Supporting Information

Møller et al. 10.1073/pnas.1508825112

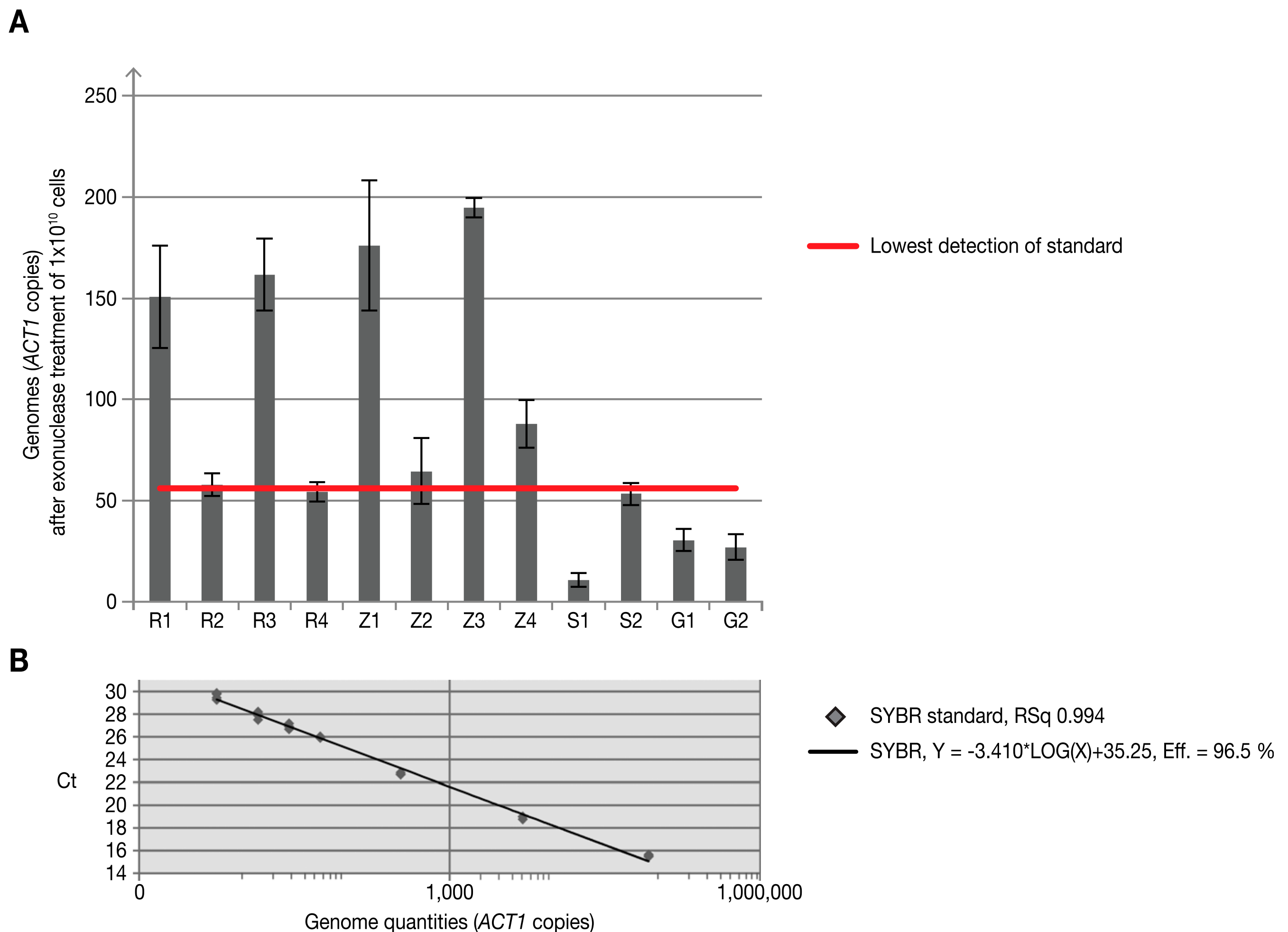


Fig. S1. Elimination of linear DNA in exonuclease-treated DNA. (*A*) Quantitative PCR for *ACT1* on chromosome VI:53,260.0.54,696 as a marker of the number of genomes in DNA after 130 h of exonuclease treatment. The mean average of linear DNA fragment was estimated to be 210 kb after NotI treatment, and no NotI restriction site is present on chromosome VI. *ACT1* is 215 kb from the farthest telomere, and the signal from *ACT1* was reduced between 5×10^7- and 4×10^8-fold. Shown are quadruplicate reference samples (R1, R2, R3, R4) from mixed populations of ~4,000 congenic single-gene deletion mutants from the haploid S288C *S. cerevisiae* reference genome; an identical set of quadruplicate reference samples grown with 0.03 mM Zeocin (Z1, Z2, Z3, Z4); and duplicate samples of clonal isogenic haploid S288C populations (S1, S2). G1 and G2, two populations that contain [$GAP1^{circles}$] (1). Bars, SDs of triplicate measurements; red line, level of *ACT1* in the most diluted control sample (genomic DNA). (*B*) Standard curves were made using genomic DNA from a haploid yeast strain as reference and serial diluted to very low quantities, indicated by the number of *ACT1* copies. Ct (cycle threshold), number of cycles required to cross threshold background level. RSq, R-square value of the fitted standard curve; Eff., reaction efficiency. Quantitative PCR was confirmed by melting curve analysis showing a single peak of specific primer binding (data not shown) and detection of 100 bp DNA in all samples by agarose gel electrophoresis, except for negative controls (data not shown).

1. Gresham D, et al. (2010) Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal element formation of the *GAP1* locus. *Proc Natl Acad Sci USA* 107(43): 18551–18556.

Fig. S2. Distribution of contiguous mapped reads. (*A*) All samples. *x* axis, region length; *y* axis, density of contiguous regions by length (black) and density of mapped reads to specific contiguous regions (gray). MtDNA, 86-kb circular mitochondrial DNA; 2μ, 6.3-kb endogenous plasmid covered by contiguous mapped reads. Dashed line, cutoff for eccDNA detection was more than seven contiguous mapped reads corresponding to fragments >1 kb. (*B*) Distributions and features as in *A*, for individual samples. Blue, R1–R4; yellow, Z1–Z4; green, S1 and S2; purple, G1 and G2 (described in legend to Fig. S1).

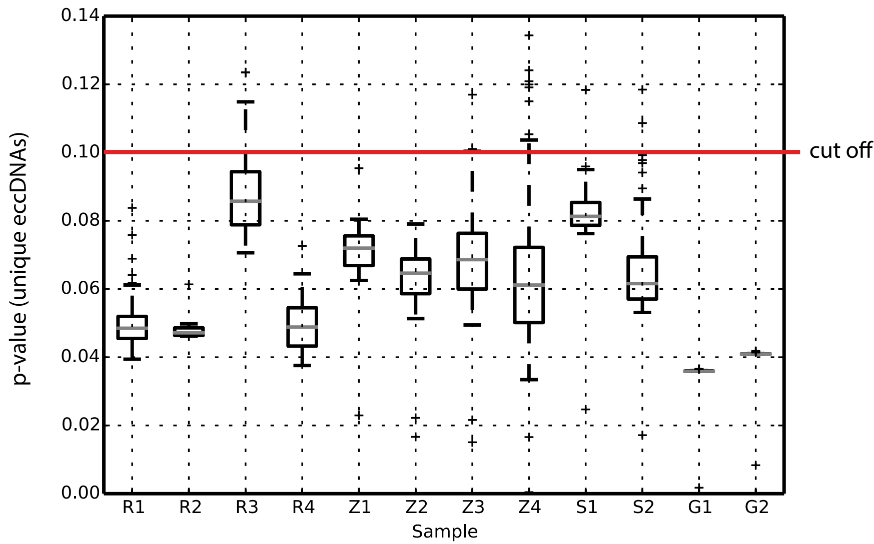


Fig. S3. Significance of contiguous reads >1 kb. To test for the likelihood that contiguous reads >1 kb were detected by chance, *P* values were generated for each recorded region with contiguous read >1 kb with at least one uniquely mapped read. Box plot, *P* value distribution after 10,000 Monte Carlo simulations for 12 samples, R1–R4, Z1–Z4, S1 and S2, and G1 and G2 (described in legend to Fig. S1). Conservative cutoff focused on only contiguous reads >1 kb (Fig. S2). All regions with contiguous reads >1 kb and *P* values <0.1 were considered positive, excluding 44 contiguous reads >1 kb with *P* values >0.1.

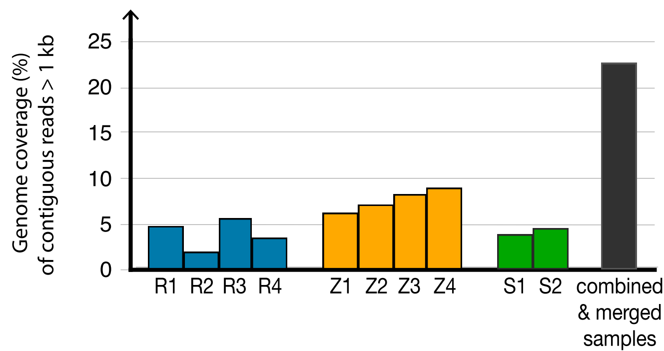


Fig. S4. Genome coverage. Percentage of the genome covered by contiguous mapped reads >1 kb DNA after Circle-Seq of 10 populations. Blue, R1–R4; yellow, Z1–Z4; green, S1 and S2 (described in the legend to Fig. S1). Black, genome coverage of samples after merging overlapping contiguous mapped reads >1 kb across all samples.

A)

B)

C)

Fig. S5. EccDNA detection limit. (*A*) Quantification of added plasmids in the 10 populations: R, Z, and S (as described in the legend to Fig. S1) relative to other known circular DNA elements. Spiked plasmids were added after cell lysis but before column purification at 1:1 per cell (pBR322), 1:50 per cell (pUC19), and 1:2,500 per cell (pUG72). By normalization to uniquely mapped reads on pBR322 (1 plasmid:1 cell), the ratio per cell of the other two plasmids was as follows: 1:25 per cell for pUC19 and 1:16,000 for pUG72. (*B*) Confirmation of plasmids by inverse PCR and agarose gel electrophoresis. Samples as in Fig. S1 legend; R1–R4, Z1–Z4, S1–S4 (S3 and S4 were not analyzed further), G1 and G2. B (non template control); (−) negative and (+) positive plasmid template; L, 1-kb ladders. (*C*) Limit for eccDNA detection by Circle-Seq on double-logarithmic plot using three plasmids (pBR322, pUC19, and pUG72) as the internal standard for the number of cells per plasmid. *x* axis, number of cells per plasmid by spike-in ratio; *y* axis, number of mapped reads to plasmids and eccDNAs normalized to fragments per kilobase from a million mapped reads (FPKM units) using only uniquely mapped reads. Shown are plasmid distribution for R1–R4, Z1–Z4, and S1 and S2 samples. Red line, lowest detection of unique eccDNA.

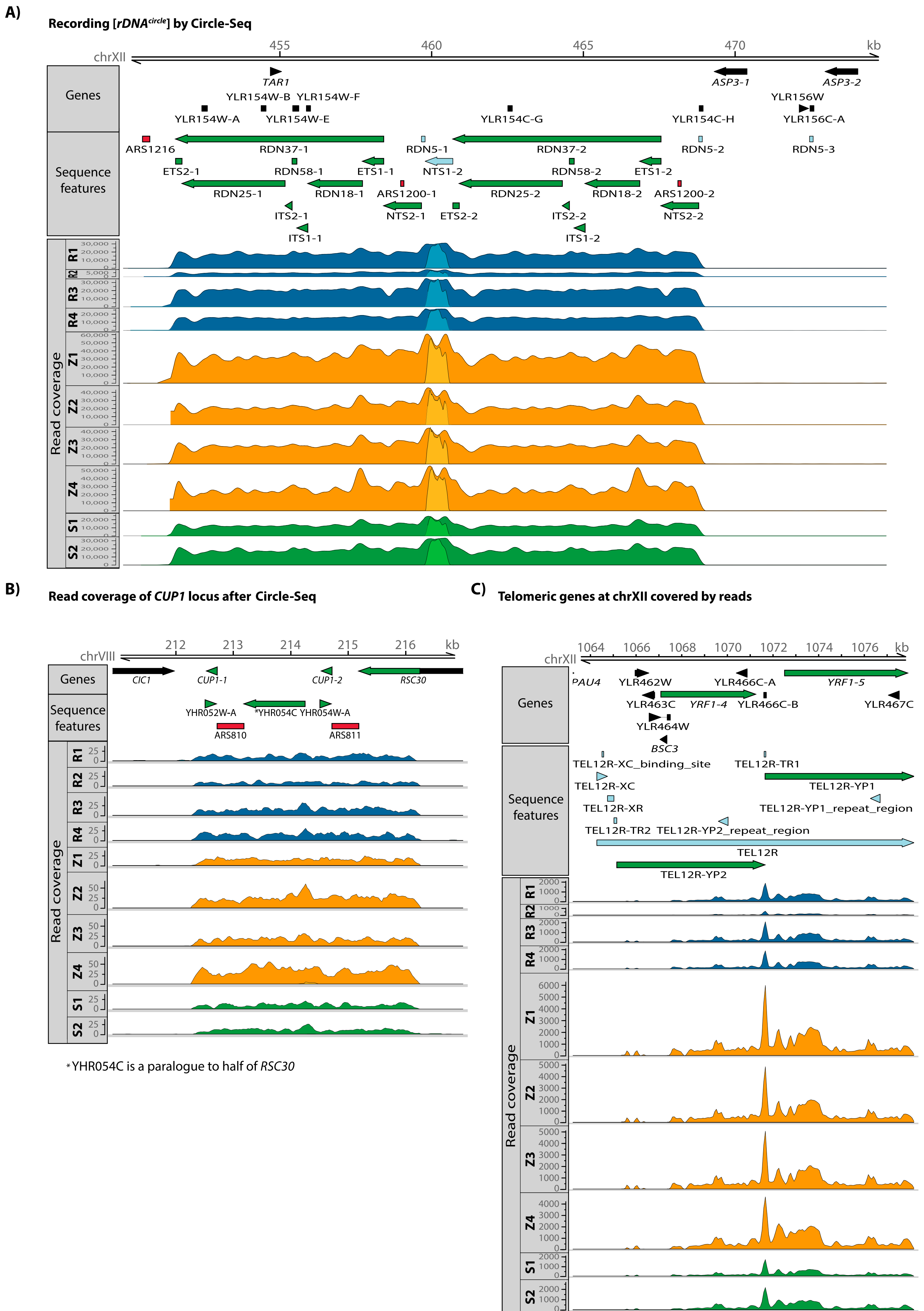


Fig. S6. Identification of [*rDNA*^{circle}], [*TEL*^{circle}] and [*CUP1*^{circle}] by Circle-Seq. Read coverage of A) ribosomal genes on chromosome XII, light-color of blue, yellow and green indicate uniquely mapped reads, B) metallothionein *CUP1-1* and *CUP1-2* genes on chromosome VIII, C) telomeric genes on chromosome XII, C). Read coverage represents mapped reads from 10 populations (Dataset S1). R, reference of mixed populations of single-gene deletion mutants; Z as R but grown with mutagenic Zeocin; S, standard WT populations.

Fig. S7. Detection of [*GAP1*^{circle}] by Circle-Seq. (*A*) Read coverage across the *GAP1* locus on chromosome XI of 12 populations: R1–R4, Z1–Z4, S1 and S2, and G1 and G2. G1 and G2 were previously described to contain [*GAP1*^{circle}]. Light purple, uniquely mapped reads; dark purple, reads mapped to several loci. (*B*) Inverse PCR detection of [*GAP1*^{circles}] as a 0.6-kb DNA band (arrow) by gel electrophoresis. Samples after 130-h exonuclease treatment, both pre- or post-ϕ29 amplification, were used as templates for PCR. Positive control was genomic DNA from a [*GAP1*^{circle}] mutant (1). GD, control genomic DNA from S288C. Blue arrows, primers.

1. Gresham D, et al. (2010) Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal element formation of the *GAP1* locus. *Proc Natl Acad Sci USA* 107(43): 18551–18556.

Mapping of *ENA5/2/1* genes after Circle-Seq

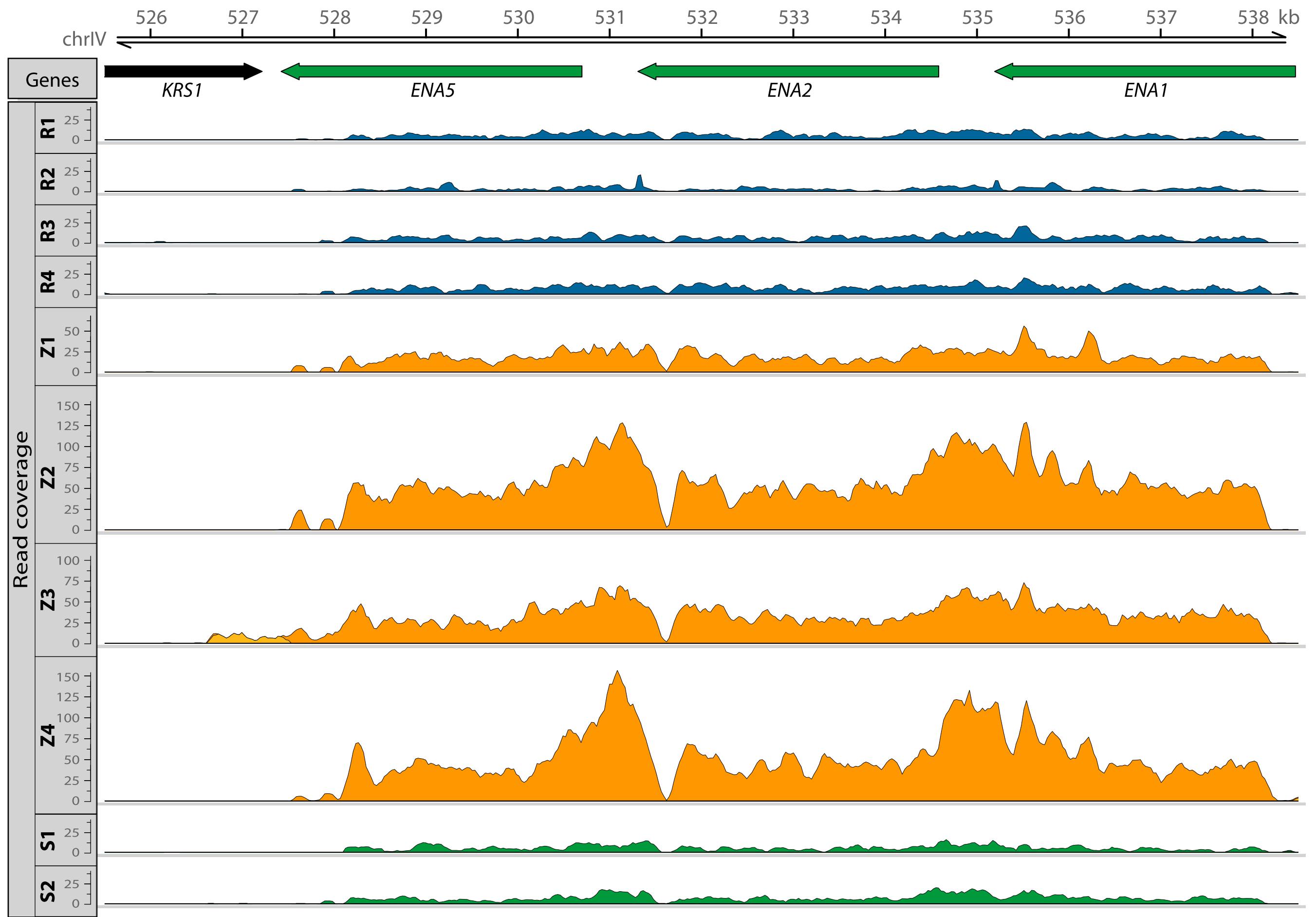
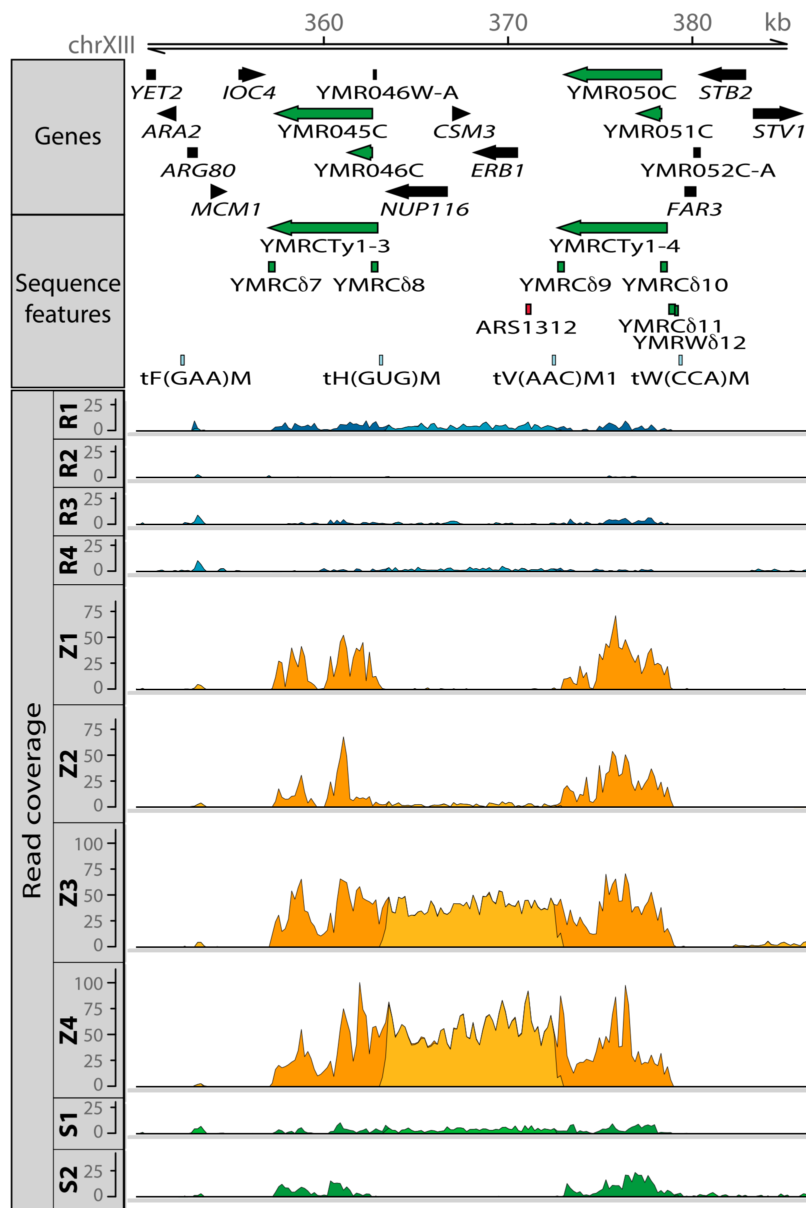


Fig. S8. Recording [*ENA1/2/5*^{circles}] by Circle-Seq. Read coverage of the P-type ATPase sodium pump genes on chromosome IV from 10 populations. R, reference of mixed populations of single-gene deletion mutants; S, standard nonmutant populations of S288C; Z, as R, but grown with mutagenic Zeocin.

A)



B)

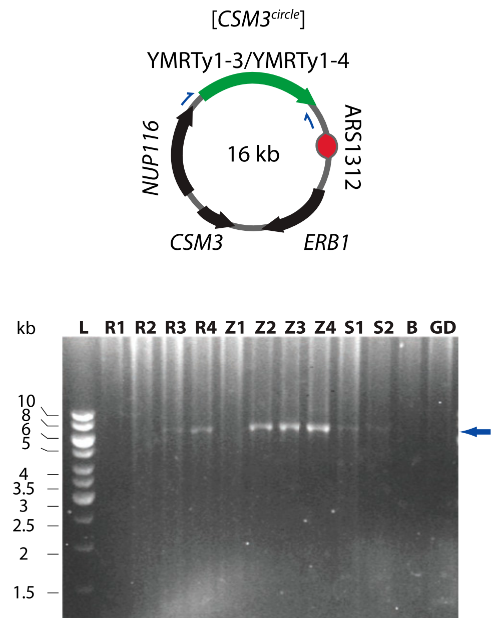


Fig. S9. [*CSM3*^{circle}] detected by Circle-Seq and inverse PCR. (*A*) Read coverage across the *CSM3* locus on chromosome XIII of 10 populations. (*B*) Inverse PCR for [*CSM3*^{circle}]. Blue arrows, specific inverse PCR primers. Negative control was no DNA template (B) and genomic DNA (GD) from S288C. R, reference of mixed populations of single-gene deletion mutants; S, standard nonmutant populations of S288C; Z, as R, but grown with mutagenic Zeocin.

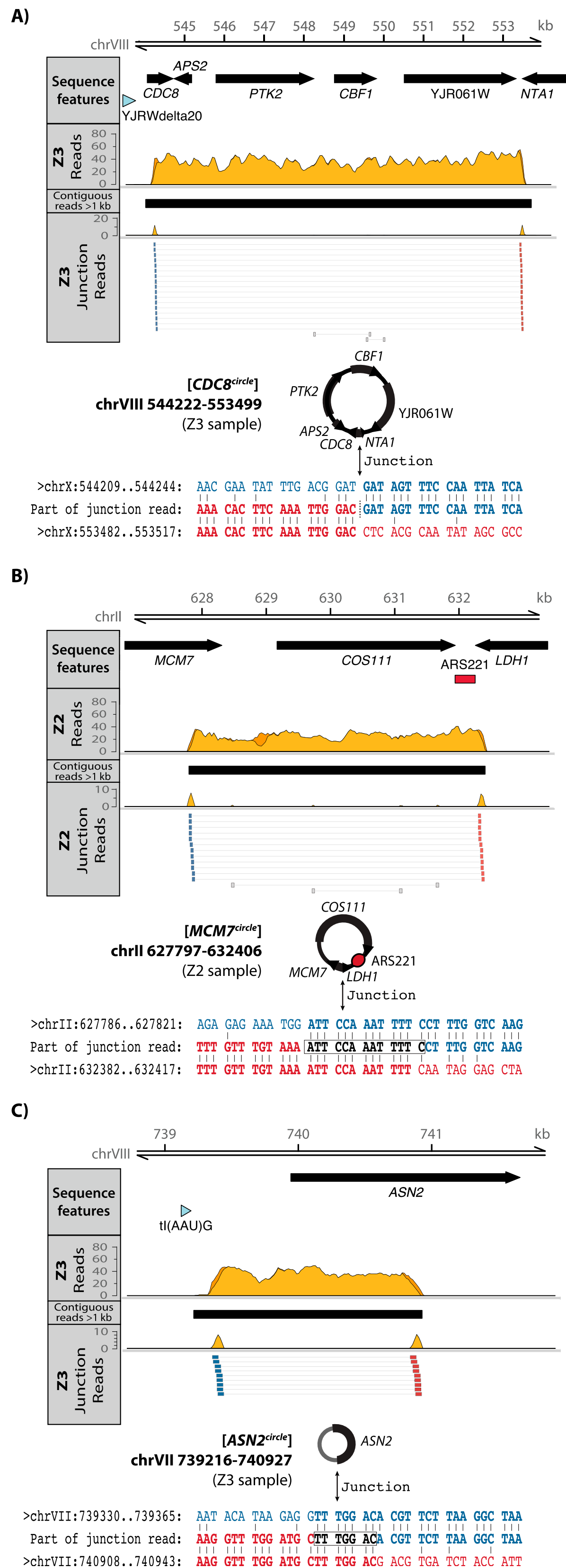


Fig. S10. Junction reads at [*CDC8*^{circle}], [*MCM7*^{circle}], and [*ASN2*^{circle}]. (*A*–*C*) Read coverage and mapping of split reads at three specific eccDNAs. Minimal five junction reads supported the junction marked with a dashed line or a square box. (*A*) The [*APS2_CBF1_PTK2*_YJR061W_*CDC8_NTA1*^{circle}], in short [*CDC8*^{circle}], and (*B*) the [*MCM7*^{circle}], both detected in the Z2 sample. (*C*) The [*ASN2*^{circle}] detected in the Z3 sample. Z, mixed populations of single-gene deletion mutants grown with mutagenic Zeocin.

Dataset S1. List of potential DNA circularization regions

Dataset S1

Dataset S1 represents sequence data and analyses of 1,862 regions, divided into 38 columns: EccDNA mapping (A–D); A (first column from left), sample where putative eccDNA is identified; B, chromosome; C and D, start and end coordinates of putative eccDNAs. EccDNA content (E–K): E, genes within the region; F, part of genes included in the region; G, number of proposed replication origins and/or autonomous replication sequences (ARS) based on refs. 1–4; H, number of ARS consensus sequences (ACS). Sequences flanking the potential eccDNA junction (I–K): I, long terminal repeats (LTRs), 100-bp windows within potential junction; J and K, features at start and end coordinates, 100-bp windows within potential junction. EccDNA coverage and *P* values (L–R): L, number of all mapped reads; M, number of uniquely mapped reads; N, coverage of all mapped reads by fragments per kilobase from a million mapped reads (FPKM); O, *P* value of putative eccDNA compared with occurrence by chance, Monte Carlo simulations; P, longest region with a uniquely annotated sequence; Q and R; as N and O, using only uniquely mapped reads (UFPKM). BLAST+ analyses (S–AG): S, queryId denotes the query sequence coordinates (first half part of eccDNA region); T, subjectId denotes the subject sequence coordinates (second half part of eccDNA region); U, base pair alignment length of queryId versus subjectId, reporting only top hit from BLAST+; V, percent homology of alignment; W, number of mismatches; X, number of open gaps; Y, alignment position from query start (first nucleotide set to 1); Z, alignment position from query end; AA and AB, as Y and Z, but from subject start and end; AC, e-value of region with sequence homology; AD, alignment quality normalized to the query length (bitscore) of which higher numbers correspond to higher similarity. Based on BLAST+ data in column S to AD; the presence or absence of homology sequences at coordinate flanks (500-bp windows around putative eccDNA junctions) is sorted in AE–AG, according to homology alignment length: i.e., AE, homology <15 nucleotides; AF, homology between 15 and 50 nucleotides; AG, homology >50 nucleotides. Discordant reads mapping and analyses at putative eccDNA junctions (AH–AL): AH, number of discordant split reads mapping to both sides of the putative junction; AI, chromosomal coordinates of merged regions, covered by minimal two discordant split reads with split read counts in brackets; AJ–AL, homologies at putative eccDNA junctions, separated after alignment length: AJ, <7 nucleotides; AK, between 7 and 14 nucleotides; AL, ≥15 nucleotides. Parameters for Monte Carlo simulation and BLAST are described in *Materials and Methods*.

1. Breier AM, Chatterji S, Cozzarelli NR (2004) Prediction of *Saccharomyces cerevisiae* replication origins. *Genome Biol* 5(4):R22.
2. Raghuraman MK, et al. (2001) Replication dynamics of the yeast genome. *Science* 294(5540):115–121.
3. Wyrick JJ, et al. (2001) Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: High-resolution mapping of replication origins. *Science* 294(5550):2357–2360.
4. Nieduszynski CA, Knox Y, Donaldson AD (2006) Genome-wide identification of replication origins in yeast by comparative genomics. *Genes Dev* 20(14):1874–1879.

Dataset S2. Common genes on eccDNA

Dataset S2

Dataset S2 includes numbers of copies of eccDNA genes identified in all replicate samples. All samples, R-Z-S; replicates, R1–R4, Z1–Z4, and S1 and S2. Complete and partial genes were counted. R1–R4, reference of mixed populations of single-gene deletion mutants; Z1–Z4, as R, but grown with mutagenic Zeocin; S1 and S2, standard nonmutant populations. Gene names can be found in Dataset S1.

Dataset S3. Small repeats at flanks of putative eccDNA regions

Dataset S3

Dataset S3 includes data for 10 samples, R-Z-S; replicates, R1–R4, Z1–Z4, and S1 and S2. In column A, sample name; B, chromosome; C and D, start and end coordinates of putative eccDNA. EccDNA content (E and F): E, genes within the region; F, part of genes included in the region; G, coordinates and base pair length of direct repeats (DRs) at sequences flanking the potential eccDNA junction, 200-bp windows; H, counts of DRs; I, *P* value estimation of DRs (Monte Carlo simulation), J–L, as G, H, and I, just for inverted repeats (IRs). Repeat analyses at coordinates mapped by split reads (M–Q): M, counts of split reads; N, coordinates and base pair length of DRs at split read regions; O, counts of DRs in split reads; P and Q, as N and O, but for IRs. Repeat analyses at merged split read regions (R–V) mapped by minimum two reads: R, coordinates of merged split read regions with read counts in brackets; S and T and U and V, similar to N and O and P and Q, respectively, but repeat analysis on merged split read regions. For more info, see legend of Dataset S1. Parameters for Monte Carlo simulation and BLAST are described in *Materials and Methods*.