

RESEARCH ARTICLE SUMMARY

EPIGENETICS

Sequence-dependent activity and compartmentalization of foreign DNA in a eukaryotic nucleus

Léa Meneu[†], Christophe Chapard[†], Jacques Serizay^{†*}, Alex Westbrook, Etienne Routhier, Myriam Ruault, Manon Perrot, Alexandros Minakakis, Fabien Girard, Amaury Bignaud, Antoine Even, Géraldine Gourgues, Domenico Libri, Carole Lartigue, Aurèle Piazza, Agnès Thierry, Angela Taddei, Frédéric Beckouët, Julien Mozziconacci*, Romain Koszul*

INTRODUCTION: The composition of genomic sequences, such as GC content, nucleotide motifs, and repeats, varies from one species to another and within the same genome. Composition correlates with gene transcriptional activity and chromosome organization, and all genome sequences have coevolved with the chromatin-associated complexes they encode to precisely regulate these two features. However, when foreign DNA—including exogenous mobile elements and natural or artificial genes—invades or integrates a host nucleus, it encounters regulatory mechanisms and rules under which it has not evolved. How host cells process and eventually adopt these unfamiliar exogenous sequences remains largely unexplored.

RATIONALE: We investigated chromatin assembly on foreign DNA sequences by integrating two types of DNA into *Saccharomyces cerevisiae*: megabase-long bacterial chromosomes and shorter eukaryotic sequences. We profiled chromatin composition, activity, and folding in these foreign DNA sequences and investigated the molecular mechanisms governing their adaptation to the host environment. Finally, we used machine learning models to identify sequence features that influence chromatin formation and activity on these integrated DNA molecules.

RESULTS: We found that bacterial DNA sequences form chromatin and are covered by nucleosomes, but that their transcriptional activity depends on sequence composition. Exogenous sequences with a GC content mirroring the host sequence become actively transcribed and intermingle with transcriptionally active yeast chromosomes in the nuclear space. Transcription of these sequences follows the original bacterial gene orientation, suggesting an intrinsic directional bias that may facilitate their adaptation to the host environment. By contrast, AT-rich chromosomes adopt an inactive chromatin state characterized by longer nucleosome spacing and a globular tridimensional shape segregated from yeast chromosomes. These two chromatin states spontaneously emerged, partitioning the hybrid genome into spatially distinct nuclear compartments that mirror the euchromatin-heterochromatin compartmentalization of metazoans. Transcriptional inactivation and compartmentalization occur independently of both canonical H3K9me3 mediated heterochromatin—absent in *S. cerevisiae*—and of yeast SIR-mediated heterochromatin. Compartmentalization disappears when transcription is reduced, suggesting that this phenomenon may involve mechanical constraints or biochemical changes in the environment of

active genes. During mitosis, compartmentalization is lost as chromatin is compacted by cohesin-mediated loops, paralleling the transient compartment disruption observed in metazoans at this stage.

Deep learning models trained only on yeast chromosomes accurately predict chromatin composition and activity on foreign DNA based on the DNA sequence alone. This implies that the behavior of any DNA in a host cell follows deterministic sequence-based rules, which could help to predict the behavior of exogenous DNA not only during natural gene transfer events but also in synthetic genome engineering projects.

CONCLUSION: The spontaneous formation of transcriptionally active and inactive chromatin compartments in a eukaryotic genome lacking such structures provides insights into their emergence during evolution. When foreign DNA sequences are integrated into the yeast genome, their underlying sequence composition determines their chromatin state and ability to recruit the host transcriptional machinery. These predictable sequence-specific properties highlight the fundamental links between DNA sequence composition, chromatin organization, and nuclear architecture. Such sequence-dependent features could have contributed to the emergence of bipartite compartmentalization of genomes during evolution, either by sequestering inactive genetic material or by facilitating the functional integration of active sequences, thus potentially creating reservoirs of genetic novelty. ■

The list of authors and their affiliations is available in the full article online.

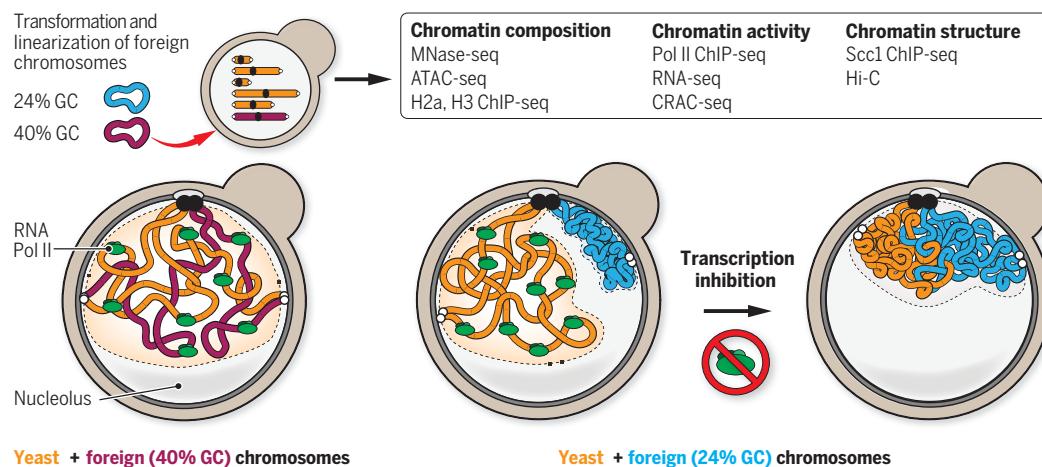
*Corresponding author. Email: romain.koszul@pasteur.fr (R.K.); julien.mozziconacci@mnhn.fr (J.M.); jacques.serizay@pasteur.fr (J.S.)

Cite this article as L. Meneu et al., *Science* **387**, eadm9466 (2025). DOI: 10.1126/science.adm9466

S READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.adm9466>

Different fates of foreign chromosomes in a yeast nucleus

nucleus. Bacterial or eukaryotic chromosomes with a GC content similar to yeast sequences (in purple) become actively transcribed and intermingle with host chromosomes. By contrast, foreign chromosomes with a lower GC (in blue) form a distinct, transcriptionally silent globular compartment within the nucleus. This segregation requires active transcription but is independent of canonical heterochromatin formation.



RESEARCH ARTICLE

EPIGENETICS

Sequence-dependent activity and compartmentalization of foreign DNA in a eukaryotic nucleus

Léa Meneu^{1,2†}, Christophe Chapard^{1†‡}, Jacques Serizay^{1*†}, Alex Westbrook^{2,3}, Etienne Routhier^{2,3,4}, Myriam Ruault⁵, Manon Perrot^{1,2}, Alexandros Minakakis⁶, Fabien Girard¹, Amaury Bignaud^{1,2}, Antoine Even⁵, Géraldine Gourguet⁷, Domenico Libri⁶, Carole Lartigue⁷, Aurèle Piazza^{1§}, Agnès Thierry¹, Angela Tadel⁵, Frédéric Beckouët⁸, Julien Mozziconacci^{3,4,9*}, Romain Koszul^{1*}

In eukaryotes, DNA-associated protein complexes coevolve with genomic sequences to orchestrate chromatin folding. We investigate the relationship between DNA sequence and the spontaneous loading and activity of chromatin components in the absence of coevolution. Using bacterial genomes integrated into *Saccharomyces cerevisiae*, which diverged from yeast more than 2 billion years ago, we show that nucleosomes, cohesins, and associated transcriptional machinery can lead to the formation of two different chromatin archetypes, one transcribed and the other silent, independently of heterochromatin formation. These two archetypes also form on eukaryotic exogenous sequences, depend on sequence composition, and can be predicted using neural networks trained on the native genome. They do not mix in the nucleus, leading to a bipartite nuclear compartmentalization, reminiscent of the organization of vertebrate nuclei.

Genome sequence composition, broadly defined by its content in GC base pairs, polynucleotide frequencies, and DNA motifs and repeats, varies considerably between species and within individual genomes (1, 2). In eukaryotes, sequence composition is known to correlate with (i) chromatin composition, which includes nucleosome formation and the binding of structural and functional proteins to DNA (3); (ii) chromatin activity, such as transcription and replication (4); and (iii) the three-dimensional (3D) functional organization of the genome into loops and compartments (5, 6). For instance in mammals, GC-rich regions are enriched in actively transcribed sequences and in chroma-

tin loops mediated by cohesin—a structural maintenance of chromosomes complex—and coalesce into a specific compartment (7). These relationships between sequence and chromatin composition, activity, and folding reflect their continuous coevolution over time.

Disruptive variations in sequence composition can occur naturally during evolution, for example when genetic material is transferred from one species to another by horizontal gene transfer or introgression, or in the case of viral infection (8–11). They can also emerge artificially—e.g., by introducing long DNA molecules into chassis microbial strains or cell lines (12–14). Such transfer can lead to the long-term integration of foreign DNA whose sequence composition diverges sharply from that of the host genome (e.g., the introgression in *Lachancea kluyveri* of a 1-Mb sequence with GC content 12% higher than that of the rest of the genome) (15). Once integrated, these sequences are organized and processed by host genome chromatin-associated proteins, obeying new rules under which they have not coevolved. How the eukaryotic host packages and regulates the activity of long exogenous DNA sequences and folds them, as well as the importance of the sequence composition in this process, remain largely unknown.

We investigated the behavior of natural, chromosome-sized bacterial and eukaryotic sequences with different sequence composition, artificially introduced into the *S. cerevisiae* genome (16, 17). We profiled nucleosome, RNA polymerase II, and G2/M cohesin landscapes, transcriptional activity, and 3D organization

of these supernumerary chromosomes during the cell cycle. We found that highly divergent bacterial or eukaryotic chromosomes, with different GC content, possessed different chromatin composition and activities. This eventually led to the spontaneous formation of two chromatin archetypes, one active and one inactive, each with different physical properties and segregating into distinct chromosomal compartments, reminiscent of those observed in multicellular organisms such as mammals. This partitioning occurred independently of heterochromatin formation but was driven by transcriptional activity. Sequence determinants, computationally learned on yeast sequences, were sufficient to predict the chromatin composition and activity of exogenous chromosomes integrated into the yeast nuclear environment. This result suggests that the behavior of a DNA molecule introduced into a given cellular context, from nucleosome positioning to 3D folding and transcriptional activity, is governed by deterministic and predictable rules.

Adaptation of supernumerary bacteria chromosomes integrated into yeast

To investigate large sequences that have not evolved in a eukaryotic context, we exploited *S. cerevisiae* strains carrying an extra 17th circular chromosome from either the *Mycoplasma mycoides* subspecies *mycoides* (referred to as “Mmyco”) or *Mycoplasma pneumoniae* (“Mpneumo”), containing a yeast centromeric sequence and an autonomous replication sequence (18) (see methods and table S1). The GC content of *S. cerevisiae* is 38%, whereas the GC content of the Mmyco chromosome is 24% (GC-poor) and the GC content of the Mpneumo chromosome is 40% (GC-neutral) (Fig. 1A). Dinucleotide composition is also different between yeast and bacterial chromosomes (19) (fig. S1A). We linearized these chromosomes, added yeast telomeres to their ends (see methods and Fig. 1A and fig. S1B), and studied their replication using marker frequency analysis (see methods). Mpneumo replicated from multiple replication origins firing in early to mid-S phase, whereas Mmyco appeared to replicate later and did not display discrete origin firing (see supplementary text and fig. S1C). DNA replication and sister chromatid cohesion (SCC) are closely linked, as cohesion is established during S phase by the trapping of sister DNA molecules by cohesin as the replication fork progresses (20). We investigated SCC by image analysis of chromosome pairing (methods and fig. S1D). Mpneumo SCC was similar to that of yeast chromosomes whereas Mmyco SCC appeared substantially reduced (see supplementary text). Overall, exogenous bacterial chromosomes did not impose a substantial fitness cost on their eukaryotic host and had a segregation rate similar to that of a centromeric plasmid (fig. S1, E and F), making them

¹Institut Pasteur, CNRS UMR 3525, Université Paris Cité, Unité Régulation Spatiale des Génomes, Paris, France.

²Sorbonne Université, College Doctoral, ³Laboratoire Structure et Instabilité des génomes, UMR 7196, Muséum National d'Histoire Naturelle, Paris, France.

⁴Laboratoire de Physique Théorique de la Matière Condensée, Sorbonne Université, CNRS, Paris, France.

⁵Institut Curie, PSL University, Sorbonne Université, CNRS UMR 3664 Nuclear Dynamics, Paris, France.

⁶Institut de Génétique Moléculaire de Montpellier, Univ Montpellier, CNRS, Montpellier, France.

⁷Univ. Bordeaux, INRAE, Biologie du Fruit et Pathologie, UMR 1332, Villenave d'Ornon, France.

⁸Molecular, Cellular, and Developmental biology unit (MCD), Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, UPS, Toulouse, France.

⁹UAR 2700 2AD, Muséum National d'Histoire Naturelle, Paris, France.

*Corresponding author. Email: roman.koszul@pasteur.fr (R.K.); julien.mozziconacci@mnhn.fr (J.M.); jacques.serizay@pasteur.fr (J.S.)

†These authors contributed equally to this work.

‡Present address: Molecular, Cellular and Developmental biology department (MCD), Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, UPS, Toulouse, France.

§Present address: Univ Lyon, ENS, UCBL, CNRS, INSERM,

Laboratory of Biology and Modelling of the Cell, Lyon, France.

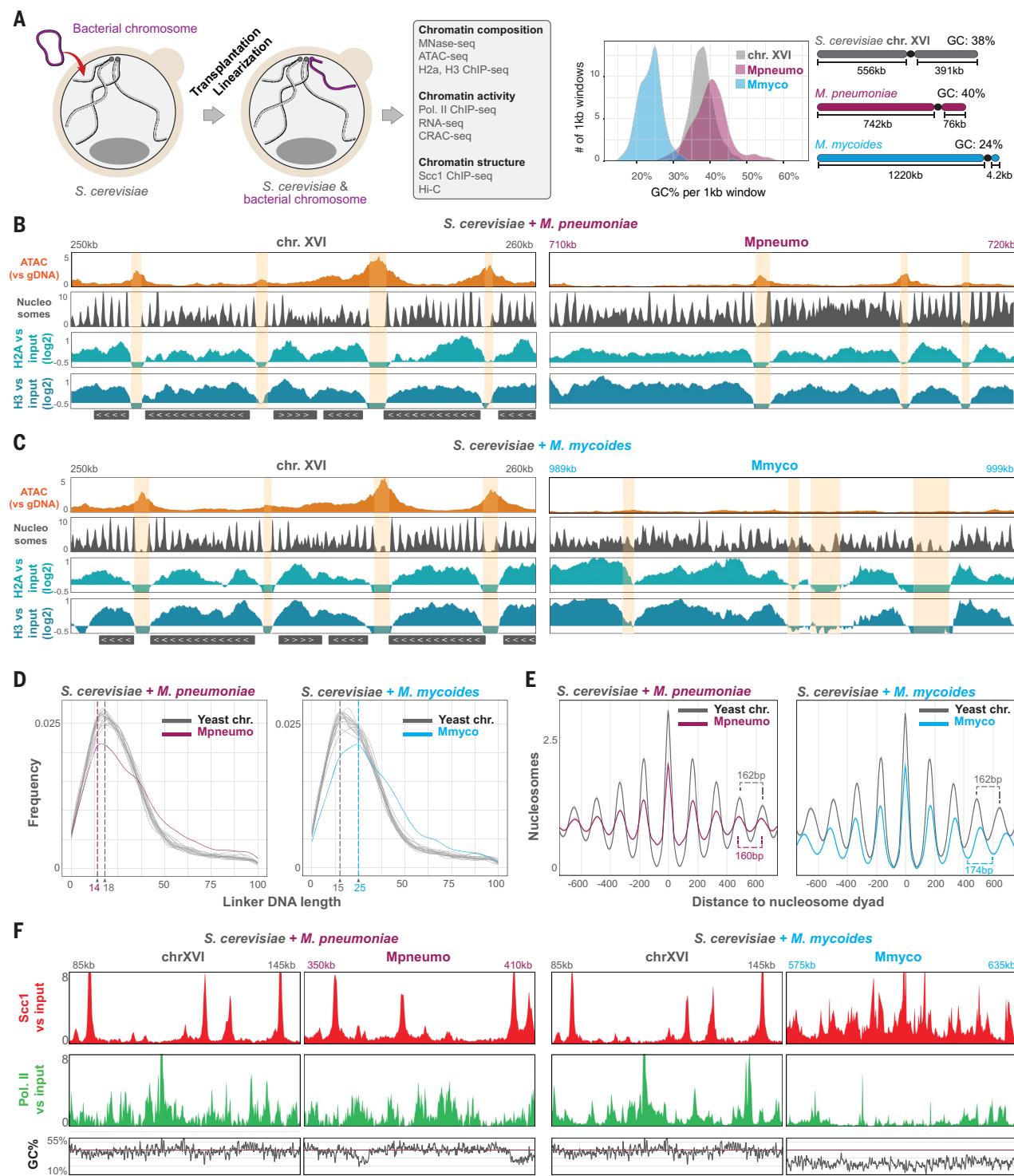


Fig. 1. Chromatin composition of bacterial chromosomes integrated in yeast. (A) Schematic of the conversion from circular to linear chromosomes integrated in yeast. Purple and blue colors represent the *M. pneumoniae* (Mpneumo) and *M. mycoides* (Mmyco) bacterial sequences in all figures, respectively. (Right) distribution of GC percentage for 1-kb windows over yeast chromosome XVI, *M. pneumoniae*, and *M. mycoides* chromosomes. (B) ATAC-seq (orange, CPM), nucleosomal track (gray; see methods) and H2A and H3 ChIP-seq (shades of blue, IP versus input, log2) profiles obtained in the Mpneumo strain (*S. cerevisiae* + *M. pneumoniae*). 10-kb-long genomic windows from

chromosome XVI [(left) 250 to 260 kb] and the Mpneumo chromosome [(right) 710 to 720 kb] are shown at the same scale. Nucleosome-depleted regions are highlighted in yellow. (C) Same as (B) in the Mmyco strain (*S. cerevisiae* + *M. mycoides*). 10-kb-long genomic windows from chromosome XVI (left, 250 to 260 kb) and the Mmyco chromosome (right, 989 to 999 kb) are shown at the same scale. (D) Frequency of nucleosome linker DNA length in Mpneumo and Mmyco strains. For each strain, the distribution is calculated for each chromosome separately. The dashed and dotted lines indicate 14 bp and 25 bp, respectively. The x-axis only shows linker DNA lengths in the 0- to 100-bp range. (E) Nucleosomal

track centered on nucleosome peaks for *S. cerevisiae*, Mpneumo and Mmyco. The y-axis represents the average nucleosomal track (130- to 165-bp MNase-seq fragments, resized to 40 bp, piled up and normalized to sequencing depth). (F) Scc1 (red) and RNA Pol II (green) ChIP-seq profiles (IP versus input, log2) obtained in

the Mpneumo (left) or in the Mmyco (right) strains. For each strain, 60-kb-long genomic windows from the bacterial chromosome and yeast chromosome XVI are shown at the same scale. GC percentage in sliding 1-kb windows is shown below the ChIP-seq profiles.

powerful tools for studying chromatin assembly on foreign sequences.

Spontaneous chromatinization of bacterial chromosomes in a eukaryotic context

Although Mmyco and Mpneumo chromosomes evolved in a prokaryotic context, both sequences formed nucleosome arrays and local nucleosome-depleted regions (NDR) in yeast nuclei as assessed by time course MNase-seq, and H3 and H2A ChIP-seq experiments (Fig. 1, B and C, and fig. S2, A and B). We annotated nucleosome positions following an approach adapted from Brogaard *et al.* (21) and found that nucleosome arrays along the Mpneumo chromosome were similar to those of the yeast chromosome, with a linker length of ~14 to 18 base pairs (bp) and a nucleosome repeat length (NRL) of 160 bp (Fig. 1, D and E). By contrast, nucleosome arrays on the Mmyco chromosome had a longer (~25 bp) linker and an NRL of 174 bp (Fig. 1, D and E, and fig. S2D). Sequence features favoring nucleosome organization [short poly(dA) or poly(dT) tracks and strong WW 10-bp periodicity, where W is either A or T base] were enriched in Mpneumo, while sequence features limiting nucleosome assembly [long poly (dA) or poly(dT) tracks and weak WW 10-bp periodicity] were enriched in the Mmyco chromosome (fig. S2, C and D).

Chromatin accessibility profiling by ATAC-seq confirmed that the Mpneumo chromosome had a density and breadth of accessible loci similar to those of yeast chromosomes and that these accessible sites were devoid of nucleosomes, thus forming NDRs (Fig. 1B and fig. S3, A, B, and D to E). In comparison, we detected only 16 weaker ATAC peaks on the Mmyco chromosome, which harbored persistent nucleosomes (Fig. 1C and fig. S3, C to E).

Cohesin (Scc1) in G2/M and RNA polymerase II (Pol II) profiles along the Mpneumo chromosome also appeared similar to that of the yeast chromosomes, with discrete Scc1 peaks preferentially located in nucleosome-depleted and Pol II-enriched regions (Fig. 1F and fig. S2, F to H). Scc1 and Pol II binding profiles in the Mmyco strain were different. Although broadly enriched along the entire Mmyco chromosome, Scc1 was not enriched locally at individual loci, as observed along G2/M yeast chromosomes (Fig. 1F and fig. S2, F and G). This broad enrichment of Scc1 on the Mmyco chromosome was accompanied by a strong reduction in Scc1 occupancy at the centromeres of *S. cerevisiae* chromosomes (fig. S2E). This result suggests that the Mmyco chromosome was competing with yeast centromeres for the recruitment of

the dynamic pool of cohesin enriched at these locations (22, 23). Pol II occupancy was, on the other hand, greatly reduced along the Mmyco chromosome compared with yeast chromosomes (Fig. 1F).

These results show that a large exogenous bacterial chromosome placed in a eukaryotic context spontaneously adopts eukaryotic chromatin features, since histones, Pol II, and cohesins can bind bacterial DNA. However, the chromatin landscapes of the two bacterial chromosomes delineate two different chromatin archetypes: (i) The yeast-like (Y) chromatin landscape, found on the Mpneumo chromosome (whose 40% GC content is close to that of *S. cerevisiae*), and (ii) the unconventional (U) chromatin, found on the Mmyco chromosome (with a low GC content of 24%), featuring less packed nucleosomes, reduced Pol II coverage, and broad binding of cohesins across the entire chromosome in G2/M.

Transcriptional activity of bacterial genomes in a yeast context

Consistent with Pol II ChIP-seq profiles (Fig. 1F), we found that the Y chromatin type on the Mpneumo chromosome was transcribed at levels similar to those of yeast chromosomes (Fig. 2A). Transcription tracks on the Mpneumo chromosome were significantly longer than that of yeast genes (4.9 kb versus 3.4 kb, P -value $< 2e^{-4}$, two-sided Student's *t* test) and did not consistently show clear boundaries, with more loci where both strands were transcribed compared with yeast (Fig. 2B, black arrowheads). Transcription was not specifically increased over bacterial gene bodies nor initiated at bacterial promoters, consistent with the lack of Pol II enrichment at these loci (fig. S4, A to C). By contrast, the Mmyco U chromatin type was only sparsely and weakly transcribed (Fig. 2, A and B), consistent with the reduced levels of Pol II deposition (Fig. 1F).

For the two bacterial chromosomes integrated in yeast, we found that strand-specific transcriptional coverage preferentially followed the orientation of annotated bacterial genes along these chromosomes, despite the fact that bacterial and eukaryotic transcription machineries evolved independently over billions of years (Fig. 2C, see gene orientation and strand-specific RNA-seq tracks in Fig. 2, A and B). This preference correlates with the overrepresentation of A versus T and G versus C on the coding strand of both yeast and bacterial genes (see discussion and fig. S4D). To assess whether this apparent strand-specific transcriptional orientation was the result of active degradation of overlapping

antisense transcripts, we profiled nascent transcripts by CRAC-seq, which quantifies RNA molecules still bound to Pol II (fig. S5A) (24). If post transcriptional degradation was responsible for strand-specific orientation, then one would expect CRAC-seq reads to be enriched on both strands irrespective of orientation. This was not the case, with CRAC-seq reads correlating strongly with strand specific orientation of RNA-seq reads for both the Mpneumo and Mmyco chromosomes (fig. S5B). We validated these observations by performing RNA-seq in *Δupf1*, *Δrrp6*, or *Δupf1/rrp6* Mpneumo strains, respectively, incompetent for nonsense-mediated RNA decay (NMD), nuclear RNA surveillance by the exosome, or both. This showed that, although some overlapping antisense transcripts emerged sporadically in mutant strains, transcription on a bacterial chromosome in yeast cells followed the direction of bacterial genes (fig. S5C). The CRAC-seq experiment further confirmed that transcription could initiate at the hundreds of NDRs accessible along the Mpneumo chromosome (Fig. 2, D and E) and that transcription was extremely sparse over the Mmyco chromosome (fig. S5A).

Inactive U chromatin forms a distinct compartment in the interphase nucleus

High-throughput chromatin conformation capture (Hi-C) contact maps showed that the centromeres added to the Mmyco and Mpneumo chromosomes clustered with those from the yeast chromosomes (Fig. 3A), as expected by the nuclear organization of budding yeast (25, 26). However, the two bacterial chromosomes exhibited distinct structural characteristics. The Y-type Mpneumo chromosome behaved similarly to an endogenous *S. cerevisiae* chromosome (Fig. 3A), with a comparable frequency of interchromosomal contact (fig. S6A), and a slope of the p(s) curve [i.e., the contact probability (p) as a function of the genomic distance (s), also called genomic contact decay curve, close to -1.5] (Fig. 3C). This value is consistent with a typical random coil structure observed in simulations (27). Fluorescent in situ hybridization (FISH) labeling of Mpneumo DNA also revealed an extended, convoluted structure in the intranuclear space, suggesting mixing of the bacterial chromosome with the actively transcribed yeast chromosomes (Fig. 3E and fig. S6, C and D).

By contrast, few contacts were detected between the U-type Mmyco chromosome and the yeast chromosomes (Fig. 3A and fig. S6A). In addition, these contacts predominantly occurred with the other 32 yeast subtelomeric

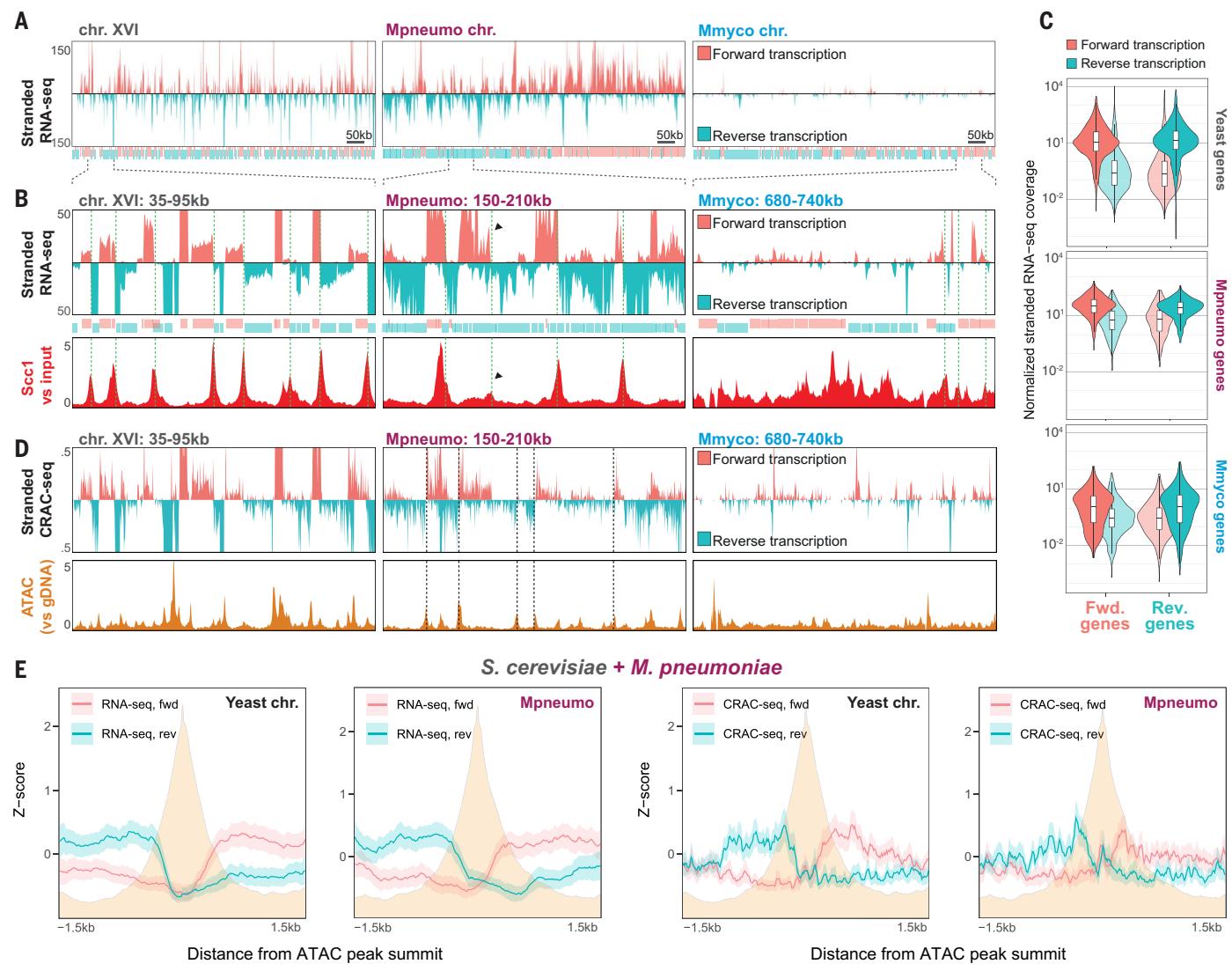


Fig. 2. Transcriptional activity of exogenic bacterial sequences in budding yeast. (A) Stranded RNA-seq profiles along either yeast chromosome XVI or bacterial chromosomes Mpneumo and Mmyco. Forward (pink) and reverse (turquoise) genes along yeast or bacterial sequences are indicated as transparent segments under the tracks. (B) (Top) stranded RNA-seq profiles along a 60-kb window along yeast chromosome XVI, Mpneumo, or Mmyco. (Bottom) Scc1 (cohesin) deposition profiles of the corresponding loci. Green dotted lines indicate identified loci of convergent transcription (see methods).

Black arrowheads indicate overlapping bidirectional transcription. (C) Forward and reverse RNA-seq coverage of forward- and reverse-oriented yeast or bacterial genes. Scores are normalized by each genomic feature length.

(D) (Top) stranded CRAC-seq profiles along the region shown in (B) (Bottom) Corresponding ATAC-seq profile. Dotted lines point at sites of bidirectional initiation along Mpneumo. (E) Metaplot of 3-kb regions centered on ATAC peaks over yeast or Mpneumo chromosomes, of stranded RNA-seq (left panels), or CRAC-seq (right panels) profiles.

regions (Fig. 3A, dotted rectangles; fig. S6B). The slope of the p(s) curve also differed substantially with a value of -1 at shorter distances, corresponding to a crumpled globule (Fig. 3C) (28, 29). DNA-FISH labeling of the Mmyco DNA confirmed the globular conformation of the chromosome and revealed that it was preferentially positioned at the nuclear periphery, where yeast telomeres were also located (Fig. 3E and fig. S6, C and D) (30). These results show that inactive, U chromatin spontaneously forms a distinct compartment in the nuclear space segregated from the active endogenous yeast chromosomes.

Cohesin compacts both exogenous bacterial chromosomes in G2/M

In yeast cells arrested at the G2/M stage of the cell cycle, chromatin is folded into arrays of cohesin-anchored loops (31, 32). This compaction leads to a reduction in interchromosomal contacts and in a shifted p(s) curve, with an inflection point around the average loop length (25, 33). These structural features were observed for both bacterial chromosomes, indicating similar mitotic compaction of Y and U chromatin types (Fig. 3, B and D, and fig. S6, A, E, and F). We identified 59 loops across the Mpneumo chromosome in G2/M, bringing

together pairs of loci located on average 28 kb apart, a distance slightly longer than that of the loops found along the yeast chromosomes (22 kb) (see methods and fig. S6G). Strong Scc1 peaks were positioned close (within 500 bp) to the loop anchors (Fig. 3, F and H), at sites of convergent transcription (Fig. 2B, green dotted lines; fig. S6H), with strong cohesin peaks found at genomic loci displaying stronger transcriptional convergence (Fig. 2B, arrowheads; fig. S6, I and J). The dotted grid pattern in the Mpneumo chromosome was reminiscent of the multiple DNA loops observed along mammalian interphase chromosomes (34), suggesting

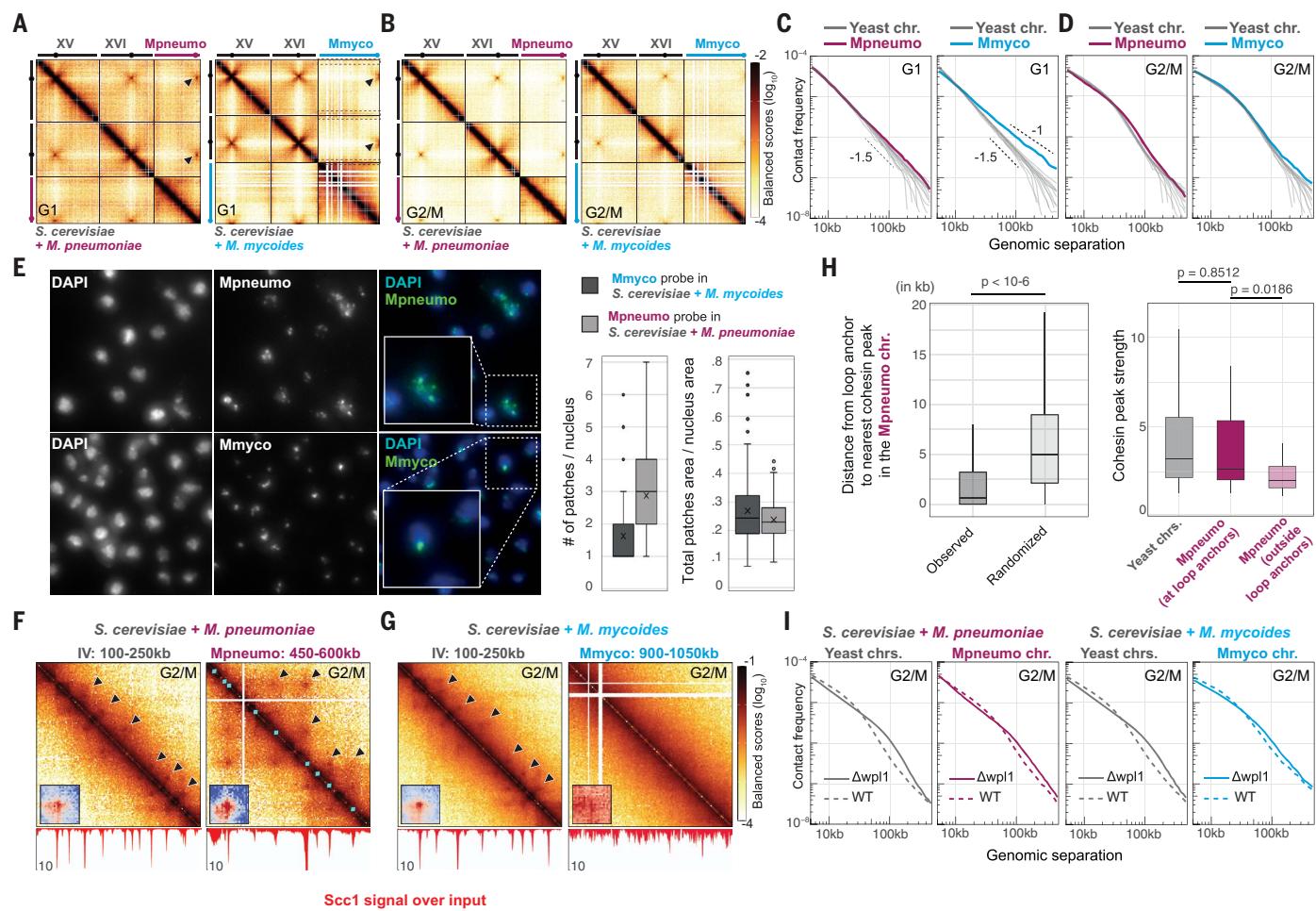


Fig. 3. Spatial folding of exogenous bacterial sequences within the yeast nucleus.

(A) and (B) Hi-C contact maps of representative endogenous and of Mmyco and Mpneumo bacterial chromosomes in G1 (A) and G2/M (B) (4-kb resolution). (C) and (D) Contact frequency (p) as a function of genomic distance (s) plots of endogenous yeast chromosomes (long arms) and of Mmyco and Mpneumo bacterial chromosomes in G1 (C) and G2/M (D). (E) (Left) FISH imaging. Representative field of either (top) Mpneumo or (bottom) Mmyco fixed cells labeled with DAPI (left panel) and hybridized with a fluorescent probe generated from either the Mpneumo or Mmyco chromosome, respectively. (Right) For each probe, number of patches detected per nucleus and surface occupied by these patches relative to the whole nucleus surface (see methods). (F) and (G) (Top) magnification of 150-kb windows from Hi-C contact maps in G2/M from

either an endogenous chromosome or the bacterial chromosome in Mpneumo (F) and Mmyco (G) strains (1-kb resolution). (Bottom) Scc1 ChIP-seq deposition profile. Black arrowheads indicate loops. Cyan diamonds in the Mpneumo Hi-C contact map indicate Scc1 peaks positions reported on the contact map diagonal. Inset in each map shows Chromosight pileup of contacts between cohesin enrichment peaks along either *S. cerevisiae* or bacterial chromosomes (see methods). (H) (Left) distance between chromatin loop anchors and the nearest Scc1 peak in Mpneumo (with and without a random shuffle of peak positions). (Right) Scc1 peak strengths in yeast or Mpneumo chromosome, near (< 1 kb) or outside loop anchors (P -values from two-sided Student's t -test). (I) Distance-dependent contact frequency in endogenous yeast and in bacterial chromosomes (left, Mpneumo; right, Mmyco), in the wild type (dashed) and in $\Delta wpl1$ mutants.

that some loop anchors were involved in loops of different sizes, a phenomenon not observed across *S. cerevisiae* chromosomes (Fig. 3F).

On the other hand, no visible discrete loops were observed along the Mmyco chromosome where cohesins uniformly covered inactive U chromatin regions without clear peaks, probably reflecting the absence of transcription (Figs. 3G, 1F, and 2B). Cohesin-mediated loops along yeast chromosomes are formed through loop extrusion, a process by which cohesin complexes organize DNA by capturing and gradually enlarging small loops (20, 35). In absence of the cohesin unloading factor Wpl1

(Wapl in human), the $p(s)$ curve is shifted toward long-range contacts reflecting the formation of longer loops in yeast and human (31, 36). The depletion of Wpl1 also led to shifted $p(s)$ for both Mmyco and Mpneumo chromosomes (Fig. 3I and fig. S6K), suggesting that active cohesin-mediated loop extrusion can proceed on both Y and U chromatin.

Mosaic chromosomes display spontaneous, chromatin type-dependent DNA compartmentalization

The structural and functional features of Y and U-type chromatin are reminiscent of the

euchromatin and heterochromatin compartments described along metazoan chromosomes (37). To test whether and how they can coexist on a single chromosome, we fused the Mmyco chromosome with yeast chromosome XVI (XVIIfMmyco chromosome) and induced translocations to generate two additional strains with chromosomes harboring regions of alternating U and Y chromatin over hundreds of kilobases (kb) (XVIIfMmycot1), 50 kb (XVIIfMmycot2), or as small as 15 kb (XVIIfMmycot2') (Fig. 4A, and fig. S7, A and B, and methods). Hi-C contact maps of XVIIfMmyco strain synchronized in G1 showed little difference compared

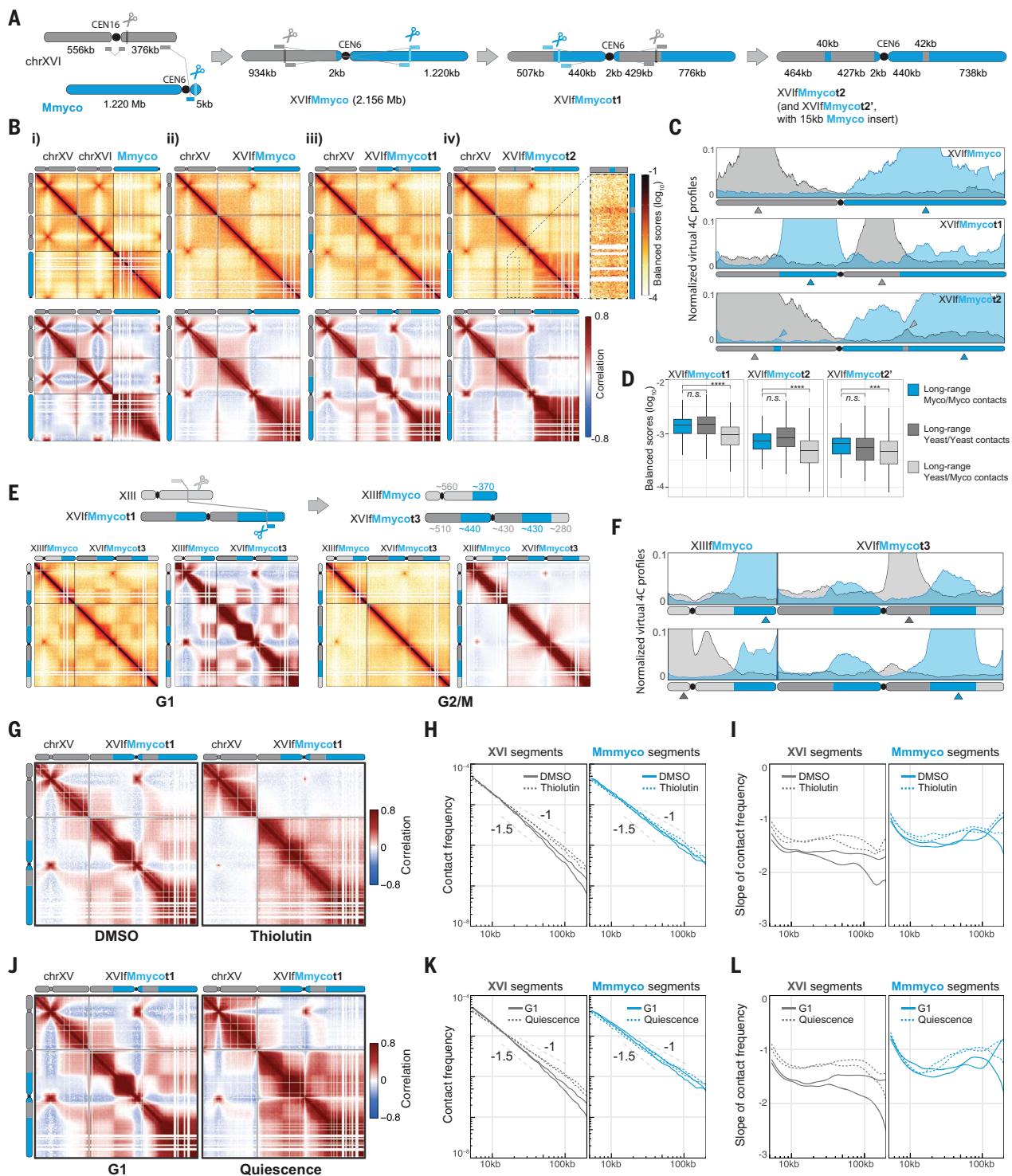


Fig. 4. Compartmentalization of mosaic chromosomes composed of Y- and U-type chromatin. (A) Schematic representation of the CRISPR strategy used to generate the Mmyco mosaic chromosomes with alternating Y and U chromatin regions. (B) (Top) G1 Hi-C contact maps of chromosomes XV and XVI in addition to bacterial chromosomes in the Mmyco strain, and in translocated derivatives from (A) (4-kb resolution). (Bottom) correlation matrices of the corresponding contact maps. (C) Virtual 4C profiles of viewpoints (indicated as gray/blue arrowheads) located within yeast segments (gray) or Mmyco chromosome segments (blue) of the chimeric chromosome XVI, in XVIIfMmyco (top), XVIIfMmycot1 (middle), and XVIIfMmycot2 (bottom) strains. (D) Quantification of long-range interactions (800 kb to 1.5 Mb) between pairs of distant Mmyco segments (blue), pairs of distant yeast segments (gray), and interactions between a Mmyco segment and a yeast segment (light gray), in double-translocation XVIIfMmyco (top), XVIIfMmycot1 (middle), and XVIIfMmycot2 (bottom) strains. (E) (Left) Schematic representation of the CRISPR strategy used to generate the mosaic XVIIfMmycot3 strain with alternating yeast and Mmyco segments in chromosomes XIII and XVI. (Middle) G1 Hi-C maps of mosaic chromosomes XIII and XVI in XVIIfMmycot3 strain. Color scale same as in (B). (F) Virtual 4C profiles of viewpoints located within yeast segments (gray arrowheads) or Mmyco segments (blue arrowheads) of the chimeric

chromosome XVIIfMmycot3 strain. (G) G1 Hi-C contact maps of chromosomes XV and XVI in the Mmyco strain (left) and the XVIIfMmycot1 strain (right) under DMSO and Thiolutin treatment. (H) Contact frequency and (I) slope of contact frequency vs. genomic distance (10kb to 1.5 Mb) for XVI segments (left) and Mmmyco segments (right). DMSO (solid lines) and Thiolutin (dotted lines) are compared. (J) G1 Hi-C contact maps of chromosomes XV and XVI in the Mmyco strain (left) and the XVIIfMmycot1 strain (right) under Quiescence. (K) Contact frequency and (L) slope of contact frequency vs. genomic distance (10kb to 1.5 Mb) for XVI segments (left) and Mmmyco segments (right). G1 (solid lines) and Quiescence (dotted lines) are compared.

chromosomes XIII and XVI, in the XVIIfMmycot3 strain. (G) Correlation matrices of the contacts in chromosomes XV and XVIIfMmycot1 in G1, after the addition of DMSO (left) or thiolutin (right). Color scale same as in (B). (H) Contact frequency (p) as a function of genomic distance (s), for contacts in yeast segments (gray) or Mmyco segments (blue) of the chimeric chromosome

XVIIfMmycot1, in G1 after the addition of DMSO (solid) or thiolutin (dotted). (I) Derivatives of curves from (H). (J) Chromosomes XV and XVIIfMmycot1 correlation maps in G1 or in quiescence. Color scale same as in (B). (K) Same $p(s)$ as in (H), but in G1 and quiescence. (L) Derivatives of curves from (K).

with the parental Mmyco strain (except for deleted and fused regions, Fig. 4B, compare panels i and ii). The Y-type chr. XVI arm mixed with the other 15 yeast chromosomes, whereas the Mmyco U-type arm remained isolated and in contact with subtelomeric regions (Fig. 4B, panel ii). In translocated strains, the alternation of chromatin-type regions resulted in checkerboard contact patterns within XVIIfMmycot1 and XVIIfMmycot2 chromosomes (Fig. 4B, panels iii and iv), with U-type regions making specific contacts with each other over long distances, bypassing the Y-type regions found in between (Fig. 4, C and D, and fig. S7, E and G1). Similarly, the Y-type regions of chromosome XVI are also involved in specific Y-Y contacts over longer distances (Fig. 4, C and D). Contact decay profiles in intra-U and -Y regions were different in the two chromatin types, with U chromatin being prone to longer range contacts in cis (fig. S7E). Translocation of a segment of U chromatin to the end of chromosome XIII also resulted in a sharp increase in interchromosomal contacts with U chromatin segments in trans (Fig. 4, E and F), a finding that broadens our understanding of chromatin compartmentalization and that further relates Y and U chromatin to the eu- and heterochromatin compartments found in higher eukaryotes (37).

Pol II ChIP-seq in strain XVIIfMmycot2 revealed that 50-kb-long translocated Y or U chromatin segments showed no change in Pol II occupancy profiles, demonstrating that Pol II binding is independent of the wider chromatin context (fig. S7C). We also performed RNA-seq profiling in strain XVIIfMycot2', which harbored a 15-kb segment of Mmyco translocated within the euchromatic yeast chromosome XVI and vice versa. Compared with the XVIIfMmycot2 strain, the right end of this 15-kb Mmyco segment was directly adjacent to yeast chromatin (fig. S7D). In this context, we observed that transcription at the junction with yeast chromatin progressed over ~1 kb in the Mmyco segment (fig. S7D, black arrowhead). By contrast, transcription of yeast genes translocated close to the Mmyco chromatin was unaffected. This suggests that, although active transcription can spread from Y chromatin to Mmyco U chromatin, U-type chromatin locally exerts a strong inhibitory effect on the transcription initiation machinery.

In larger genomes, euchromatin and heterochromatin compartments are abolished following metaphase chromosome compaction mediated by loop extrusion (35, 38, 39). Similarly, in our chimeric strains, Hi-C maps of mosaic chromo-

somes in G2/M revealed that all the intra- and interchromosomal U-type compartments disappeared upon cohesin-mediated compaction (Fig. 4E and fig. S7F), with cohesin being over-enriched along inactive Mmyco regions irrespective of the proximity to the centromere (fig. S7C). At this stage, all chimeric chromosomal regions showed similar contact frequencies as a function of distance (fig. S7E, G2/M). In addition, chromatin loops could extend across the Y/U chromatin junction, bridging the nearest cohesin enrichment sites on each side (fig. S7G, green arrowheads) (40, 41).

Compartmentalization of U-type chromatin depends on transcriptional activity of the yeast genome

We investigated the molecular mechanisms that may be responsible for the formation of this U-type heterochromatin-like chromatin compartment. In *S. cerevisiae*, heterochromatin formation is not based on the canonical H3K9me3 modification found in most eukaryotes (42). Instead, heterochromatin is formed and maintained at telomeres, mating type, and rDNA loci by the silent information regulator (SIR) complex, which consists of the histone deacetylase Sir2 (HDAC class III); Sir3, a structural chromatin-binding protein that stabilizes deacetylated histones; and Sir4, another structural protein bridging Sir2 and Sir3 and interacting with nuclear envelope-associated proteins (43). We found that (i) H4K16ac levels were globally reduced on the Mmyco chromosome (fig. S8A); (ii) the Sir2 inhibitor nicotinamide (NAM) increased H4 acetylation levels only locally in the telomeric regions of both yeast and Mmyco chromosomes (fig. S8B); and (iii) Sir3 was not enriched along the Mmyco chromosome (fig. S8C). These observations reveal that U chromatin is hypoacetylated independently of SIR-driven telomere heterochromatinization. We then investigated whether generalized deacetylation by other histone deacetylases could lead to Mmyco heterochromatinization. After treatment with trichostatin A (TSA), an HDAC I/II inhibitor, we found that H4 acetylation levels increased globally on yeast segments but not on Mmyco segments (fig. S8D) and that the Mmyco U compartment and $p(s)$ contact decay curves were unaffected (fig. S8, F to H). This indicates that neither U-chromatin hypoacetylation nor compartmentalization depends on histone deacetylation activity.

The unusual 10-bp increase in DNA linker length in the Mmyco chromosome compared

with that of yeast (Fig. 1, D and E) suggests that Histone 1 (H1) may bind DNA linkers between consecutive nucleosomes, leading to the formation of U-type chromatin. We tested this hypothesis in an yeast H1 (Hho)-deleted mutant and found that the longer NRL, the distance-dependent contact frequency, and the Mmyco compartment remained unchanged (fig. S8, E and I to K), showing that H1 is not required for U-type chromatin formation.

Finally, we assessed whether active transcription of yeast chromatin could be responsible for segregating the inactive Mmyco chromosome into a separate compartment. We treated cells with thiolutin, an inhibitor of RNA polymerase that leads to its rapid dissociation from chromatin. ChIP-seq profiling of Pol II confirmed its dissociation from yeast chromatin after treatment (fig. S8L), and calibrated RNA-seq revealed that transcription of yeast genes was reduced overall (fig. S8M, 23 to 34% reduction in steady-state transcription). Concomitantly, Hi-C revealed that long-range interactions within the yeast genome increased and that Y/U compartments were strongly affected (Fig. 4, G to I), suggesting that chromatin compartmentalization depends on yeast transcriptional activity. To test this hypothesis under physiological conditions, we performed Hi-C analysis in quiescent yeast cells, where transcription is largely silenced (44). Hi-C revealed that the Mmyco U-type compartment disappeared upon entry into quiescence (Fig. 4, J to L), and that Mmyco chromatin segregated away from the yeast telomere hypercluster appearing in quiescence (45) (fig. S8N), supporting the independence between U-type chromatin and the telomeric compartment.

Taken together, these results show that the formation and maintenance of the Mmyco compartment are independent of histone deacetylation or linker histone H1, and that transcription is required to segregate the U and Y chromatin compartments.

Chromatin features of bacterial, eukaryotic, and random exogenous chromosomes can be predicted by their sequence composition

We investigated the extent to which intrinsic DNA sequence composition can be predictive of specific chromatin features of foreign DNA integrated into yeast. We trained convolutional neural networks (CNNs) using yeast DNA sequences from chromosomes I to XV as independent variables to predict coverage tracks of nucleosome, cohesin, and Pol II along

them (see methods and fig. S9A). These models were then validated on the sequence of yeast chromosome XVI (correlation with experimental signals of 0.63, 0.82, and 0.68 for the nucleosome, Scc1 ChIP, and Pol II ChIP predictions respectively; see methods and Fig. 5A), confirming that the genomic sequence is predictive of chromatin composition and activity in yeast (46–48).

Using our CNN models, we then predicted coverages over Mpneumo and Mmyco chromosomes and found that the rules learned from the yeast sequences were sufficient to accurately predict many experimentally characterized features on these chromosomes. The predictions recapitulated the features of Y chromatin on Mpneumo, but also the U chromatin features on Mmyco sequences, including (i) increased linker length (Fig. 5A, insert i, and Fig. 5, B and C); (ii) Scc1 over-enrichment without discrete peaks (Fig. 5A, insert ii), and (iii) lower unstructured Pol II coverage, comparable to the background signal over yeast chromosomes (Fig. 5A, insert iii).

To broaden the spectrum of exogenous DNA origin, we experimentally characterized the nucleosome coverage and spatial organization of two additional yeast artificial chromosomes (YACs) integrated into yeast, one containing two sequences from the genome of the eukaryote *Plasmodium falciparum* (91 kb and 58 kb, 18% GC content; see methods) and the other containing a 284-kb-long, 21% GC YAC from *Candidatus Phytoplasma vitis*, a non-mycoides bacterial species (fig. S10 and table S1). Exogenous AT-rich chromosomes, whether bacterial or eukaryotic, exhibited U-type chromatin characteristics, including (i) a NRL longer than that of yeast chromosomes (~182 bp) and (ii) segregation into a distinct compartment, with reduced trans-chromosomal contacts and increased contacts with yeast telomeres (fig. S10, A to E). Notably, transchromosomal contacts between the two *P. falciparum* YACs were frequent (fig. S10A), reminiscent of transchromosomal contacts between translocated Mmyco segments (Fig. 4E). We found that the increased NRL in these two foreign chromosomes was correctly predicted by the CNN model (fig. S10, F and G). The CNN model also accurately predicted nucleosome coverage on an 18-kb-long GC-rich artificial sequence with 50% GC, for which MNase-seq data in yeast have already been generated (49).

To further validate the accuracy of the CNN model, we compared Pol II coverage predictions over several natural and artificial sequences that had been integrated and characterized in yeast (fig. S10H). In particular, the human sequence HPRT1, its reverse sequence (HPRT1r) (~41% GC), and an 18-kb-long artificial sequence (~50% GC) were experimentally found to be more transcriptionally active than yeast chromosomes, when integrated in yeast (49, 50). On these

sequences, we predict a higher Pol II occupancy compared with yeast chromosomes, consistent with published experimental results (49, 50).

GC content, dinucleotide content, and more complex sequence features influence chromatin composition

To better characterize the relationship between GC content and chromatin composition, we predicted the average coverage of the nucleosome, Scc1, and Pol II on thousands of 1-kb artificial random sequences with varying GC content (see methods and Fig. 5D). We observed different behaviors as a function of GC content: (i) Nucleosome signal was highest for intermediate GC percentage (i.e., between 30 and 50%) and decreased sharply outside this range; (ii) cohesin signal decreased continuously with increasing GC%; (iii) Pol II signal was minimal up to 25% GC, constant up to 45% GC, and then increased up to 85% GC. These predictions on random sequences accurately recapitulated experimental measurements on 1-kb segments from yeast, Mmyco, and Mpneumo sequences (Fig. 5D), suggesting that these predicted variations did indeed reflect the intrinsic properties of chromatin assembled on DNA with different GC content.

Comparing the performances of CNNs with two simpler linear regression models, based on either GC percentage or dinucleotide composition as predictor variables, we found that all models accurately capture dinucleotide signatures for nucleosome, cohesin, or Pol II coverage (e.g., AG, AC, CT, CA, GT, and TG in the nucleosome) (fig. S9B), but we also found that the CNN approach performs better in predicting actual tracks (fig. S9C). Motif analysis based on the saliency computed from the trained networks also enabled us to identify some of the DNA motifs involved in sequence-dependent chromatin composition (fig. S9D). These results show that, in addition to GC percentage and dinucleotide signatures, more complex sequence features influence chromatin composition, irrespective of their evolutionary origin.

Discussion

Chromatin composition and activity of exogenous chromosomes is based on their underlying DNA sequence

Several studies have shown that exogenous or random DNA segments in the yeast nucleus with a GC percentage relatively similar to that of yeast chromosomes are actively transcribed (46, 49–51). Here, we show that this is not always the case, and that DNA sequences with different GC content adopt one of two archetypes of chromatin that we called Y and U for yeast-like and unconventional, respectively. Y chromatin displays nucleosome arrays with a canonical 160-bp NRL, recruits Pol II, and is transcriptionally active, whereas U chromatin is characterized by a longer (174 bp) NRL, and

a lower Pol II occupancy and transcriptional activity (Fig. 5E).

Using machine learning models solely trained on yeast chromosomes, we show that the chromatin composition experimentally measured over these exogenous chromosomes can be estimated from their DNA sequence only. This demonstrates that the fate of any DNA molecule introduced into a given cellular context, including its nucleosomal packaging and its transcriptional activity, is influenced by sequence-based rules that are both deterministic and predictable (Fig. 5A and fig. S11). CNN-based methods can thus be useful to predict the behavior of exogenous DNA in natural gene transfer events or in synthetic genome engineering.

Biased orientation of RNA-seq signal along prokaryotic sequences in eukaryotic context

Transcriptomics signal on bacterial chromosomes in yeast follows, on average, bacteria gene orientation (Fig. 2, B and C). Sequence determinants, including GC and AT skews, predate the divergence of eukaryotes and prokaryotes, and are found in both Mpneumo and Mmyco chromosomes. Such determinants are known to influence polymerase directionality (52, 53). Eventually, the resulting conserved orientation could facilitate the domestication of exogenous sequences during horizontal transfer/introgression events between distant species. Whether these RNA molecules are translated, and peptides of bacterial origins exist in the yeast cell, remains to be determined. If so, they could provide a source of diversity and adaptation.

Cohesin-mediated chromatin folding along non-transcribed DNA templates

Cohesins have been proposed to actively extrude DNA loops until they encounter an obstacle and/or a release signal (6, 35). In yeast and other species, these anchors are determined by a combination of convergent transcription, replication fork progression during S phase and/or the presence at these positions of stably bound cohesin promoting sister chromatid cohesion (54–56). Here, we show that the barely transcribed Mmyco chromosome is compacted by cohesin at G2/M, without focal loop anchoring, suggesting that transcription is neither necessary for loading nor a primary driver for translocation. We propose that cohesins can move freely along this template without encountering significant obstacles, making it a suitable model for studying potentially blocking sequences or molecules.

Compartmentalization of transcriptionally inactive foreign DNA in host's nucleus

We show that the introduction of foreign DNA with lower GC content spontaneously promotes the formation of transcriptionally inactive chromatin, which folds into an isolated crumpled globule compartment at the periphery of

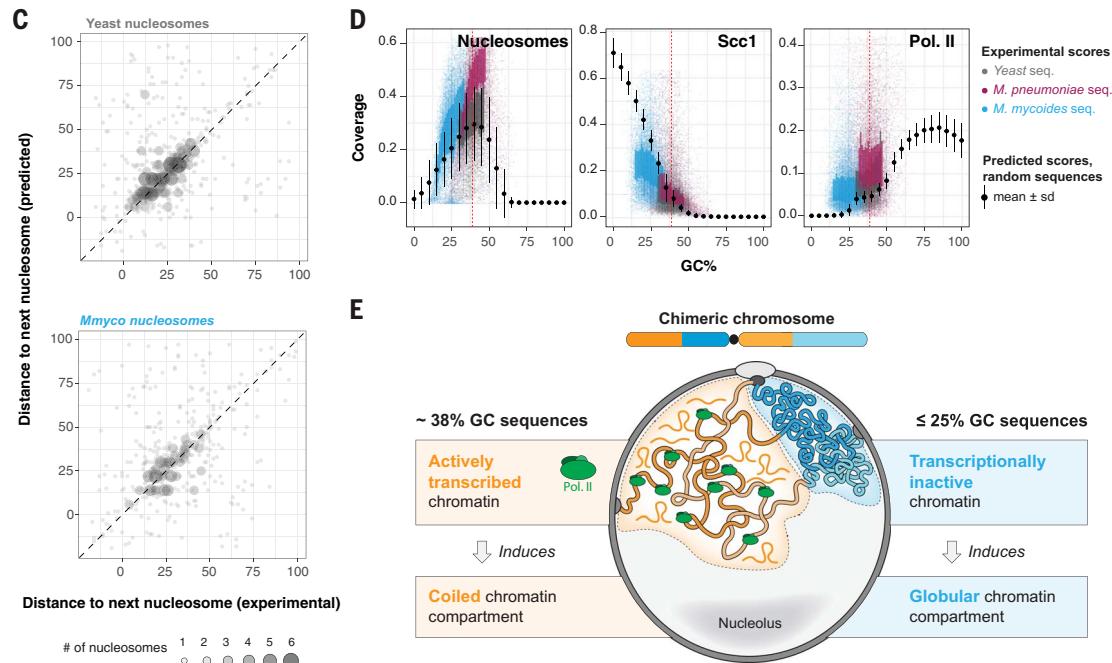
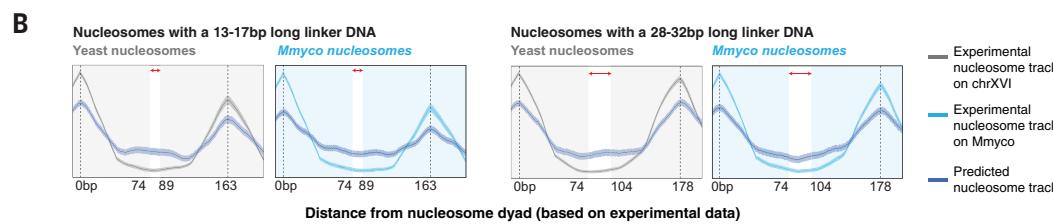
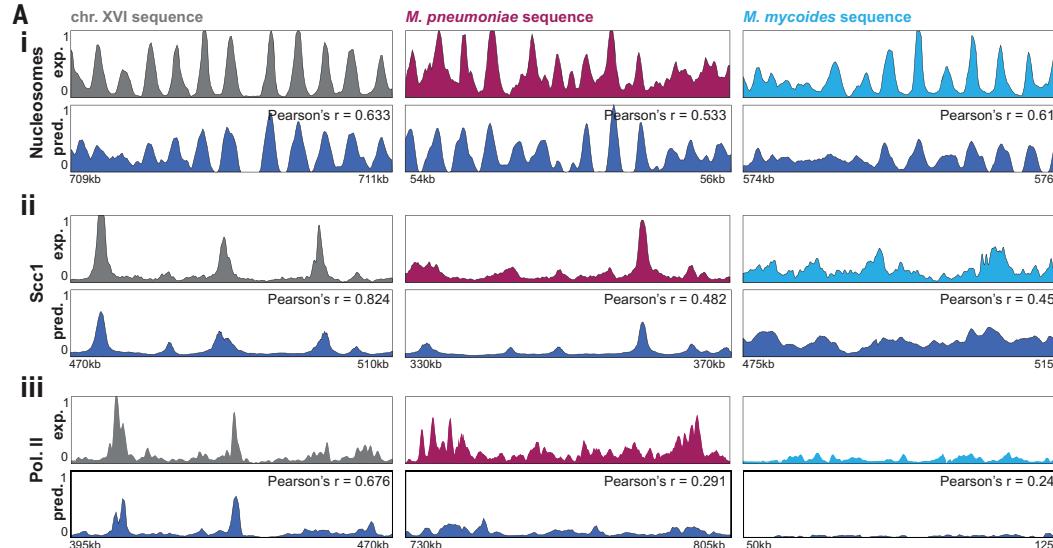
Fig. 5. DNA sequence is sufficient to predict Y/U chromatin composition.

(A) Experimental (top) and NN prediction (bottom) (dark blue) of nucleosome tracks or Scc1 and Pol II ChIP-seq coverage tracks, over yeast chromosome XVI or over Mmyco and Mpneumo bacterial chromosomes in yeast. Signals have been smoothed using a sliding genomic window of 10-bp (nucleosome tracks) or 500-bp (ChIP-seq). For each chromosome, the Pearson correlation score was computed between experimental and predicted scores, using averaged scores over non-overlapping 10-bp bins for nucleosomes or 500-bp bins for ChIP-seq. Bins with an average score lower than 0.01 were excluded.

(B) Nucleosome tracks (gray or light blue, experimental; dark blue, predicted), aligned at experimental dyads of yeast or Mmmcyo nucleosomes (position 0 on the x-axis). Nucleosomes are grouped by the length of their linker DNA (experimentally computed and represented by a red double-helix separating consecutive nucleosomes). Dotted lines indicate the position of nucleosome dyads (centered at 0 bp and 163 or 178 bp), and shaded areas represent nucleosome-occupied regions (\pm 74 bp around dyads).

(C) Correlation between predicted and experimental linker DNA lengths, for yeast (left) and Mmmcyo (right) nucleosomes. Only nucleosomes whose dyads are aligned $\leq \pm$ 2 bp between experimental and predicted nucleosome track are considered.

(D) Nucleosome, Scc1, and Pol II ChIP-seq average predicted scores in 2 kb (nucleosome) or 30-kb-long (ChIP-seq) sequences. Scores predicted from random sequences with varying GC content are shown as mean \pm sd in black, and average scores predicted from chromosome sequences of individual genomes are shown as colored points. Average experimental scores in 100-bp sequences along yeast; Mmyco and Mpneumo chromosomes are shown as gray, blue, or purple dots (scores of the middle half for each GC percentage unit are in bold).



the budding yeast nucleus (Fig. 3E and fig. S6H). This behavior mirrors the metazoan compartment “B” formed by inactive H3K9me3 heterochromatin, which is absent in yeast.

Lamin proteins are required for the peripheral localization of this B compartment (57), but the molecular mechanisms at work for its coalescence are still under investigation. Recent

studies have highlighted a critical role for HP1 in nuclear compartmentalization in *Drosophila* embryos (58), but not during mammalian mitosis-to-G1 transition (59). In the emerging

picture, multiple types of active and repressive compartments coexist, driven by diverse mechanisms, including gene activity (60). Our work provides a simplified model to explore how transcription drives compartmentalization in eukaryotes, hinting that this process may represent one of the earliest evolutionary mechanisms. This mechanism operates based on the genomic sequence itself, as high or low levels of Pol II occupancy can be accurately predicted in both chromatin archetypes on the basis of their sequence alone (Fig. 5).

Inhibiting transcription by dissociating Pol II from chromatin reduces the segregation of U and Y chromatin, suggesting that recruitment of the transcriptional machinery directly contributes to the physical segregation of adjacent Pol II-depleted chromatin into a distinct, globular, inactive compartment (Fig. 5E). This behavior may be driven by differences in chromatin composition (e.g., acetylated histone tail residues or intrinsically disordered proteins composing the transcription machinery) in each compartment.

These divergent fates for two foreign chromosomal sequences raise interesting evolutionary considerations in the context of the invasion of the genome by exogenous mobile elements. This could lead either to the spontaneous isolation of inactive foreign DNA or to the co-option of a set of active sequences that could represent a reservoir of genetic innovations. We show that an active chromatin region as small as 15 kb is not sensitive to the surrounding silenced chromatin environment. This suggests that small regions, tuned to the genome, could readily be harnessed by the host during introgressions or HGT events. This sequence-dependent mechanism may have contributed to the heterochromatinization of AT-rich transposable elements integrated in mammalian genomes.

Methods and Materials

Strains and medium culture conditions

All yeast strains used in this study are derivatives of W303 or VL6-48N and are listed in the table “strain list” (table S1). The original Mpneumo and Mmyco yeast strains carry the circular genomes of *Mycoplasma mycoides* subsp. *mycoides* strain PG1 (16) and *Mycoplasma pneumoniae* strain M129 (17), respectively. Cells were grown in a synthetic complete medium deprived of histidine (SC -His) (0.67% yeast nitrogen base without amino acids) (Difco), supplemented with a mix of amino acids, 2% glucose) or in rich medium (YPD): 1% bacto peptone (Difco), 1% bacto yeast extract (Difco), and 2% glucose. Yeast cells were synchronized in G1 by adding α-factor (Proteogenix, WY-13) in the media every 30 min during 2h30 (1 μg/mL final). To arrest cells in metaphase, cells were washed twice in fresh YPD after G1 arrest and released in rich medium (YPD) containing

Nocodazole (Sigma-Aldrich, M1404-10MG) during 1h30. Cell synchronization was verified by flow cytometry.

Isolation of quiescent cells from stationary phase cultures

Quiescent cells were obtained after isolation from density gradients as described in (61) with a few modifications. Cells were first precultured in YPD overnight and diluted the next day at 0.1 OD_{600 nm} /ml in an aerated flask with a ratio 1/10 of the flask volume. Cells were grown for 7 days at 30°C shaking at 250 rpm. To establish the density gradients, 4,5 mls of a premixed 150 mM NaCl / 90% Percoll (Sigma-Aldrich #P1644) solution were added to a 5 mls open-top thinwall ultraclear Beckman Coulter tube and centrifuge in the MSL-50 rotor at 25000 g for 18 min (Accel 4, Decel 9). To isolate the Q cells, 500 ODs of stationary phase cells were washed in 10 mls of 50 mM Tris buffer pH7.5, resuspended in 500 μl before loading on 4 preestablished density gradients. The gradients were then run at 1000 g for 30 min. The lower fraction (750 μl) was collected for each gradient, pooled and washed twice with 10 mls of 50mM Tris buffer pH7.5 before resuspension in a 1 ml final volume. 100 ODs Q cells when then fixed for Hi-C experiments.

Drug treatments

For HDAC inhibition, nicotinamide (NAM) (Sigma-Aldrich, N3376) or trichostatin A (Sigma-Aldrich, TSA) (T8552) were added to cell cultures at a final concentration of 5 mM final and 10 μM final during 4 hours. We then synchronized the cells in G1, and cultures were fixed for Hi-C and ChIP-seq.

For transcription inhibition, after synchronization in G1, thiolutin (Abcam, ab143556) was added to cell cultures at a final concentration of 20 μg mL⁻¹ for 30 min. We collected 5 mL of culture for the RNA-seq and the rest was fixed for Hi-C. *Candida glabrata* strain is used for spike-in normalization in RNA-seq.

CRISPR-Cas9 engineering

We used a CRISPR–Cas9 strategy to linearize the circular bacterial chromosome present in the parental yeast strains. Note that the highly acrocentric structure of Mmyco corresponds to the only position that was cleaved among several tested to linearize the chromosome. We suspect that due to the high AT content the CRISPR-Cas9 targeting remains relatively poorly efficient. Plasmids pML107 (62) or PAEF5 (63) carrying gRNA were co-transformed with 200 bp DNA repair recombinant donor sequences carrying telomeric repeats seeds (Genscript Biotech Netherlands) (fig. S1B).

We also applied a CRISPR-Cas9 approach to concomitantly fuse chromosome XVI with the linear bacterial chromosomes and remove chro-

mosome XVI centromere, as described in Luo *et al.* (64). The resulting chromosome (XVIIfMmyco) has one bacterial DNA arm carrying U chromatin and one yeast DNA arm made of Y chromatin (Fig. 4A). Strain XVIIfMmyco’s karyotype was verified using pulsed-field gel electrophoresis (fig. S6A), and that it grows normally was verified by growth assay (fig. S6, B and C). Briefly, we used three gRNAs inserted in pGZ110 and pAEF5 (63) co-transformed with recombinant DNA donors sequences (Twist Biosciences). These donor sequences are designed so that upon recombination 1) centromere XVI is removed and 2) fusion takes place between subtelomeres of chromosome XVI and either Mmyco and Mpneumo right arms (Fig. 4A). The same strategy was applied to generate reciprocal translocations between chromosome XVI and Mmyco sequences of the XVIIfMmyco chromosome (Fig. 4A). We generated alternating domains of Y and U chromatin along the XVIIfMmyco chromosome using CRISPR-induced reciprocal translocations between the two arms, resulting in strains XVIIfMmycot1 and t2 carrying alternating regions of U and Y type chromatin along a chromosome (Fig. 4A). The chromatin composition of these chimeric chromosome consists of [TEL - 500kb Y - 700kb U - CEN - 400kb Y - 500kb U - TEL] for XVIIfMmycot1 and of [TEL - 500kb Y - 50kb U - 350kb Y - CEN - 650kb U - 50kb Y - 500kb U - TEL] for XVIIfMmycot2.

gRNA sites were chosen to optimize CRISPR targeting specificity using *chopchop.cbu.uib.no* (65) or CRISPOR (66). For all experiments, 100 ng of gRNA expression plasmids and ~400 ng of recombinant donors were co-transformed. After co-transformation, yeast cells were immediately plated on the corresponding selective media. Yeast cells transformed with antibiotic resistance plasmide were allowed to recover in fresh YPD for 2h before plating.

Linearization, fusion and translocations of chromosomes were verified by PCR, pulse-field gel electrophoresis, and Hi-C.

All gRNA sequences, DNA donor sequences, and PCR primers used are provided in tables S2 to S4.

Liquid growth assays and segregational plasmid stability

Growth rate of parental, Mmyco, and Mpneumo strains carrying linear or fused chromosomes were assessed for independent clones in triplicates. Briefly, independent clones were grown overnight in selective medium (SC -HIS) to saturation at 30°C. The following morning the cultures were diluted to OD₆₀₀ = 0.01 and inoculated in 96 well plates containing 100 μl fresh SC -HIS. The cells were grown under agitation in a Tecan Sunrise plate reader at 30°C (Tecan). Optical densities at 600 nm were recorded every 10 min to generate the growth curves for

individual wells. The average of triplicates was plotted to compute the doubling time.

Segregational plasmid stability was measured as described in (67). Three individual transformants for each strain were inoculated into SC-HIS. P1 is the percentage of bacterial chromosome-carrying cells in selective media. It was determined from the ratio of viable colonies obtained by plating on YPD plates then replicated in selective medium (SC-HIS). To measure the chromosome stability, cultures were diluted into nonselective media and grown for X generations (g). After approximately 12 generations, the percentage of cells containing the bacterial chromosome (P2) was also determined from the ratio of viable colonies on rich medium (YPD) then replicated in selective medium. We then used P1, P2, and (g) number to calculate the segregation rate (m), which is defined as the percent of plasmid-free segregants appearing in the final population after a single doubling.

Pulsed-field Gel Electrophoresis and Southern blot hybridization

Agarose plugs containing yeast chromosomes were prepared as described (68) and separated by clamped homogeneous electric field gel electrophoresis Rotaphor (Biometra) using the following parameters. Gel: 1% (SeaKemGTG); t = 12 °C; buffer: 0.25 × Tris-Borate-EDTA; Program: [140 V, switch time: 300 s to 100 s, run time: 70 hours].

DNA from the PFGE was transferred on a membrane. Southern blot was performed using digoxigenin-labeled DNA probes as described (69). We used the HIS3 gene (YOR202w) present near the Mmyco and Mpneumo centromere was used as target to validate the linearization and the fusion. Digoxigenin-labeled DNA probes were synthesized from the pRS413 plasmid with the following oligos: 3'CTACATAAGAACACCTTTGG5' and 3'ATGACAGAGCAGAAAGCCCT5'.

3'ATGACAGAGCAGAAAGCCCT5'.

Hi-C procedure and sequencing

Cell fixation with 3% formaldehyde (Sigma-Aldrich, Cat. F8775) was performed as described in Dauban *et al.* (37). Quenching of formaldehyde with 300 mM glycine was performed at 4°C for 20 min. Hi-C experiments were performed with a Hi-C kit (Arima Genomics) with a double DpnII + HinfI restriction digestion following manufacturer instructions. Samples were purified using AMPure XP beads (Beckman A63882), recovered in 120 μl H2O and sonicated using Covaris (DNA 300bp) in Covaris microTUBE (Covaris, 520045). Biotinylated DNA was loaded on Dynabeads Streptavidin C1 (FISHER SCIENTIFIC, I0202333). Preparation of the samples for paired-end sequencing on an Illumina NextSeq500 (2x35 bp) was performed using Invitrogen TM Collibri TM PS DNA Library Prep Kit for Illumina and following manufacturer instructions.

ChIP-seq

The ChIP-seq protocol is described in (70). Experimental replicates (x3) were made for each condition. Briefly, cells of either *S. cerevisiae* or *Candida glabrata* were grown exponentially to OD600 = 0.5. 15 OD600 units of *S. cerevisiae* cells were mixed with 3 OD600 units of *C. glabrata* cells to a total volume of 45 mL for Scc1 calibration only. Cells were fixed using 4 mL of fixative solution (50 mM Tris-HCl, pH 8.0; 100 mM NaCl; 0.5 mM EGTA; 1 mM EDTA; 30% (v/v) formaldehyde) for 30 min at room temperature (RT) with rotation. The fixative was quenched with 2 mL of 2.5 M glycine (RT, 5 min with rotation). The cells were then harvested by centrifugation at 3,500 rpm for 3 min and washed with ice-cold PBS. The cells were then resuspended in 300 mL of ChIP lysis buffer [50 mM HEPES KOH, pH 8.0; 140 mM NaCl; 1 mM EDTA; 1% (v/v) Triton X-100; 0.1% (w/v) sodium deoxycholate; 1 mM PMSF; 2X Complete protease inhibitor cocktail (Roche)] and transferred in 2 mL tubes containing glass beads (ozyme, P000913-LYSKO-A0) before mechanical cells lysis. The soluble fraction was isolated by centrifugation at 2000 rpm for 3 min then transferred to sonication tubes (Covaris milliTUBE 1 ml, 520135) and samples were sonicated to produce sheared chromatin with a size range of 200 to 1000 bp using a Covaris sonicator. After sonication the samples were centrifuged at 13,200 rpm at 4°C for 20 min and the supernatant was transferred into 700 μL of ChIP lysis buffer. 80 μL (27 μL of each sample) of the supernatant was removed [termed 'whole cell extract (WCE) sample'] and stored at -80°C. For Scc1-PK, 5 μg of antibody (ab27671, Abcam) was added to the remaining supernatant. For Pol II, 5 μL (05-952-I, Merck Millipore), H2A 2 μL (AB_2687477, Active motif), H4K16ac 2 μL (07-352, Sigma-Aldrich), and 4 μg of Sir3 (a polyclonal ab gift from L. Pillus, University of California, San Diego). The supernatant is then incubated overnight at 4°C (wheel cold room). 50 μL of protein G Dynabeads or protein A Dynabeads was then added and incubated at 4°C for 2 h. Beads were washed 2 times with ChIP lysis buffer, 3 times with high salt ChIP lysis buffer (50 mM HEPES-KOH, pH 8.0; 500 mM NaCl; 1 mM EDTA; 1% (v/v) Triton X-100; 0.1% (w/v) sodium deoxycholate; 1 mM PMSF), 2 times with ChIP wash buffer (10 mM Tris-HCl, pH 8.0; 0.25M LiCl; 0.5% NP-40; 0.5% sodium deoxycholate; 1 mM EDTA; 1 mM PMSF) and 1 time with TE pH7.5. The immunoprecipitated chromatin was then eluted by incubation in 120 μL TES buffer (50 mM Tris-HCl, pH 8.0; 10 mM EDTA; 1% SDS) for 15 min at 65°C and the supernatant is collected termed 'IP sample'. The WCE samples were mixed with 40 μL of TES3 buffer (50 mM Tris-HCl, pH 8.0; 10 mM EDTA; 3% SDS). All (IP and WCE) samples were de-crosslinked by incubation at 65°C overnight.

RNA was degraded by incubation with 2 μL RNase A (10 mg/mL) for 1 h at 37°C. Proteins were removed by incubation with 10 μL of proteinase K (18 mg/mL) for 2 h at 65°C. DNA was purified by a phenol/Chloroform extraction. The triplicate IP samples were mixed in 1 tube and libraries for IP and WCE samples were prepared using Invitrogen TM Collibri TM PS DNA Library Prep Kit for Illumina and following manufacturer instructions. Paired-end sequencing on an Illumina NextSeq500 (2x35 bp) was performed. Libraries were performed in one or two biological replicates. When available, the duplicates were averaged for visualization. We calculated pairwise Pearson correlation scores between replicates and all showed high concordance.

RNA-seq

RNA was extracted using MN Nucleospin RNA kit and following manufacturer instructions. Directional mRNA library (rRNA removal) was prepared by the Biomics platform of Institut Pasteur, Paris and Paired-end sequencing (PE150) was performed by Novogene. Libraries were performed in three biological replicates and the triplicates were averaged for visualization. We calculated pairwise Pearson correlation scores between replicates and all showed high concordance.

CRAC-seq

CRAC-seq was performed essentially as described (24, 71). Briefly, 2 L of yeast cells containing Rpb1 tagged with a his6-TEV-proteinA (HTP) tag were grown in CSM-Trp medium to an OD600 of 0.6. 5% of *S. pombe* cells also containing an HTP tagged version of Rpb1 were added to the culture before cross-linking. Cells were exposed to UV light using a W5 UV crosslinking unit (UVO3 Ltd) for 50 s, harvested and resuspended in TN150 buffer (50 mM Tris pH 7.8, 150 mM NaCl, 0.1% NP-40 and 5 mM beta mercaptoethanol, 2.4 ml/g of cells) with protease inhibitors. Cells were broken using a Mixer Mill MM 400. Extracts were treated for one hour at 25°C with DNase I (165 U/g of cells) to solubilize chromatin and then clarified by centrifugation (20 min at 20000 g at 4°C). Rpb1-RNA complexes were purified by a two-step procedure using the protein A and the his-6 portion of the tag, with TEV cleavage for eluting complexes at the protein A-dependent purification step. Adaptors were added during the purification steps. The protein-RNA adduct was further purified by denaturing PAGE and the RNA recovered by proteinase K degradation and phenol extraction. The library was amplified by 10 cycles of PCR after reverse transcription and sequenced using Illumina technology.

MNase-seq

Each strain was grown to 10⁷ cells (OD600 of 0.8~) in 150 mL of SC-his medium at 30°C in

exponential phase. Cells were fixed with 4 mL of formaldehyde 37% (1% final) for 20 min at room temperature. Cross-linking was stopped by adding 8 mL of 2.5M glycine (125 mM final) for 30 min. The fixed cells were centrifuged, washed twice with cold phosphate-buffered saline (PBS 1X) and stored at -80°C. Once thawed, 900 μL of cells lysate obtained with a Precellys (Bertin Technologies) were recovered. 100 μL of 1X Micrococcal Nuclease Reaction Buffer (NEB, M0247S) and 5 μL of Bovine Serum Albumin (BSA, 20mg/mL) were added. The mix was divided into 10 tubes × 100 μL. 1 μL of Mnase enzyme (NEB, M0247S) at a concentration of 2,000,000 gel units/ml was added (2,000 units/sample final) in each tube. Samples were incubated at 37°C for varying times (0, 1, 2, 3, 5, 10, 15, 20, 40, and 60 min). 300 μL of Stop solution (10 μL of 0.5M EGTA pH 8.0, 30 μL of 20% SDS, 240 μL of H2O, and 20 μL Proteinase K 20mg/mL) was added to each tube. The tubes were then gently mixed and incubated at 65°C overnight. Samples were extracted with phenol/chloroform, and DNA was ethanol precipitated treated with DNase-free RNase. To evaluate the MNase digestion kinetics, DNA samples were analyzed using an Agilent TapeStation. DNA was purified using 2.2X volume of AmpureXP beads and sequencing libraries prepared with Invitrogen TM Collibri TM PS DNA Library Prep Kit for Illumina following manufacturer instructions. Paired-end sequencing on an Illumina NextSeq2000 (2 × 50 bp) was performed.

ATAC-seq

Independent clones for each strain were inoculated into SC-HIS for overnight culture at 30°C. Saturated overnight cultures were diluted to an OD600 of 0.1 and cultured for 6 hours at 30°C, until OD600 reached ~0.6. Around 1 to 5×10^6 cells were taken from each culture, pelleted at 3,000 g for 5 min, washed twice with spheroplasting buffer (1 M sorbitol, 40 mM HEPES-KOH pH 7.5, 10 mM MgCl₂), resuspended in 200 μL spheroplasting buffer with 10 μL of Zymolyase (10 mg/mL), then incubated for 30 min at 37°C. Spheroplasts were washed twice with 500 μL spheroplasting buffer then resuspended in 50 μL 1 × TD buffer with 2.5 μL of TDE1 (Illumina 20034197). Fragmentation was performed for 30 min at 37°C, 800 rpm with a thermomixer, and DNA was purified using the DNA Clean and Concentrator 5 kit (Zymo Research D4004). PCR was performed using Phusion Master Mix, 11 total cycles. Paired-end sequencing on an Illumina NextSeq2000 (2 × 50 bp) was performed.

Replication MFA experiment

Genomic DNA was prepared from asynchronous and G1 arrested cells in triplicates using Qiagen DNeasy Kit. Pellets were recovered,

washed with cold 70% ethanol, air dried and dissolved in 50 μL 1xTE. 100ng of soluble gDNA was transferred to sonication tubes and samples were sonicated to produce sheared chromatin with a size of about 300 bp using a Covaris sonicator. Samples were purified using AMPure XP beads (Beckman A63882). Preparation of the samples for paired-end sequencing on an Illumina NextSeq500 (2x35 bp) was performed using Invitrogen TM Collibri TM PS DNA Library Prep Kit for Illumina and following manufacturer instructions.

Processing of reads

Hi-C processing

Reads were aligned and contact maps generated and processed using Hicstuff (<https://github.com/koszullab/hicstuff>). Briefly, pairs of reads were aligned iteratively and independently using Bowtie2 (72) in its most sensitive mode against their reference genome. Each uniquely mapped read was assigned to a restriction fragment. Quantification of pairwise contacts between restriction fragments was performed with default parameters: uncuts, loops and circularization events were filtered as described in (73). PCR duplicates (defined as multiple pairs of reads positioned at the exact same position) were discarded. Pairs were binned at 1kb resolution and multiresolution balanced contact maps (in mcool format) were generated using cooler (74).

MNase-seq processing

For each time point, bowtie2 was used to align paired-end MNase-seq data on the appropriate genome reference. Only concordant pairs were retained, and fragments with a mapping quality lower than 10 were discarded. PCR duplicates were removed using samtools. Fragment coverage was normalized by library depth (CPM) and converted into a genomic track (bigwig) using deepTools (75).

RNA-seq processing

Bowtie2 was used to align paired-end RNA-seq data on the appropriate genome reference. Only fragments with an insert size shorter than 1 kb were retained. Only concordant pairs were retained, and fragments with a mapping quality lower than 10 were discarded. PCR duplicates were removed using samtools (76). Fragment coverage was normalized by library depth (CPM) and converted into stranded or unstranded genomic tracks (bigwig) using deepTools.

CRAC-seq processing

CRAC-seq datasets were processed and aligned on the appropriate genome reference using the pyCRAC script as described (77). Only reads longer than 20 nt were mapped on the *M. mycoides* and *M. pneumoniae* chimeric genomes and only reads longer than 40 nt were mapped to the *S. pombe* genome for the spike-

in to avoid inter-species mapping. Signals mapping to the chimeric genomes were normalized using the *S. pombe* spike-in. Homopolymers (fragments with 6+ identical nt) and polyadenylated reads were discarded.

ATAC-seq processing

Bowtie2 was used to align paired-end ATAC-seq data on the appropriate genome reference. Only fragments with an insert size shorter than 1 kb were retained. Only concordant pairs were retained, and fragments with a mapping quality lower than 10 were discarded. PCR duplicates were removed using samtools. Fragment coverage was normalized by library depth (CPM) and converted into a genomic track (bigwig) using deepTools. MACS2 2.2.7.1 was used to call peaks from ATAC-seq data.

ChIP-seq processing

For standard ChIP-seq, bowtie2 was used to align paired-end data on the appropriate genome reference. Only fragments with an insert size shorter than 1 kb were retained. Only concordant pairs were retained, and fragments with a mapping quality lower than 10 were discarded. PCR duplicates were removed using samtools. When available, the input was similarly processed. Fragment coverage was normalized by library depth (CPM) and converted into a genomic track (bigwig) using deepTools. Input-normalized fragment coverage was also generated, when possible, using bamCompare from deepTools with “-scaleFactorsMethod readCount.” For histone ChIP-seq experiments, IP tracks were divided by input and log2-scaled. For calibrated Scc1 ChIP-seq, CPM-normalized fragment coverages were also multiplied by the ORi factor for calibration (WCE_{glabrata} × IP_{cerevisiae} / WCE_{cerevisiae} × IP_{glabrata}) in which WCE_{glabrata} and IP_{glabrata} correspond to the number of paired reads that mapped uniquely on *C. glabrata* genome and same for *S. cerevisiae* reads). MACS2 2.2.7.1 was used to call peaks using the input alignment files as control.

Analysis of genome-wide assays

All downstream analysis steps were performed in R ≥ 4.1 / Bioconductor ≥ 3.16, using in-house scripts, unless mentioned otherwise.

Data visualization

Linear tracks were plotted in R using tidyCoverage (77).

Coverage heatmaps and aggregated profiles (average ± 95% CI and heatmap) over windows centered at genomic features (e.g., ChIP-seq peaks or TSSs) were plotted after averaging coverage over 1 bp-moving 200bp-wide rolling windows, using tidyCoverage.

All Hi-C contact, ratio and correlated maps were plotted with a log10, a log2, or a linear scaling respectively, using the “plotMatrix” function from HiContacts (78).

RNA-seq analysis

To estimate changes in transcript abundance of yeast genes (or bacterial Mmmyco genes) between two yeast strains with double translocations (eg XVIffMycot2 and XVIffMycot2'), *featureCounts* was first used to count raw number of stranded RNA-seq reads mapping to these genomic features of interest, using the corresponding rearranged gene annotation files (79). *DESeq2* was then used to identify yeast/Mmmyco genes that were differentially expressed between two chimeric strains (80).

MNase-seq analysis

To generate nucleosome tracks, “nucleosomal” fragments (between 130 and 165 bp) from the different MNase-seq timepoints were merged together then resized to a fixed 40 bp length centered at the dyad. The resulting coverages were normalized to account for different numbers of nucleosomal fragments sequenced in the different MNase-seq timecourse experiments.

Nucleosome tracks were subsequently used to identify positioned nucleosomes using an approach adapted from (21). Briefly, we used a greedy algorithm to identify nucleosome dyads sequentially based on the magnitude of the nucleosome track scores. Each chromosome was segmented into 50 bp-overlapping, 147 bp-long bins. For each bin, the position that had the greatest score from the nucleosome track (see above) was annotated as a nucleosome dyad. The greatest score from the nucleosome track was then iteratively recalculated within genomic bins recentered to that position. The algorithm stopped when the local maximum was found for every genomic bin. All the nucleosome dyads were then recovered as the central position of the set of unique genomic bins. The linker DNA length was inferred as the distance between consecutive nucleosome dyads - 147bp.

ChIP-seq analysis

Median Scc1 coverage was calculated over 20 kb windows centered at yeast centromeres or over 300 bp nonoverlapping windows tiling the entire yeast chromosomes (excluding centromeres or Scc1 peaks).

Correlation scores between ChIP-seq replicates were calculated from the average coverage scores over 100 bp tiled windows of yeast chromosomes.

Sequence biases

10-bp periodicity of dA, dT or dW dinucleotides in sequences up to 160 bp was computed over windows centered at yeast TSSs or tiling bacterial chromosome sequences, using “getPeriodicity” function from *periodicDNA* (81), using ushuffle (82) to maintain a constant dinucleotide frequency in shuffled control sequences. K-mer occurrences were estimated over 147 bp windows centered at yeast TSSs or tiling bacterial

chromosome sequences. AT (and GC) skews were calculated using the sequence of each DNA strand, in 10bp-wide bins.

Transcription convergence and Scc1 binding

At every position i along the genome (every 10 bp), a Dir_i directionality score was calculated as follows:

$$Dir_i = \frac{\sum_{k=i}^{i+200} (RNA_{fwd,k} - RNA_{rev,k})}{\sum_{k=i-200}^i (RNA_{fwd,k} - RNA_{rev,k})}$$

Genomic positions j for which $Dir_j-101 > 0$ & $Dir_j-100 < 0$ were then recovered and correspond to convergent or divergent transcription positions. For every transcription switch position k , a convergence score $Conv_k$ was subsequently computed as follows:

$$Conv_k = \left[\sum_{l=k-2000}^k (RNA_{fwd,l} > 20) - \sum_{l=k-2000}^{k-2000} (RNA_{rev,l} > 20) + \sum_{l=k}^{k+2000} (RNA_{rev,l} > 20) - \sum_{l=k}^{k+2000} (RNA_{fwd,l} > 20) \right] \times 100 \div (4000 \times 4)$$

Thus, $Conv_k$ scores range between -50 and 50 and a positive (negative) $Conv_k$ corresponds to a local genomic position k of convergent (divergent) transcription. The relationship between $Conv$ convergence scores and average Scc1 coverage scores (over a 500bp window centered at the convergent position) was then computed.

Replication MFA profile

Replication profiles were analyzed using Repliscope version 1.1.1 as in (83)

Loop detection

Chromosight 1.3.1 (83) was used to call loops de novo from contact maps binned at 1 kb and balanced with Cooler (74). Matrices were subsampled to contain the same total number of contacts. De novo loop calling was computed using the “detect” mode of Chromosight, with minimum loop length set at 2kb, percentage undetected set at 25 and pearson correlation threshold set at 0.315. Loop strength was quantified for each loop using the quantify mode of Chromosight and the mean loop score was calculated for each condition. Loop pile-ups of averaged 17kb windows were generated with Chromosight.

Cis-trans ratio and P(s) analysis

Cis-trans ratios were calculated using the “cisTransRatio” function from HiContacts. Contact probability as a function of genomic distance $P(s)$ were determined using pairs files (generated by hicstuff) and the “distanceLaw” function from HiContacts with default parameters, averaging the contact data by individual chromosomes. Briefly, the $P(s)$ were computed by binning interactions according to their genomic distance, using logarithmic-sized distance bins. The number of interactions within each bin was then divided by the bin width and by the total number of interactions, to normalize for varying bin widths and for sequencing depth.

Virtual 4C profiles

Virtual 4C profiles for 20kb-wide viewpoints were computed from 2kb-binned contact maps, using the “virtual4C” function from HiContacts (78). Contacts of entire chromosomes across the entire genome were manually computed using subsetting functions from HiContacts.

Deep-Learning analysis**Models architectures and training**

Three different models were trained to learn either on the nucleosome, Pol II and Scc1 profiles from the underlying genomic sequence. The signals were pretreated as follows. For ChIP profiles we first filtered outlier values, converted the signal in count per million and aggregated the different replicates. We computed the ratio between IP and INPUT and discarded all regions for which either profile was equal to zero. Then for all data, we truncated the experimental profiles to a threshold corresponding to the 99th percentile of the profile distribution. We then divided all values in the truncated profile by the maximum value to get a normalized signal between 0 and 1. For each position along the genome, a DNA sequence of length W was associated with a subset of n_{out} values from the corresponding profile to make the three datasets used for our deep learning framework. For each of these datasets, we used the yeast chromosomes I to XIII for training, XIV and XV for validation and XVI for test.

We implemented all the CNNs using the Keras library (84) and Tensorflow (85) as backend. A RTX 2080 Ti GPU was used to improve the training speed. We used the adaptive moment estimation (ADAM) optimization method to compute adaptive learning rates for each parameter. The batch size was set to 512 (Scc1) or 1024 (nucleosome, Pol II).

For the nucleosome prediction task, our CNN architecture was similar to the one used in (86). It consists of three convolutional layers with respectively 64, 16 and 8 kernels of shape (3x1x4), (8x1x64) and (80x1x16). A max pooling layer of size 2 and a ReLu activation function

was applied after each of these three convolutions. Batch normalization and a dropout of 0.2 was applied after each convolution and the convolution stride was set to 1. Our model takes inputs of shape (2001, 1, 4), the last dimension representing the four nucleotides, and outputs a single value (i.e., n_out =1) corresponding to the value of the nucleosome profile in the middle of the input sequence.

For the Pol II and Scc1 tasks, the architecture was modified as follows to take into account longer range influences (87). It consists of three convolutional layers all with kernels of shape (12x1x4), (8x1x64) and (80x1x16). Max Pooling layers of respective lengths of 8, 4 and 4 followed by ReLu activation were applied after each convolution. Four dilated convolution layers were then applied using 16 kernels of length 5 and dilatation values of respectively 2, 4, 8 and 16. For Pol II, the input sequence length was set to 2048 bp and the output length n_out to 16, corresponding to the 16 values of the signal found every 128 bp over the 2,048 bp input sequence. For Scc1, the input sequence length was set to 32768 bp and the output length n_out was 256, corresponding to the 256 values of the signal found every 128bp over the 32,768 bp input sequence.

The loss function used is the sum of the Pearson's dissimilarity (1-correlation) between the prediction x and the target y and the mean absolute error (MAE) between them ($\text{loss} = \text{MAE}(x, y) + 1 - \text{corr}(x, y)$). This loss function has been previously shown to enable both a faster and a more accurate convergence (87).

An early stopping procedure was applied during training to prevent models from overfitting. The loss function was calculated on the validation set at every epoch to evaluate the generalisability of the model. The training procedure was stopped if the validation loss did not decrease at all for 6 epochs and the model parameters were set back to their best performing value. The training procedure usually lasted for 15 to 20 epochs.

Predictions

Predictions were done on all chromosomes, including the chromosomes from the training, validation and test sets as well as exogenous bacterial chromosomes.

In order to characterize the influence of the GC content of the sequences on the overall profile heights, we generated 10000 random 2kb (nucleosome, Pol II) or 32kb (Scc1) sequences and predicted each of the three profiles on these sequences.

Comparison with experimental data

For each chromosome, the Pearson correlation score was calculated between experimental measurements and predictions, using averaged scores over non-overlapping 10 bp bins for

nucleosomes or 500 bp bins for ChIP-seq. Bins with an average experimental or predicted score lower than 0.01 were excluded from the correlation analysis.

When comparing nucleosome repeat length (NRL) obtained from experimental or predicted nucleosomes tracks, only the reference nucleosomes whose dyads are aligned $\leq \pm 2$ bp between experimental and predicted nucleosome track were considered.

Identification of nucleotide motifs influencing CNN-based predictions

To identify relevant nucleotide motifs for prediction, we computed the gradients of the model outputs with respect to the input (i.e., the saliency). In the case of models with multiple outputs, corresponding to different locations in a genomic window, we computed the gradient only for one output, corresponding to the center of the genomic window. We used the gradient correction method from Majdanovic *et al.* to compute attribution maps for all possible windows of the *S. cerevisiae* genome (88). Then, for each bp we averaged the absolute values of their attributions computed in all the genomic windows that contain this bp. This resulted in a genome-wide signal of unsigned attribution score for every bp. We computed z-scores for the attribution score and selected bases with a z-score of 4 or higher, to identify regions of high importance in the genome. We then grouped these bases into regions when the genomic distance between these bp was strictly smaller than 20, discarded any remaining solitary bases and extended selected regions by ± 10 bp on their side. As such there were no overlaps between regions, which varied in length from 22bp to 200bp-1000bp (depending on the model). We got 12125, 4446 and 11295 regions of interest respectively for nucleosome, polymerase and cohesin models. We then extracted the sequences in each region and used them as input to the meme suite tools (v5.5.5) (89). We ran XSTREME with default parameters, against 5 databases (Yeast-tract, SwissRegulon Scer, macisaac, SCPD and uniprobe yeast) using the dinucleotide distribution of the W303 genome as background. This resulted in 6 to 8 motifs discovered per model, most of them matching existing motifs in yeast databases.

Imaging and analysis

Two-dots assay (SCC)

Strains yLD126-36c, FB176 and FB200 were inoculated and grown overnight in SC-MET medium. The next day cultures were diluted to OD600=0.2 in SC-MET. After 3h of exponential growth, alpha factor (10 μ l at 5mg/ml) was added every 30 min for 2 hours. G1-arrested cells were released into YPD (plus 2 mM methionine), and aliquots sampled for FACS and imaging analysis.

FISH experiments

FISH experiments were performed as described in Gotta *et al.* (90), with some modifications (91). The probes were obtained by direct labeling of the bacterial DNA (1.5 μ g) using the Nick Translation kit from Jena Bioscience (Atto488 NT Labeling Kit), the labeling reaction was performed at 15°C for 90 min. The labeled DNA was purified using the Qiaquick PCR purification kit from Qiagen, eluted in 30 μ l of water. The purified probe was then diluted in the probe mix buffer (50% formamide, 10% dextran sulfate, 2 \times SSC final). 20 OD of cells (1 OD corresponding to 107 cells) were grown to mid-logarithmic phase (1-2 \times 107 cells/ml) and harvested at 1,200 g for 5 min at RT. Cells were fixed in 20 ml of 4% paraformaldehyde for 20 min at RT, washed twice in water, and resuspended in 2 ml of 0.1 M EDTA-KOH pH 8.0, 10 mM DTT for 10 min at 30°C with gentle agitation. Cells were then collected at 800 g, and the pellet was carefully resuspended in 2 ml YPD - 1.2 M sorbitol. Next, cells were spheroplasted at 30°C for 10 min with Zymolyase (60 μ g/ml Zymolyase-100T to 1 ml YPD-sorbitol cell suspension). Spheroplasting was stopped by the addition of 40 ml YPD - 1.2 M sorbitol. Cells were washed twice in YPD - 1.2 M sorbitol, and the pellet was resuspended in 1 ml YPD. Cells were put on diagnostic microscope slides and superficially air dried for 2 min. The slides were plunged in methanol at -20°C for 6 min, transferred to acetone at -20°C for 30 s, and air dried for 3 min. After an overnight incubation at RT in 4 \times SSC, 0.1% Tween, and 20 μ g/ml RNase, the slides were washed in H2O and dehydrated in ethanol 70%, 80%, 90%, and 100% consecutively at -20°C for 1 min in each bath. Slides were air dried, and a solution of 2 \times SSC and 70% formamide was added for 5 min at 72°C. After a second step of dehydration, the denatured probe was added to the slides for 10 min at 72°C followed by a 37°C incubation for 24h in a humid chamber. The slides were then washed twice in 0.05 \times SSC at 40°C for 5 min and incubated twice in BT buffer (0.15 M NaHCO3, 0.1% Tween, 0.05% BSA) for 30 min at 37°C. For the DAPI staining, the slides were incubated in a DAPI solution (1 μ g/ml in 1 \times PBS) for 5 min and then washed twice in 1 \times PBS without DAPI.

FISH image analysis

Images were acquired on a wide-field microscopy system based on an inverted microscope (Nikon TE2000) equipped with a 100 \times /1.4 NA immersion oil objective, a C-mos camera and a Spectra X light engine lamp for illumination (Lumencor, Inc). The system is driven by the MetaMorph software (Molecular Devices). The axial (z) step is 200 nm and images shown are maximum intensity projection of z-stack images (MIP). Quantifications were done on the MIP images using ilastik for segmentation and Fiji for analyses of the particles.

REFERENCES AND NOTES

1. J. Romiguier, V. Ranwez, E. J. P. Douzery, N. Galtier, Contrasting GC-content dynamics across 33 mammalian genomes: Relationship with life-history traits and chromosome sizes. *Genome Res.* **20**, 1001–1009 (2010). doi: 10.1101/gr.104372.109; pmid: 20530252
2. J. Tajbakhsh et al., Spatial distribution of GC- and AT-rich DNA sequences within human chromosome territories. *Exp. Cell Res.* **255**, 229–237 (2000). doi: 10.1006/excr.1999.4780; pmid: 10694438
3. A. Valouev et al., Determinants of nucleosome organization in primary human cells. *Nature* **474**, 516–520 (2011). doi: 10.1038/nature10002; pmid: 21602827
4. G. P. Holmquist, Evolution of chromosome bands: Molecular ecology of noncoding DNA. *J. Mol. Evol.* **28**, 469–486 (1989). doi: 10.1007/BF02602928; pmid: 2549255
5. L. Mirny, J. Dekker, Mechanisms of Chromosome Folding and Nuclear Organization: Their Interplay and Open Questions. *Cold Spring Harb. Perspect. Biol.* **a040147** (2022). doi: 10.1101/cshperspect.a040147; pmid: 34518339
6. I. F. Davidson, J.-M. Peters, Genome folding through loop extrusion by SMC complexes. *Nat. Rev. Mol. Cell Biol.* **22**, 445–464 (2021). doi: 10.1038/s41580-021-00349-7; pmid: 33767413
7. A. S. Belmont, Nuclear Compartments: An Incomplete Primer to Nuclear Compartments, Bodies, and Genome Organization Relative to Nuclear Architecture. *Cold Spring Harb. Perspect. Biol.* **a041268** (2022). doi: 10.1101/cshperspect.a041268; pmid: 34400557
8. A. Crisp, C. Boschetti, M. Perry, A. Tunnicliffe, G. Micklem, Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol.* **16**, 50 (2015). doi: 10.1186/s13059-015-0607-3; pmid: 25785303
9. N. B. Edelman, J. Mallet, Prevalence and Adaptive Impact of Introgression. *Annu. Rev. Genet.* **55**, 265–283 (2021). doi: 10.1146/annurev-genet-021821-020805; pmid: 34579539
10. J. Van Etten, D. Bhattacharya, Horizontal Gene Transfer in Eukaryotes: Not if, but How Much? *Trends Genet.* **36**, 915–925 (2020). doi: 10.1016/j.tig.2020.08.006; pmid: 33012528
11. J. Peter et al., Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature* **556**, 339–344 (2018). doi: 10.1038/s41586-018-0030-5; pmid: 29643504
12. V. Baby et al., Cloning and Transplantation of the Mesoplasma florum Genome. *ACS Synth. Biol.* **7**, 209–217 (2018). doi: 10.1021/acssynbio.7b00279; pmid: 28893065
13. A. L. V. Coradini, C. B. Hull, I. M. Ehrenreich, Building genomes to understand biology. *Nat. Commun.* **11**, 6177 (2020). doi: 10.1038/s41467-020-19753-2; pmid: 33268788
14. A. Currin et al., The evolving art of creating genetic diversity: From directed evolution to synthetic biology. *Biootechnol. Adv.* **50**, 107762 (2021). doi: 10.1016/j.biotechadv.2021.107762; pmid: 34000294
15. C. Payen et al., Unusual composition of a yeast chromosome arm is associated with its delayed replication. *Genome Res.* **19**, 1710–1721 (2009). doi: 10.1101/gr.090605.108; pmid: 19592681
16. F. Labroussa et al., Impact of donor-recipient phylogenetic distance on bacterial genome transplantation. *Nucleic Acids Res.* **44**, 8501–8511 (2016). doi: 10.1093/nar/gkw688; pmid: 27488189
17. E. Ruiz et al., CREasPy-Cloning: A Method for Simultaneous Cloning and Engineering of Megabase-Sized Genomes in Yeast Using the CRISPR-Cas9 System. *ACS Synth. Biol.* **8**, 2547–2557 (2019). doi: 10.1021/acssynbio.9b00224; pmid: 31663334
18. C. Lartigue et al., Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science* **325**, 1693–1696 (2009). doi: 10.1126/science.1173759; pmid: 19696314
19. S. Karlin, J. Mrázek, Compositional differences within and between eukaryotic genomes. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10227–10232 (1997). doi: 10.1073/pnas.94.19.10227; pmid: 9294192
20. K. Nasmyth, Disseminating the genome: Joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu. Rev. Genet.* **35**, 673–745 (2001). doi: 10.1146/annurev.genet.35.102401.09134; pmid: 11700297
21. K. Brogaard, L. Xi, J.-P. Wang, J. Widom, A map of nucleosome positions in yeast at base-pair resolution. *Nature* **486**, 496–501 (2012). doi: 10.1038/nature11142; pmid: 22722846
22. M. T. Ocampo-Hafalla, Y. Katou, K. Shirahige, F. Uhlmann, Displacement and re-accumulation of centromeric cohesin during transient pre-anaphase centromere splitting. *Chromosoma* **116**, 531–544 (2007). doi: 10.1007/s00412-007-0118-4; pmid: 17763979
23. K.-L. Chan et al., Cohesin's DNA exit gate is distinct from its entrance gate and is regulated by acetylation. *Cell* **150**, 961–974 (2012). doi: 10.1016/j.cell.2012.07.028; pmid: 22901742
24. D. Challal et al., General Regulatory Factors Control the Fidelity of Transcription by Restricting Non-coding and Ectopic Initiation. *Mol. Cell* **72**, 955–969.e7 (2018). doi: 10.1016/j.molcel.2018.11.037; pmid: 3076657
25. L. Lazar-Stefanita et al., Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle. *EMBO J.* **36**, 2684–2697 (2017). doi: 10.15252/embj.201797342; pmid: 28729434
26. Z. Duan et al., A three-dimensional model of the yeast genome. *Nature* **465**, 363–367 (2010). doi: 10.1038/nature08973; pmid: 20436457
27. H. Wong et al., A predictive computational model of the dynamic 3D interphase yeast nucleus. *Curr. Biol.* **22**, 1881–1890 (2012). doi: 10.1016/j.cub.2012.07.069; pmid: 22940469
28. A. Y. Grosberg, S. K. Nechaev, E. I. Shakhnovich, The role of topological constraints in the kinetics of collapse of macromolecules. *J. Phys. (Paris)* **49**, 2095–2100 (1988). doi: 10.1051/jphys:0198800490120209500
29. L. A. Mirny, The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res.* **19**, 37–51 (2011). doi: 10.1007/s10577-010-9177-0; pmid: 21274616
30. A. Taddei, S. M. Gasser, Structure and function in the budding yeast nucleus. *Genetics* **192**, 107–129 (2012). doi: 10.1534/genetics.112.140608; pmid: 22964839
31. L. Dauban et al., Regulation of Cohesin-Mediated Chromosome Folding by Eco1 and Other Partners. *Mol. Cell* **77**, 1279–1293.e4 (2020). doi: 10.1016/j.molcel.2020.01.019; pmid: 32032532
32. L. Costantini, T. S. Hsieh, R. Lamothé, X. Darzacq, D. Koschland, Cohesin residency determines chromatin loop patterns. *eLife* **9**, 59889 (2020). doi: 10.7554/eLife.59889; pmid: 33170773
33. S. A. Schalbetter et al., SMC complexes differentially compact mitotic chromosomes according to genomic context. *Nat. Cell Biol.* **19**, 1071–1080 (2017). doi: 10.1038/nccb3594; pmid: 28825700
34. S. S. P. Rao et al., A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014). doi: 10.1016/j.cell.2014.11.021; pmid: 25497547
35. A. Goloborodko, M. V. Imakaev, J. F. Mariko, L. Mirny, Compaction and segregation of sister chromatids via active loop extrusion. *eLife* **5**, e14864 (2016). doi: 10.7554/eLife.14864; pmid: 27192037
36. J. H. I. Haahrhus et al., The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* **169**, 693–707.e14 (2017). doi: 10.1016/j.cell.2017.04.013; pmid: 28475897
37. E. Lieberman-Aiden et al., Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009). doi: 10.1126/science.1181369; pmid: 19815776
38. N. Naumova et al., Organization of the Mitotic Chromosome. *Science* **342**, 948–953 (2013). doi: 10.1126/science.1236083; pmid: 24200812
39. J. H. Gibcus et al., A pathway for mitotic chromosome formation. *Science* **359**, eaao6135 (2018). doi: 10.1126/science.aaao6135; pmid: 29348367
40. G. Spraklin et al., Diverse silent chromatin states modulate genome compartmentalization and loop extrusion barriers. *Nat. Struct. Mol. Biol.* **30**, 38–51 (2023). doi: 10.1038/s41594-022-00892-7; pmid: 36550219
41. E. P. Nora, Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930–944.e22 (2017). doi: 10.1038/s41594-022-00892-7; pmid: 28525758
42. M. Kabi, G. J. Filion, Heterochromatin: Did H3K9 methylation evolve to tame transposons? *Genome Biol.* **22**, 325 (2021). doi: 10.1186/s13059-021-02250-5; pmid: 34857038
43. M. R. Gartenberg, J. S. Smith, The Nuts and Bolts of Transcriptionally Silent Chromatin in *Saccharomyces cerevisiae*. *Genetics* **203**, 1563–1599 (2016). doi: 10.1534/genetics.112.145243; pmid: 27516616
44. L. L. Breedon, T. Tsukiyama, Quiescence in *Saccharomyces cerevisiae*. *Annu. Rev. Genet.* **56**, 253–278 (2022). doi: 10.1146/annurev-genet-080320-023632; pmid: 36449357
45. M. Guidi et al., Spatial reorganization of telomeres in long-lived quiescent cells. *Genome Biol.* **16**, 206 (2015). doi: 10.1186/s13059-015-0766-2; pmid: 26399229
46. K. Struhl, E. Segal, Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* **20**, 267–273 (2013). doi: 10.1038/nsmb.2506; pmid: 23463311
47. E. Routhier, E. Pierre, G. Khodabandehlu, J. Mozziconacci, Genome-wide prediction of DNA mutation effect on nucleosome positions for yeast synthetic genomics. *Genome Res.* **31**, 317–326 (2021). doi: 10.1101/gr.264416.120; pmid: 33355297
48. R. V. Chereji, D. J. Clark, Major Determinants of Nucleosome Positioning. *Biophys. J.* **114**, 2279–2289 (2018). doi: 10.1016/j.bpj.2018.03.015; pmid: 29628211
49. Z. Gvozdenov, Z. Bartucean, K. Struhl, Functional analysis of a random-sequence chromosome reveals a high level and the molecular nature of transcriptional noise in yeast cells. *Mol. Cell* **83**, 1786–1795.e5 (2023). doi: 10.1016/j.molcel.2023.04.010; pmid: 37137302
50. B. R. Camellato, R. Brosh, H. J. Ashe, M. T. Maurano, J. D. Boeke, Synthetic reversed sequences reveal default genomic states. *Nature* **628**, 373–380 (2024). doi: 10.1038/s41586-024-07128-2; pmid: 38448583
51. I. Luthra et al., Regulatory activity is the default DNA state in eukaryotes. *Nat. Struct. Mol. Biol.* **31**, 559–567 (2024). doi: 10.1038/s41594-024-01235-4; pmid: 38448573
52. P. A. Ginn, P. L. Lott, H. C. Christensen, I. Korf, F. Chédin, R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **45**, 814–825 (2012). doi: 10.1016/j.molcel.2012.01.017; pmid: 22387027
53. J. R. Lobry, Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* **13**, 660–665 (1996). doi: 10.1093/oxfordjournals.molbev.a025626; pmid: 8676740
54. N. Bastié et al., Sister chromatid cohesion halts DNA loop expansion. *Mol. Cell* **84**, 1139–1148.e5 (2024). doi: 10.1016/j.molcel.2024.02.004; pmid: 38452765
55. E. J. Banigan et al., Transcription shapes 3D chromatin organization by interacting with loop extrusion. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2210480120 (2023). doi: 10.1073/pnas.2210480120; pmid: 36897969
56. B. J. H. Dequeker et al., MCM complexes are barriers that restrict cohesin-mediated loop extrusion. *Nature* **606**, 197–203 (2022). doi: 10.1038/s41586-022-04730-0; pmid: 35585235
57. M. Falk et al., Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570**, 395–399 (2019). doi: 10.1038/s41586-019-1275-3; pmid: 31168090
58. F. Zenk et al., HP1 drives de novo 3D genome reorganization in early *Drosophila* embryos. *Nature* **593**, 289–293 (2021). doi: 10.1038/s41586-021-03460-z; pmid: 33854237
59. H. Zhao et al., Genome folding principles uncovered in condensin-depleted mitotic chromosomes. *Nat. Genet.* **56**, 1213–1224 (2024). doi: 10.1038/s41588-024-01759-x; pmid: 38802567
60. T. S. Hsieh et al., Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol. Cell* **78**, 539–553.e8 (2020). doi: 10.1016/j.molcel.2020.03.002; pmid: 32213323
61. C. Allen et al., Isolation of quiescent and nonquiescent cells from yeast stationary-phase cultures. *J. Cell Biol.* **174**, 89–100 (2006). doi: 10.1083/jcb.b.200604072; pmid: 16818721
62. M. F. Laughery et al., New vectors for simple and streamlined CRISPR-Cas9 genome editing in *Saccharomyces cerevisiae*. *Yeast* **32**, 711–720 (2015). doi: 10.1002/yea.3098; pmid: 26305040
63. N. Agier, A. Fleiss, S. Delmas, G. Fischer, A Versatile Protocol to Generate Translocations in Yeast Genomes Using CRISPR/Cas9. *Methods Mol. Biol.* **2196**, 181–198 (2021). doi: 10.1007/978-0-7162-0868-5_14; pmid: 32889721
64. J. Luo, X. Sun, B. P. Cormack, J. D. Boeke, Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast. *Nature* **560**, 392–396 (2018). doi: 10.1038/s41586-018-0374-x; pmid: 30069047
65. K. Labun et al., CHOPCHOP v3: Expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* **47** (W1), W171–W174 (2019). doi: 10.1093/nar/gkz365; pmid: 31106371
66. J.-P. Concordet, M. Haueussler, CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* **46**, W242–W245 (2018). doi: 10.1093/nar/gky354; pmid: 29762716
67. T. W. Christianson, R. S. Sikorski, M. Dante, J. H. Sher, P. Hietter, Multifunctional yeast high-copy-number shuttle vectors. *Gene* **110**, 119–122 (1992). doi: 10.1016/0378-119(92)90445-W; pmid: 1544568
68. R. Koszul, S. Caburet, B. Dujon, G. Fischer, Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. *EMBO J.* **23**, 234–243 (2004). doi: 10.1038/sj.emboj.7600024; pmid: 14685272
69. D. Viterbo et al., A fast, sensitive and cost-effective method for nucleic acid detection using non-radioactive probes. *Biol.*

- Methods Protoc.* **3**, bpo006 (2018). doi: [10.1093/biomet/bpo006](https://doi.org/10.1093/biomet/bpo006); pmid: [32161800](https://pubmed.ncbi.nlm.nih.gov/32161800/)
70. B. Hu *et al.*, Biological chromodynamics: A general method for measuring protein occupancy across the genome by calibrating ChIP-seq. *Nucleic Acids Res.* **43**, e132 (2015). doi: [10.1093/nar/gkv670](https://doi.org/10.1093/nar/gkv670); pmid: [26130708](https://pubmed.ncbi.nlm.nih.gov/26130708/)
71. S. Granneman, G. Kudla, E. Petfalski, D. Tollervey, Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 9613–9618 (2009). doi: [10.1073/pnas.0901997106](https://doi.org/10.1073/pnas.0901997106); pmid: [19482942](https://pubmed.ncbi.nlm.nih.gov/19482942/)
72. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi: [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923); pmid: [22388286](https://pubmed.ncbi.nlm.nih.gov/22388286/)
73. A. Courzac, H. Marie-Nelly, M. Marbouty, R. Koszul, J. Mozziconacci, Normalization of a chromosomal contact map. *BMC Genomics* **13**, 436 (2012). doi: [10.1186/1471-2164-13-436](https://doi.org/10.1186/1471-2164-13-436); pmid: [22935139](https://pubmed.ncbi.nlm.nih.gov/22935139/)
74. N. Abdennur, L. A. Mirny, Cooler: Scalable storage for Hi-C data and other genetically labeled arrays. *Bioinformatics* **36**, 311–316 (2020). doi: [10.1093/bioinformatics/btz540](https://doi.org/10.1093/bioinformatics/btz540); pmid: [31290943](https://pubmed.ncbi.nlm.nih.gov/31290943/)
75. F. Ramírez *et al.*, deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–5 (2016). doi: [10.1093/nar/gkw257](https://doi.org/10.1093/nar/gkw257); pmid: [27079975](https://pubmed.ncbi.nlm.nih.gov/27079975/)
76. P. Danecek *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021). doi: [10.1093/gigascience/giab008](https://doi.org/10.1093/gigascience/giab008); pmid: [33590861](https://pubmed.ncbi.nlm.nih.gov/33590861/)
77. J. Serizay, R. Koszul, Epigenomics coverage data extraction and aggregation in R with tidyCoverage. *Bioinformatics* **40**, btae487 (2024). doi: [10.1093/bioinformatics/btae487](https://doi.org/10.1093/bioinformatics/btae487); pmid: [39073878](https://pubmed.ncbi.nlm.nih.gov/39073878/)
78. J. Serizay, C. Matthey-Doret, A. Bignaud, L. Baudry, R. Koszul, Orchestrating chromosome conformation capture analysis with Bioconductor. *Nat. Commun.* **15**, 1072 (2024). doi: [10.1038/s41467-024-44761-x](https://doi.org/10.1038/s41467-024-44761-x); pmid: [38316789](https://pubmed.ncbi.nlm.nih.gov/38316789/)
79. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014). doi: [10.1093/bioinformatics/btt656](https://doi.org/10.1093/bioinformatics/btt656); pmid: [24227677](https://pubmed.ncbi.nlm.nih.gov/24227677/)
80. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi: [10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8); pmid: [25516281](https://pubmed.ncbi.nlm.nih.gov/25516281/)
81. J. Serizay, J. Ahringer, periodicDNA: An R/Bioconductor package to investigate k-mer periodicity in DNA. *F1000 Res.* **10**, 141 (2021). doi: [10.12688/f1000research.511431](https://doi.org/10.12688/f1000research.511431); pmid: [33953908](https://pubmed.ncbi.nlm.nih.gov/33953908/)
82. M. Jiang, J. Anderson, J. Gillespie, M. Mayne, uShuffle: A useful tool for shuffling biological sequences while preserving the k-let counts. *BMC Bioinformatics* **9**, 192 (2008). doi: [10.1186/1471-2105-9-192](https://doi.org/10.1186/1471-2105-9-192); pmid: [18405375](https://pubmed.ncbi.nlm.nih.gov/18405375/)
83. C. Matthey-Doret *et al.*, Computer vision for pattern detection in chromosome contact maps. *Nat. Commun.* **11**, 5795 (2020). doi: [10.1038/s41467-020-19562-7](https://doi.org/10.1038/s41467-020-19562-7); pmid: [33199682](https://pubmed.ncbi.nlm.nih.gov/33199682/)
84. Keras 3: A new multi-backend Keras. Keras (2023); <https://github.com/keras-team/keras>.
85. M. Abadi *et al.*, “TensorFlow: a system for large-scale machine learning” in *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation* (USENIX Association, 2016), pp. 265–283.
86. E. Routhier, A. B. Kamruddin, J. Mozziconacci, keras_dna: a wrapper for fast implementation of deep learning models in genomics. *Bioinformatics* **37**, 1593–1594 (2021). doi: [10.1093/bioinformatics/btaa929](https://doi.org/10.1093/bioinformatics/btaa929); pmid: [33135730](https://pubmed.ncbi.nlm.nih.gov/33135730/)
87. D. R. Kelley *et al.*, Sequential regulatory activity prediction across chromosomes with convolutional neural networks. *Genome Res.* **28**, 739–750 (2018). doi: [10.1101/gr.227819.117](https://doi.org/10.1101/gr.227819.117); pmid: [29588361](https://pubmed.ncbi.nlm.nih.gov/29588361/)
88. A. Majdandzic, C. Rajesh, P. K. Koo, Correcting gradient-based interpretations of deep neural networks for genomics. *Genome Biol.* **24**, 109 (2023). doi: [10.1186/s13059-023-02956-3](https://doi.org/10.1186/s13059-023-02956-3); pmid: [37161475](https://pubmed.ncbi.nlm.nih.gov/37161475/)
89. T. L. Bailey, J. Johnson, C. E. Grant, W. S. Noble, The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015). doi: [10.1093/nar/gkv416](https://doi.org/10.1093/nar/gkv416); pmid: [25953851](https://pubmed.ncbi.nlm.nih.gov/25953851/)
90. M. Gotta, T. Larache, S. M. Gasser, Analysis of nuclear organization in *Saccharomyces cerevisiae*. *Methods Enzymol.* **304**, 663–672 (1999). doi: [10.1093/nar/gkv416](https://doi.org/10.1093/nar/gkv416); pmid: [10372389](https://pubmed.ncbi.nlm.nih.gov/10372389/)
91. J. G. Henikoff, J. A. Belsky, K. Krassovsky, D. M. MacAlpine, S. Henikoff, Epigenome characterization at single base-pair resolution. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18318–18323 (2011). doi: [10.1073/pnas.1110731108](https://doi.org/10.1073/pnas.1110731108); pmid: [22025700](https://pubmed.ncbi.nlm.nih.gov/22025700/)
92. J. Serizay, Sequence-dependent activity and compartmentalization of foreign DNA in a eukaryotic nucleus, Version 1.1.0, Zenodo (2024); <https://zenodo.org/records/14037364>.

ACKNOWLEDGMENTS

We are grateful to M. Lanzer (Heidelberg University) and C. Sanchez for providing us with *P. falciparum* YACs, and to S. Baumgartner for long-read sequencing of these YACs. We also thank B. Dujon, M. Fromont-Racine, A. Jacquier, G. Liti, B. Llorente, M. Nollmann, C. Saveanu, B. Le Tallec, and all members of the laboratory Régulation Spatiale des Génomes for fruitful comments on the work and the manuscript. We thank C. Matthey-Doret and G. Mercy for helping during the earlier steps of the project, and E. Turc and L. Lemée from the Institut Pasteur Biomics Platform; the PICT-IBISA Imaging Facility of the Institut Curie, a member of the France Biomaging National Infrastructure (ANR-10-INBS-04), and particularly M. Garnier for his help with FISH quantification. Strains RSG_Y711, RSG_Y681, and RSG_Y1392 (and derivatives) are available from C. Lartigue under a material transfer agreement with the INRAE. **Funding:** This work was supported by the European Research Council under the Horizon 2020 Program (ERC grant agreement 771813) and Agence Nationale pour la Recherche (ANR-19-CE13-0027-02) to R.K. R.K., F.B., J.M., and An. T. also received support from Agence Nationale pour la Recherche (ANR-22-CE12-0013-01). D.L. received support from ANR (ANR-21-CE12-0040). A.M. is supported by an FRM fellowship. C.C. was supported by a Pasteur-Roux-Cantarin fellowship. J.S. was supported by an ARC fellowship. Biomics Platform, C2RT, Institut Pasteur, Paris, France, is supported by France Génomique (ANR-10-INBS-09-09) and IBISA for processing and

sequencing RNA samples. **Author contributions:** Conceptualization: C.C., L.M., J.S., J.M., and R.K. Methodology: L.M., C.C., C.L., J.M., and R.K. Software: J.S. Validation: C.C., J.S., and L.M. Investigation: L.M. and C.C. with contributions from M.P., F.G., An.P., A.B., A.E., Ag.T., M.R. and F.B. Formal analysis: J.S. (all data processing and integration), A.W. and E.R. (CNN models), with contributions from C.C., L.M., F.B., M.R., An.T. Data Curation: J.S., with contributions from C.C. and L.M.. Resources: L.M., R.K., G.G. and C.L. Visualization: J.S. Writing - original draft preparation: C.C., J.S., L.M., J.M., and R.K. Writing - editing: all authors. Writing - revisions: J.S., L.M., A.W., J.M., and R.K. Supervision: An.T., D.L., J.M., R.K. (C.C. co-supervised a student.) Funding acquisition: An.T., D.L., F.B., J.M., and R.K. Project Administration: R.K. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Sample description and raw sequences for all figures are accessible on the GEO database through accession number GSE217022. Rearranged genome reference sequences for all strains generated in this study are provided via Zenodo (<https://doi.org/10.5281/zenodo.14024599>). Micrographs obtained after FISH imaging and used in Figs. 3E and S6, C and D, are also provided via Zenodo. All custom-made code of the analysis of sequencing data is available online (<https://github.com/koszullab/>) and is archived on Zenodo (<https://doi.org/10.5281/zenodo.14037364> (92)). Open-access versions of the programs and pipelines are available online on the github account of the Koszul lab: Hicstuff (<https://github.com/koszullab/hicstuff>, version 3.1.2) and Chromosight (version 1.4.1, <https://github.com/koszullab/chromosight>) (79). Bowtie2 (version 2.4.5) is available online at <https://bowtie-bio.sourceforge.net/bowtie2/>, SAMtools (version 19) is available online at <http://www.htslib.org/>, Bedtools86 (version 2.29.1) is available online at <https://bedtools.readthedocs.io/en/latest/content/installation.html> and Cooler (version 0.8.7–0.8.11) is available online at <https://cooler.readthedocs.io/en/latest/> (69). tidyCoverage (version 1.1.2) is available at <https://github.com/j32264/tidyCoverage>, HiContacts (version 1.5.0) is available at <https://github.com/j32264/HiContacts> (73). All deep learning codes are available at https://github.com/Alexwestbrook/bacterial_genome. Any additional information, including custom-made code required to reanalyze the data reported in this paper, is available from the lead contact upon request. **License information:** Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. <https://www.science.org/about/science-licenses-journal-article-reuse>

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adm9466

Supplementary Text

Figs. S1 to S11

Tables S1 to S5

References (93–97)

MDAR Reproducibility Checklist

Submitted 22 November 2023; resubmitted 26 September 2024

Accepted 21 November 2024

10.1126/science.adm9466