ACS SyntheticBiology
Cite This: ACS Synth. Biol. 2017, 6, 2028-2034
Letter
pubs.acs.org/synthbio

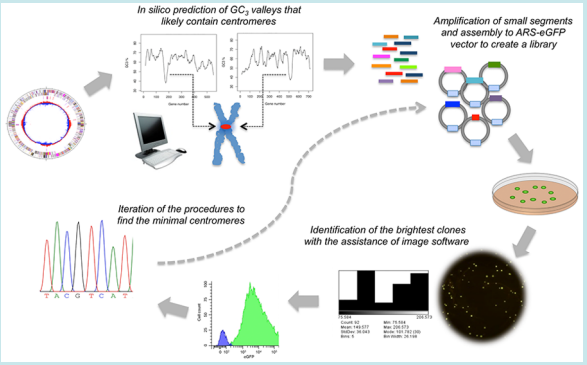
Rapid Isolation of Centromeres from *Scheffersomyces stipitis*

Mingfeng Cao,[†,‡] Arun Somwarpet Seetharam,[§] Andrew Josef Severin,[§] and Zengyi Shao*[,†,‡,∥,⊥]

[†]Department of Chemical and Biological Engineering, [‡]NSF Engineering Research Center for Biorenewable Chemicals (CBiRC), [§]Genome Informatics Facility, Office of Biotechnology, [∥]Interdepartmental Microbiology Program, and [⊥]The Ames Laboratory, 4140 Biorenewables Research Laboratory, Iowa State University, Ames, Iowa 50011, United States

Supporting Information

ABSTRACT: Centromeres (CENs) are the chromosomal regions promoting kinetochore formation for faithful chromosome segregation. In yeasts, CENs have been recognized as the essential elements for extra-chromosomal DNA stabilization. However, the epigeneticity of CENs makes their localization on individual chromosomes very challenging, especially in many not well-studied nonconventional yeast species. Previously, we applied a stepwise method to identify a 500-bp CEN5 from *Scheffersomyces stipitis* chromosome 5 and experimentally confirmed its critical role on improving plasmid stability. Here we report a library-based strategy that integrates *in silico* GC_3 chromosome scanning and high-throughput functional screening, which enabled the isolation of all eight *S. stipitis* centromeres with a 16 000-fold reduction in sequence very efficiently. Further identification of a 125-bp CEN core sequence that appears multiple times on each chromosome but all in the unique signature GC_3-valley indicates that CEN location might be accurately discerned by their local GC_3 percentages in a subgroup of yeasts.



KEYWORDS: *nonconventional yeasts, Scheffersomyces stipitis, episomal plasmids, centromeres, GC_3 chromosome scanning, epigeneticity*

Centromeres (CENs) are the genomic elements recognized by centromeric proteins to form kinetochores that interact with spindle microtubules during cell division to enable chromosome stable segregation.[1] In the yeast phyla, some species (e.g., *Candida albicans* and *Schizosaccharomyces pombe*) contain "regional CENs" spanning from several kb up to 110 kb that provide a large array of binding sites for centromeric proteins,[2,3] whereas other species (e.g., *Saccharomyces cerevisiae*) have "point CENs" that are only a couple of hundred bp long but provide sufficient anchoring points to form kinetochores that connect to microtubules when cells divide.[4] Point CENs in combination with autonomously replicating sequence (ARS) are the essential genetic elements in the development of stable episomal expression technology to manipulate various nonmodel yeasts with desired biotechnology potentials.

In the absence of a functional CEN, the ARS-containing vector can be replicated, but suffers from severe instability. For example, our recent study showed that when enhanced green-fluorescence protein (eGFP) was expressed in the CEN-lacking ARS vector in *Scheffersomyces stipitis* (a species renowned for its high xylose assimilation capacity[5]), the expression of eGFP exhibited an unusually broad range of signal and the copy number of plasmids varied from 0 to 140 copies per cell (Figure S1 in the Supporting Information).[6] This heterogeneity was caused by the plasmid unfaithful segregation during cell division. The plasmids after replication tended to stay in the parental cells, which led to progressive accumulation of plasmids in those cells. Consequently, more than 70% of the population completely lost the plasmids after 48 h of cultivation. To solve this plasmid instability issue, we introduced the bioinformatics tool previously developed by Lynch *et al.* that used *in silico* GC_3 chromosome scanning to analyze the GC% of the wobble positions of the codons for each chromosome with a sliding window of 15 genes.[7] With a stepwise functional validation strategy, one CEN was successfully identified from the single signature "GC_3-valley" of chromosome 5 (CEN5). The incorporation of the 500-bp CEN5 into the plasmid backbone only containing ARS improved the eGFP-positive population from 28% to 93%. The vector was stable with 3−5 copies per cell for at least 7 days, enabled homogeneous protein expression, and increased the titer of a commercially relevant compound by 3-fold.[6]

CEN formation is generally modulated in an epigenetic manner and no *cis*-acting conserved sequences could be used to predict the identity of CENs across species.[8] CEN specification depends on the presence of CenH3 nucleosomes, whose post-translational modifications (e.g., methylation, acetylation, and ubiquitylation) contribute to CEN function. Even for the same species (e.g., *S. cerevisiae*, *Kluyveromyces lactis*, and *Candida glabrata*), the CENs originated from different chromosomes

Received: May 18, 2017
Published: August 24, 2017

DOI: 10.1021/acssynbio.7b00166
ACS Synth. Biol. 2017, 6, 2028−2034

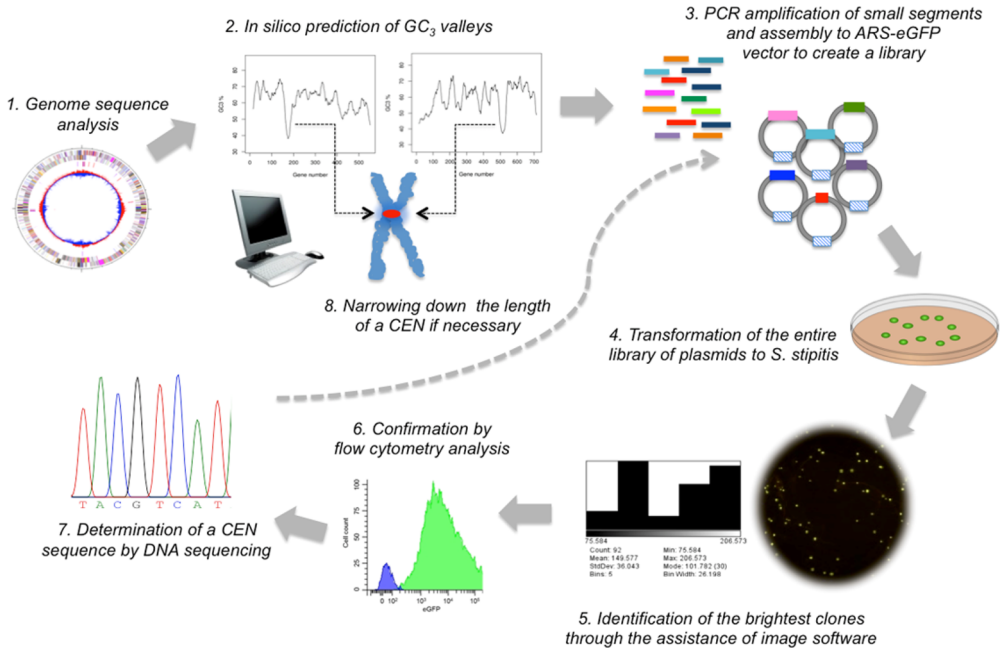


Figure 1. Scheme describing the rapid library-based CEN identification procedures.

Table 1. Summary of the Library-Based and the Stepwise CEN Identification Approaches

	library approach							stepwise approach[a]
	CEN1	CEN2	CEN3	CEN4	CEN6	CEN7	CEN8	CEN5
size of each chromosome (Mbp)	3.51	2.74	1.84	1.80	1.72	1.11	0.98	1.73
length of the GC_3-valley (bp)	194 092	119 392	130 977	93 962	128 639	118 669	91 786	105 828
length of the longest intergenic region located in each GC_3-valley (bp)	14 594	38 042[b]	24 208	26 877	30 037[b]	15 698[b]	36 077	17 264
average length of the remaining intergenic regions in each GC_3-valley (bp)	930	1976	1070	1163	1027	1230	1529	732
number of 3-kb fragments used to create the corresponding library	5	13	8	9	10	5	12	n.a.
number of clones giving the desired eGFP peak among 10 randomly picked colonies	8	2	10	8	10	2	9	n.a.
repeat number of the 125-bp CEN core in the entire predicted GC_3-valley	3	2	4	2	4	4	3	2
additional repeat number of the 125-bp CEN in each chromosome outside the GC3-valley/ distance to the boundary of GC3 valley (bp)/	0	0	1/2929/	1/900/	0	1/15606/	0	1/5056/
chromosomal location of the longest intergenic region identified in each GC_3-valley	2293090–2307683	1668602–1706643	1426116–1450323	1030956–1057832	886494–916530	273467–289164	290537–326613	652653–669916

[a]n.a., not applicable. [b]The longest intergenic regions were reported to be 38 214, 30150, and 16 278 in the work of Lynch *et al.*[7]

only share a CDEI/II/III-block structure, despite that the actual sequences comprising the individual blocks are highly variable.[3] To date, very little information is available regarding common features of yeast CENs. Encouraged by the previous success in isolating the functional 500-bp CEN5 in *S. stipitis*, we were intrigued by examining the accuracy of tagging a GC_3 valley to a CEN neighbor on a broader scale. To streamline the identification process so this CEN identification method could be applied more effectively, here we replaced the previous stepwise approach by a library-based identification strategy for the remaining seven chromosomes belonging to *S. stipitis* (Figure 1). Lynch *et al.*'s GC_3 chromosome scanning indicated that a uniquely sharp GC_3 valley arose on each chromosome,[7] but in fact the valleys spanned on chromosomes with a size variation from 92 kb to 194 kb in length (Figure S2). Within each valley, an unusually long intergentic region spanning from 14.6 kb to 38.0 kb was found (in contrast to the average length of the other intergenic regions ranging from 0.7 to 2.0 kb) (Table 1). We hypothesized that CENs are located in these abnormally long intergenic regions. Depending on the lengths of the seven predicted sequences, 5−13 segments of approximately 3-kb from each GC_3 valley were amplified. Seven libraries were subsequently created with the 3-kb amplicons inserted into the backbone plasmid containing ARS and eGFP expression cassette (hereafter named as ARS-eGFP). When the seven libraries were transformed to *S. stipitis*, the segments from each library endowed their transformants on the Petri dishes with different fluorescence levels, with the mean fluorescence value for each plate varying from 141 ± 29 to 175 ± 37, whereas that observed with the control ARS-eGFP plasmid was 107 ± 21 (Figure 2). Next, 15 colonies, including 10 colonies with high fluorescence values and 5 colonies with

DOI: 10.1021/acssynbio.7b00166
ACS Synth. Biol. 2017, 6, 2028−2034

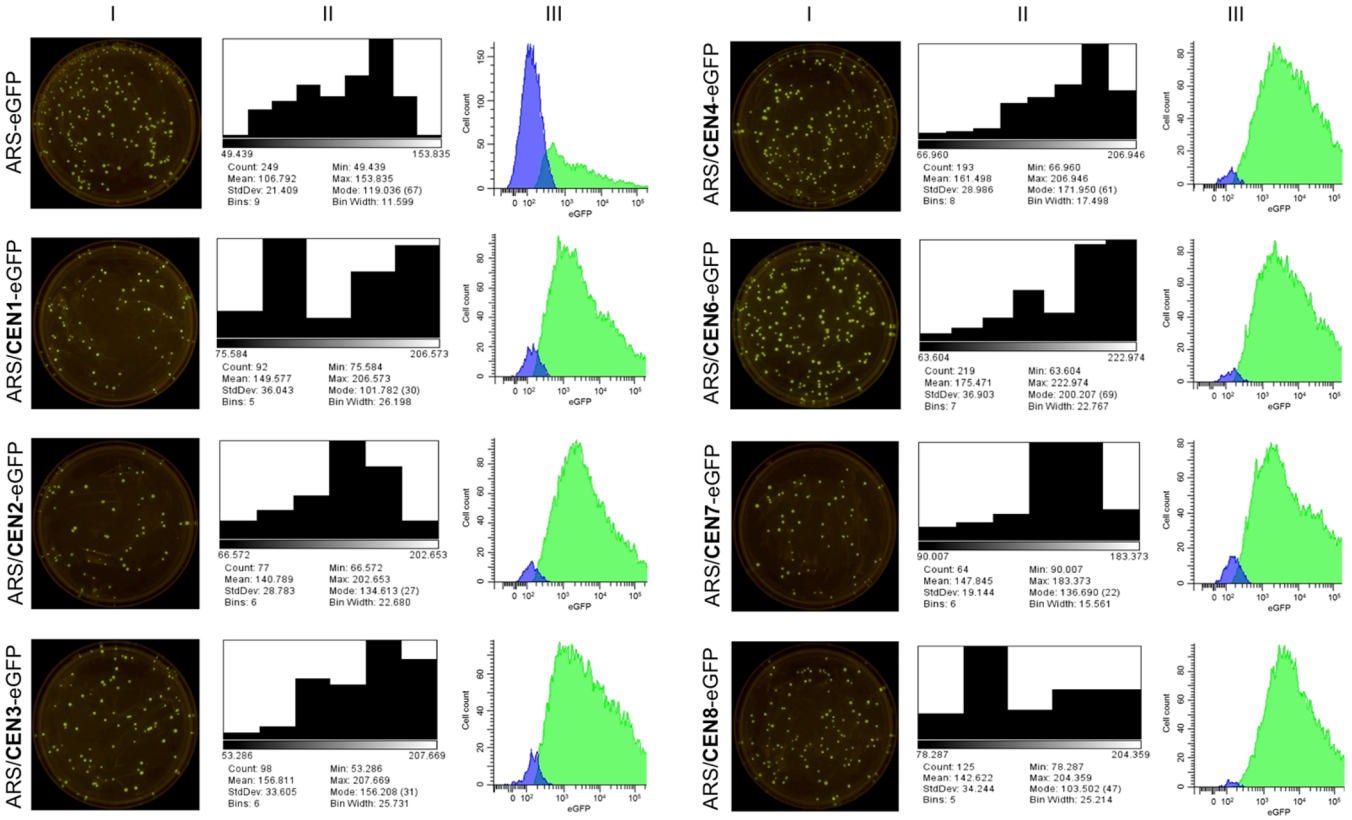


Figure 2. Analysis of the fluorescence displayed by each colony to identify clones carrying CENs. I. Imaging of the fluorescent colonies on Petri dishes. II. Processing of each picture and quantification of the fluorescence intensity of individual colonies using the software ImageJ. III. Flow cytometric analysis of the brightest clones to confirm CEN function.

low fluorescence values, were picked for further screening by flow cytometry. The results revealed that only the colonies with maximum values (~200, for CEN7:183) showed the desired symmetric fluorescence peaks, with more than 90% of the entire population being eGFP-positive, whereas the colonies picked from low-signal groups invariably showed broad peaks similar to those of cells transformed with the control ARS-eGFP vector.

The likelihood of obtaining a functional CEN based on the fluorescence intensity of each colony was estimated. For CEN1, CEN3, CEN4, CEN6, and CEN8, at least 8 of the 10 selected colonies with high fluorescence showed the desired symmetric peaks (Table 1). The associated inserts were subsequently sequenced to ascertain the CEN positions in the corresponding GC_3 valleys (Figure S2), which were designated as CEN1−3kb to CEN8−3kb (see Table S1 for sequences). The selected region for CEN2 has the longest length (38-kb), which might contribute to the low efficiency of isolating the functional CEN. For the selected region for CEN7, although we designed 3-kb segments for library creation, each of the confirmed clones that displayed desired GFP signal peaks turned out to carry a 1.6-kb segment. Apparently, the corresponding 3-kb segment was prone to misassembly and truncation, which might disrupt CEN activity and presumably led to a low CEN identification efficiency. Nevertheless, this library-based method integrated with GC_3 chromosome scanning, was highly effective for pinpointing CENs, requiring much less time than the stepwise approach to determine all functional CENs in the *S. stipitis* genome.

Although the 3-kb CEN sequences from individual chromosomes differed markedly from each other, a common core sequence of 125-bp (98.4% similarity) was found in seven out of the eight functional 3-kb regions (Figure 3). This sequence was also previously identified as the minimal length in CEN5 that rendered a symmetric eGFP expression peak.[6] The only exception was found in chromosome 8 to have one region of 139 bp sharing low similarity with the 125-bp CEN core. This result was not expected because the vast majority of the point CENs that have been identified from the same species are only arranged in the signature CDEI/II/III blocks but do not share a high sequence identity.[3] The centromeric function of these highly similar 125-bp fragments was subsequently validated by recloning them into the ARS-eGFP backbone followed by flow cytometry analysis, all of which showed symmetrically shaped eGFP fluorescence peaks. These results support that the 125-bp sequence is the minimal CEN core, which supports the classification of *S. stipitis* CENs into the "point CENs" category.

Beyond CEN sequence identification, we surprisingly observed 2−4 repeats of the 125 bp core CEN in each of the eight GC_3 valleys (Figure 4 and Table 1), leaving a very interesting question as to why most 3-kb segments containing this core sequence were not identified either in the stepwise shortening process (for CEN5)[6] or the 3-kb library screening (for the other seven CENs). To find an answer, we recloned one 3-kb segment from each chromosome that was not identified in library screening but contained the core 125-bp CEN sequence. We found that only the segment from CEN1 showed the symmetric fluorescence peak (95% eGFP-positive cells), whereas the 3-kb segments from CEN2−8 did not display well-shaped eGFP expression peaks (Figure S3), suggesting that the genome contexts where the 125-bp core

DOI: 10.1021/acssynbio.7b00166
ACS Synth. Biol. 2017, 6, 2028−2034

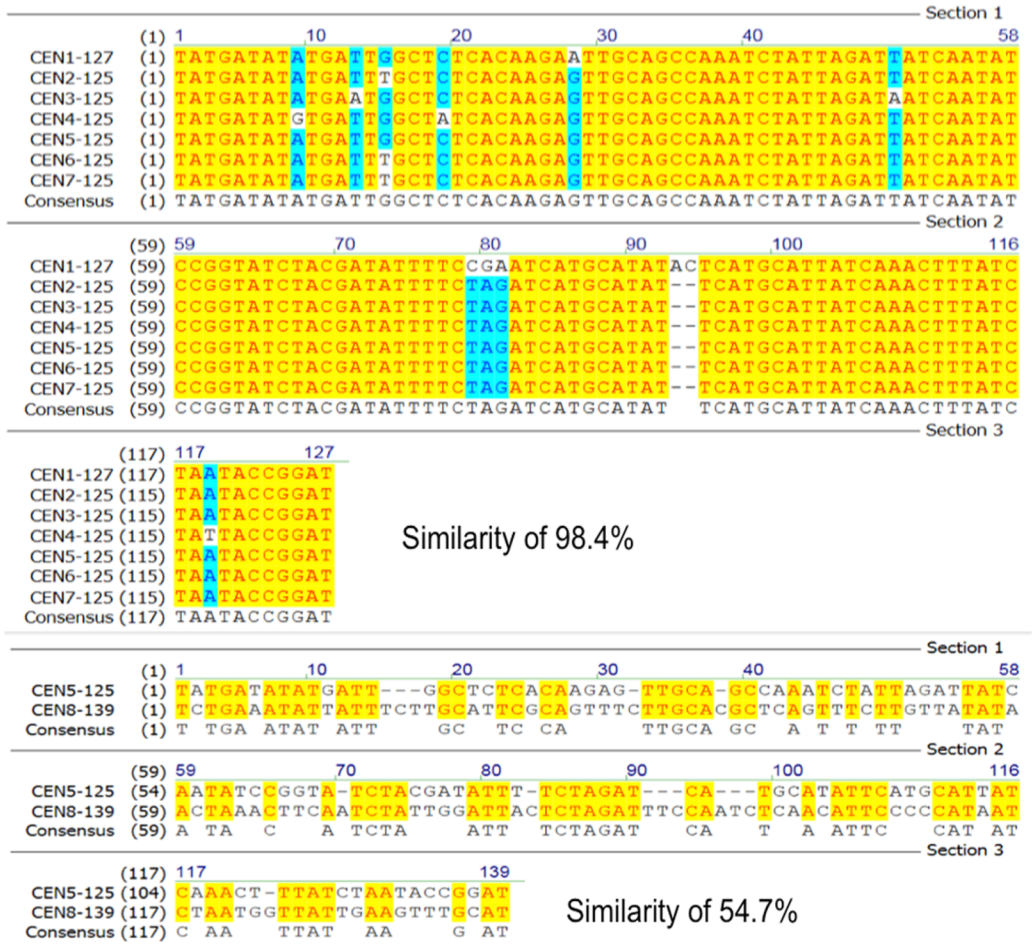


Figure 3. Alignment of the identified CENs enabled confirmation of the 125-bp sequence as the core CEN.

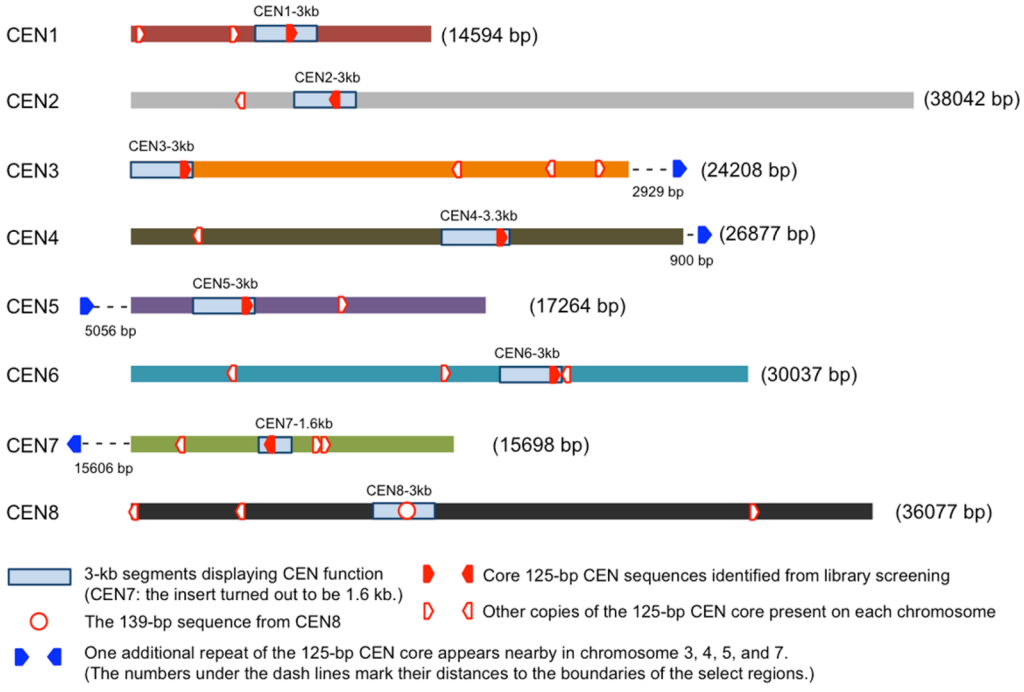


Figure 4. Locations and the repeats of the 125-bp core CEN in the corresponding predicted GC_3 valley neighborhood. The point angle of a pentagon marks that the 125-bp core CEN was found from the sense strand or the antisense strand. The numbers shown in the parentheses illustrate the lengths of the abnormally long intergenic regions presumably contain the CENs.

DOI: 10.1021/acssynbio.7b00166
ACS Synth. Biol. 2017, 6, 2028−2034

CENs sat played an essential role in defining CEN function epigenetically, and probably in the natural genome context, only one sequence served as the anchoring point for kinetochore assembly. Another very interesting observation was that the 125-bp core sequence is absolutely unique to these GC_3-valleys. Scanning through each chromosome showed that for chromosomes 1, 2, 6, and 8, no additional 125-bp sequences are present outside of the target long intergenic regions in the GC_3-valleys, and for chromosomes 3, 4, 5, and 7, only one additional copy of the 125-bp core sequence can be found from the individual chromosome, but still in the close proximity to the selected region within the GC_3-valley on each chromosome (Figure 4). This is consistent with the observation that in *Yarrowia lipolytica*, multiple CEN-like regions existed in relatively close proximity and promoted accurate chromosomal segregation when the real CEN was deleted.[9] Also, in both *S. pombe* and *C. albicans*, neocentromeres (neoCENs) were formed in close proximity with the deleted CENs because the pericentric regions could facilitate the formation of unique 3D structures that favored neoCEN formation.[10−12] Generation of a neoCEN when the original CEN undergoes a detrimental change might be the means utilized by yeasts to guarantee the integrity of their chromosomes in evolution. Collectively, these observations convey the important information regarding the aforementioned CEN epigeneticity,[8] indicating the indispensability of integrating *in silico* prediction and library screening for a rapid and accurate CEN search.

The traditional CEN isolation methods include chromatin immunoprecipitation followed by next-generation sequencing (ChIP-Seq)[2,13] and functional selection based on toxic gene lethality.[14,15] A recent method that couples chromosome conformation capture with next-generation sequencing, named Hi-C, has been applied for CEN identification based on the fact that just prior to chromosome segregation, CENs from individual chromosomes are functionally aggregated and colocalized in a higher-order architecture interacting with kinetochores.[16−18] All these methods follow a "bottom-up" strategy, in which wet experiments were performed on the entire genomes. In contrast, our method employed a unique sequence-to-function order, zooming into narrowed regions on the chromosomes and thus representing a "top-down" approach to rapidly pinpoint the minimal genetic sequences that function as stabilizing partitioning elements for extrachromosomal DNA. This method enabled the isolation of all *S. stipitis* CENs with a 16 000-fold reduction in sequence (from an average of 2 Mbp per chromosome to 125 bp) with a much higher efficiency than the bottom-up strategies.

Since the first CEN was isolated 37 years ago by Clarke and Carbon from budding yeast,[19] the puzzling, broad range of CEN complexity and diversity has defied explanation. A key question is what constitutes a fully functional CEN in the absence of defining *cis*-acting sequences. One observation based on comparing centromeric proteins from diverse eukaryotic lineages indicated that the presence of CenH3 nucleosomes in centromeric chromatin is ubiquitous among eukaryotes,[20] and therefore the positions of centromeric nucleosome formation indicate CEN locations.[21,22] Second, the point CENs belonging to a number of budding yeasts share the CDEI/II/III-blocks structure, in which the CDEI (∼8 bp) and CDEIII (∼25 bp) regions are separated by a nonconserved AT-rich CDEII region (78−86 bp).[3] In this study, we propose that a third piece of genetic commonality could occur in a subgroup of yeasts whose CEN locations might be accurately discerned by their local GC_3 percentages. We performed GC_3 chromosome scanning among 73 yeast species whose full genome sequences have been deposited in one of the five public databases (Candida Genome Database, FungiDB, National Center for Biotechnology Information, Ensembl Genome Browser, and Yeast Gene Order Browser). Very interestingly, in addition to *S. stipitis*, 29 species had a single pronounced GC_3 valley occurring on at least one of the chromosomes, among which, *Hansenula polymorpha*, *Candida lusitaniae*, and *Y. lipolytica* showed a unique GC_3 valley on each of the chromosomes (Figure S4). Supporting evidence was found in *Y. lipolytica*, for which five experimentally confirmed CENs fell in the unique GC_3 valleys.[7] Future work will target at examining the generalizability of our CEN identification strategy in a broader scope of yeasts that have attractive biotechnological potentials.

METHODS

Strains and Chemicals. *S. stipitis* FLP-UC7 (*ura*3−3, NRRL Y-21448) was gifted from Dr. Thomas W. Jeffries (University of Wisconsin & Xylome Corporation, Madison, WI).[23] *S. cerevisiae* YSG50 (MATa, *ade*2−1, *ade*3D22, *ura*3−1, *his*3−11, 15, *trp*1−1, *leu*2−3, 112, and *can*1−100) was used as the host for plasmid assembly. *E. coli* strain BW25141 (*lacI*^{q} *rrnB*$_{T14}$ Δ*lacZ*$_{WJ16}$ Δ*phoBR580 hsdR514* Δ*araBAD*$_{AH33}$ Δ*rhaBAD*$_{LD78}$ *galU95 endA*$_{BT333}$ *uidA*(Δ*MluI*)::*pir*$^+$ *recA1*; derived from *E. coli* K-12 strain BD792) was provided by Dr. William W. Metcalf (University of Illinois, Urbana, IL) and used for plasmid augmentation. All primers and genes were synthesized by Integrated DNA Technologies (Coralville, IA). Q5 High-Fidelity DNA Polymerase was purchased from NEB. FastDigest restriction enzymes were purchased from Thermo Scientific (Waltham, MA). The Wizard Genomic DNA Purification Kit was purchased from Promega (Madison, WI). The QIAprep Spin Plasmid Mini-prep Kit and QIAquick PCR Purification Kit were purchased from Qiagen (Valencia, CA). Zymolyase, Zymoprep Yeast Plasmid Miniprep II Kit, Zymoclean Gel DNA Recovery Kit, and DNA Clean & Concentrator-5 Kit were purchased from Zymo Research (Irvine, CA). Yeast nitrogen base without amino acids, yeast extract, peptone, agar, and other reagents required for cell culture were obtained from Difco (Franklin Lakes, NJ). Luria−Bertani (LB) broth, ammonium sulfate, sugars, and other chemicals were obtained from Fisher Scientific (Waltham, MA). CSM-URA and CSM-TRP amino acid mixtures were purchased from MP Biomedicals (Santa Ana, CA). YPAD medium containing 1% yeast extract, 2% peptone, and 2% dextrose supplemented with 0.01% adenine hemisulfate was used to grow *S. stipitis* FLP-UC7 and *S. cerevisiae* YSG50. Synthetic complete dropout medium lacking uracil or tryptophan (SC-URA or SC-TRP) was used to select transformants containing the assembled plasmids of interest.

GC_3 Chromosome Scanning. The whole genome sequence of *S. stipitis* was downloaded from National Center for Biotechnology Information (NCBI), along with their annotations (in fasta and GFF format, respectively). The coding sequences (CDS) were then extracted from the genome using BEDTools (v2.20.1).[24] CodonW (v1.4.4)[25] was used to calculate the GC_3 percentage for each coding sequence and a line graph was generated with a moving average of 15 genes corresponding to each chromosome. The protocol was previously developed in the work of Lynch *et al.*[7]

Plasmid Construction and Yeast Transformation. The majority of the plasmids used in this study were constructed

DOI: 10.1021/acssynbio.7b00166
ACS Synth. Biol. 2017, 6, 2028−2034

using the DNA assembler method developed previously.[26] The *S. stipitis* expression vector ARS-eGFP was constructed previously in the stepwise CEN identification method.[6] In brief, the PCR-amplified fragments with overlapping ends were cotransformed with a digested plasmid backbone into *S. cerevisiae* YSG50 for plasmid assembly *via* electroporation or lithium acetate-mediated methods. The isolated yeast plasmids were then transformed into *E. coli* BW25141 for enrichment, and their identities were verified by restriction digestion or sequencing. The correctly assembled plasmids were subsequently transformed into *S. stipitis* for target gene expression. CEN-containing sequences and key primer sequences and are summarized in Supporting Information Table S1 and S2.

Flow Cytometry Analysis. The transformed *S. stipitis* cells were cultured in SC-URA medium for ∼36−48 h and then centrifuged for 2 min at 2000*g* to remove the supernatant. The cell pellets were resuspended in 10 mM phosphate-buffered saline (pH 7.4) to an optical density at 600 nm ($OD_{600\ nm}$) between 0.1 and 0.2, and then analyzed by flow cytometry at 488 nm on a FACSCanto flow cytometer (BD Biosciences, San Jose, CA). The fluorescence-intensity distribution of each clonal population was calculated by BD FACSCanto Clinical Software.

Library-Based Screening of *S. stipitis* CENs. Primers were designed to amplify 5−13 segments of ∼3-kb, covering the longest intergenic region corresponding to each GC_3 valley. To ensure high assembly efficiency, 400-bp overlaps between adjacent fragments were maintained through carefully designing primer-annealing positions.[27,28] For each library, the gel-purified 3-kb bands were mixed at a molar ratio of 1:1, and cotransformed with the linearized ARS-eGFP backbone into *S. cerevisiae via* electroporation for library creation. Cells were spread on SC-TRP agar plates, and $>10^3$ colonies appeared after 2 days of incubation at 30 °C. All colonies were washed from the plates using 2 mL of SC-TRP liquid medium, and the plasmid library in the resulting mixture was isolated using the Zymoprep Yeast Plasmid Miniprep II Kit (Zymo Research, Irvine, CA). Next, 2 μL of each plasmid library was electroporated into *E. coli* ElectroMAX DH5α-E Competent Cells (Life Technologies, Carlsbad, CA), and the transformants were selected on Luria−Bertani (LB) agar plates supplemented with 100 μg/mL ampicillin. *E. coli* transformants were subsequently washed off from the plates by 2 mL of LB medium, and plasmids were then isolated from the pooled colonies. Finally, 3 μg of the isolated plasmid mixture was electroporated into *S. stipitis* strain FLP-UC7, which was then spread on SC-URA agar plates and incubated at 30 °C for 3−4 days.

After incubation, the colonies on each plate were examined for eGFP fluorescence intensity using a DR46B transilluminator (Clare Chemical Research, Dolores, CO). The images were analyzed using ImageJ software (http://imagej.nih.gov/ij/). Ten colonies with high fluorescence and five colonies with low fluorescence were propagated individually in SC-URA medium for further flow cytometry analysis. Only the cultures giving a symmetric eGFP peak with the eGFP-positive population exceeding 80% were considered to carry a functional CEN, which was then analyzed by sequencing.

CEN Alignment. The 125-bp sequence identified from CEN5 was used to search against the other seven 3-kb regions identified through library screening, in both sense and antisense directions using the align function of software Vector NTI 11 (Life Technologies, Carlsbad, CA). The regions showing high sequence similarity with CEN5−125bp were assembled into the ARS-eGFP backbone to verify their plasmid-stabilizing function.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.7b00166.

Tables S1−S2: The lists of the eight identified *S. stipitis* CEN sequences and the key primer sequences. Figures S1−S4: The eGFP expression profiles when being cloned in the ARS-only and ARS-CEN plasmids, the GC_3 chromosome scanning profile of *S. stipitis*, the flow cytometry analysis of the additional clones carrying 3-kb segments, and the GC_3 chromosome scanning of the additional yeast species with full genome sequences deposited in public databases (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: 515-294-1132. E-mail: zyshao@iastate.edu.

ORCID

Zengyi Shao: 0000-0001-6817-8006

Author Contributions

M.C. and Z.S. conceived the idea and wrote the manuscript. M.C. performed the experiments to identify and analyze *S. stipitis* CENs. A.S.S. and A.J.S. wrote the code for GC_3 chromosome scanning.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We would like to thank Dr. Thomas W. Jeffries for providing the strain *S. stipitis* FLP-UC7 (*ura*3-3, NRRL Y-21448). We also thank Dr. Shawn M. Rigby for performing flow cytometry analysis, and Dr. Fuyuan Jing for ImageJ analysis. This work was supported in part by the Iowa Energy Center (4782024) and the National Science Foundation Grants EEC-0813570, MCB 1716837, and EPS-1101284.

■ REFERENCES

(1) Furuyama, S., and Biggins, S. (2007) Centromere identity is specified by a single centromeric nucleosome in budding yeast. *Proc. Natl. Acad. Sci. U. S. A. 104*, 14706−14711.

(2) Padmanabhan, S., Thakur, J., Siddharthan, R., and Sanyal, K. (2008) Rapid evolution of Cse4p-rich centromeric DNA sequences in closely related pathogenic yeasts, Candida albicans and Candida dubliniensis. *Proc. Natl. Acad. Sci. U. S. A. 105*, 19797−19802.

(3) Meraldi, P., McAinsh, A. D., Rheinbay, E., and Sorger, P. K. (2006) Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. *Genome Biology 7*, R23−R23.

(4) Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., de Montigny, J., Marck, C., Neuveglise, C., Talla, E., Goffard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.-M., Beyne, E., Bleykasten, C., Boisrame, A., Boyer, J., Cattolico, L., Confanioleri, F., de Daruvar, A., Despons, L., Fabre, E., Fairhead, C., Ferry-Dumazet, H., Groppi, A., Hantraye, F., Hennequin, C., Jauniaux, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R., Lemaire, M., Lesur, I., Ma, L., Muller, H., Nicaud, J.-M., Nikolski, M., Oztas, S., Ozier-Kalogeropoulos, O., Pellenz, S., Potier, S., Richard, G.-F., Straub, M.-L., Suleau, A., Swennen, D., Tekaia, F., Wesolowski-Louvel, M., Westhof, E., Wirth, B., Zeniou-Meyer, M., Zivanovic, I., Bolotin-Fukuhara, M., Thierry, A.,

DOI: 10.1021/acssynbio.7b00166
ACS Synth. Biol. 2017, 6, 2028−2034

Bouchier, C., Caudron, B., Scarpelli, C., Gaillardin, C., Weissenbach, J., Wincker, P., and Souciet, J.-L. (2004) Genome evolution in yeasts. *Nature 430*, 35−44.
(5) Jeffries, T. W., Grigoriev, I. V., Grimwood, J., Laplaza, J. M., Aerts, A., Salamov, A., Schmutz, J., Lindquist, E., Dehal, P., Shapiro, H., Jin, Y. S., Passoth, V., and Richardson, P. M. (2007) Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*. *Nat. Biotechnol. 25*, 319−326.
(6) Cao, M., Gao, M., Lopez-Garcia, C. L., Wu, Y., Seetharam, A. S., Severin, A. J., and Shao, Z. (2017) Centromeric DNA facilitates nonconventional yeast genetic engineering. *ACS Synth. Biol. 6*, 1545.
(7) Lynch, D. B., Logue, M. E., Butler, G., and Wolfe, K. H. (2010) Chromosomal G + C content evolution in yeasts: systematic interspecies differences, and GC-poor troughs at centromeres. *Genome Biol. Evol. 2*, 572−583.
(8) McKinley, K. L., and Cheeseman, I. M. (2016) The molecular basis for centromere identity and function. *Nat. Rev. Mol. Cell Biol. 17*, 16−29.
(9) Lefrancois, P., Auerbach, R. K., Yellman, C. M., Roeder, G. S., and Snyder, M. (2013) Centromere-like regions in the budding yeast genome. *PLoS Genet. 9*, e1003209.
(10) Ketel, C., Wang, H. S., McClellan, M., Bouchonville, K., Selmecki, A., Lahav, T., Gerami-Nejad, M., and Berman, J. (2009) Neocentromeres form efficiently at multiple possible loci in *Candida albicans*. *PLoS Genet. 5*, e1000400.
(11) Cleveland, D. W., Mao, Y., and Sullivan, K. F. (2003) Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell 112*, 407−421.
(12) Ishii, K. (2009) Conservation and divergence of centromere specification in yeast. *Curr. Opin. Microbiol. 12*, 616−622.
(13) Sanyal, K., Baum, M., and Carbon, J. (2004) Centromeric DNA sequences in the pathogenic yeast *Candida albicans* are all different and unique. *Proc. Natl. Acad. Sci. U. S. A. 101*, 11374−11379.
(14) Hieter, P., Pridmore, D., Hegemann, J. H., Thomas, M., Davis, R. W., and Philippsen, P. (1985) Functional selection and analysis of yeast centromeric DNA. *Cell 42*, 913−921.
(15) Hieter, P., Mann, C., Snyder, M., and Davis, R. W. (1985) Mitotic stability of yeast chromosomes - a colony color assay that measures nondisjunction and chromosome loss. *Cell 40*, 381−392.
(16) Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y. J., Lee, C., Shendure, J., Fields, S., Blau, C. A., and Noble, W. S. (2010) A three-dimensional model of the yeast genome. *Nature 465*, 363−367.
(17) Marie-Nelly, H., Marbouty, M., Cournac, A., Liti, G., Fischer, G., Zimmer, C., and Koszul, R. (2014) Filling annotation gaps in yeast genomes using genome-wide contact maps. *Bioinformatics 30*, 2105−2113.
(18) Varoquaux, N., Liachko, I., Ay, F., Burton, J. N., Shendure, J., Dunham, M. J., Vert, J. P., and Noble, W. S. (2015) Accurate identification of centromere locations in yeast genomes using Hi-C. *Nucleic Acids Res. 43*, 5331−5339.
(19) Clarke, L., and Carbon, J. (1980) Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature 287*, 504−509.
(20) Malik, H. S., and Henikoff, S. (2009) Major evolutionary transitions in centromere complexity. *Cell 138*, 1067−1082.
(21) Cui, F., Chen, L., LoVerso, P. R., and Zhurkin, V. B. (2014) Prediction of nucleosome rotational positioning in yeast and human genomes based on sequence-dependent DNA anisotropy. *BMC Bioinf. 15*, 313.
(22) Cole, H. A., Howard, B. H., and Clark, D. J. (2011) The centromeric nucleosome of budding yeast is perfectly positioned and covers the entire centromere. *Proc. Natl. Acad. Sci. U. S. A. 108*, 12687−12692.
(23) Lu, P., Davis, B. P., Hendrick, J., and Jeffries, T. W. (1998) Cloning and disruption of the beta-isopropylmalate dehydrogenase gene (LEU2) of *Pichia stipitis* with URA3 and recovery of the double auxotroph. *Appl. Microbiol. Biotechnol. 49*, 141−146.
(24) Quinlan, A. R. (2014) BEDTools: the Swiss-Army tool for genome feature analysis. *Curr. Protoc. Bioinformatics 47*, 11.12.11−11.12.34.
(25) Peden, J. (2014) CodonW: Correspondence Analysis of Codon Usage, http://codonw.sourceforge.net/ (accessed 2017).
(26) Shao, Z., and Zhao, H. (2013) Construction and engineering of large biochemical pathways *via* DNA assembler. *Methods Mol. Biol. 1073*, 85−106.
(27) Shao, Z., and Zhao, H. (2012) Exploring DNA assembler: a synthetic biology tool for characterizing and engineering natural product gene clusters. *Methods Enzymol. 517*, 203−224.
(28) Shao, Z., and Zhao, H. (2014) Manipulating natural product biosynthetic pathways *via* DNA assembler. *Curr. Protoc Chem. Biol. 6*, 65−100.

DOI: 10.1021/acssynbio.7b00166
ACS Synth. Biol. 2017, 6, 2028−2034