**Supplemental Data**

**Methods**

**rDNA Southern blotting**

*C. elegans:*

For Southern blotting, CHEF gels were prepared and probed using protocols outlined in Tsuchiyama *et al.* 2013 (Tsuchiyama *et al.* 2013). In short, each gel was washed twice for 10min each in 0.25N HCl to nick and depurinate DNA, followed by two washes for 15min each in 0.5N NaOH, 1M NaCl. These incubations were followed by two washes for 15min each in 0.5M Tris, 3M NaCl. DNA was then transferred from the gel to a nylon membrane (Perkin Elmer GeneScreen Hybridization Transfer Membrane) and crosslinked using a Stratagene Stratalinker UV Crosslinker in preparation for radioactive probe hybridization.

The probe for Southern blotting *C. elegans* rDNA was created from an 850bp PCR product that overlaps the first 300bp of *rrn-1* (18S) (Table S14), purified with the Zymo Clean & Concentrator kit (D4013) before radioactive labeling. Please note that we observed another rDNA probe to behave anomalously during standard gel electrophoresis. Our initial PCR product intended for Southern blotting was designed for an 806bp region within the 26S gene and ran at the correct size following PCR amplification, but ran incorrectly (at ~500bp) following purification and elution in water. Addition of buffer (Invitrogen Y02028) returned the product to an 800bp run size. The 850bp PCR product (Table S14) behaved more consistently and thus was the one used for probing in this study.

Band measurement: To determine rDNA copy number from CHEF gel followed by Southern blot, the distance between the bottom of the well and the middle of a sample band was measured manually for the presented Southern blots. The distance between the bottom of the well and each of the ladder bands was measured from the ethidium bromide stain of the gel before it was transferred to the blot. The relationship between band size and gel distance was plotted and used to determine band size for each rDNA band (Table S15). There were two distinct, approximately equal-intensity bands for MY16 in two of the CHEF gels (Table S15). For CHEF average calculation, the two bands were averaged first to produce that CHEF replicate’s rDNA estimation for MY16. ED3040 and MY6 displayed minor bands; for these strains, only the major band was considered.

**Analysis software and version**

bowtie2/2.2.3 (Langmead and Salzberg 2012; Langmead *et al.* 2019)

bwa/0.7.15 (Li 2013)

samtools/1.4 and samtools/0.1.18 (Li *et al.* 2009)

picard/2.14.0 (“Picard Tools - By Broad Institute”)

trim\_galore/0.4.1 (Krueger 2015)

cutadapt/1.8.3 (Martin 2011)

fastqc/0.11.7 (Andrews 2010)

java/8u25

wget

python/2.7.3

R version 3.5.1 (R Core Team 2018)

Data were visualized with ggplot2 (Wickham 2009). The colorspace package (Zeileis *et al.* 2019) and Color Universal Design palette (Okabe and Ito 2008) were used for some visualizations.

**Statistics**

Pearson’s correlations and p-values were calculated using the R cor.test function. To assess if ddPCR and CHEF values were significantly different, t-tests (two-tailed, unequal variance) were performed on the rDNA copy number of the six yeast strains for which we had three replicates each for the two methods.

**Single copy region copy number estimation**

Twenty-nine 7.2kb regions of the *C. elegans* genome were selected for use in library quality control analysis (Table S5). An original list of 32 regions was generated by extracting 7.2kb regions of non-masked sequence from the masked version of the *C. elegans* genome. One of the regions on this list was eliminated for multiple alignment. Another two were eliminated for absence in one or more of our wild isolates. The remaining 29 were analyzed for estimated copy number by read counting, in the same manner described above for rDNA copy number estimation. A custom Perl script counted how many reads aligned to each of the 29 regions. (Table S6). Copy number for each region was calculated by the ratio of reads aligning for that region to total aligned reads.

**Maximum likelihood estimation GC content correction (GCC) method of rDNA copy number determination**

Method is based on Parks and Blanchard 2018 (Parks *et al.* 2018) and Benjamini and Speed (Benjamini and Speed 2012).

*C. elegans:* demultiplexed, paired FASTQ files were aligned to the unmasked WS230 genome (“all reads”) and to a single copy 45S rDNA sequence (“rDNA reads”) with bowtie2/2.2.3 (Langmead and Salzberg 2012; Langmead *et al.* 2019), retaining only mapped reads. Resulting all-reads .bam files were sorted with samtools/1.4 (Li *et al.* 2009). Median fragment length of all reads was determined with Picard CollectInsertSizeMetrics (picard/2.14.0, java/8u25) (“Picard Tools - By Broad Institute”). The .bam files were reduced to text files containing the chromosome mapped to and the leftmost mapping position for each properly aligned read on the forward strand for both all reads and rDNA reads. Maximum likelihood estimates of rDNA copy number were determined with the following equations using a custom python script. Briefly, for each sample the GC content-specific fragmentation rate ($λG$) for a fragment of the median fragment length was calculated based on methods described by Benjamini and Speed (Benjamini and Speed 2012). $λG$was calculated with all properly aligned reads, excluding those that mapped to the 5S or 45S rDNA or the telomeres.

GCC Equation (based on Parks and Blanchard (Parks *et al.* 2018)):

$$Copy Number= \frac{\# Fragments mapping to rDNA}{\sum\_{i=1}^{n}λG(pi)}$$

Where the region of interest is defined by positions [pi,…,pn].

*S. cerevisiae:* Split, paired FASTQ files were aligned to the unmasked *S. cerevisiae* S288CR57-1-1 genome (“all reads”) and to a single copy 45S rDNA sequence (“rDNA reads”) with bowtie2/2.2.3, retaining only mapped reads (Langmead and Salzberg 2012; Langmead *et al.* 2019). .bam files were processed as for *C. elegans*. Maximum likelihood estimates of rDNA copy number were determined with a custom python script. Reads mapping to or overlapping telomeres, centromeres, Ty elements, and long terminal repeats were excluded from the analysis (Rienzi *et al.* 2012). Reads mapping to the rDNA were excluded from the GC fragmentation calculation.

**Alignment with BWA-MEM**

For the data represented in Table S4, the demultiplexed *C. elegans* reads described above were aligned to the WS230 and single copy 45S rDNA sequence with BWA-MEM (Li 2013). Sequential analysis was performed with the GCC metric as described above. Reads were not trimmed for this analysis.

**Trimming**

Reads for the sequencing data in the main text were not trimmed. For the trimming analysis in demonstrated in Table S4, the demultiplexed *C. elegans* reads described above were trimmed (maintaining read pairs) with Trim Galore, with the settings --paired -phred33 -q 20 (Krueger 2015). Trimmed reads were aligned with bowtie2 analyzed with the GCC metric as described above.

**Downsampling**

Whole genome sequencing data from MY1 and JU775 .bam files from (Thompson *et al* 2013, aligned to WS230) were downsampled to 90%, 50% or 5% of the original bam, using samtools/1.4 view –b –s (Li *et al.* 2009). Downsampling was done five times for each strain and each percent, using a different seed for each. rDNA copy number of the 45S (7197bp repeat unit, in WS230 coordinates ChrI 15060288-15071022), as well as the 5S (976bp repeat unit, coordinates ChrV 17115879-17131432) was calculated by relative read coverage method:

(rDNA\_counts\*100286070)/(total\_counts\*7197) = 45S rDNA copy number

Mitochondrial genome copies were also counted as number of reads aligning to the mitochondrial genome relative to average total genome coverage:

(mtDNA\_counts\*100286070)/(total\_counts\*13794) = mitochondrial copy number

See Table S3.

**smMIP design**

Repeat region inversion probes were designed using custom scripts made in perl and R (Mok *et al.* 2017). rDNA-specific MIPs were designed using the sequences from *rrn-1*, *rrn-2*, and *rrn-3* with an additional 250bp of sequence at the beginning and end of each gene. Each rDNA sequence was analyzed using a sliding window of 501bp to design probes with the highest likelihood of success based on empirical experimentation parameters (Turner *et al.* 2009; O’Roak *et al.* 2012). A total of 53 candidate probes were synthesized for follow-up investigation. An additional set of four MIPs to target single-copy regions of the genome was chosen from a prior set of well-behaved MIPs (Mok *et al.* 2017). MIP sequence information can be found in Table S14.

**Normalization plasmid design**

The normalization plasmid pEM48 was generated by amplifying the 7.2 kb rDNA locus (primers in Table S14) and inserting the product into the pDONR221 backbone using BP clonase (Invitrogen #11789-020). Five mutations were introduced into the rDNA sequence to distinguish plasmid from genomic DNA (Tables S14), using Q5 mutagenesis (NEB #E0554S). Four control regions from the genome, also containing distinguishing SNVs, were cloned into the vector by Gibson assembly (NEB #E2611S) (Table S14).

**MIP library preparation**

smMIPibraries were prepared using the same protocol as previously described (Mok *et al.* 2017). Briefly, equimolar amounts of MIPs from 100 µM concentrations were pooled and 85 µl of this pool was treated with 50 units of polynucleotide kinase (NEB) for 45 minutes at 37˚ and then 20 minutes at 80˚C in a 100 µl reaction. The 5’-phosphorylated probes were diluted to 330 nM for use in later steps. MIP libraries were based on (Hiatt *et al.* 2013). Annealing reactions containing 0.5 ng (~500 genomes) to 500 ng (5x106 genomes) of target genomic DNA, an amount of plasmid equivalent to 1X to 19X molar whole genome input, 330 fmoles of MIP pool, and 1X Ampligase buffer (Epicentre) in 10 µl total were treated for 3 minutes at 98˚C, 30 minutes at 85˚C, 60 minutes at 60˚ and 120 minutes at 56˚C. MIP pools were composed of 80% repeat region inversion probe pool and 20% wild isolate-specific MIPs (set of 113 probes). The purpose of this mixture was two-fold: to ensure basepair diversity in the sequencing process and to confirm the identity of wild-isolate genomic DNA samples. To gap fill the product, 300 pmoles dNTPs, 7.5 µmoles Betaine (Sigma), 20 nmoles NAD+, 1X Ampligase buffer (Epicentre), 5 units Ampligase, and 2 units Phusion DNA polymerase (NEB) were added to the 10 µl anneal reaction to a final volume of 20μL and incubated for 120 minutes at 56˚C, and 20 minutes at 72˚C. To degrade genomic template and any remaining linear MIPs, 20 units Exonuclease I (NEB) and 50 units Exonuclease III (NEB) were added and incubated for 45 minutes at 37˚C, and 20 minutes at 80˚C. 10 µl of this capture reaction was then amplified by 18 rounds of PCR (15 seconds at 98˚C, 15 seconds at 65˚C, 45 seconds at 72˚) with 1 unit Kapa Hifi Hotstart TAQ, 10 nmoles dNTPs, and 25 pmoles each of forward and reverse primers in a 50 µl reaction. Libraries generated with an initially high genomic input amount were size-selected between 250-450 bp and purified with Agencourt AMPure XP beads before sequencing with Illumina sequencing technology.

DNA input conditions were based on molecular ratios of available oligo probes per reaction (330 fmoles total) versus potential rDNA template sites. We aimed to have a minimum of 1000 available probes per potential target site in a range of rDNA copy numbers from 50 to 400 copies. Empirically, we found that lower DNA input amounts were also more accurate in estimating rDNA copy number than the amounts used for whole genome sequencing. Presented in Fig 3 and S3 are 5ng and 10ng genomic DNA input amounts. In cases of low genomic input, smMIP libraries did not amplify enough to be easily size-separated by acrylamide gel. Instead, libraries were size-selected using a two-step Ampure bead purification process. PCR-amplified libraries were pooled in equivolume amounts to a total of N μl (N=200-400) before adding 0.5N μl of Ampure bead solution. Samples were gently mixed and beads were separated on magnetic racks. The supernatant was then transferred to tubes with 0.4N μl bead solution. Beads were mixed and cleaned with 80% ethanol, then briefly dried before eluting in 25-50μl EB.

**smMIP analysis**

Analysis of smMIP reads was completed using a custom R-script modified from (Mok *et al.* 2017). Briefly, for rDNA calculation in each library, MIPS were sorted and analyzed separately based on the following formula:

*rDNA estimate =  ×* 

Although four single copy loci were present in the normalization plasmid, only the two most consistently-behaving smMIPs were used in the presented analysis (VC20019-MIPS-IV-4 and VC20019-MIPS-V-16b, Table S14). The exception was strain JU775, for which only one single copy probe locus was used (VC20019-MIPS-IV-4) because the VC20019-MIPS-V-16b target genomic locus is deleted in JU775.

**Supplemental Tables**

Supplemental tables are available on the GSA FigShare platform.

|  |  |
| --- | --- |
| Table S1 | Worm WGS metrics |
| Table S2 | Worm Percent Error |
| Table S3 | Worm Downsampling WGS |
| Table S4 | Aligner comparison |
| Table S5 | Worm Single Copy Region Coordinates |
| Table S6 | Worm Single Copy Region Counts |
| Table S7 | Yeast Strain SRA Numbers |
| Table S8 | Yeast Diploid Strain rDNA Estimates |
| Table S9 | Yeast Strain SNV Genotyping |
| Table S10 | ddPCR Raw Data |
| Table S11 | smMIP Data |
| Table S12 | FISH Summary Cell Counts |
| Table S13 | FISH Data |
| Table S14 | Oligo Sequences |
| Table S15 | CHEF Measurements |

**REFERENCES**

Andrews, S., 2010 FastQC A Quality Control tool for High Throughput Sequence Data.

Benjamini, Y., and T. P. Speed, 2012 Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res. 40: e72–e72.

Hiatt, J. B., C. C. Pritchard, S. J. Salipante, B. J. O’Roak, and J. Shendure, 2013 Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome Res. 23: 843–854.

Krueger, F., 2015 Trim Galore.

Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nat. Methods 9: 357–359.

Langmead, B., C. Wilks, V. Antonescu, and R. Charles, 2019 Scaling read aligners to hundreds of threads on general-purpose processors. Bioinformatics 35: 421–432.

Li, H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arxiv.org.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.

Martin, M., 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17: 10–12.

Mok, C. A., V. Au, O. A. Thompson, M. L. Edgley, L. Gevirtzman *et al.*, 2017 MIP-MAP: High Throughput Mapping of Caenorhabditis elegans Temperature-Sensitive Mutants via Molecular Inversion Probes. Genetics genetics.300179.2017.

Okabe, M., and K. Ito, 2008 Color Universal Design (CUD): How to Make Figures and Presentations That Are Friendly to Colorblind People.

O’Roak, B. J., L. Vives, W. Fu, J. D. Egertson, I. B. Stanaway *et al.*, 2012 Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science 338: 1619–1622.

Parks, M. M., C. M. Kurylo, R. A. Dass, L. Bojmar, D. Lyden *et al.*, 2018 Variant ribosomal RNA alleles are conserved and exhibit tissue-specific expression. Sci. Adv. 4: eaao0665.

Picard Tools - By Broad Institute.

R Core Team, 2018 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rienzi, S. C. D., K. C. Lindstrom, T. Mann, W. S. Noble, M. K. Raghuraman *et al.*, 2012 Maintaining replication origins in the face of genomic change. Genome Res. 22: 1940–1952.

Thompson, O., M. Edgley, P. Strasbourger, S. Flibotte, B. Ewing *et al.*, 2013 The million mutation project: A new approach to genetics in Caenorhabditis elegans. Genome Res. 23: 1749–1762.

Tsuchiyama, S., E. Kwan, W. Dang, A. Bedalov, and B. K. Kennedy, 2013 Sirtuins in yeast: phenotypes and tools. Methods Mol. Biol. Clifton NJ 1077: 11–37.

Turner, E. H., S. B. Ng, D. A. Nickerson, and J. Shendure, 2009 Methods for Genomic Partitioning. Annu. Rev. Genomics Hum. Genet. 10: 263–284.

Wickham, H., 2009 *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.

Zeileis, A., J. C. Fisher, K. Hornik, R. Ihaka, C. D. McWhite *et al.*, 2019 colorspace: A Toolbox for Manipulating and Assessing Colors and Palettes.