–1723453
CdCENR-3(R)	ACGTTGCTTACTGGTGGCTATGCGG	1721710–1721686
CdCENR-4(F)	AAGCTTTTATTGCGGTGAACTGGGG	1721461–1721485
CdCENR-5(R)	ACATATAATAGCCTACCACACGCCTTGC	1719373–1719346
CdCENR-6(F)	TGACATTGTGGAAAGTTAATCGCGG	1719202–1719226
CdCENR-7(R)	TGAAATTGGAGACTAAGTGTTGCATTCG	1717531–1717504
CdCENR-8(F)	ACAGTTTCCACACAACTCAGCAAGACA	1717330–1717356
CdCENR-9(R)	TTTGCCGGGATAAGCTTTTATTGCG	1715642–1715618
CdCENR-10(F)	TTTCAGGACACCAGAAGATGGCCAC	1715409–1715433
CdCENR-9′(F)	CCCCCGCCGTGAAAAACA	1713200–1713217

Primer	Sequence	Chromosomal locations
CdCENR-10′(R)	CTACAAACGCCACACCCGAAACT	1713426–1713404
CdCENR-11(R)	ACCTCAACATCGACACAGTCGCACC	1712709–1712185
CdCENR-12(F)	AGCAGAAACCTCGATGTTTGAGCCG	1712487–1712511

The coordinates of the *C. dubliniensis* (Cd) chromosomes correspond to the present version of the *Candida dubliniensis* Genome Database (May 16, 2007).

List of PCR primers used for Cse4 complementation experiments

Primer	Sequence
FCaCse4	CCCGAGCTCCAATTAACAAATATTAATTACAAATG
RCaCse4	TGCTCTAGACCAAAATCCCTCTTTCTGTATTTG
FCdCse4	CCCGAGCTCCAAGTGTATTTTTCATCTTTGGTAG
RCdCse4	CCCAAGCTTCTATTTTGCCACCAAAACCCATCTT

Table S3. Sequence coordinates of the Cse4p-binding and the pericentric regions in all the chromosomes of *C. albicans* and *C. dubliniensis*

Chr no.	Regions	*C. albicans* coordinates	*C. dubliniensis* coordinates
R	Region from left ORF	1748819–1750873	1713450–1716138
	Cse4-binding region	1750874–1755348	1716139–1720954
	Region from right ORF	1755349–1757404	1720955–1722609
1	Region from left ORF	1561872–1564187	1594163–1596130
	Cse4-binding region	1564188–1567117	1596131–1600697
	Region from right ORF	1567118–1580116	1600698–1611889
2	Region from left ORF	1924364–1928514	1939609–1943699
	Cse4-binding region	1928515–1931474	1943700–1946867
	Region from right ORF	1931475–1931569	1946868–1947202
3	Region from left ORF	821379–823848	866084–867273
	Cse4-binding region	823849–826997	867274–870883
	Region from right ORF	826998–827104	870884–871878
4	Region from left ORF	1000559–1002628	1027127–1029834
	Cse4-binding region	1002629–1006266	1029835–1034637
	Region from right ORF	1006267–1009311	1034638–1036395
5	Region from left ORF	467209–469044	494073–495323
	Cse4-binding region	469045–472074	495324–499155
	Region from right ORF	472075–473740	499156–500591
6	Region from left ORF	975879–976872	993828–1003043
	Cse4-binding region	976873–980625	1003044–1006692
	Region from right ORF	980626–981389	1006693–1009568
7	Region from left ORF	423632–426037	425994–435239
	Cse4-binding region	426038–428938	435240–438230
	Region from right ORF	428939–430172	438231–439177

The DNA sequences present on either side of the *C. dubliniensis* Cse4-binding regions, orthologous to *C. albicans CEN*-containing ORF-free regions, are considered as pericentric regions. The region considered for pericentric sequence analysis in chromosome 6 in *C. dubliniensis* is the intergenic region between an ORF with no known *C. albicans* homolog (Cd36_65030) and the homolog of Orf19.4553 (Cd36_64990) (see supporting Fig. S1).

Table S4. Homology between the repeats in the pericentric region of *C. albicans* and *C. dubliniensis*

Chr no.	Repeat	Coordinates in *C. dubliniensis*	% homology between the inverted repeats*
R	IRR	1720958–1721270 (D)	100
	IRR	1716158–1715822 (R)	
1	IR1	1595932–1595989 (D)	96
	IR1	1602853–1602907 (R)	
5	IR5	493690–494369 (D)	99
	IR5	500277–500974 (R)	

A BLAST search was done to identify the repeats flanking the *CEN* region against the *C. dubliniensis* genome database with *C. albicans CEN* flanking repeats as the query sequences (10). The inverted repeats (IRR, IR1, and IR5) were observed in the chromosomes R, 1, and 5 of *C. albicans* and *C. dubliniensis*, respectively. The Epsilon repeats were observed in chromosomes 1, 2, 3, and 6 of *C. dubliniensis* but their homology is $<$25%. The Episemon repeats flank the *CEN6* region in both *C. albicans* and *C. dubliniensis*. *ClustalW scores

Table S5. List of strains

Yeast strains	Genotype	Source
Candida dubliniensis		
Cd36	Clinical isolate	13
CdUM4B	*ura3D1*::*FRT*/ *ura3D2*::*FRT*	9
CdM1	*ura3D1*::*FRT*/ *ura3D2*::*FRT MIF2*/*MIF2-TAP* (*URA3*)	This study
C. albicans		
BWP17	Δ*ura3*::*imm434*/ Δ*ura3*::*imm434* Δ*his1*::*hisG*/ Δ*his1*::*hisG* Δ*arg4*::*hisG*/ Δ*arg4*::*hisG*	14
CAKS1b	Δ*ura3*::*imm434*/ Δ*ura3*::*imm434* Δ*his1*::*hisG*/ Δ*his1*::*hisG* Δ*arg4*::*hisG*/ Δ*arg4*::*hisG CSE4*/ *cse4*::*hisG:URA:hisG*	This study
CAKS2b	Δ*ura3*::*imm434*/ Δ*ura3*::*imm434* Δ*his1*::*hisG*/ Δ*his1*::*hisG* Δ*arg4*::*hisG*/ Δ*arg4*::*hisG CSE4*/ *cse4*::*hisG*	This study
CAKS3b	Δ*ura3*::*imm434*/ Δ*ura3*::*imm434* Δ*his1*::*hisG*/ Δ*his1*::*hisG* Δ*arg4*::*hisG*/ Δ*arg4*::*hisG* *cse4*::*PCK1pr-CSE4*(*URA3*)/ *cse4*::*hisG*	This study