Science
AAAS

Supplementary Materials for

Efficient Formation of Single-copy Human Artificial Chromosomes

Craig W. Gambogi, Gabriel J. Birchak, Elie Mer, David M. Brown, George Yankson, Kathryn Kixmoeller, Janardan N. Gavade, Josh L. Espinoza, Prakriti Kashyap, Chris L. Dupont, Glennis A. Logsdon, Patrick Heun, John I. Glass, and Ben E. Black
Corresponding author: blackbe@pennmedicine.upenn.edu

The PDF file includes:

Materials and Methods

Cell lines

The *H. sapiens* fibrosarcoma-derived HT1080$^{\text{Dox-inducible mCherry-LacI-HJURP}}$ cell line (*16*) was cultured in DMEM with 4.5 g/L D-Glucose and L-Glutamine (Gibco) supplemented with 10% FBS, 100 U/mL penicillin, and 100 µg/ml streptomycin at 37 ºC in a humidified incubator with 5% CO_2.

The *H. sapiens* osteosarcoma-derived U2OS $^{\text{Dox-inducible mCherry-LacI-HJURP}}$ cell line was generated via recombinase-mediated cassette exchange (RMCE) using the U2OS RMCE system (*47*). This system allowed us to insert the dox-inducible mCherry-LacI-HJURP transgene cassette at a single genomic locus. Briefly, a monoclonal U2OS acceptor cell line with loxP and lox2272 recombination sites at a single chromosomal locus was co-transfected with a donor plasmid containing an mCherry-LacI-HJURP gene under the control of a tetracycline-responsive element (TRE) (all flanked by loxP and lox2272 sites) and a second plasmid expressing Cre recombinase (pEM784 (*47*)). The cells were co-transfected at a 100:1 ratio (990 ng mCherry-LacI-HJURP donor plasmid and 10 ng Cre recombinase plasmid) using FuGENE 6 (Promega). 2 days after transfection, 2 µg/mL puromycin was added to the growth medium for selection of the stable cell line. 2 µg/mL doxycycline was added to the growth medium for 48 h to induce expression of mCherry-LacI-HJURP from the TRE promoter.

YAC construction

A total of 6 fragments were prepared for TAR cloning to make a HAC forming YAC construct (Fig. 1A). Two linker fragments were ordered from Integrated DNA Technologies (IDT). The remaining two were PCR amplified from 4q21 BAC$^{\text{LacO}}$ (note, the LacO array is the classic 10 kb fragment containing 256 tandem repeats of LacO (*48*)) and a vector containing the orotidine 5'-phosphate decarboxylase gene (URA3) gene, which enables yeast to grow on media lacking uracil, and mCherry under a CMV promoter. The 4q21 BAC$^{\text{LacO}}$ was cut into two fragments with Mre1 and Nru1 prior to insertion into yeast.

S. cerevisiae (strain VL6-48N) cells containing a YAC with *M. mycoides* synthetic genomic DNA (JCVI-syn3B) were transformed with a plasmid, pDB18-cas9-CRISPR, containing an expression cassette for guide RNAs to cut the yeast construct and Cas9. Yeast were grown in 30 ml SD-HIS overnight at 30 ºC. Yeast cells were centrifuged at 1800 × g for 3 min. Cells were resuspended to an OD_{600} of 0.4 in yeast extract-peptone-glycerol (YPG) medium. Yeast cells were grown for 6 h to an OD_{600} of 1. Then, 1.5 ml of the culture was centrifuged at 20817 × g for 15 min with an Eppendorf 5425 desktop centrifuge. The supernatant was removed and resuspended in 1 ml 0.1 M lithium acetate (LiOAc). Cells were centrifuged again at 20817 × g for 15 min before resuspending in 1 ml 0.1 M LiOAc. Cells were incubated at 30 ºC for 30 min. Cells were centrifuged at 2655 × g for 3 min. Cells were resuspended in 50 µl 0.1 M LiOAc in 1x 10 mM Tris-HCl (pH 8.0), 1 mM EDTA (TE), and add 5 µl denaturated carrier DNA (10 µg/ml sheared salmon sperm DNA; ThermoFisher) and 10 µl DNA insert mix. With mixing between each addition, 500 µl of 40% PEG 4000 (BioUltra, Sigma-Aldrich) and 56 µl of DMSO were added to the solution. The solution was incubated for 30 min at 30 ºC and then 25 min at 42 ºC. The solution was spun at 2655 × g for 3 min before resuspending in 100 µl dH_2O and plating on SD-HIS-URA plates.

Junction PCR to assess YAC

Genomic DNA was prepared as described (*49*) by first growing yeast overnight in a 5 ml culture to an OD_{600} > 0.4. Then, 200 µl of yeast was centrifuged at 2000 × g for 3 min at 4 °C. Yeast was resuspended in 100 µl 200 mM LiOAc with 1% SDS and incubated for 5 min at 70 °C. Yeast DNA was first extracted with phenol-chloroform-isoamyl alcohol (25:24:1, ThermoFisher) before precipitating with isopropanol and resuspending in elution buffer (EB; Qiagen). YAC containing yeast were assessed by PCR using Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs: M0494S). The primers sets used were: (1: Fwd: 5'-GTACCACCGCAACTTTCTTG-3', Rev: 5'- CGGCGCAGTTTCTGAGAAG-3', 2: Fwd: 5'-TATTGGTGAACCAGTGGG-3', Rev: 5'-CCTTGTTCAACACGTAATACTG-3', 3: Fwd: 5'-CCGTAATATCCAGCTGAACG-3', Rev: 5'- CAGCCAAGATATCAGCATCA-3'). Thirty-five colonies were assessed and nine were clearly positive for all three PCR reactions (fig. S1, B and C).

Sequencing to assess the YAC

For short-read sequencing, 3 µg yeast genomic DNA was digested with NEBNext dsDNA Fragmentase (New England Biolabs: M0348S) in Fragmentase Reaction Buffer v2 at 30 °C for 30 min before stopping the reaction with 5 µl .5 M EDTA. Sequencing libraries were generated and barcoded for multiplexing according to Illumina recommendations with minor modifications. Briefly, 10 ng DNA was end-repaired and A-tailed. Illumina TruSeq adaptors were ligated, libraries were size-selected to exclude polynucleosomes, and the libraries were PCR-amplified using KAPA DNA polymerase (Sigma-Aldrich). All steps in library preparation were carried out using New England BioLabs enzymes except as noted. Resulting libraries were submitted for 75-bp, single-end Illumina sequencing on a NextSeq 500 instrument.

For Oxford Nanopore Technologies (ONT) long-read sequencing of yeast harboring YAC-*Mm*-4q21^{LacO}, genomic DNA was first incubated in 25 µg/ml RNase for 1 h at 37 °C and then extracted as before phenol-chloroform-isoamyl alcohol before precipitating with isopropanol. Once the DNA was fully resuspended in EB the following day, DNA was sheared with a g-tube (Covaris) we prepared the DNA for ONT long-read sequencing using the ONT ligation sequencing kit (ONT; # SQK-LSK112), following the manufacturer's instructions. The library was loaded onto a primed FLO-MIN106 R9.4.1 flow cell for sequencing on the MinION. All ONT data was basecalled with Guppy 3.6.0 (GitHub) with the HAC model.

To generate the sequence of the YAC, our expected input sequence, short (because of the high read accuracy) and long (to identify any major sequence rearrangements) read sequencing data was input into the EPI2ME pipeline wf-bacterial-genomes v0.2.12 (GitHub). The draft assembly output was used as a template for two kinds of manual revisions. First, in several places long reads spanned a gap with no coverage, the regions with no coverage were deleted to allow continuous coverage of those long reads. Second, in two locations, there were insertions not represented by the existing sequence map. Thus, reads with alignment to those regions were identified and the sequence from those reads not on the existing map was added to the assembly. Alignments were performed via the EPI2ME pipeline wf-alignment v0.3.3 (GitHub).

Spheroplast fusions

Yeast harboring YAC-*Mm*-4q21^{LacO} were grow overnight to saturation in a 5 ml culture of SD-URA. This culture was diluted to 50 ml in SD-URA and grown for 7-8 h at 30 °C to an OD_{600} of 0.8-1.0. Yeast were centrifuged at 1741 × g for 3 min in an A-4-62 swinging bucket rotor (Eppendorf), resuspended in 20 ml 1 M sorbitol and incubated at 4 °C overnight (<18 h).

Yeast cells were spun down at 1741 × g for 3 min and resuspended in 20 ml SPEM (1 M sorbitol, 10 mM EDTA, 10 mM sodium phosphate at pH 7.4). 40 µl of beta-mercaptoethanol and 60 µl of Zymolyase 20-T (stock solution of 200 mg Zymolase 20-T (MP Biomedicals: 320921) resuspended in 9 ml H_2O, 1 ml 1 M Tris pH 7.5, 10 ml 50% glycerol and stored at -20 °C) were added to the yeast solution. Yeast were incubated for 1 h at 37 °C to digest the cell wall. The success of spheroplasting was assessed by measuring the OD_{600} of solution diluted 1:10 in 1 M sorbitol (suspension A) and 1:10 in 2% SDS (suspension B). The 37 °C incubation was continued until the OD_{600} ratio of suspension A to suspension B was >10. At that point, 30 ml of 1 M sorbitol at 4 °C was added to the spheroplasts before spinning at 627 × g for 8 min at 4 °C. The spheroplast pellet was gently resuspended in 20 ml 1 M sorbitol before adding an additional 30 ml of 1 M sorbitol. The solution was centrifuged at 627 × g for 8 min at 4 °C. The pellet was resuspended in 1 ml of STC (1 M sorbitol, 10 mM $CaCl_2$, 10 mM Tris pH 7.5) and incubated at RT for 10-60 min.

Tissue culture cells were processed in parallel to the yeast. The day of fusion, 10 µl of 50 mg/ml S-trityl-L-cysteine (STLC) (Fisher: AAL1438406) was added to 70-80% confluent plate of HT1080 or U2OS cells for 6 h. Cells were trypsinized and neutralized with DMEM with 4.5 g/L D-Glucose and L-Glutamine (Gibco) before counting on a hemacytometer and spinning at 495 × g for 5 min at RT. Cells were resuspended in PBS to a concentration of 6×10^5 cells/ml. The concentration of yeast cells was determined by assuming 2×10^7 cells per OD_{600} per ml. 3×10^5 mammalian cells and 9×10^7 yeast cells were mixed in a 1.8 ml Eppendorf tube and incubated for 5 min at RT. The mixture was spun at 1699 × g for 30 s on a tabletop centrifuge. The pellet was resuspended in 45% PEG, 10% DMSO in 75 mM HEPES at pH 8.0 and incubated for 5 min at RT. The reaction was quenched by adding 1 ml of DMEM to solution before spinning at 4000 rpm for 30 s. The mixture was resuspended in 1 ml of DMEM before adding the mixture to a 6-well plate containing 2 ml of DMEM with 4.5 g/L D-Glucose and L-Glutamine (Gibco) supplemented with supplemented 10% FBS, 100 U/mL penicillin, and 100 µg/ml streptomycin. Cell lines were maintained at 37 °C in a humidified incubator with 5% CO_2 (all HT1080 and U2OS cells were cultured in these conditions unless otherwise stated). After 3-4 h and cells have adhered to the plate, the media was replaced with fresh media containing 2 µg/ml doxycycline. The media was replaced again the following morning.

After fusion, cells were incubated in DMEM with 2 µg/ml doxycyline for 48 h. Next, cells were trypsinized and moved to a 10 cm plate and cultured in DMEM supplemented with 10% FBS, 100 U/mL penicillin, and 100 µg/mL streptomycin with 333 µg/ml G418-S for 8 days. Cells were then moved to a lower concentration of G418-S (150 µg/ml) and, after three days, were processed further for IF-FISH, frozen down, or isolated into single clones.

Isolating monoclonal cell lines harboring HACs

Polyclonal HAC lines were first trypsinized before quenching with DMEM. Cells were centrifuged at 1,500 rpm for 5 min before being washed once with PBS. Cells were counted using a hemacytometer and centrifuged at 1,000 rpm for 3 min before resuspending in 10 ml PBS with 1 mM EDTA. The cells were centrifuged and resuspended once more to wash them. Cells were centrifuged once again at 1,000 rpm for 3 min before resuspending in PBS supplemented with 1 mM EDTA and 1% BSA to a final concentration of 1×10^6 cells/ml before being transferred to a 5 ml sterile polystyrene tube. Single cells were sorted in wells of a 96-well plate using a FacsJAZZ sorter. Cells were cultured for ~2-3 weeks in 50% DMEM and 50% conditioned media. Conditioned media was made by culturing HT1080 or U2OS cells overnight,

collecting media and filtering with a 0.22 μm filter. After colonies were visible, cells were scaled up to a 24-well, 6-well and then 10 cm plate while culturing in DMEM supplemented with 150 μg/ml G418-S. Clones were assessed for the presence of HACs via IF-FISH on metaphase spreads.

IF-FISH on metaphase spreads

IF-FISH was performed as described (*50*) with some modifications. HT1080 cells were treated with 50 μM STLC for 2-4 h to arrest cells during mitosis. Mitotic cells were blown off using a transfer pipette and swollen in a hypotonic buffer consisting of a 1:1:1 ratio of 75 mM KCl : 0.8% NaCitrate : 3 mM $CaCl_2$ and 1.5 mM $MgCl_2$ for 15 min. 3×10^4 cells were cytospun at 1500 rpm on high acceleration in a Shandon Cytospin 4 onto an ethanol-washed positively charged glass slide and allowed to adhere for 1 min before permeabilizing with KCM buffer for 15 min. Cells were blocked for 20 min in IF block buffer (2% FBS, 2% BSA, 0.1% Tween-20, and 0.02% NaN_3) before incubating for 45 min at RT with a monoclonal α-CENP-A antibody (Enzo; ADI-KAM-CC006-E) diluted 1:1000 in IF block buffer. Slides were washed 3 × 5 min in KCM buffer. Slides were incubated for 25 min at RT with Cy3 conjugated to donkey anti-mouse diluted 1:200. Slides were washed 3x in KCM for 5 min at RT. Slides were fixed in 4% formaldehyde in PBS, before washing 3x in dH_2O for 1 min each. Slides were incubated with 5 μg/ml RNAseA in 2x SSC at 37 °C for 5 min. Cells were subjected to an ethanol series to dehydrate the cells and then denatured in 70% formamide/2x SSC at 77 °C for 2.5 min. Cells were dehydrated again with an ethanol series.

Biotinylated DNA probe was generated using purified *M. mycoides* DNA with a Nick Translation Kit (Roche; 10976776001) according to the manufacturer's instructions, purified with a G-50 spin column (Illustra), and ethanol-precipitated with salmon sperm DNA and Cot-1 DNA. Precipitated BAC^{LacO} DNA or LacO plasmid was suspended in 50% formamide/10% dextran sulfate in 2x SSC and denatured at 77 °C for 5-10 min before being placed at 37 °C for at least 20 min. 300 ng DNA probe was incubated with the cells on a glass slide at 37 °C overnight in a dark, humidified chamber. The next day, slides were washed 2x with 50% formamide in 2x SSC for 5 min at 37 °C (45 °C for repetitive LacO FISH probe). Next, slides were washed 2x with 2x SSC for 5 min at 37 °C (45 °C and 0.1x SSC for repetitive lacO FISH probe). Slides were blocked with 2.5% milk in 4x SSC with 0.1% Tween-20 for 10 min. Cells were incubated with NeutrAvidin-FITC (ThermoFisher Scientific; 31006) diluted to 25 μg/mL in with 2.5% milk in 4x SSC with 0.1% Tween-20 for 10 min for 1 h at 37 °C in a dark, humidified chamber. Cells were washed 3x with 4x SSC and 0.1% Tween-20 at 45 °C, DAPI-stained, and mounted on a glass coverslip with Vectashield (Vector Labs). Slides were imaged on an inverted fluorescence microscope (Leica DMI6000B) equipped with a charge-coupled device camera (Hamamatsu Photonics ORCA AG) and a 100x 1.4 NA objective lens.

A "small HAC" designation was given if the cell contained a chromosome in which the FISH signal colocalized with CENP-A signal, was not overlapping an endogenous centromere, and the maximum diameter was < 1.0 μm. A "large HAC" designation was given if the cell contained a chromosome in which the FISH signal colocalized with CENP-A signal, was not overlapping an endogenous centromere, and the maximum diameter was > 1.0 μm. An ''integration'' designation was given if the cell contained a chromosome in which FISH probe signal localized to the DAPI-stainable region on the chromosome but did not colocalize with CENP-A signal; and a ''no signal'' designation was given if the cell did not contain a BAC probe signal on any DAPI-stainable region or colocalized with CENP-A signal.

For polyclonal HAC lines, 50 spreads were counted for each experimental condition and each HAC assay was performed in triplicate. The fraction of HACs with each designation was determined by dividing by 50. For isolated clones, 20 spreads were imaged and a clone was considered a HAC line if >20% of spreads contained a "small HAC" and no integrations or large HACs were present. The fraction of HACs in the isolated clone was determined by dividing the total number of "small HACs" by 20.

Lentivirus production

HA-LacI or EGFP-LacI lentivirus was produced by co-transfecting the HA-LacI or EGFP-LacI lentiviral plasmid and two packaging plasmids, pMD2.G and psPax2 (Addgene plasmids #12259 and #12260, respectively), into 293GP cells (*51*) and harvesting the media 48 h later. Specifically, a 10 cm plate of 50%–80% confluent 293GP cells was transfected with 6 µg of DNA (3 µg of the HA-LacI lentiviral vector, 750 ng pMD2.G, and 2.25 µg psPax2) and 18 µL of FuGENE 6 (Promega). The culture medium was changed 6-24 h later. 48 h post-transfection, the culture medium was harvested, filtered through a 0.45 µm filter, and stored at -80 ºC.

IF of mitotic cells

HAC-containing cells were plated in a 6-cm plate (in the presence of 150 µg/ml G418-S) and allowed to adhere to the bottom of the plate. The next day (when cells were 20%–30% confluent), the culture medium was replaced with fresh medium containing 500 µl of eGFP-LacI lentiviral supernatant and 18 µg polybrene (Specialty Media, TR-1003-G). 24 h later, the culture medium was changed to remove the lentiviral particles and polybrene. 48 h after transduction cells were seeded on an 18 x 18 mm^2 polystyrene coated coverslip. The following day, coverslips were transferred to a 6-well plate containing preheated PBS. PBS was removed via aspiration and 1 ml of 4% formaldehyde, 60 mM PIPES, 25 mM HEPES, 10 mM EGTA, and 4 mM $MgSO_4 \cdot 7H_20$ at pH 6.9 with 1% Triton-X-100 was added to each well, and cells were incubated for 20 min at 37 ºC. Formaldehyde solution was removed and fixing quenched with 2 ml 100 mM Tris pH 7.5 for 5 min. Coverslips were washed 3x in 2 ml PBS + 0.1% Tween for 5 min. Coverslips were placed in IF Block (2% FBS, 2% BSA, 0.1% Tween-20, and 0.02% NaN_3) for 20 min at RT. Coverslips were then incubated for 45 min at RT in a human ACA (Antibodies Inc.; 15-235) that we prepared by affinity purifying with recombinant CENP-A/H4 heterotetramers (*52*) and used at 0.74 µg/ml, mouse Aim-1 antibody (BD Transduction Laboratories; 611082) diluted 1:1,000 (serum), and rabbit anti-GFP antibody (made in-house) (*53*) used at 0.1 µg/ml in IF Block. Coverslips were washed 3x in 2 ml PBS supplemented with 0.1% Tween-20 for 5 min. Coverslips were then incubated for 25 min at RT in IF Block with Cy5 conjugated to donkey anti-human diluted 1:200, Cy3 conjugated to goat anti-mouse, and FITC conjugated to anti-rabbit. Coverslips were washed in 2 ml PBS + 0.1% Tween for 5 min before incubating in DAPI diluted 1:10,000 in PBS + 0.1% Tween-20 for 10 min. Coverslips were washed in PBS + 0.1% Tween-20, PBS, and then dH_2O before mounting coverslips on slides with vectashield. Slides were imaged on an inverted fluorescence microscope (Leica DMI6000B) equipped with a charge-coupled device camera (Hamamatsu Photonics ORCA AG) and a 40x 1.4 NA objective lens. HACs were identified via the presence of GFP signal. Each HAC was determined to be Aurora B positive if Aurora B signal was at least 50% above background. The total fraction of Aurora B positive HACs was measured across three independent experiments.

To calculate HAC distance compared to each endogenous centromere from the center of DNA mass in each mitotic bundle, we first determined the center of DNA mass by finding the longest distance across each mitotic bundle and determining the midpoint. Next, HACs were identified by the presence of GFP signal colocalized with ACA signal and endogenous centromeres were determined by the presence of ACA signal. We measured the coordinates of each HAC and endogenous centromere and calculated the distance of each from the center. HAC distance from the center was normalized based on the total length across (i.e. the diameter) of mitotic chromosomes and plotted in a radial plot.

Immunoblotting

Samples prepared from whole cell lysates were separated by SDS-PAGE and then transferred to nitrocellulose membrane. Blots were probed using the affinity purified human anti-centromere antibody at .74 µg/ml and horseradish peroxidase-conjugated donkey anti-human IgG (1:10,000, Jackson ImmunoResearch Laboratories #709-035-149). Antibodies were detected by enhanced chemiluminescence (ThermoFisher Scientific).

HAC retention assay

Four isolated HAC clones were cultured in the absence of G418-S selection for 30 days in triplicate. IF-FISH was performed at Day 0 and Day 30, and at least 20 cells were assessed for the presence of a HAC in each cell line at both time points. A daily HAC loss rate was determined using the following equation: $N_{30} = N_0 (1\text{-}R)^{30}$, where R is the daily HAC loss rate and N_0 and N_{30} are the number of metaphase chromosome spreads containing a HAC at Day 0 and Day 30, respectively (*12*, *29*).

Interphase IF-FISH

Clone 8 cells were washed in PBS and incubated in trypsin to resuspend. The suspension was quenched in DMEM, centrifuged, and then resuspended to a 2.5×10^7 cells/ml concentration. 100 µl of the cell suspension was pipetted into the center of a positively charged slide bordered by a hydrophobic barrier using a PAP pen. Cells were allowed to adhere for two hours before washing twice with PBS and fixing with 4% formaldehyde for 10 minutes at room temperature. Fixation was quenched with 100 mM Tris-HCl, pH 7.5 and then cells were permeabilized with KCM buffer (10 mM Tris-HCl, pH 7.7, 120 mM KCl, 20mM NaCl, 0.1% Triton X-100). Cells were washed in KCM 3x prior to incubating cells in IF block (2% FBS, 2% BSA, 0.1% Tween-20, 0.02% NaN_3 in 1X PBS). From here, CENP-A IF-*M. mycoides* FISH was conducted as described, above, for IF-FISH on metaphase spreads.

Enriching HACs via a sucrose gradient

8 15-cm plates of cells harboring HACs were cultured to a confluence of 80-95%. Cells were centrifuged at 1,500 rpm for 5 min at 4 ºC and resuspended in 30 ml of PBS. Cells were counted using a hemacytometer. Cells were centrifuged at 1,500 rpm at 4 ºC. Keeping cells on ice, the cell pellet was resuspended in 0.32 M sucrose in 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EGTA, 0.5 mM DTT, 0.1 mM PMSF, 1 mM leuptatin/pepstatin, 1 mM aprotinin, and 15 mM Tris pH 7.5. 2 ml of 0.32 M sucrose in 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, .1 mM EGTA, 0.5 mM DTT, 0.1 mM PMSF, 1 mM leupeptin/pepstatin, 1 mM aprotinin, and 15 mM Tris pH 7.5 with 0.1% IGEPAL were added to 2 ml of cells and incubated for 10 min. The mixture was added onto a Sarstedt tube containing 8 ml of 1.2 M sucrose in 60 mM

KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EGTA, 0.5 mM DTT, 0.1 mM PMSF, 1 mM leuptatin/pepstatin, 1 mM aprotinin, and 15 mM Tris pH 7.5. The mixture was added slowly to avoid mixing the two layers of differing sucrose concentration. The sucrose gradient was centrifuged at 10,000 × g, 20 min, 4 °C at acceleration setting 9 and deceleration setting 5 with an SS-34 rotor in a Sorvall RC-6 Plus centrifuge. Individual 1 ml fractions were collected for qPCR analysis and the top ~2.5 ml of solution (with care to avoid collecting cell debris) were collected for IF-FISH experiments.

qPCR of sucrose gradient-enriched HACs

DNA collected from sucrose gradient was first extracted with Phenol-Chloroform-Isoamyl alcohol before precipitating with isopropanol. DNA concentrations were determined via a Nanodrop. qPCR was performed in triplicate with 10 ng of initial DNA used in each reaction. qPCR amplification was detected using a 2x SYBR green master mix. Two primer sets were used, one with amplification of the CENP-A gene (present on endogenous chromosome) and another with amplification of the NeoR gene (present on the HAC). Nucleic acid amount was determined by an A^{260} measurement via a NanoDrop 2000 spectrophotometer. HAC enrichment was determined by the following calculation: fold enrichment = 1.81^([Ct sucrose fraction CENP-A gene – Ct sucrose fraction NeoR gene] – [Ct genomic DNA CENP-A gene – Ct genomic DNA NeoR gene]. HAC content in each fraction was determined by the following calculation: HAC DNA = 1.81^[Ct sucrose fraction NeoR- Ct genomic DNA NeoR]*[total DNA] * 760/6,270,000. Note that this calculation assumes 1 HAC per cell and a diploid genome.

IF-FISH on HACs isolated via sucrose gradient

HACs collected from the top of a sucrose gradient were cytospun onto slides and IF-FISH was performed as described above but with the following modifications. Before cytospinning, 50 µl HAC solution was diluted in 450 µl H_2O and incubated for 15 min. Next, during IF, slides were incubated at RT with CENP-A antibody for 2.5 h. HACs were identified on the slide via colocalization of DAPI (with size of the HAC DNA <2.5 µm and >0.5 µm), CENP-A IF signal, and LacO FISH signal. The number of foci containing CENP-A signals and LacO signals were counted from two separate experiments.

Quantification of IF-FISH on HACs isolated via sucrose gradient

Three-dimensional Z-stacks were processed using a maximum intensity projection algorithm into a two-dimensional image containing pixels of highest intensity from every slice of the Z-stack. Distances were measured between CENP-A and LacO from the pixels that had the maximum quantified intensity for each respective fluorescent signal (Mean = 0.13 µm, SD = 0.07 µm, N = 48). A one sample t-test using a hypothetical mean of 0, denoting that CENP-A and LacO are not distinctly localized, was performed.

Further enrichment of HACs for sequencing and H3K9me3 IF-FISH

The top 2.5 ml from the initial sucrose gradient step, described above, was collected using a cut pipette tip. This fraction was then loaded onto a second sucrose gradient comprised of 22% and 30% v/v sucrose both in 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EGTA, 0.5 mM DTT, 0.1 mM PMSF, 1 mM leuptatin/pepstatin, 1 mM aprotinin, and 15 mM Tris (pH 7.5). This solution was centrifugated at 20,000 × g for 20 min at 4 °C using an SW 60 Ti Swinging-Bucket Rotor in a Beckman Coulter Ultracentrifuge. At the 22-30% interface, 500 µL of solution

was collected using a cut pipette tip. For sequencing, a phenol-chloroform DNA extraction was performed on the collected solution. IF-FISH was performed, as above, but here α-H3K9me3 (1:500; Abcam; ab8898) was used followed by a Cy5 donkey anti-rabbit secondary diluted 1:200 (Jackson ImmunoResearch: 711-175-152).

Southern blots

Genomic DNA from the indicated cell lines (the multimerized HAC line used was Clone 11 from (*16*) and was diluted 20-fold in HT1080 HiLo cells) was prepared in agarose plugs by resuspending 5×10^6 cells/ml in 0.8% agarose and digested overnight with *Fse*I (NEB; R0588L) at 37 ºC. Digested DNA was separated via CHEF electrophoresis (Bio-Rad, CHEF DR II System) at 3 V/Cm, 250 to 900 s, for 50 h. The blot was transferred to a membrane (Amersham Hybond-N+) and blot-hybridized with a 100 bp probe that binds to the LacO sequence (5'-TTGTTATCCGCTCACAATTCCACATGTGGCCACAAATTGTTATCCGCTCACAATTCCACATGTGGCCACAAATTGTTATCCGCTCACAATTCCACATGTG-3'). The LacO-specific probe was end labeled with ^{32}P-γ-ATP for 1 h at 37 ºC before cleaning with illustra ProbeQuant G-50 micro column (GE Healthcare; 28-9034-08). The blot was incubated for 2 h at 42 ºC in hybridization buffer (ULTRAhyb™ Ultrasensitive Hybridization Buffer [Invitrogen; AM8669]). The probe was added to hybridization buffer and hybridized to the blot overnight at 38 ºC. The blot was washed twice with 2x SSC with 0.5% SDS for 30 min at 42 ºC. Finally, the blot was exposed to a phosphorimager screen for 2 weeks before imaging with an Amersham Typhoon.

FISH on stretched chromatin fibers

Extended chromatin fibers from human cells were prepared and FISH was performed as described (*40*) with some modifications. The modified steps include the following: 5×10^4 of HAC-containing cells were pelleted by centrifugation at 1000 × g for 5 min at RT. The cell pellet was resuspended in 500 µl of hypotonic buffer (75 mM KCl) and incubated for 10 min at RT. Slides were then cytospun for 4 min at 800 rpm on high acceleration in a Shandon Cytospin 4 onto a poly-lysine coated glass slide. Slides were transferred quickly into a falcon tube containing freshly prepared salt-detergent lysis (SDL) buffer composed of 25 mM Tris-HCl (pH 9.5), 500 mM NaCl, 1 mM PMSF, and 1% Triton X-100. After 20 min of incubation at RT, slides were washed for 15 min in PBS supplemented with 0.1% Triton X-100 and again in SDL buffer for 15 min before fixation with 3.7% formaldehyde. For experiments with mitotic enrichment for chromatin fiber stretching, colcemid was added to cell cultures at a final concentration of 0.1 µg/ml and incubated at 37 ºC in the presence of 5% CO_2 for 3-4 h. Growth flasks were then gently tapped with the palm of the hand, dislodging mitotic cells from the surface. Mitotic cells were harvested and transferred into a 15 ml falcon tube and centrifuged at 1500 rpm for 5 min at RT. The pellet was resuspended in 0.5 ml PBS and cells were counted. An aliquot of cell suspension of concentration 1×10^5 cells/ml was centrifuged at 1500 rpm for 5 min at RT. The pellet was resuspended in 1 ml of hypotonic buffer and incubated at RT for 15 min. An aliquot of 500 µl of cell suspension was loaded into cytospin funnel with poly-lysine coated slide and centrifuged at 1500 rpm for 5 min with cytospin set to high acceleration. One slide was carried through the fiber preparation protocol and the other slide proceeded through the mitotic spread protocol described above as a control slide to confirm successful mitotic chromosome enrichment.

Chromatin fibers from budding yeast were made from spheroplasts using an established protocol (*54*). Here, the yeast strain harboring YAC-*Mm*-4q21^{LacO} was grown overnight in SD-

URA media to ~1 × 10^7 cells/ml. Cells were harvested at 1200 × g for 5 min at RT in a pre-weighed 50 ml falcon tube. The supernatant was decanted and the mass of the pellet was measured. The pellet was then resuspended in a mixture of 0.1 M EDTA-KOH (pH 8.0) and 10 mM DTT at 1 ml/0.1 g of cells. The cell suspension was then incubated at 30 °C for 10 min with gentle agitation. After incubation, cells were pelleted at 800 × g for 5 min at RT. The cell pellet was carefully resuspended in a mixture of SD-URA and 1.2 M sorbitol at 1 ml/0.1 g of cells (first resuspended with 500 µl of mixture before adjusting to the final volume). Lyticase (Sigma #L2524) and zymolyase-20T (SEIKAGAKU #120491) were added at 1000 U/ml and 300 µg/ml respectively. The cell suspension was transferred into the Erlenmeyer flask used for the overnight culture and then incubated at 30 °C for 1 h for maximum spheroplast formation of 60-70%. Spheroplasts were centrifuged at 1000 × g for 5 min. Chromatin fiber-FISH was performed as described, above.

Short-read sequencing of cells harboring HACs

Total genomic DNA or sucrose gradient fractions from the HAC enrichment procedure described above were extracted and subsequently ethanol precipitated. Library preparation was performed using Illumina 20060060 following the manufacturer instructions. Libraries were indexed during amplification using Illumina 20027213. Subsequently, libraries were loaded onto an S2 flow cell for 2x150bp sequencing with the NovaSeq 6000. Demultiplexed, adapter-removed reads were aligned against a concatenated genome consisting of T2Tv2, YBB6v3, and Saccer3 using Bowtie2 with the following arguments: -p 8 -D 20 -R 3 -N 1 -L 20 -q -t. Samtools was used to generate BAM files of the alignments. Reads aligning to YBB6v3 were extracted and converted into a bedcoverage file for view in the UCSC genome browser. In order to view the alignment track, the YBB6v3 sequence was uploaded as a custom hub.

Snipgenie v0.6.0 (*55*) was used to assess the sequence integrity of covered regions and inspection of potential variants. The following sequences were extracted from YBB6_manual_revisions_v3: 378324-380510, 555068-558944, and 751339-757911) into a single reference FASTA. Said reference was used for aligning enriched library reads to assess for SNP and indel variants.

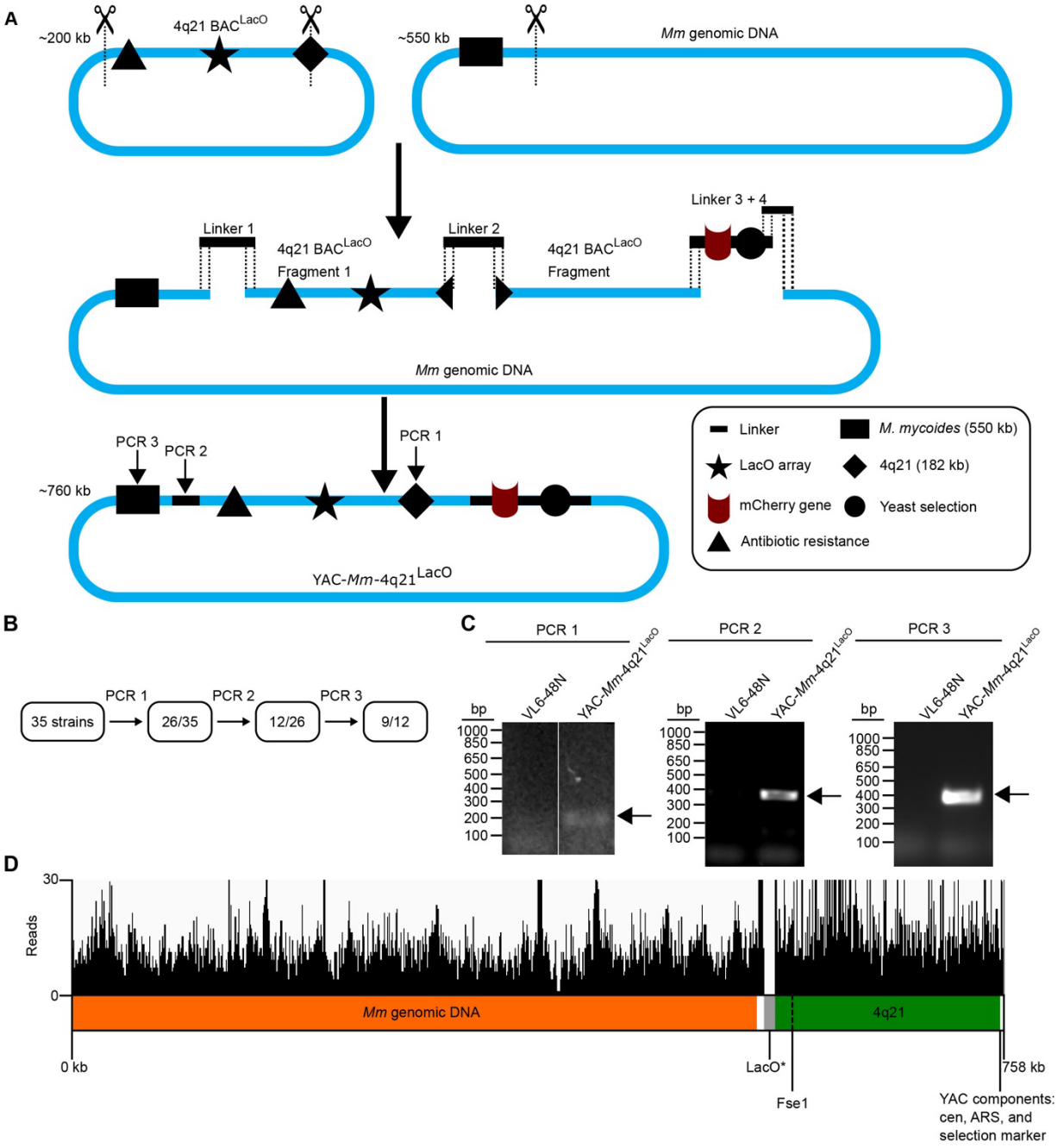


Fig. S1: A 760 kb YAC construct with the necessary components for HAC formation was generated via TAR cloning. A) Schematic of YAC construct, *Mm*-4q21^{LacO}, that was formed to generate single copy HACs. B) PCR used to validate the construct was performed on a total of 35 yeast strains of which 9 had positive bands in all three PCR reactions. C) Junction PCR of the yeast strain containing YAC-*Mm*-4q21^{LacO} D) Draft assembly (see Methods) of YAC-*Mm*-4q21^{LacO}. Sequencing reads aligned to the YAC construct confirm presence of all components of

the YAC except for the LacO array. A separate alignment of reads to the LacO array while allowing for multiple alignments confirmed the presence of the LacO array.

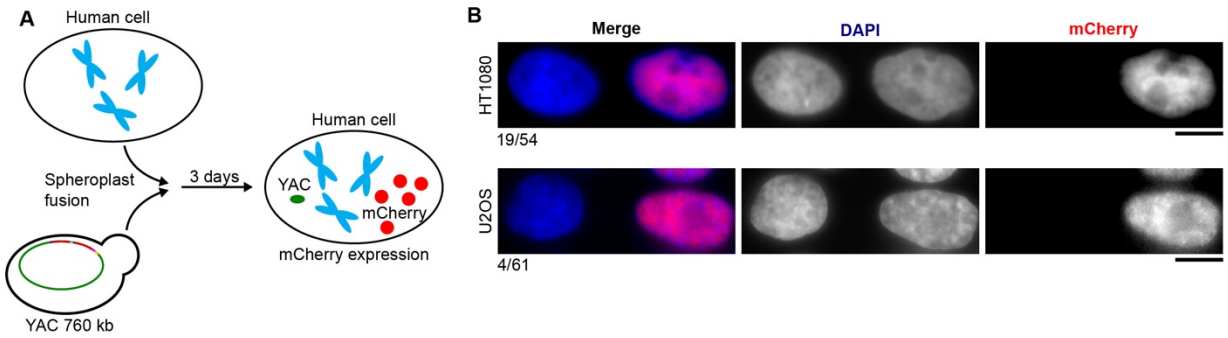


Fig. S2: Yeast fusion approach is effective for delivering large DNA constructs in U2OS and HT1080. A) Schematic of approach to test for successful yeast fusion. B) Examples of successful yeast fusion with mCherry expression as well as the proportion of cells showing mCherry expression. The proportion of cells that were mCherry positive for each cell line is noted. Bar, 10 µm.

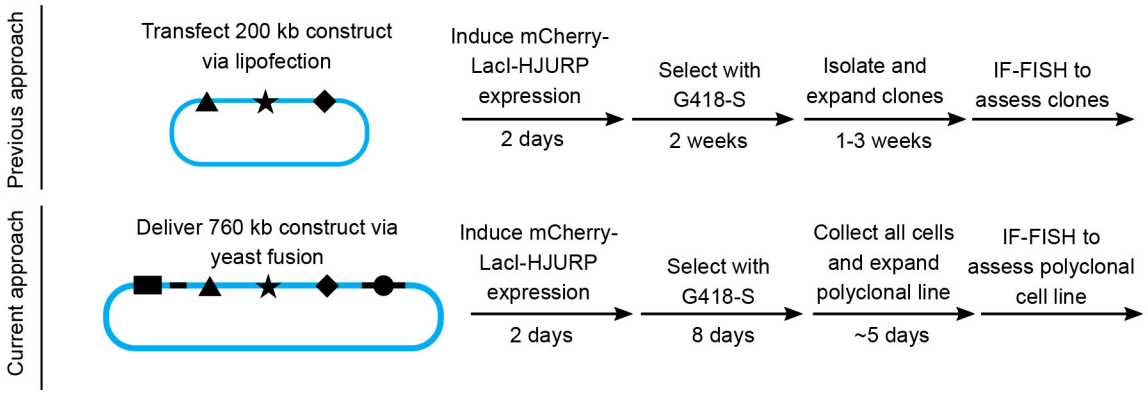


Fig. S3: Schematic juxtaposing the approach to generate HACs in prior generations (*16*) and the approach reported here.

440

Figure S4

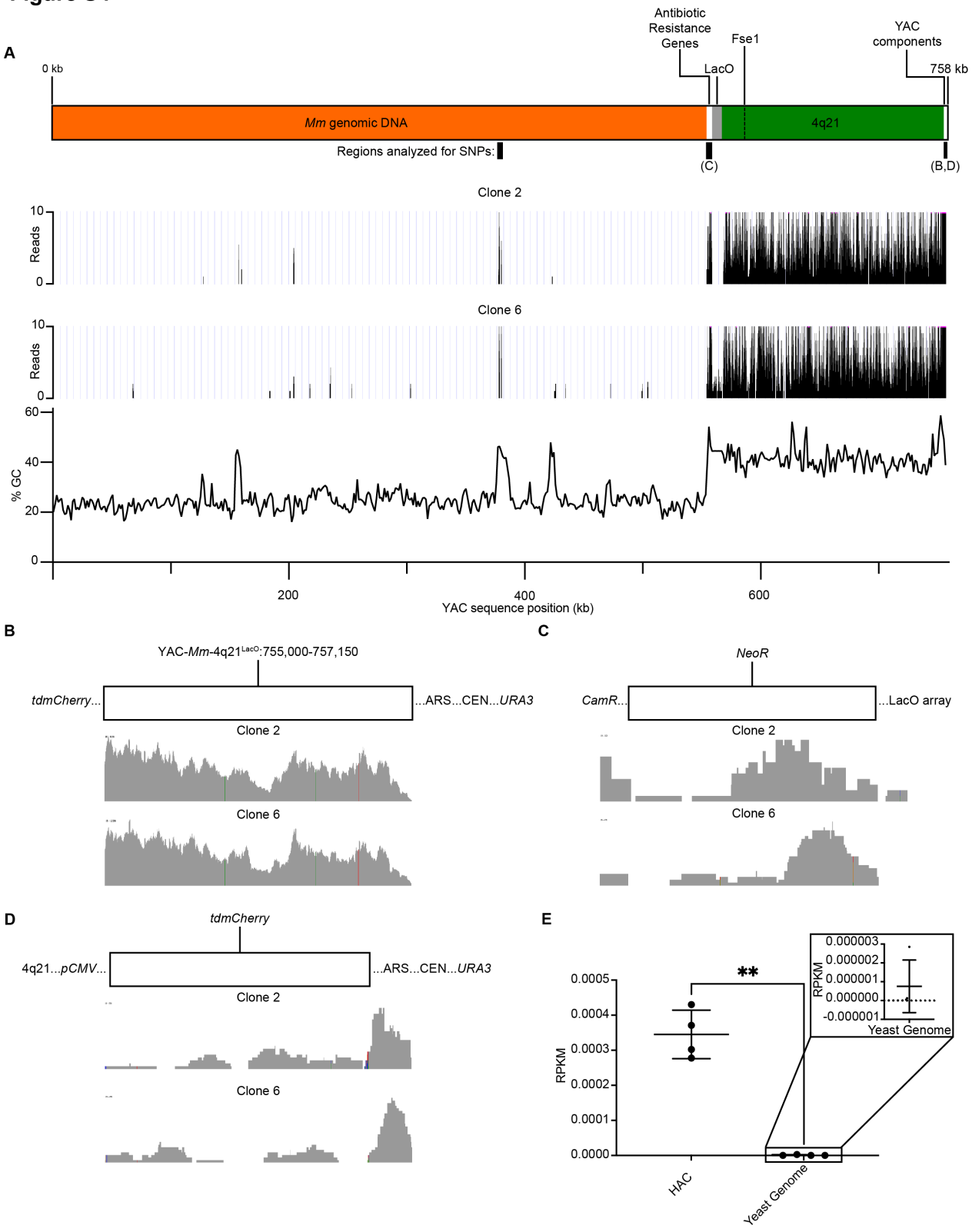
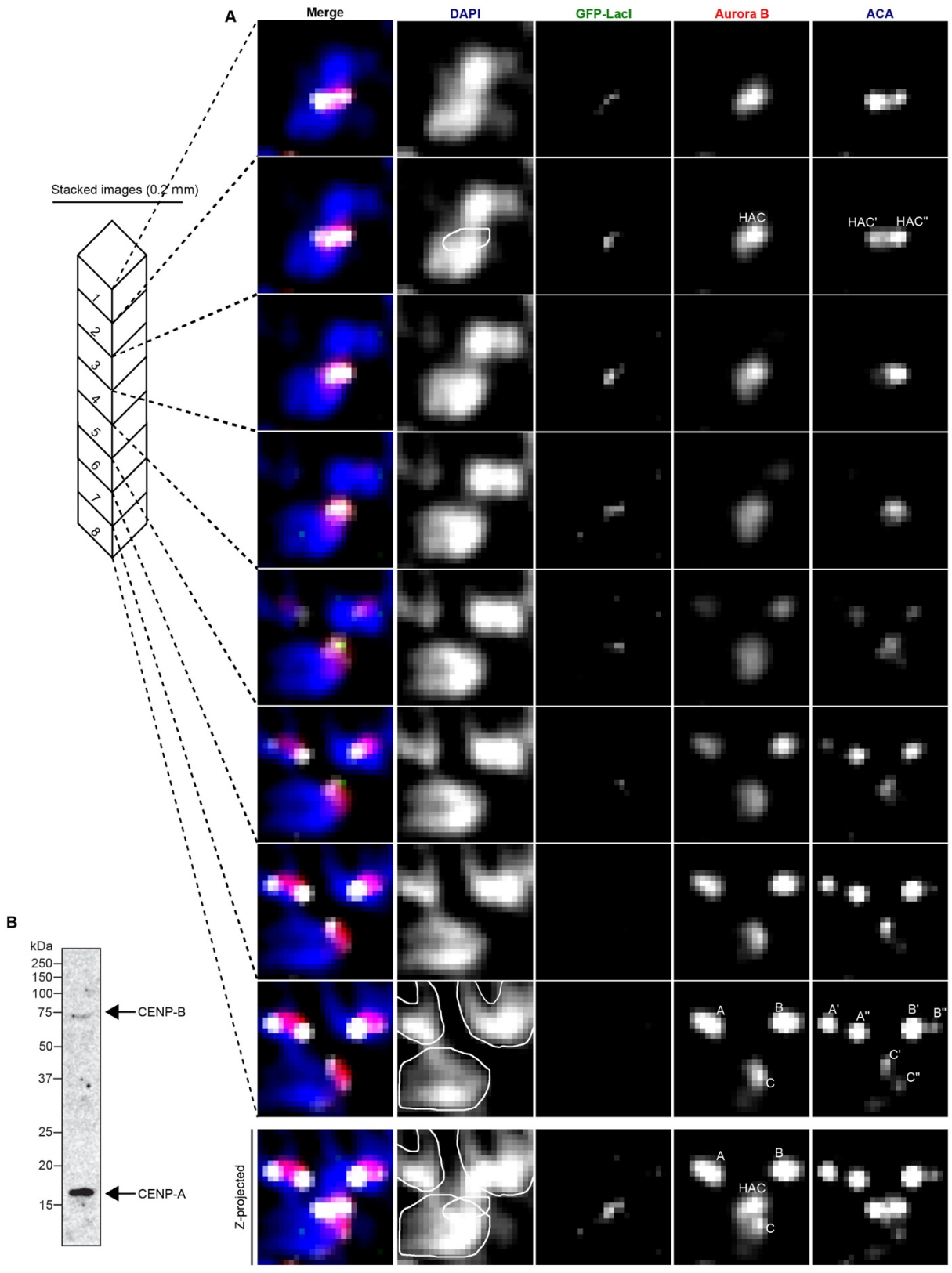

A
Antibiotic Resistance Genes
Fse1
YAC components
0 kb
LacO
758 kb
Mm genomic DNA
4q21
Regions analyzed for SNPs:
(C)
(B,D)
Clone 2
Reads
Clone 6
% GC
YAC sequence position (kb)
B
YAC-Mm-4q21LacO:755,000-757,150
tdmCherry...
...ARS...CEN...URA3
C
NeoR
CamR...
...LacO array
D
tdmCherry
4q21...pCMV...
E
RPKM
HAC
Yeast Genome

Fig. S4: Sequencing reveals elements incorporated into the *M. mycoides* stuffer, integrity at the single nucleotide level, a lack of detectable integration of natural yeast chromosomes into the human genome, but coverage biased against the AT-rich stuffer DNA. A) Illumina sequencing reads derived from Clone 2 and Clone 6 genomic DNA subjected to a serial enrichment strategy (see Materials and Methods) were aligned to the YAC-*Mm*-4q21^{LacO} sequence and visualized in the UCSC Genome Browser. Analysis confirms that the two genes incorporated into the HAC (NeoR and mCherry) are present. Coverage of the HAC was incomplete, omitting significant portions of the extremely AT-rich *M. mycoides* stuffer DNA. The low degree of sequence coverage observed throughout the *M. mycoides* in combination with the robust signal observed in our cytological experiments, both in yeast (as a YAC) and human cells (as a HAC) (Figs. 4 and S8), supports the notion that a very strong sequencing bias influences coverage. Indeed, sequence coverage closely coincide with GC content. The coverage paucity is, thus, most likely due to high sequencing bias resulting from its average GC content of 25%. For future designs of HAC constructs where complete sequence coverage in the context of a complex genome (such as a mammalian genome) is a goal, stuffer with GC content that is in the realm of most endogenous chromosome sequences, which is ~60%, would likely offer a benefit. Snipgenie was used to assess clone-specific SNPs and indels located within the contiguous regions denoted by black bars, resulting in no calls and further suggesting sequence integrity at the single nucleotide level. Note the unique regions of the HAC can be unambiguously assigned; whereas, the 4q21 reads could be derived from the native locus as well as the HAC locus. B) A contiguous stretch of high sequence coverage from the same data in panel A, which was selected for close analysis with Snipgenie, visualized in the Integrated Genomics Viewer. Light grey denotes a perfect match; whereas, an alternately colored line denotes mismatch relative to the reference. Note that the three mismatches appear in both clones, suggesting inheritance from the parent YAC-*Mm*-4q21^{LacO}, followed by sequence stability through YAC-*Mm*-4q21^{LacO} HAC formation and maintenance. C and D) Coverage mapping to the YAC-*Mm*-4q21^{LacO} reference from 556,800-557,800 (C) and 752,220-753,920 (D). We chose the region in panel C for display since it contains the neomycin resistance gene (NeoR) coding sequence and the region in panel D for display since it contains the tandem dimer mCherry (tdmCherry) coding sequence (ORFs denoted with rectangles above each plot). Coloring is as in panel B. The apparent mismatches are a consequence of lower coverage in these regions relative to that shown in panel B and do not meet criteria that Snipgenie requires to call SNPs. E) Read counts of unenriched libraries derived from Clone 1, Clone 2, Clone 6, and Clone 8 mapping to the YAC-*Mm*-4q21^{LacO} reference (minus 4q21) and to the Saccer3 yeast genome (minus the rDNA locus and an adjacent region that maps to a 28*S* human rDNA gene) per kilobase of queried reference per million mapped reads (RPKM). Statistical analysis was performed with a one-tailed, Welch's unpaired T-test. The mean (+/- SD) is shown. ** denotes $p<0.01$.


A
Merge
DAPI
GFP-LacI
Aurora B
ACA
Stacked images (0.2 mm)
1
2
3
4
5
6
7
8
HAC
HAC'
HAC"
A
B
C
A'
A"
B'
B"
C'
C"
Z-projected
B
kDa
250
150
100
75
50
37
25
20
15
CENP-B
CENP-A

480

Fig. S5: 3-dimensional localization of HACs in monopolar mitotic cells. A) Mitotic chromosomes commonly overlap upon z-dimensional projections but are separable upon analysis of individual z-stacked images. This is especially important since DAPI staining is so heavily dominated by the natural chromosomes that are ~100-fold larger than single-copy HACs. ACA signal brightness was increased relative to 2D to increase visibility without z-projection. Three native chromosomes (labeled A, B and C) are immediately adjacent or overlapping to the HAC in the x and y dimensions but are ~1 μm from the HAC in the z dimension. The close proximity of these chromosomes can account for the DAPI staining seen near to the HAC. In the z-stack of maximal Aurora B intensity, Aurora B from the HAC or natural chromosomes is labeled HAC, A, B or C and centromere double dots are labeled with HAC', HAC'', A', A'' B', B'', C' or C''. Additionally, the outlines of these chromosomes based on DAPI staining are shown. A z-projected image of all 8 stacks and all centromeres labeled is shown below the individual z-stacks. The peak Aurora B fluorescence is found in the following z-stack images: HAC, 2; A, 8; B, 8; C, 7. The peaks of the two ACA foci for each centromere are found in the following z-stack images: HAC', 1; HAC'', 3; A', 8, A'', 8; B', 8, B'', >8; C', 7; C'', 8. B) Immunoblot of whole cell lysate using affinity purified ACA antibody shows two bands at the expected size for CENP-B and CENP-A.

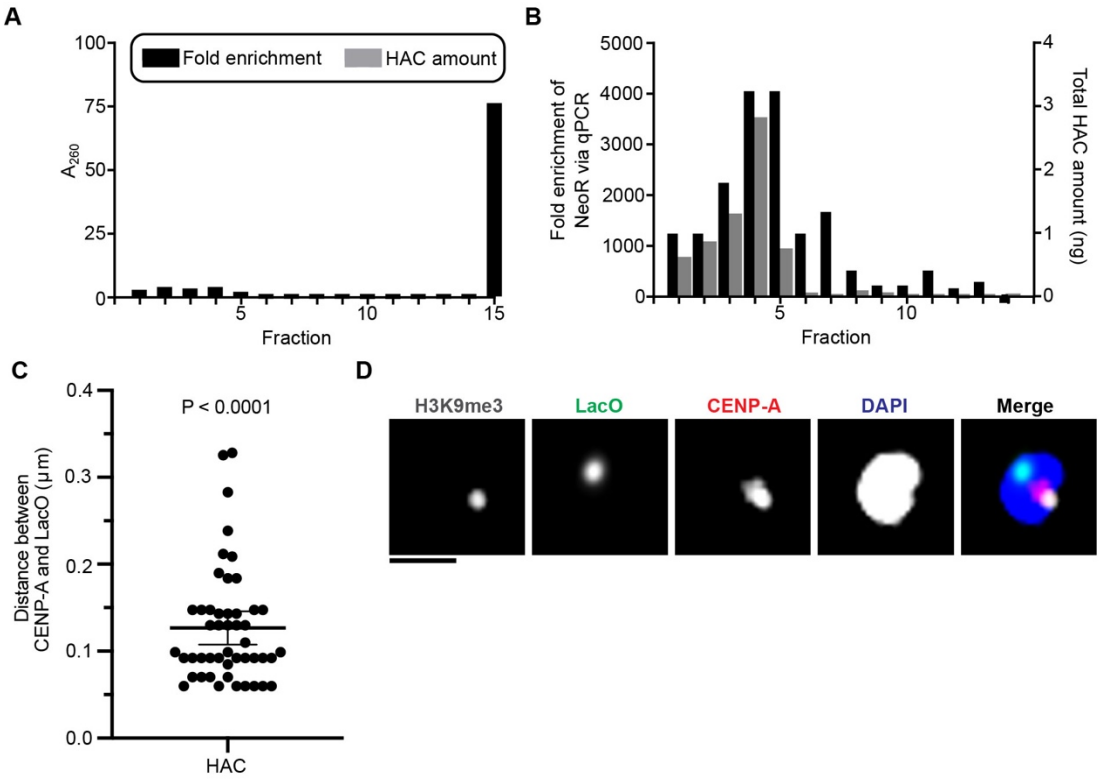
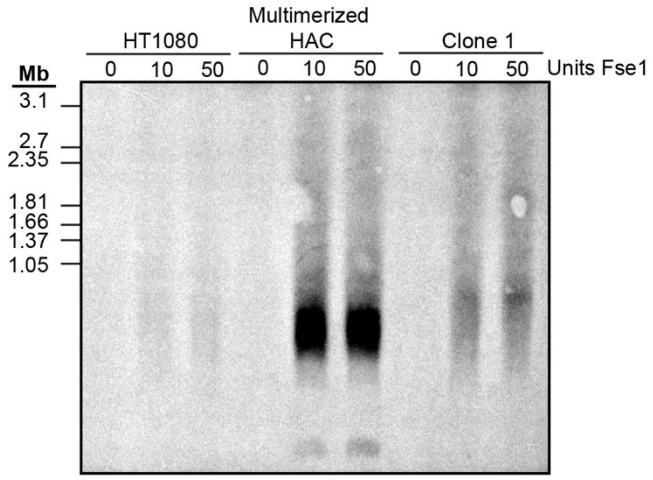


Fig. S6: Quantification of CENP-A and LacO distances show distinct localization and H3K9Me3 is present on HACs. A) A_{260} measurements of fractions collected from a sucrose gradient from top (fraction 1) to bottom (fraction 14) as well as pelleted nuclei (fraction 15). Fraction 15 was diluted 33.3x relative to other samples to acquire a reading in the measurable range (dilution corrected values are plotted). B) Enrichment of HAC DNA compared to endogenous chromosomal DNA. The HAC DNA concentration is also shown. C) Three-dimensional Z-stacks were processed using a maximum intensity projection algorithm into a two-dimensional image containing pixels of highest intensity from every slice of the Z-stack. Distances were measured between CENP-A and LacO from the pixels that had the maximum quantified intensity for each respective fluorescent signal (Mean = 0.13 µm, SD = 0.07 µm, N = 48). A one sample t-test using a hypothetical mean of 0, denoting that CENP-A and LacO are not distinctly localized, was performed. The t-test result indicates that the mean is significantly different from 0 ($P < 0.0001$), suggesting distinct localization. D) H3K9Me3 was present on 21/26 HACs, further confirming the presence of the inner centromere on the HAC. Bar, 1 µm.



515

Fig. S7: Southern analysis comparing HACs to the parental cell line as a control. The multimerized HAC line was diluted 20-fold in HT1080 cells to reduce signal relative to the YAC-*Mm*-4q21^{LacO}-based HAC.

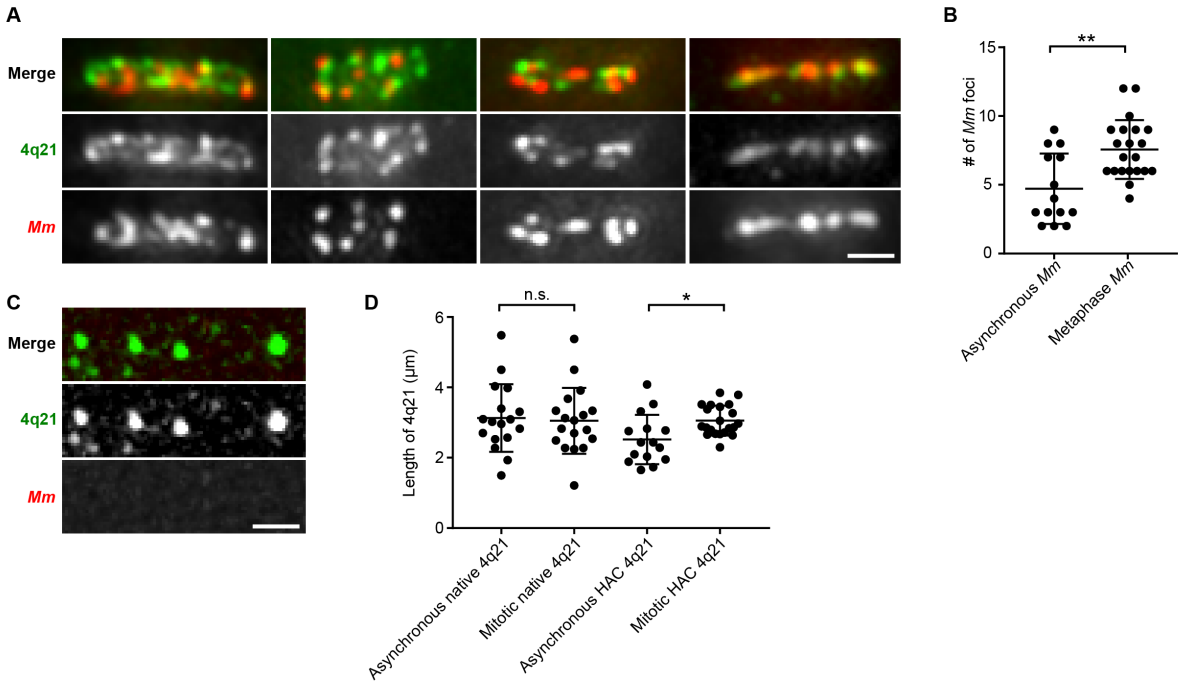


Fig. S8: Representative examples of the HAC after physical stretching. A) 4q21 signal is typically more diffuse compared to *M. mycoides* signal suggesting that the two types of DNA have distinct chromatin properties. B) Quantification of *Mm* FISH foci in experiment shown in Fig. 4, C and G. C) Representative images of stretched endogenous 4q21 labeled via FISH after enriching for cells in metaphase. D) Quantification of length of 4q21 in experiment shown in Fig. 4, D and G.

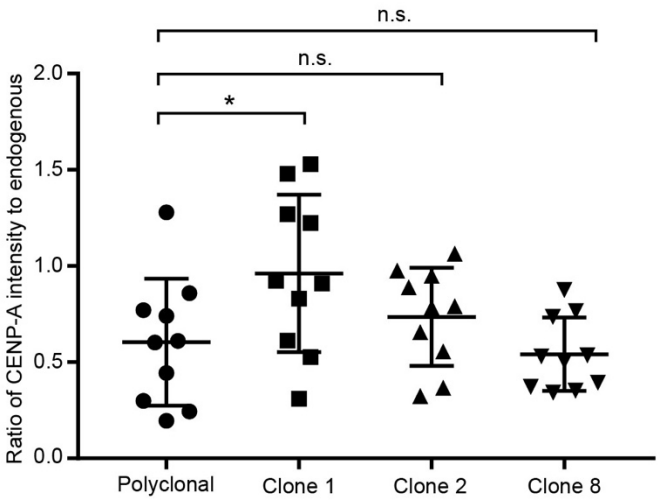


Fig. S9: Quantification of CENP-A intensity in HAC metaphase chromosome spreads in a polyclonal line compared to 3 isolated HAC lines. Quantification of the intensity of a CENP-A at HACs relative to the intensity at endogenous centromeres. Each point represents one HAC. The measured intensities for each HAC line were: polyclonal: 0.20, 0.24, 0.61, 0.74, 0.44, 0.77, 0.60, 0.86, 0.30, 1.28; clone 1: 1.48, 1.27, 0.83, 1.53, 0.53, 0.92, 0.61, 0.91, 0.31, 1.22; clone 2: 0.98, 0.78, 0.89, 0.66, 0.95, 1.06, 0.79, 0.32, 0.56, 0.37; clone 8: 0.54, 0.51, 0.37, 0.88, 0.34, 0.35, 0.74, 0.77, 0.53, 0.39. The mean (+/- SD) is shown. $p < 0.05$, n.s, and n.s. based on an unpaired, two-tailed t-test compared isolated HAC lines to the polyclonal line.

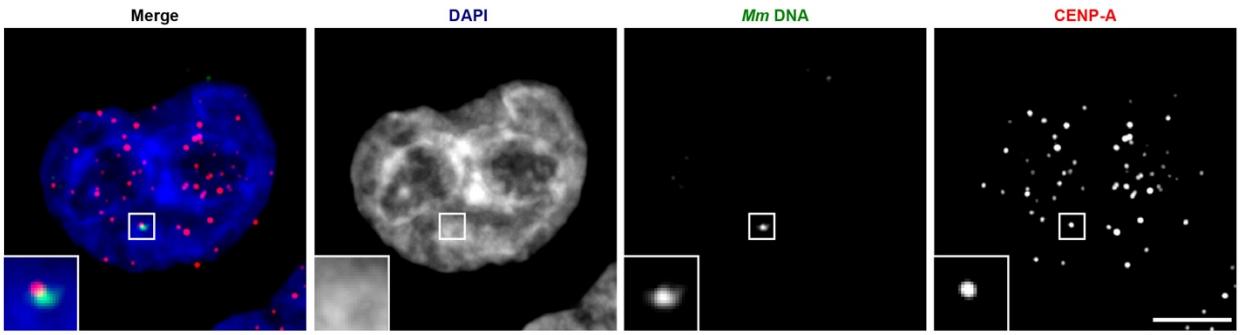


Fig. S10: HACs are found in the nucleus in interphase cells. Representative image from clone 8 IF-FISH showing HAC inside of nucleus. Inset: 3x magnification. Bar, 5 μm.

Table S1. Summary of the HAC clones generated in this study.
The relevant figures that the HAC clones are described in is listed.

Clone	Figures referenced	% of cells with HACs	Identifier
1	2B-C, 4A, 4D-G, S4, S7, S8	0.70	yBB6 HAC 1-7
2	2A-E, 3B, 4A, S4, S5, S6	0.90	yBB6 HAC 1-15
3	2B-C	0.87	yBB6 HAC 1-16
4	2B	0.60	yBB6 HAC 1-19
5	2B-C	0.60	yBB6 HAC 1-23
6	2B, S4	0.65	yBB6 HAC 4-1-1
7	2B	0.65	yBB6 HAC 4-2-1
8	2B, S4, S10	0.67	yBB6 HAC 4-3-4