Exogenous chromosomes reveal how sequence composition drives chromatin assembly, activity, folding and compartmentalization

Christophe Chapard[1,Ψ], Léa Meneu[1,2,Ψ], Jacques Serizay[1,Ψ] , Etienne Routhier[4], Myriam Ruault[3], Amaury Bignaud[1,2], Géraldine Gourgues[5], Carole Lartigue[5], Aurèle Piazza[1,#], Angela Taddei[3], Frédéric Beckouët[6], Julien Mozziconacci[4,*] and Romain Koszul[1,*]

Affiliation

[1] Institut Pasteur, CNRS UMR 3525, Université Paris Cité, Unité Régulation Spatiale des Génomes, 75015 Paris, France

[2] Sorbonne Université, College Doctoral

[3] Institut Curie, PSL University, Sorbonne Université, CNRS, Nuclear Dynamics, 75005 Paris, France

[4] Laboratoire Structure et Instabilité des génomes UMR 7196, Muséum National d'Histoire Naturelle, Paris 75005, France

[5] Univ. Bordeaux, INRAE, Biologie du Fruit et Pathologie, UMR 1332, F-33140 Villenave d'Ornon, France

[6] Molecular, Cellular and Developmental biology department (MCD), Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, UPS, 31062, Toulouse, France

[#] Present address: Univ Lyon, ENS, UCBL, CNRS, INSERM, Laboratory of Biology and Modelling of the Cell, UMR5239, U 1210, F-69364, Lyon, France

Ψ These authors contributed equally

* Corresponding authors: romain.koszul@pasteur.fr, julien.mozziconacci@mnhn.fr

Abstract

Genomic sequences co-evolve with DNA-associated proteins to ensure the multiscale folding of long DNA molecules into functional chromosomes. In eukaryotes, different molecular complexes organize the chromosome's hierarchical structure, ranging from nucleosomes and cohesin-mediated DNA loops to large scale chromatin compartments. To directly explore the causal relationships between the DNA sequence composition and the spontaneous loading and activity of these DNA-associated complexes in the absence of co-evolution, we characterized chromatin assembly and activity in yeast strains carrying exogenous bacterial chromosomes that diverged from eukaryotic sequences over 1.5 billion years ago. Combining this synthetic approach with a deep learning-based analysis, we show that, in this cellular context, the sequence composition drives nucleosome assembly, transcriptional activity, cohesin-mediated looping, and chromatin compartmentalization. These results are a step forward in understanding the molecular events at play during natural horizontal gene transfers as well as synthetic genomics engineering projects.

Introduction

Genomic sequences coevolve with molecular complexes to regulate processes such as gene expression, DNA replication and the functional organization of the genome in 3D. This interdependence raises intriguing questions in light of transfer of genetic material across species such as horizontal gene transfers (HGT), introgression, and invasion by DNA viruses or retrotransposons [1–4], since the base composition (e.g.GC%) of these elements can significantly diverge from their host's genome (e.g. the introgression of a 12% GC-divergent 1Mb sequence in *Lachancea kluyveri* [5]). In the synthetic genomics field, large DNA molecules carrying entire exogenous metabolic and genetic pathways are introduced into chassis microbial strains or cell lines [6–12]. How the host genome copes with these exogenous sequences remains unclear, with some studies reporting variable levels of transcription by the host cell [6,13], outlining the need to better understand how an exogenous DNA molecule is processed by a host' molecular machinery. Addressing this question also holds significant promise for our understanding of host genome metabolism, as it means better characterizing the molecular processes involved. The sequence composition of genomes can vary between different species (e.g. 20 to 65% GC), but also within genomes. Local sequence composition is related to chromatin activity, with high GC regions on average enriched in actively transcribed sequences [14]. Chromatin structure also depends on underlying sequence composition at the nucleosomal level [15], and on the activity of structural maintenance of chromosome (SMCs) ring complexes (e.g. cohesin) [16–18] at higher-order levels. In GC-rich gene-rich regions, CTCF binding to chromatin leads to the formation of cohesin-mediated loops and topologically associating domains (TADs). In contrast, in AT-rich, inactive regions, CTCF binding sites are often methylated leading to the absence of these structures [19]. In addition, variations in sequence composition along metazoan and plant chromosomes also correlate with the global genome 3D fold, as chromatin organized on sequences of similar composition tends to coalesce in distinct compartments [20,21]. In contrast, the shorter, compact genomes of eukaryotic microorganisms such as budding yeast do not exhibit large GC variations, nor compartments.

Here we explore the causal relationships between the DNA sequence and the structure and activity of chromatin, including transcription, replication, and multiscale 3D organization. To this end, we explore the behavior of entire bacterial chromosomes with highly divergent GC contents introduced in the *S. cerevisiae* genome. We map nucleosome assembly, polymerase activity and cohesin binding and show that regions with different GC contents lead to different regional

chromatinization profiles. These regions self-associate in 3D leading to the spontaneous formation of active / inactive chromosomal compartments similar to the ones observed in complex multicellular organisms. In parallel, we trained neural networks (NN) on the chromatin profiles along native yeast chromosomes to learn their sequence determinants and predict that the sequence composition spontaneously determines the fate of any DNA molecule introduced into a given cellular context, including its organization from nucleosome positioning up to 3D folding and transcriptional activity.

Results

Adaptation of supernumerary bacteria chromosomes to yeast

To investigate the behavior of large sequences which have not evolved in a eukaryotic context, we exploited *S. cerevisiae* strains carrying an extra 17th circular chromosome made either from the *Mycoplasma mycoides* subsp. *mycoides* (referred to as "Mmyco") or *Mycoplasma pneumoniae* ("Mpneumo") genomes (**Methods**; **Table S1**). We linearised these chromosomes, which contain a yeast centromeric sequence and an autonomous replication sequence (ARS) allowing them to self-replicate and segregate [22]. (**Figure 1a; Supplementary Figure 1a**; **Method;** see also Supplementary Figure 2 and results). While the GC content of *S. cerevisiae* is 38%, the GC content of the Mmyco and Mpneumo chromosomes 24% and 40%, respectively (**Figure 1a**). These chromosomes do not impose a significant fitness cost on their eukaryotic host, since each strain displays similar growth rates compared to their parental strains in selective media (**Supplementary Figure 1b**) with an estimated segregation rate similar to a centromeric plasmid [23] (**Supplementary Figure 1c**).

Spontaneous chromatinization of bacterial chromosomes in a eukaryotic context

We profiled the chromatin landscape of both Mmyco and Mpneumo strains, and of a parental wild-type strain, by performing MNase-seq, and H3, H2A, RNA polymerase II (PolII) and cohesin (Scc1) ChIP-seq experiments (Methods; **Supplementary Figure 3a**).

Strikingly, the chromatin composition of bacterial sequences varies greatly depending on which chromosome is inserted. Genome-wide profiles along the Mpneumo chromosome appear similar to wild-type (WT) yeast chromosomes (compare **Figure 1b**, panel **i** with **ii**). Locally, Scc1 peaks across the Mpneumo chromosome display a strong signal-to-noise ratio (SNR) (Figure **Supplementary Figure 3g**), and are preferentially located in nucleosome-depleted and PolII enriched regions, as observed for Scc1 peaks located in yeast

chromosomes of the same strain (Figure **Supplementary Figure 3h**). In sharp contrast, profiles in the Mmyco strain showed several differences from the profiles found in the parental WT strain (compare **Figure 1b**, panel **iii** with **i**). Firstly, ChIP-seq profiles of histones H2A and H3 show that these two histones are present along the Mmyco chromosome, but MNase-seq signal is nevertheless strongly reduced (**Figure 1b, Supplementary Figure 3b**). We found that poly(dA) and poly(dT) tracks, which repel nucleosomes, are over-represented on Mmyco, while AA/TT 10-bp periodicity, which facilitates DNA bending around histone cores, is reduced (**Supplementary Figure 3c,d**). These results suggest that histones can bind the Mmyco chromosome but cannot form well-assembled nucleosomes. MNase has more affinity for AT-rich sequences [24] and this property may also decrease the MNase-seq coverage along Mmyco.

Secondly, the average PolII profile along Mmyco is lower than typical WT profiles (**Figure 1b**), whereas Scc1 is enriched over the whole Mmyco chromosome, without the distinct peaks observed on endogenous chromosomes. Scc1 levels also appear strongly and moderately reduced at centromeres and along the arms of the endogenous *S. cerevisiae* chromosomes, respectively (**Figure 1b**, panel **i** and panel **iii, Supplementary Figure 3e**). This suggests that the Mmyco chromosome titrates predominantly a dynamic pool of cohesins at yeast centromeres during metaphase, and less so the stable cohesins involved in sister chromatid cohesion along the arms.

These results show that large exogenous bacterial chromosomes placed in a eukaryotic context spontaneously display some eukaryotic chromatin characteristics: histones, polII and cohesins all bind bacterial DNA irrespectively of GC content. However, striking differences in the observed chromatin profiles led us to define two chromatin types. The chromatin landscape over Mpneumo, whose 40% GC content is close to the native *S. cerevisiae* GC content, is similar to the landscape found on yeast chromosomes. We will hereafter refer to this type of chromatin as “Y” for yeast-like. On the other hand, the chromatin landscape over Mmyco, with a low 24% GC content, displays flatter patterns for nucleosomes, polII and cohesins, with lower average values for both well-assembled nucleosomes and polII and an unusual high value for cohesins in G2. We will hereafter refer to this type of chromatin as “U” for unstructured.

GC content drives chromatin composition of exogenous chromosomes

The different GC content values of Mmyco and Mpneumo prompted us to investigate if the GC content could explain the establishment of Y and U chromatin. To do so, we turned to convolution neural networks (CNN), a computational approach increasingly used in genomics to learn genome annotations from the underlying DNA sequence [25,26] (**Supplementary Figure 1d)**.

We trained CNNs to learn MNase-seq, and Scc1 or PolII ChIP-seq profiles using yeast chromosomes I to XV sequences (Methods) [27]. For each strain, the trained CNNs predict these three profiles with good accuracy on the held out sequence of yeast chromosome XVI (correlation of 0.41, 0.55 and 0.45 for MNase, Scc1 ChIP and PolII ChIP signals, respectively; **Figure 2a-c**), confirming that genomic sequence is a primary determinant of the chromatin landscape [27–31]. Note that individual cohesin peaks at centromeres were not correctly predicted by the trained NN (**Figure 2b**).

Next, we applied the trained CNNs to Mmyco and Mpneumo sequences and found that the rules learned on yeast sequences were sufficient to predict the features experimentally characterized on both exogenous sequences. Not only the predictions correctly recapitulate the features of Y chromatin on Mpneumo, but also the U chromatin features on Mmyco sequences, notably: (1) a lower MNase-seq coverage; (2) a broad enrichment of Scc1, with no discrete peaks and (3) an absence of PolII peaks (**Figure 2a-c**).

To further characterize the link between the GC content and the chromatin composition and activity, we predicted the MNase-seq, PolII and Scc1 ChIP-seq profiles on artificial, randomly generated 30 kb sequences with variable GC contents ranging from 0 to 100 % (**Figure 2d**-**f**, Methods). All predicted signals display a strong variability with respect to the GC content. The Mnase average occupancy observed along endogenous chromosomes was only recovered for a GC % between 40 and 60, whereas for higher and lower values it was strongly decreased. The cohesin signal showed a gradual decrease for increasing GC %, and the PolII signal was predicted to be minimal between 20 and 40 % and increased strongly for smaller and larger values. These predictions are in good agreement with the experimental results obtained for both bacterial and yeast sequences, suggesting that the trained CNNs are indeed able to correctly predict these patterns on a broad range of sequences.

Taken together, these findings show that the GC content of any sequence, either exogenous or endogenous, is a major determinant of the spontaneous establishment of the chromatin and functional landscape of a given species.

Transcriptional activity of bacterial genomes in a yeast context

We next explored whether transcription occurs on the bacterial chromosomes harboring either Y or U chromatin types by performing total, stranded RNA-seq analysis (**Figure 3a,b; Method**). Consistent with PolII ChIP-seq profiles (**Figure 1b**), the whole Mpneumo chromosome is transcribed to levels similar to those of endogenous yeast chromosomes (**Figure 3a ii**). Mpneumo transcription tracks are significantly longer than yeast genes (4.9 kb vs 3.4 kb, p-

value < 2e-4) and do not systematically display clear boundaries (**Figure 3b**). Transcription along Mpneumo does not preferentially occur over bacterial gene bodies nor seem to initiate at bacterial TSSs, also consistent with an absence of PolII enrichment at these loci (**Supplementary Figure 4a-c**).

While the Y chromatin type found on Mpneumo is transcribed, the Mmyco U chromatin type is only sparsely and lowly transcribed (**Figure 3a iii**), again in good agreement with the poor levels of PolII deposition (**Figure 1b**). This silencing is not correlated with the recruitment of the SIR Sir2/3/4 complex (**Supplementary Figure 4d**) and likely reflects the impairment of a proper nucleosomal organization with clear nucleosome depleted regions, known to be required at TSS [32].

Surprisingly, the orientation of transcription tracks of the two bacterial chromosomes in a yeast environment follows, on average, the orientation of the bacterial genes annotated along these sequences although the transcription machinery has diverged billions years ago (**Figure 3c**, see the orientation of genes and stranded tracks in **Figure 3b**).

Interplay between transcription and cohesin localization

On yeast and Mpneumo chromosomes cohesin accumulates during G2/M at sites of convergent transcription, as expected (Lengronne et al. 2004; Glynn et al. 2004) (**Figure 3b**, green dotted lines, **Supplementary Figure 4e**). Comparison of the Scc1 ChIP-seq and stranded RNA-seq profiles over the Mpneumo chromosome suggests that in Y chromatin regions sharper cohesin peaks are associated with sharper transcriptional convergence (**Figure 3b**, arrows) and that the strongest Scc1 peaks are correlated with a stronger transcriptional convergence (**Figure 3d**, **Supplementary Figure 4f,** Methods). In sharp contrast, high cohesin levels span the entire transcriptionally silent U chromatin regions of Mmyco, with no clear peaks, most likely reflecting the lack of convergent transcription sites (**Figure 1b** and **Figure 3b**). Taken together, these results suggest that cohesin can bind DNA in absence of active transcription (and/or of proper chromatinization) *in vivo*, but accumulates at convergent transcription sites only in presence of active transcription.

Spatial organization of exogenous bacterial chromosomes in a eukaryote context

To assess the 3D organization of Y and U chromatin types, we performed capture of chromosome conformation (Hi-C) experiments [20,35] on the Mmyco or Mpneumo strains arrested either in G1 or G2/M (**Figure 4a, b**).

In G1-arrested cells, yeast chromosomes exhibit the expected Rabl configuration, with clustered centromeres and subtelomeric regions colocalizing at the nuclear periphery, resulting in a dotted pattern in Hi-C contact maps [36–38]. For both Mmyco and Mpneumo chromosomes, the added yeast centromeric sequence clusters with the native yeast centromeres.

However, the two bacterial chromosomes, being either covered with Y or U chromatin, exhibited very distinct structural characteristics (**Figure 4a**). The Y-type Mpneumo chromosome behaves as an endogenous *S. cerevisiae* chromosome (**Figure 4a**): (1) the trans-contacts (**Supplementary Figure 5a**) and long-range cis-contacts (**Figure 4c**) are similar to those found for yeast chromosomes; (2) the slope of the p(s) averaged contact probability as a function of the genomic distance (s) is close to -1.5 both for yeast and Mpneumo chromosomes, a value corresponding to the typical polymer structure observed in simulations (**Figure 4c**) [39]; (3) DNA-FISH labeling of Mpneumo DNA reveals an extended and contorted structure within the intranuclear space, confirming the intermixing of the chromosome with the actively transcribed native chromosomes (**Figure 4e, Supplementary Figure 5g,h**). In contrast, in the U-type Mmyco chromosome, (1) only few trans-contacts with the endogenous chromosomes were detected (**Supplementary Figure 5a**), and predominantly with the 32 yeast subtelomeric regions (**Figure 4a**, dotted rectangles; **Supplementary Figure 5b**); (2) the slope of the p(s) curve also changes dramatically towards a value of -1 at shorter distances, corresponding to a polymer globule (**Figure 4c**) [20]; (3) DNA-FISH labeling of the Mmyco chromosome confirmed the globular conformation of the chromosome and revealed its preferential position at the nuclear periphery, reflecting its proximity with sub/telomeres (**Figure 4e; Supplementary Figure 5g,h**).

We then generated Hi-C contact maps of both strains synchronized in G2/M. In yeast, chromosome compaction and individualization in G2/M-arrested cells reflect cohesin-mediated chromatin folding into arrays of loops. This compaction results in a typical p(s) curve that presents an inflection point around the average loop length [38,40,41]. The decrease of interchromosomal contacts (**Figure 4b, Supplementary Figure 5a)** and the change in p(s) (**Figure 4d, Supplementary Figure 5c,d**) observed for both native and bacterial chromosomes indicate mitotic compaction of both Y and U chromatin types, although they display different contact patterns at short scales. Endogenous yeast and Mpneumo chromosomes exhibit focal enrichments of medium-range contacts corresponding to anchored loops (black triangles on **Figure 4f**) [40,42,43]. 59 loops were called across the Mpneumo chromosome, spanning over slightly longer genomic distances (28 kb on average) than the loops found along yeast sequences (22 kb) (**Supplementary Figure 5e**). The dotted grid pattern in Mpneumo is

reminiscent of multiple DNA loops observed along mammalian interphase chromosomes [44]. This suggests that some loop anchors have the ability to be involved in loops of various sizes, a phenomenon not observed across native *S. cerevisiae* chromosomes (**Figure 4f**). On the other hand, Mmyco chromosome, albeit covered by cohesin, displayed a strikingly different pattern with no visible loops (**Figure 4g**), suggesting that cohesins are able to promote compaction but that the loop anchors are not preferentially positioned along U chromatin. Overall, these observations highlight two different spatial configurations of Mmyco and Mpneumo exogenous chromosomes. On the one hand, the inactive Mmyco chromosome in G1 is compacted and positioned at the periphery of the nucleus, away from the active chromosome set. Upon S-phase and G2 arrest, the DNA is further compacted in a cohesin dependent manner with no visible loops at discrete positions. On the other hand, the active Mpneumo chromosome behaves as endogenous yeast chromosomes, displaying discrete mitotic DNA loops in G2/M.

Cohesin compacts exogenous bacterial chromosomes in yeast

Mmyco and Mpneumo chromosomes display different Scc1 binding landscapes (**Figure 1b, 1c**), and we therefore investigated the relationship between cohesin deposition patterns and the corresponding contact maps. In yeast, cohesin peaks correspond to the loop anchors in metaphase Hi-C maps [40,42,43]. Along Mpneumo, strong Scc1 peaks are also close to anchors with a median spacing of 566 bp between consecutive anchors) (**Figure 4f, 4h**). This suggests that these loops are formed through the process of loop extrusion (LE) by which cohesin complexes organize DNA by capturing small loops and gradually enlarging them [45,46]. To test this hypothesis, we performed Hi-C in absence of the cohesin-holocomplex member Wpl1 (human Wapl), which impairs cohesin removal and typically results in longer loops readily visible by a shift of the inflection point in the p(s) curve [43,46]. In absence of Wpl1, we observed increased long-range contacts in all chromosomes, including Mmyco and Mpneumo (**Figure 4i; Supplementary Figure 5f**), in agreement with extended loop formation. These results show that Y and U chromatin are prone to cohesin-dependent loop extrusion in G2/M, although the generated loops are not focally stabilized along the Mmyco chromosome.

Mosaic chimeric chromosomes display spontaneous, chromatin type-dependent DNA compartmentalization

The structural and functional features of the Y- and U-type chromatin are reminiscent of the euchromatin and heterochromatin described along metazoan chromosomes [20]. We therefore wondered whether they can coexist on a single chromosome. Using CRISPR-Cas9, we first

fused the Mmyco chromosome with the native yeast chromosome XVI (XVIfMmyco chromosome). We then induced translocations to generate two other strains with chromosomes harboring alternating U and Y chromatin regions as small as 50 kb (XVIfMmycot1 and XVIfMmycot2; **Figure 5a**; **Supplementary Figure 6a-c** Methods). ChIP-seq profiles of RNA PolII on Mmyco regions were similar before and after translocation, demonstrating that PolII binding is independent of the broader regional context (**Supplementary Figure 6d**). The overall Hi-C contact maps of XVIfMmyco strain synchronized in G1 showed little differences compared to the parental Mmyco strain (but for the deleted and fused regions; e.g. Lazar-Stefanita et al. 2022; Shao et al. 2018) (**Figure 5b**, compare panel **i** with **ii**). The Y-type chr. XVI arm intermixes with the other 15 yeast chromosomes while the U-type Mmyco arm remains isolated from the rest of the genome, contacting subtelomeric regions (**Figure 5b**). In the translocated strains, however, the alternating chromatin type regions resulted in striking checkerboard contact patterns within XVIfMmycot1 and XVIfMmycot2 chromosomes (**Figure 5b**, panels **iii** and **iv**). This pattern reflects the fact that U-type regions make specific contacts over long distances, bypassing the Y-type regions found in between (**Figure 5c**; **Supplementary Figure 6e** G1). Similarly, Y-type regions of chromosome XVI also are also involved in specific Y-Y contacts over longer distances (**Figure 5c**). Intra U and Y regions contact decay is also different in the two chromatin types, U chromatin being prone to longer range contacts in cis (**Supplementary Figure 6e**). These results unambiguously show that alternate GC-regular Y and GC-poor U regions along chromosomes are spontaneously segregated within the nuclear space in G1 (**Figure 5d, Supplementary Figure 6h)**, a phenomenon clearly reminiscent of eu- and hetero chromatin compartments found in higher eukaryotes [20].

We next investigated the behavior of mosaic chromosomes in G2/M by performing Scc1 ChIP-seq and Hi-C. First, the Scc1 deposition pattern was highly conserved along native and mosaic chromosomes. Notably, the loss of cohesin at all centromeres was maintained (**Supplementary Figure 6d**). In addition, cohesin coating remained enriched over U-type Mmyco sequences although to a lower level, including over the 50 kb encompassed within 850 Mb of yeast DNA on XVIfMmycot2 chromosome. These results show that cohesin is loaded along inactive Mmyco regions independently of the proximity with the centromere (as along WT chromosomes), and independently of transcription. The larger chromatin context plays little if no role in this cohesin-mediated loading. The only noticeable differences resulting from the fusion/translocation events in G2/M Hi-C contact maps (**Supplementary Figure 6f**) were observed at the junction between Mmyco and the endogenous yeast sequences. There, a DNA

loop appears over the boundary (**Supplementary Figure 6g**). The two anchors of the loop consisted, on the one hand, in a strongly cohesin-enriched region already present on chr. XV sequence, and on the other hand, in another region slightly enriched in cohesin already present on the Mmyco sequence (**Supplementary Figure 6g**). Together with observations from the previous section, these results show that Y and U chromatin types are both prone to dynamic cohesin-dependent DNA compaction and loop formation in G2/M phase (**Figure 5d**) and further suggest that loop extrusion can proceed through the Y/U chromatin junction to form a loop bridging the nearest cohesin enrichment sites on each side [19,47].

At a broader scale, Hi-C maps of the mosaic chromosomes in G2/M show that U-type compartments disappear upon cohesin-mediated compaction (**Supplementary Figure 6f**). The analysis of distance-dependent contact frequency shows that all chromosome regions behave similarly at this stage (**Supplementary Figure 6e**, G2/M). This observation shows that, as observed in multicellular organisms, LE mediated metaphase chromosome compaction abolishes chromatin compartments to prepare chromosomes for segregation (**Figure 5d)** [45,48,49].

Discussion

Chromatin structure and activity is driven by the sequence composition

In this work, we show that bacterial DNA sequences introduced in a eukaryotic cellular context into which they have not evolved behave in a way that is largely predictable by neural networks trained on endogenous sequences. This suggests that the cellular machinery is extremely robust, and indiscriminately processes any sequence. The GC% content nevertheless appears as an important driver of chromatin activity as two different values lead to the formation of two archetypes of chromatin, that we call here Y and U for Yeast-like and Unstructured, reminiscent of the mammalian active euchromatin and inactive heterochromatin found in A and B compartments, respectively. In both chromatin types, DNA is covered with histones, as this association only depends on the attraction between positively charged proteins and the negatively charged double helix, but low levels of GC are not compatible with the wrapping of DNA into nucleosomes. We hypothesize that this could be due to the frequent occurrence of poly(A) motifs that are known to prevent the formation of nucleosomes [27]. The coating of the DNA molecule with histones may result in the presence of intra nucleosomal patches / hydrophobic interaction along the molecule that are prone to sticky interactions. The resulting self-attracting potential would lead to the collapse of the whole region into a globule [50], and to

the colocalization of dispersed U-type chromatin domains into a new compartment, in agreement with both the observed p(s) and FiSH microscopy images. We also observed a correlation between this compact, globular state and a suppression of transcription, although we cannot conclude here about causal relationship. In metaphase, the SMC cohesin complex binds to all exogenous sequences, regardless of their transcriptional status and GC content, and appears to start extruding DNA loops through an active ATPase dependent process [16,51]. Along Y chromatin, these complexes accumulate at convergent transcription loci [33,43,52]. Along U chromatin, in the absence of well-defined nucleosomes and transcription, translocation of the cohesin complex occurs uniformly all along the molecule, leading to a homogeneous enrichment pattern.

The existence of these alternative chromatin types has important implications in synthetic biology and genomics: the sequence composition should be taken into account when engineering new chromosomal segments. The performance of deep learning models to learn chromatin composition from sequences establishes these approaches as must-have steps in the validation of the compatibility of any sequence design, including data storage, prior synthesis and integration.

Prokaryotic DNA contains some determinants of transcription orientation in eukaryote

Previous works have characterized the behavior of exogenous DNA within the yeast nucleus. For instance, short stretches or small (13 kb) random DNA sequences can be used to generate yeast promoters and to decipher the regulatory logic of transcription activation [53–55]. Here we show that both normal GC% and AT-rich bacterial chromosomes are spontaneously transcribed when introduced in the yeast eukaryotic environment [6,13,56]. Active transcription follows, on average, the orientation of the bacteria genes (**Figure 3c**). Several mechanisms can account for this observation. The Nonsense-Mediated Decay (NMD) complex, which destabilizes eukaryotic transcripts with long 3' UTR, could be frequently activated by bacteria antisense RNA compared to sense RNA. Alternatively, bacteria codon usage bias may also result in enrichment in Nrd1-Nab3-Sen1 (NNS) binding sites along the ncRNA strand, leading to transcription termination [57]. Another, more speculative possibility is that sequence determinants predating the divergence of eucaryotes and procaryotes drive the preferred orientation of the RNA polII. Whatever the cause, 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 loop formation along non-transcribed DNA templates

Our results on long exogenous sequences provide new insights on the mechanisms behind cohesin-mediated loops. They suggest that cohesin (1) can be loaded on long lowly transcribed DNA sequences, independently of centromere vicinity or active promoters, and (2) compacts DNA along these regions, without forming visible DNA loops with discrete anchors. The positions of loop anchors have been proposed to be determined by the combination of convergent transcription, replication fork progression during S phase and/or stably bound cohesin at cohesin associated regions [42,43,58]. Loop positioning, stabilization and maintenance at discrete loci also involves Eco1 dependent acetylation of cohesin Smc3 [43,51,59,60]. We propose that low transcription accounts for the absence of large DNA loop positioning at discrete loci along the Mmyco chromosome. Our results also suggest that transcription may not be considered as a primary molecular motor for cohesin mediated loop extrusion, although transcription interferes with loop extrusion process [58]. In agreement with *in vitro* data [16], cohesin expands DNA loops *in vivo* independently of its chromatinization and transcription status, sequence composition and GC% content, resulting in chromosome compaction.

On a transcribed chromosome, cohesin is strikingly located at convergent transcription sites as was observed for WT yeast chromosomes [33,34]. We suggest that cohesin may be stopped and stabilized at these sites during processes of loop extrusion. Active transcription may therefore inhibit cohesin translocation and/or loading along the fiber, resulting in accumulation of the complex in non-transcribed regions.

Compartmentalization of the genome

The nucleus of WT yeast (and other fungi) is typically separated into two compartments: the active chromatin and the nucleolus containing rDNA. Here, we show that the introduction of non-transcribed AT-rich DNA is sufficient to spontaneously promote the formation of a third compartment reminiscent of the "B" compartment of inactive metazoan heterochromatin. This new compartment is positioned at the periphery of the nucleus, close to the nuclear envelope, so that the active chromatin occupies the central space of the nucleus (**Figure 5d; Supplementary Figure 6h)**. The spontaneous segregation of a long DNA molecule that has not evolved in this context, based solely on its composition and resulting activity, could be highly

significant with respect to genome invasion by exogenous mobile elements. One could predict, based on the properties revealed in our work, that a genome suddenly harboring a large amount of un-transcribed and divergent DNA will spontaneously maintain its original chromatin active while dramatically changing its 3D structure, in a way highly reminiscent of the described architecture in mammalian genomes in which heterochromatin at AT-rich transposable elements is positioned at the nuclear periphery.

Competing interests

The authors declare no competing interests.

Authors Contributions

Conceptualization: CC, LM, JS, JM and RK. Methodology: CC, JS, LM, CL, JM, RK. Software: JS. Validation: CC, JS, LM. Investigation: CC, LM, with contributions from AP, AB, MR and FB. Formal analysis: JS, ER, CC, LM, FB, MR, AT. Data Curation: JS, with contributions from CC, LM. Resources: GG, CL. Visualization: JS, with advice from CC and LM. Writing - original draft preparation: CC, JS, LM, JM and RK. Other drafts: JS, JM, RK. Writing – Review & Editing: all authors. Supervision: JM, RK. CC co-supervised a student. Funding acquisition: RK. Project Administration: RK.

Acknowledgements

We are grateful to Bernard Dujon, Micheline Fromont-Racine, Alain Jacquier, Gianni Liti, Bertrand Llorente, Marcelo Nollmann, Cosmin Saveanu and Benoit le Tallec for fruitful comments on the work and the manuscript. This research 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 RK. RK, JB, JM and AT also received support from Agence Nationale pour la Recherche (ANR-22-CE12-0013-01). CC was supported by a Pasteur-Roux-Cantarini fellowship. JS was supported by a postdoctoral ARC fellowship. We thank all our colleagues from the laboratory régulation spatiale des génomes for fruitful discussions. We especially thank Agnès Thierry and Cyril Matthey-Doret for support during the earlier steps of the project. E. Turc and L. Lemée at Biomics Platform, C2RT, Institut

Pasteur, Paris, France, are supported by France Génomique (ANR-10-INBS-09-09) and IBISA for processing and sequencing RNA samples.

Figure legends

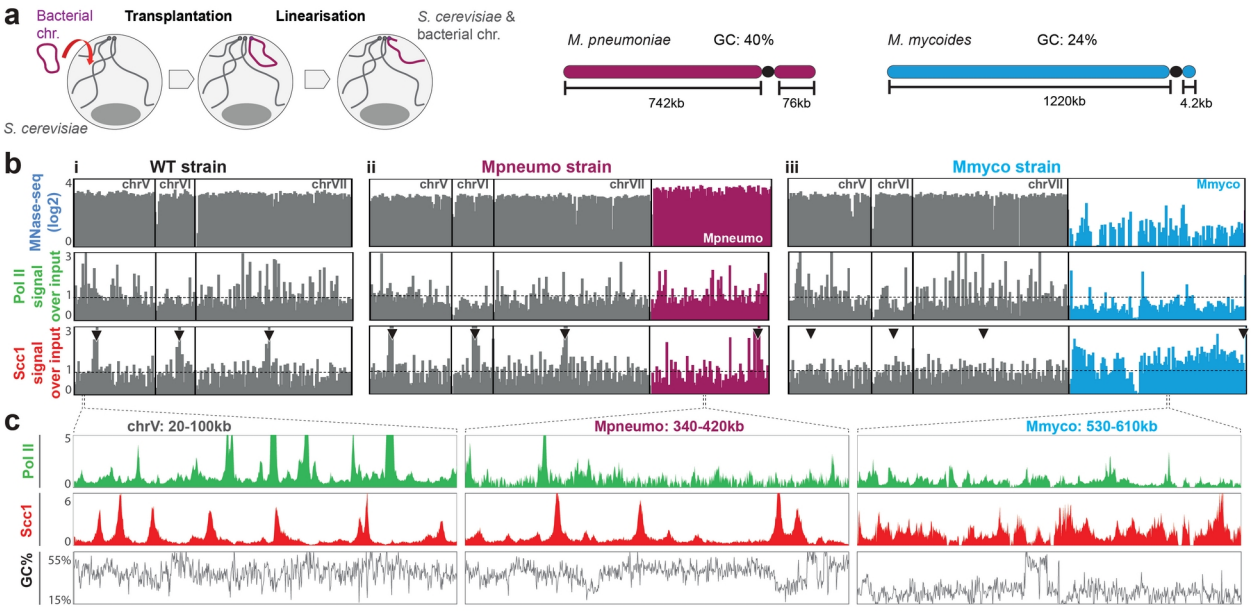


Fig. 1

Figure 1. Chromatin composition of bacterial chromosomes in yeast.

a, Schematic representation of the conversion from circular to linear chromosomes integrated in yeast. The purple and blue colors represent the *M. pneumoniae* (Mpneumo) and *M. mycoides* (Mmyco) bacterial sequence in all figures, respectively.

b, MNase-seq (top), PolII (middle) and Scc1 (bottom) ChIP-seq deposition profiles along three representative yeast chromosomes and, if present, the bacterial chromosome in i) wild-type, ii) Mpneumo and iii) Mmyco strains. Black triangles: centromere positions. Bin size: 10kb.

c, Magnification of PolII and Scc1 ChIP-seq profiles along 80 kb windows of i) a wt yeast chromosome, and ii) the Mpneumo and iii) Mmyco sequences. Bin size: 100bp.

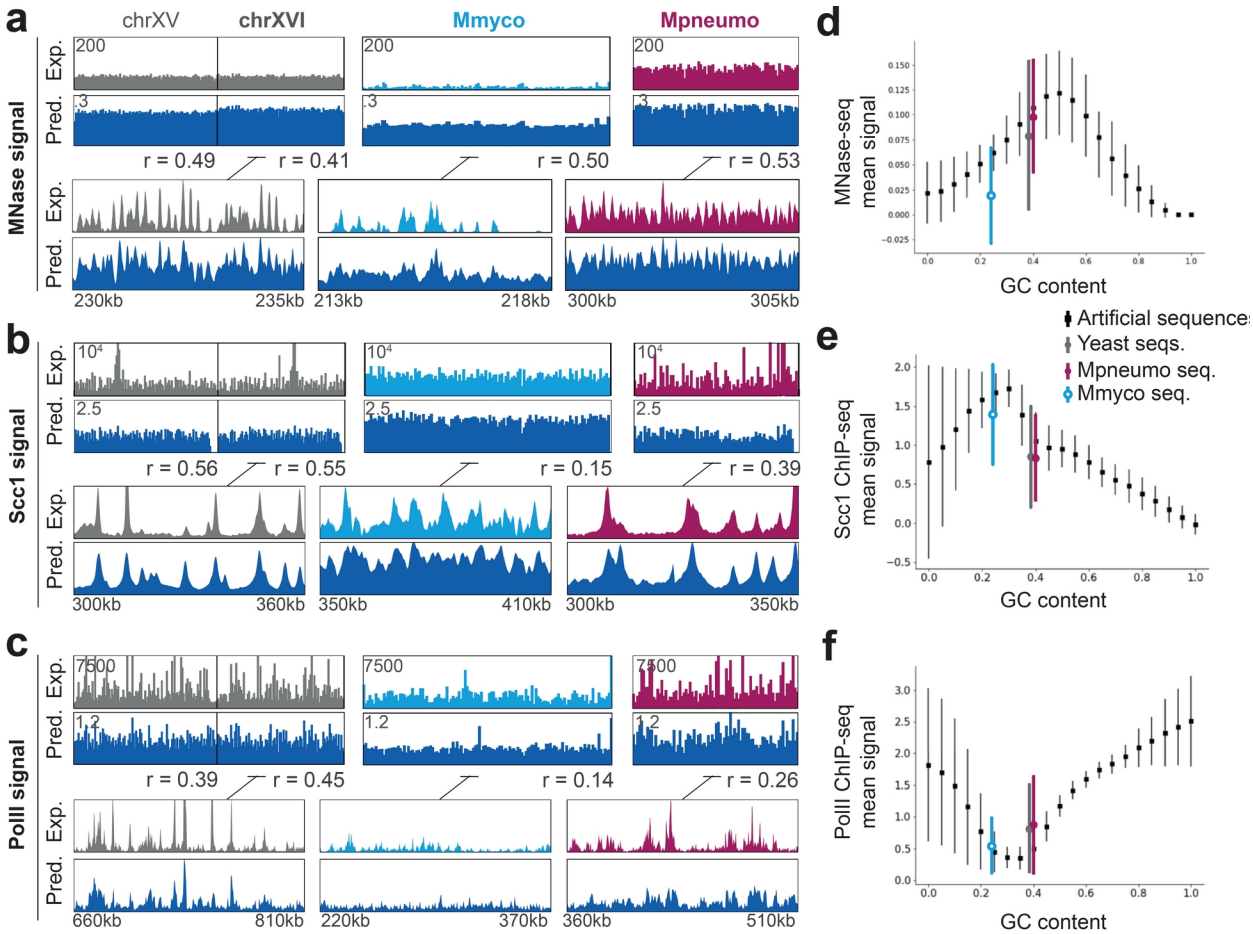


Fig. 2

Figure 2. Deep-learning analysis of chromatin composition along yeast and bacterial chromosomes.

a-c, Experimental (top) and NN prediction (bottom) of MNase-seq, Scc1 and PolII ChIP-seq coverage tracks over yeast and bacterial chromosomes Mmyco and Mpneumo in yeast, at the chromosome level and in 60kb-wide genomic loci.

d–f, Prediction of MNase-seq, Scc1 and PolII ChIP-seq average signal over random sequences with varying GC content. The average signals experimentally measured in yeast and along the Mmyco and Mpneumo bacterial chromosomes are overlaid at their corresponding GC content.

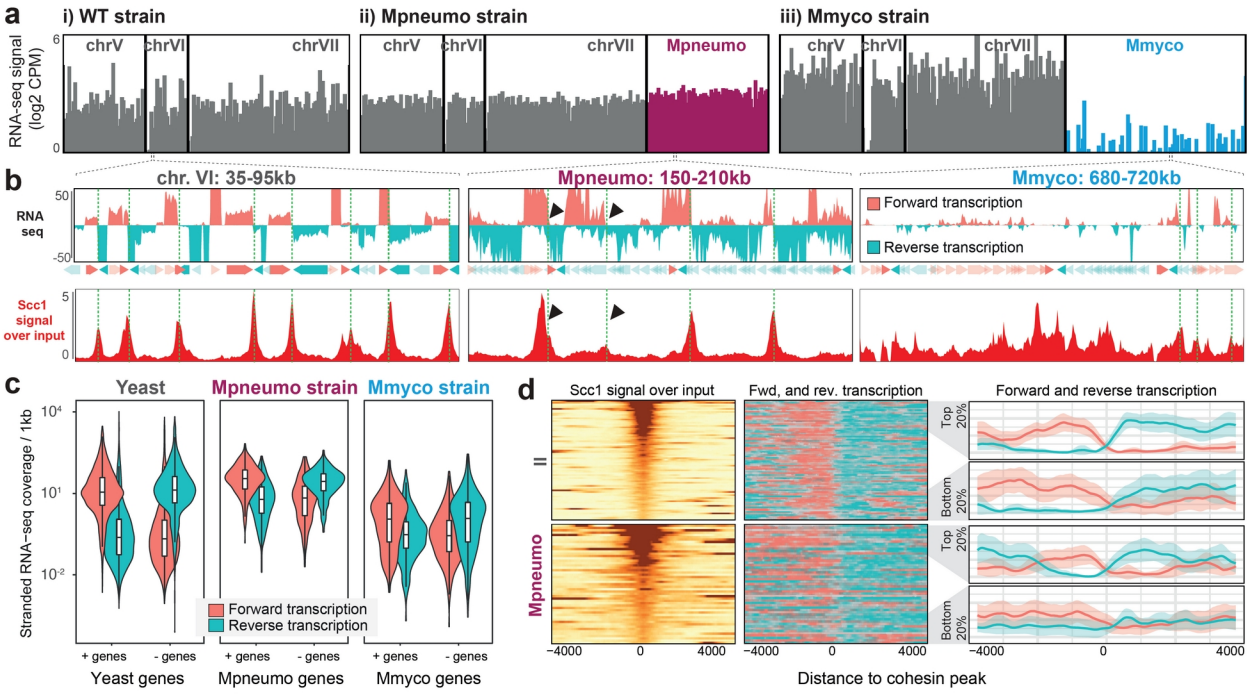


Fig. 3

Figure 3. Expression of exogenic bacterial sequences in budding yeast.

a, RNA-seq profiles along three representative yeast chromosomes and, if present, the bacterial chromosome in i) wild-type, ii) Mpneumo and iii) Mmyco strains.

b, Top: stranded RNA-seq data along a 60kb window along either wt chr. VI, Mpneumo or Mmyco. The pink and turquoise colors represent forward and reverse transcription, respectively. The orientation of the genes along the yeast or bacterial sequences are indicated under the plot. Bottom: Scc1 (cohesin) deposition profiles of the corresponding region. Green dotted lines indicate identified loci of convergent transcription (see Methods).

c, Forward and reverse RNA-seq coverage of forward- and reverse-oriented yeast or bacterial genes in WT, Mpneumo and Mmyco strains. Scores were normalized by the length of each genomic feature.

d, Left: aggregated profile of Scc1 deposition centered at Scc1 peaks (+/- 4kb) called over chr. II (top) or Mpneumo (bottom), with peaks ordered by peak strength. Middle: corresponding stranded transcription tracks, colored according to their forward or reverse orientation. Right: for chr. II or Mpneumo chromosome, average forward and reverse transcription over the 20% strongest or 20% weakest Scc1 peaks.

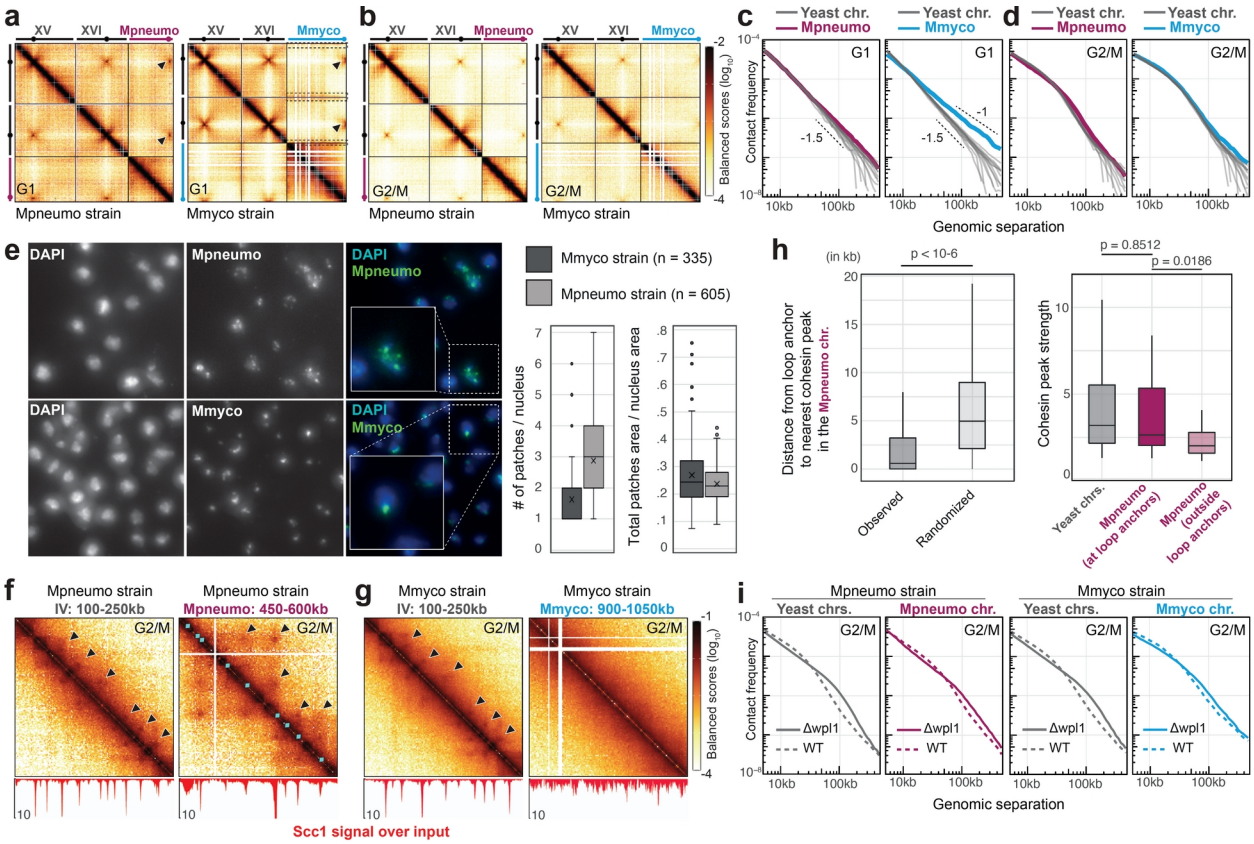


Fig. 4

Figure 4. Folding of exogenic bacterial sequences within the yeast nucleus.

a, b, Hi-C contact maps of representative endogenous and of Mmyco and Mpneumo bacterial chromosomes in G1 (**a**) and G2/M (**b**) (4kb resolution).

c, **d**, Contact frequency as a function of genomic distance 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 (Methods).

f, g, Top: magnification of 150kb windows from Hi-C contact maps in G2/M from either an endogenous or the bacterial chromosome in Mpneumo (**f**) and Mmyco (**g**) strains (1kb resolution). Bottom: Scc1 ChIP-seq deposition profile. Black arrowheads: loops. Cyan diamonds: location of Scc1 peaks indicated on the main Hi-C diagonal in the Mpneumo chromosome.

h, Left: Distance measured between chromatin loop anchors in the Mpneumo chromosome and their nearest Scc1 peak, or expected after randomly shuffling the position of the loop anchors. Right: Scc1 peak strength for peaks located in yeast WT chromosomes or over the Mpneumo chromosome, near loop anchors (< 1kb) or outside of loop anchors.

i, Distance-dependent contact frequency in endogenous yeast chromosomes and in bacterial chromosomes in Mpneumo (left) and Mmyco (right) strains, in "wild-type" (dashed) and in Δwpl1 mutants.

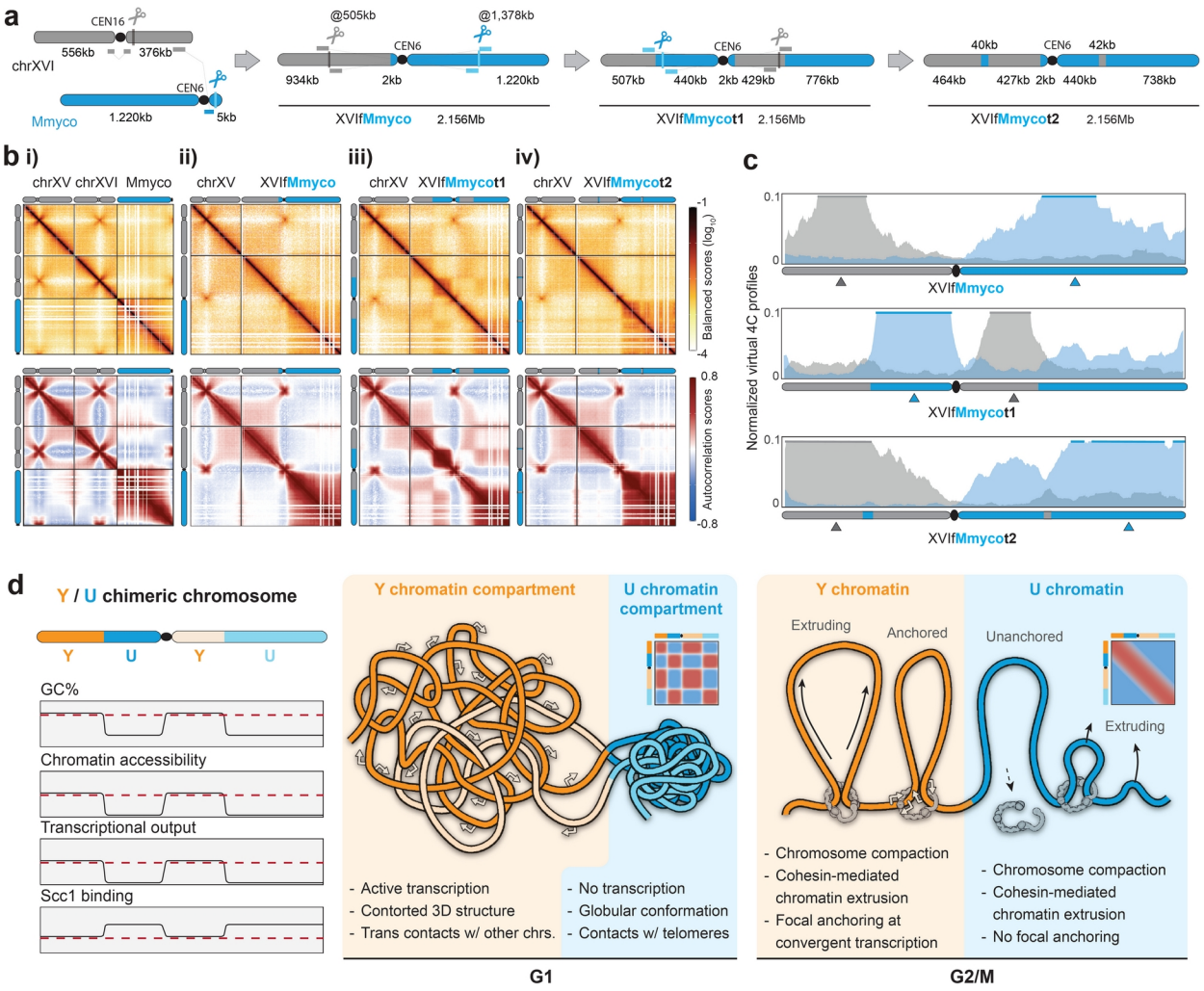


Fig. 5

Figure 5. 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 chr. XV, XVI, and bacterial chromosomes in the Mmyco strain, and in its derivative (i.e. the chr. XVI and Mmyco fusion, and the two strains with translocations resulting in Y and U chromatin alternation; Methods) (4kb resolution). Bottom: auto-correlation contact matrices in wt and mosaic chromosomes strains.

c, Virtual 4C profiles of viewpoints located within yeast chr. XVI (in grey) or Mmyco chromosome segments (in blue), in XVIfMmyco, XVIfMmycot1 and XVIfMmycot2 strains. The large grey/blue arrows indicate the location of each viewpoint. Note the increase in contacts of chr. XVI Y chromatin segments with distant segments of itself, and in alternance with decreased contacts with Mmyco U chromatin segments.

d, Schematic of a chimeric chromosome composed of alternating Y- and U-type chromatin. Left: composition of Y- and U-type chromatin. Middle: Spatial and functional organization in G1. Right: Organization of Y and U-type chromatin segments in G2/M.

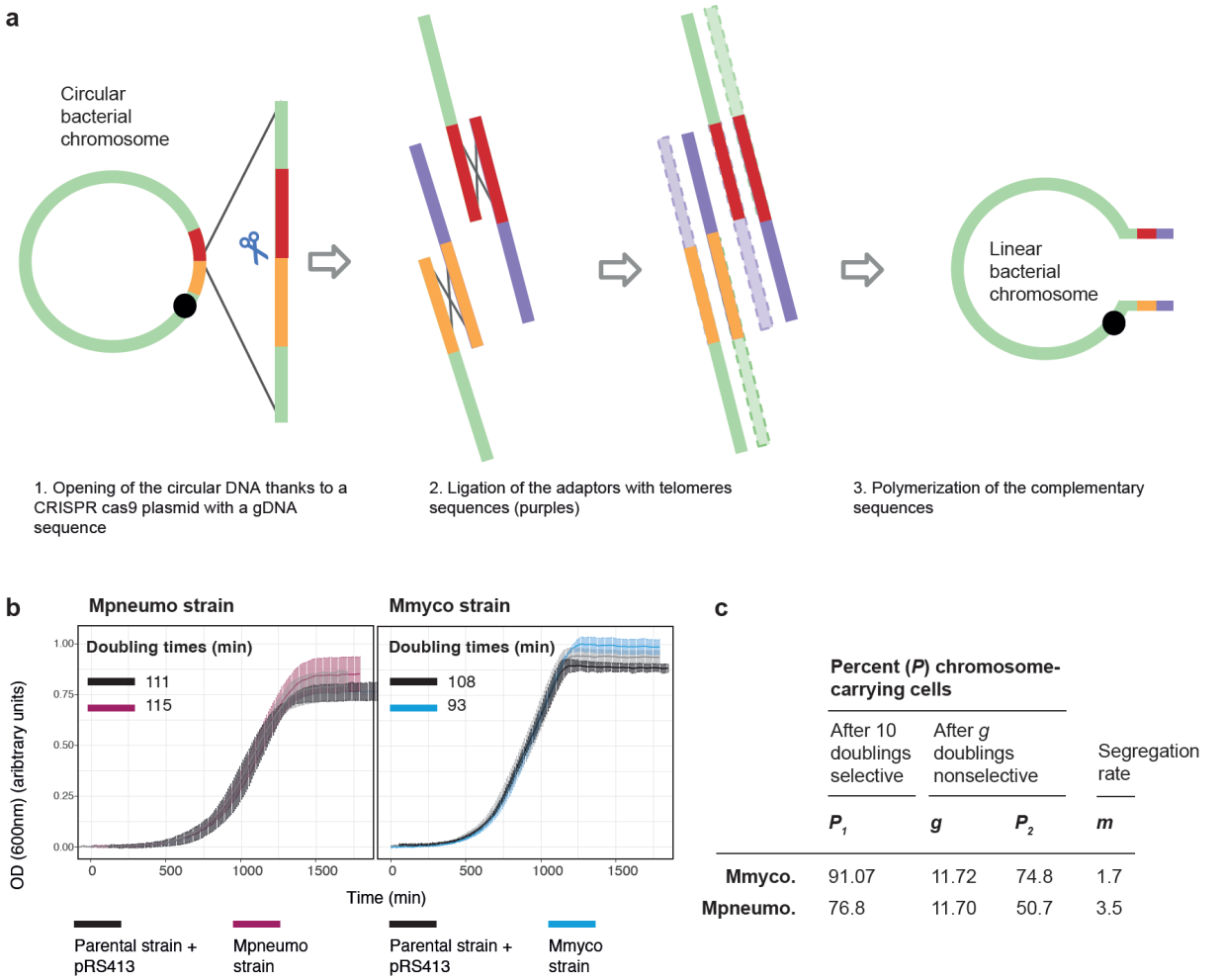
References

1. Crisp, A., Boschetti, C., Perry, M., Tunnacliffe, A. & Micklem, G. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol* **16**, 50 (2015).
2. Edelman, N. B. & Mallet, J. Prevalence and Adaptive Impact of Introgression. *Annu Rev Genet* **55**, 265–283 (2021).
3. Peter, J. *et al.* Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature* **556**, 339–344 (2018).
4. Van Etten, J. & Bhattacharya, D. Horizontal Gene Transfer in Eukaryotes: Not if, but How Much? *Trends Genet* **36**, 915–925 (2020).
5. Payen, C. *et al.* Unusual composition of a yeast chromosome arm is associated with its delayed replication. *Genome Res* **19**, 1710–1721 (2009).
6. Baby, V. *et al.* Cloning and Transplantation of the Mesoplasma florum Genome. *ACS Synth. Biol.* **7**, 209–217 (2018).
7. Coradini, A. L. V., Hull, C. B. & Ehrenreich, I. M. Building genomes to understand biology. *Nat Commun* **11**, 6177 (2020).
8. Currin, A. *et al.* The evolving art of creating genetic diversity: From directed evolution to synthetic biology. *Biotechnol Adv* **50**, 107762 (2021).
9. Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M. & Smolke, C. D. Complete biosynthesis of opioids in yeast. *Science* **349**, 1095–1100 (2015).
10. Meng, F. & Ellis, T. The second decade of synthetic biology: 2010-2020. *Nat Commun* **11**, 5174 (2020).
11. Trolle, J. *et al.* Resurrecting essential amino acid biosynthesis in mammalian cells. *Elife* **11**, e72847 (2022).
12. Pinglay, S. *et al.* Synthetic regulatory reconstitution reveals principles of mammalian Hox cluster regulation. *Science* **377**, eabk2820 (2022).
13. Camellato, P., Brosh, R., Maurano, M. & Boeke, J. Genomic analysis of a synthetic reversed sequence reveals default chromatin states in yeast and mammalian cells. *bioRxiv* (2022).
14. Holmquist, G. P. Evolution of chromosome bands: Molecular ecology of noncoding DNA. *J Mol Evol* **28**, 469–486 (1989).
15. Valouev, A. *et al.* Determinants of nucleosome organization in primary human cells. *Nature* **474**, 516–520 (2011).
16. Davidson, I. F. & Peters, J.-M. Genome folding through loop extrusion by SMC complexes. *Nat Rev Mol Cell Biol* **22**, 445–464 (2021).
17. Mirny, L. & Dekker, J. Mechanisms of Chromosome Folding and Nuclear Organization: Their Interplay and Open Questions. *Cold Spring Harb Perspect Biol* **14**, a040147 (2022).
18. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e24 (2017).
19. Spracklin, G. *et al.* Heterochromatin diversity modulates genome compartmentalization and loop extrusion barriers. *bioRxiv* 2021.08.05.455340 (2021) doi:10.1101/2021.08.05.455340.
20. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
21. Cournac, A., Koszul, R. & Mozziconacci, J. The 3D folding of metazoan genomes correlates

with the association of similar repetitive elements. *Nucleic Acids Res* **44**, 245–255 (2016).
22. Lartigue, C. *et al.* Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science* **325**, 1693–1696 (2009).
23. Christianson, T. W., Sikorski, R. S., Dante, M., Shero, J. H. & Hieter, P. Multifunctional yeast high-copy-number shuttle vectors. *Gene* **110**, 119–122 (1992).
24. Chereji, R. V., Ocampo, J. & Clark, D. J. MNase-Sensitive Complexes in Yeast: Nucleosomes and Non-histone Barriers. *Mol Cell* **65**, 565-577.e3 (2017).
25. Eraslan, G., Avsec, Ž., Gagneur, J. & Theis, F. J. Deep learning: new computational modelling techniques for genomics. *Nat Rev Genet* **20**, 389–403 (2019).
26. Routhier, E. & Mozziconacci, J. Genomics enters the deep learning era. *PeerJ* **10**, e13613 (2022).
27. Routhier, E., Pierre, E., Khodabandelou, G. & Mozziconacci, J. Genome-wide prediction of DNA mutation effect on nucleosome positions for yeast synthetic genomics. *Genome Res.* **31**, 317–326 (2021).
28. Chereji, R. V. & Clark, D. J. Major Determinants of Nucleosome Positioning. *Biophys J* **114**, 2279–2289 (2018).
29. Levo, M. *et al.* Unraveling determinants of transcription factor binding outside the core binding site. *Genome Res* **25**, 1018–1029 (2015).
30. Sahu, B. *et al.* Sequence determinants of human gene regulatory elements. *Nat Genet* **54**, 283–294 (2022).
31. Struhl, K. & Segal, E. Determinants of nucleosome positioning. *Nat Struct Mol Biol* **20**, 267–273 (2013).
32. Tirosh, I. & Barkai, N. Two strategies for gene regulation by promoter nucleosomes. *Genome Res.* **18**, 1084–1091 (2008).
33. Lengronne, A. *et al.* Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* **430**, 573–578 (2004).
34. Glynn, E. F. *et al.* Genome-wide mapping of the cohesin complex in the yeast Saccharomyces cerevisiae. *PLoS Biol* **2**, E259 (2004).
35. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
36. Bystricky, K., Heun, P., Gehlen, L., Langowski, J. & Gasser, S. M. Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc Natl Acad Sci U S A* **101**, 16495–16500 (2004).
37. Duan, Z. *et al.* A three-dimensional model of the yeast genome. *Nature* **465**, 363–367 (2010).
38. Lazar-Stefanita, L. *et al.* Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle. *EMBO J* **36**, 2684–2697 (2017).
39. Wong, H. *et al.* A predictive computational model of the dynamic 3D interphase yeast nucleus. *Curr Biol* **22**, 1881–1890 (2012).
40. Garcia-Luis, J. *et al.* FACT mediates cohesin function on chromatin. *Nat Struct Mol Biol* **26**, 970–979 (2019).
41. Schalbetter, S. A. *et al.* SMC complexes differentially compact mitotic chromosomes according to genomic context. *Nat Cell Biol* **19**, 1071–1080 (2017).
42. Costantino, L., Hsieh, T.-H. S., Lamothe, R., Darzacq, X. & Koshland, D. Cohesin residency

determines chromatin loop patterns. *Elife* **9**, e59889 (2020).
43. Dauban, L. *et al.* Regulation of Cohesin-Mediated Chromosome Folding by Eco1 and Other Partners. *Mol Cell* **77**, 1279-1293.e4 (2020).
44. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
45. Goloborodko, A., Imakaev, M. V., Marko, J. F. & Mirny, L. Compaction and segregation of sister chromatids via active loop extrusion. *Elife* **5**, e14864 (2016).
46. Haarhuis, J. H. I. *et al.* The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* **169**, 693-707.e14 (2017).
47. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944.e22 (2017).
48. Gibcus, J. H. *et al.* A pathway for mitotic chromosome formation. *Science* **359**, eaao6135 (2018).
49. Naumova, N. *et al.* Organization of the mitotic chromosome. *Science* **342**, 948–953 (2013).
50. Barbieri, M. *et al.* Complexity of chromatin folding is captured by the strings and binders switch model. *Proc Natl Acad Sci U S A* **109**, 16173–16178 (2012).
51. Bastié, N. *et al.* Smc3 acetylation, Pds5 and Scc2 control the translocase activity that establishes cohesin-dependent chromatin loops. *Nat Struct Mol Biol* 1–11 (2022) doi:10.1038/s41594-022-00780-0.
52. Ocampo-Hafalla, M., Muñoz, S., Samora, C. P. & Uhlmann, F. Evidence for cohesin sliding along budding yeast chromosomes. *Open Biol* **6**, 150178 (2016).
53. de Boer, C. G. *et al.* Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nat Biotechnol* **38**, 56–65 (2020).
54. Vaishnav, E. D. *et al.* The evolution, evolvability and engineering of gene regulatory DNA. *Nature* **603**, 455–463 (2022).
55. de Boer, C. G. *et al.* A unified model for yeast transcript definition. *Genome Res* **24**, 154–166 (2014).
56. Luthra, I. *et al.* Biochemical activity is the default DNA state in eukaryotes. *bioRxiv* (2022) doi:https://doi.org/10.1101/2022.12.16.520785.
57. Schulz, D. *et al.* Transcriptome Surveillance by Selective Termination of Noncoding RNA Synthesis. *Cell* **155**, 1075–1087 (2013).
58. Jeppsson, K. *et al.* Cohesin-dependent chromosome loop extrusion is limited by transcription and stalled replication forks. *Sci Adv* **8**, eabn7063 (2022).
59. Wutz, G. *et al.* ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesinSTAG1 from WAPL. *Elife* **9**, e52091 (2020).
60. van Ruiten, M. S. *et al.* The cohesin acetylation cycle controls chromatin loop length through a PDS5A brake mechanism. *Nat Struct Mol Biol* **29**, 586–591 (2022).
61. Matthey-Doret, C. *et al.* Computer vision for pattern detection in chromosome contact maps. *Nat Commun* **11**, 5795 (2020).

Supplementary Figure Legends



	Percent (*P*) chromosome-carrying cells			
	After 10 doublings selective	After *g* doublings nonselective		Segregation rate
	P_1	g	P_2	m
Mmyco.	91.07	11.72	74.8	1.7
Mpneumo.	76.8	11.70	50.7	3.5

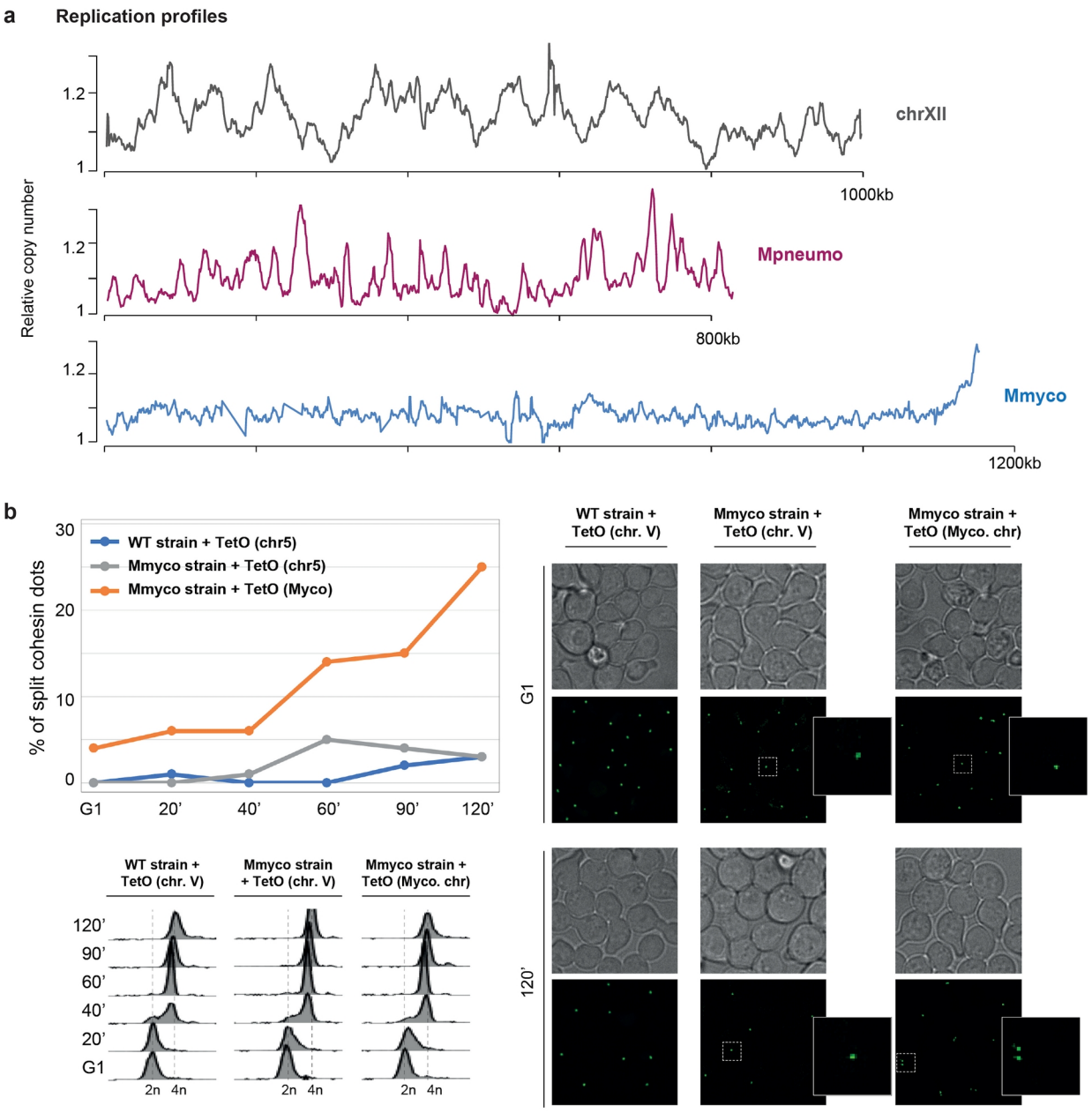
Supplementary Figure 1. Characterization of Mpneumo and Mmyco strains metabolism.

a, Schematic representation of the CRISPR strategy applied to linearize and telomerize the circular bacterial genomes.

b, Growth curves of WT, Mpneumo and Mmyco strains. For each strain, 3 independent cultures were performed. A pRS413 centromeric plasmid similar to the one onto which bacterial chromosomes were originally cloned is included in the WT strain as a control.

c, The chromosome stability and the segregation rate were measured as described in [8] (Methods). Yeast strains used are RSG_Y712 (Mmyco linear) and RSG_Y960 (Mpneumo linear). **P1**: % of bacterial chromosome-carrying cells in selective media. **g**: number of doubling. **P2**: of bacterial chromosome-carrying cells after 12 generations in non-selective media. **m**: segregation rate, i.e. % of plasmid-free segregants appearing in the final population after a single doubling.

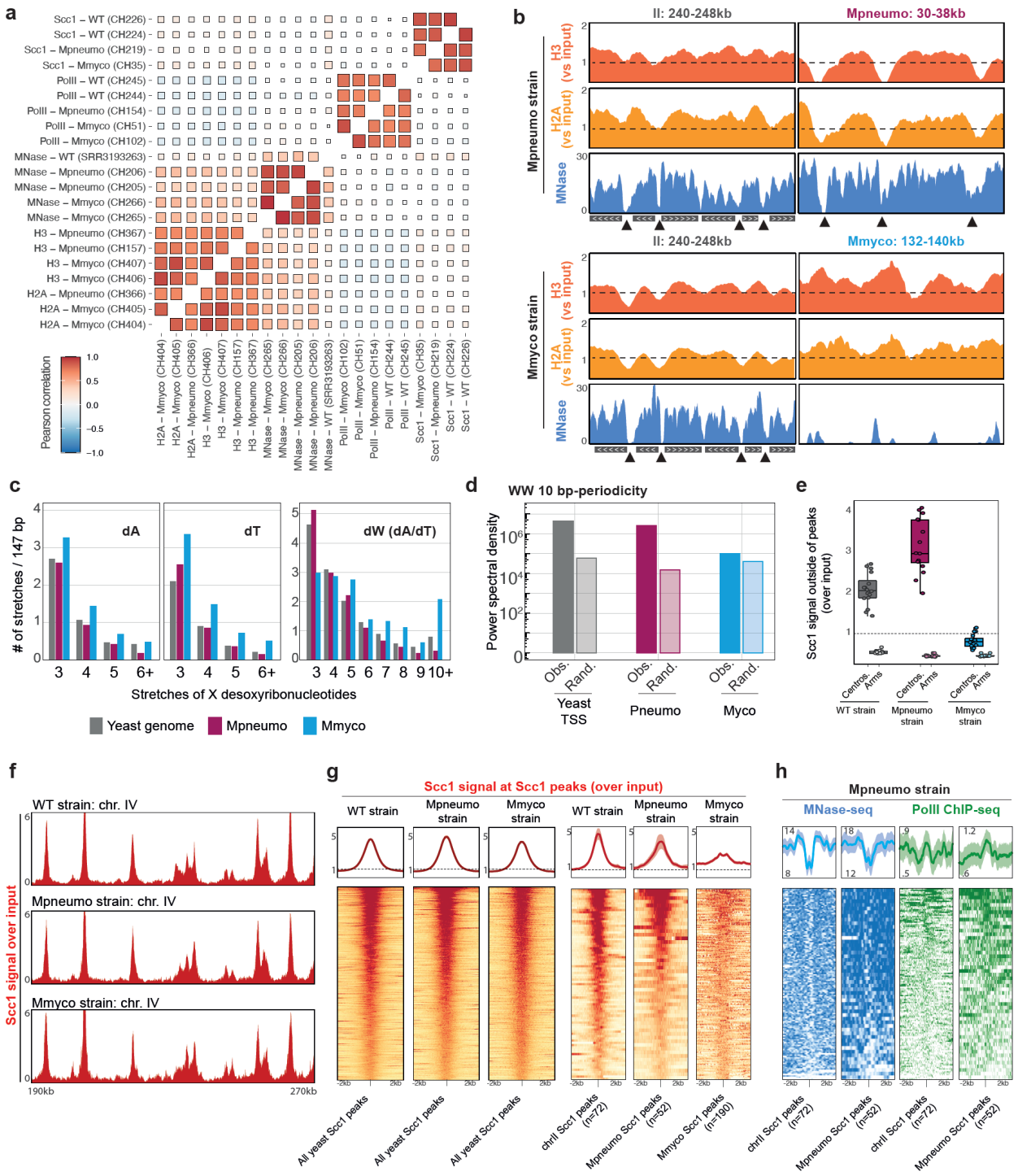
d, Left: Schematic representation of a neural network. The input of the network is a 2 to 30kb DNA sequence. The output is the corresponding value of the Mnase, Scc1 or PolII signal. Details about the size of the input/output and the architecture used are discussed in the Material and Methods. Right: Overall strategy. We train a CNN on sequences of chrI-XVI to predict (i) genome-wide MNase-seq, Scc1 or PolII ChIP-seq coverage tracks over chrXVI and bacterial chromosomes, and (ii) the average MNase-seq, Scc1 or PolII ChIP-seq signal over 10kb random sequences with varying GC%.



Supplementary Figure 2. Replication of bacterial genomes in yeast.

a, MFA (replication profile) analysis of a representative WT yeast chromosome, and of Mmcyo and Mpneumo chromosomes.

b, Cohesin split dot assays in WT strain (TetO array in chr. V), or in Myco strain (TetO array in chr. V or Myco. chromosome). Top left: % of split cohesin dots in each strain upon G1-release; bottom left: cell ploidy upon G1-release; right: representative cohesin immuofluorescence imaging in each strain following G1-release.



Supplementary Figure 3. Chromatin structuration of the bacterial chromosomes in yeast.

a, Pearson correlation between replicates of MNase and ChIP experiments. Bin size: 100bp.

b, MNase-seq, H2A and H3 ChIP-deq deposition profiles over representative 8kb windows from either (left) endogenous yeast chr. II or (right) the bacterial sequences in Mpneumo and Mmyco strains.

c, Number of poly-dA, poly-dT and poly-dW (dA/dT) stretches of various lengths in the yeast genome and over the Mpneumo and Mmyco chromosomes. Stretch numbers are scaled to 147bp.

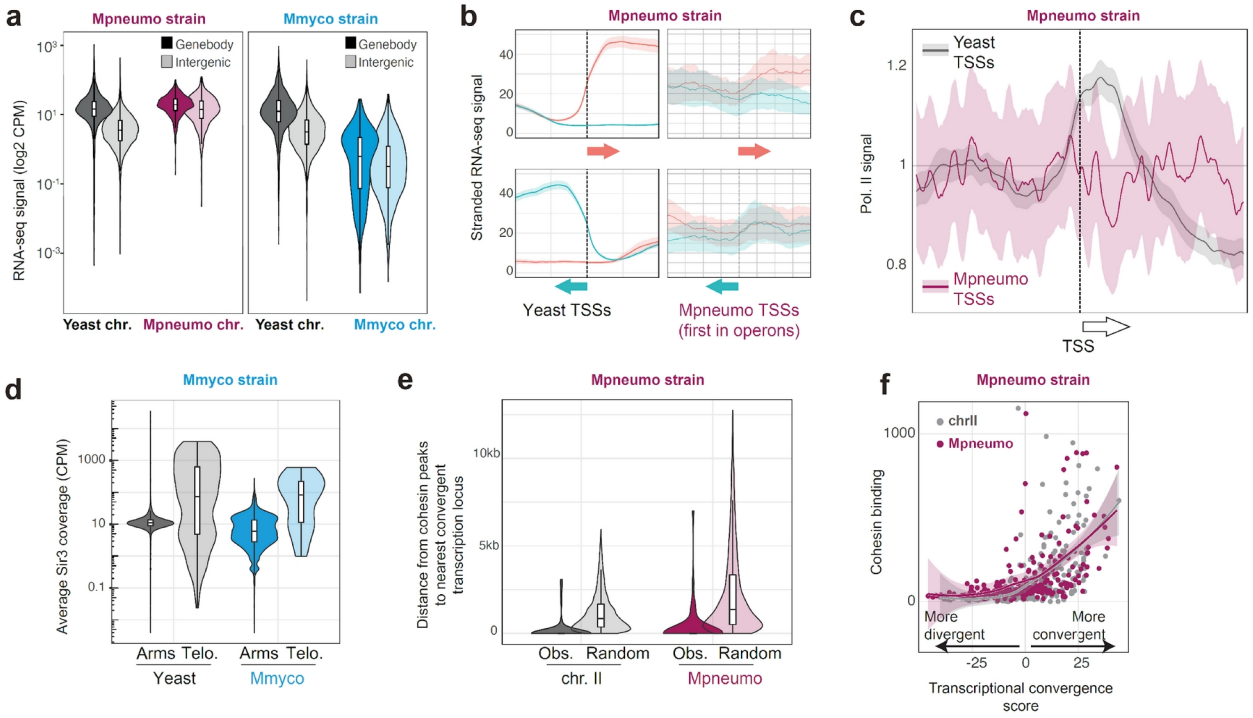
d, Power spectral density (PSD) of WW 10-bp periodicity, in 300bp-long sequences centered over yeast TSSs or along the Mpneumo or Mmyco chromosomes. Random sequences were generated by shuffling actual sequences while preserving dinucleotide frequency.

e, Cohesin (Scc1) enrichment over yeast centromeres and yeast arms in WT, Mpneumo and Mmyco strains. Scc1 signal over arms was calculated outside of any Scc1 peak.

f, Representative 80kb window of Scc1 ChIP-seq deposition signal over yeast chr. IV in the WT, Mpneumo, and Mmyco strains.

g, Aggregated profile of Scc1 deposition centered at Scc1 peaks (+/- 2kb) called over all endogenous (left panels) or only chr. II (right panels) yeast chromosomes in the WT, Mpneumo and Mmyco strains.

h, Aggregated profile of MNase-seq (blue) or PolII (green) deposition centered at Scc1 peaks (+/- 2kb) called over the yeast chromosome II (left) or the Mpneumo chromosome (right)



Supplementary Figure 4. Transcription orientation of bacterial genomes in yeast.

a, RNA-seq average signal over yeast or bacterial gene bodies and intergenic regions, in Mpneumo and Mmyco strains. Scores were normalized by the length of each genomic feature (CPM: counts per million of sequenced fragments).

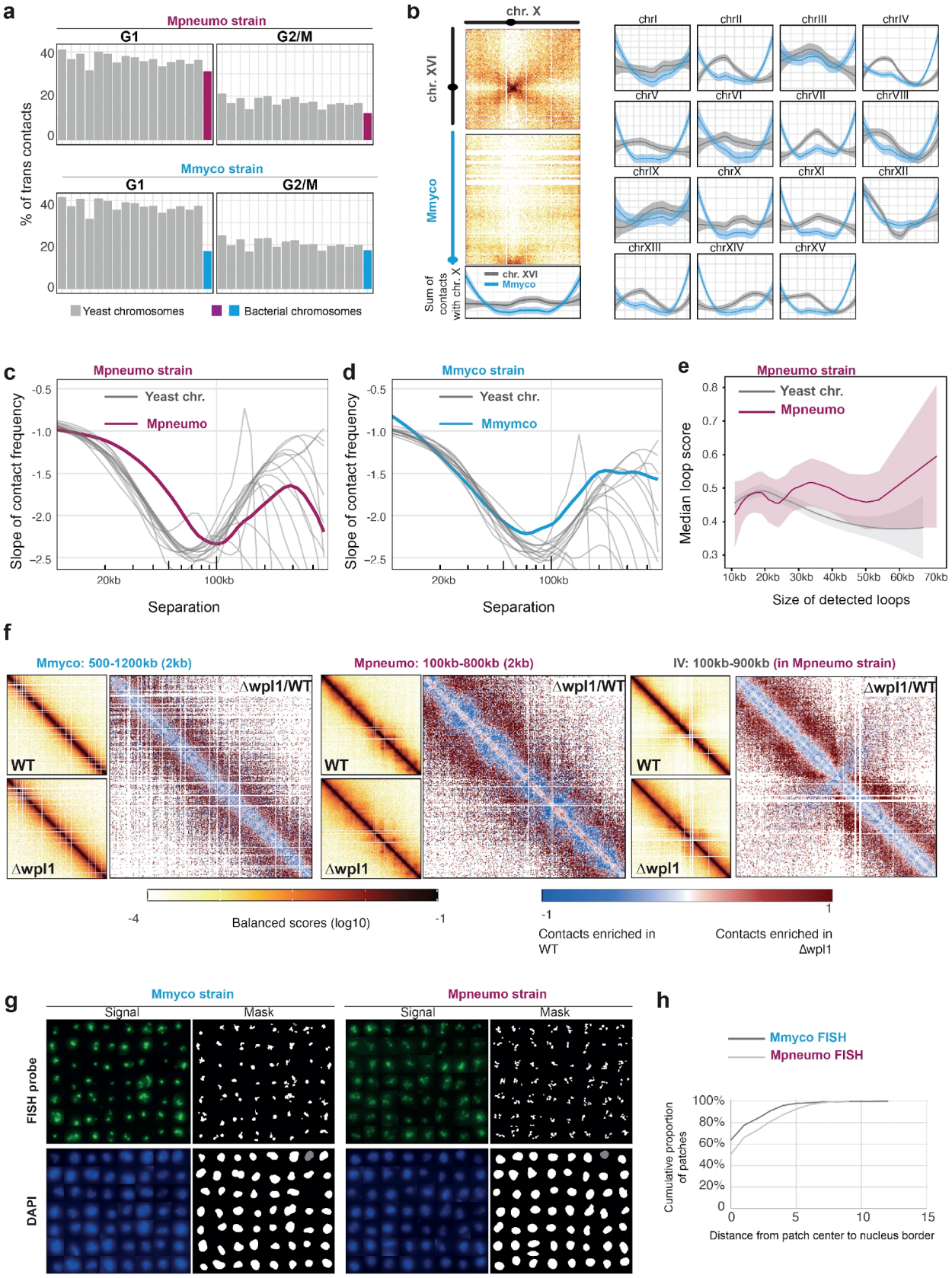
b, Stranded analysis of RNA-seq data in Pneumo. strain. Pile-up of 1kb windows centered on transcription start sites (TSS) of genes either in the forward (Top) or reverse (Bottom) orientation. Left: endogenous yeast genes. Right: TSS of the first gene of annotated operons along the *M. pneumoniae* sequence.

c, PolII ChIP-seq coverage analysis in Pneumo strain. Pile-up of 2kb windows centered on transcription start sites (TSS). Grey: endogenous yeast genes. Purple: TSS of the first gene of annotated operons along the *M. pneumoniae* sequence.

d, Average Sir3 ChIP-seq signal over chromosome arms or telomeres (up to 3kb from end of chromosomes) in endogenous yeast chromosomes or in Myco. chromosome.

e, Distance measured between cohesin peaks and their nearest convergent transcription locus, for peaks located in chr. II or in Mpneumo chromosome. Expected distances, measured after randomly shuffling the position of the cohesin peaks 100 times, are also shown.

f, Correlation between Scc1 (cohesin) binding and convergent transcription strength (see Methods) in chr. II and in Mpneumo chromosome.


a
Mpneumo strain
G1
G2/M
% of trans contacts
Mmyco strain
Yeast chromosomes
Bacterial chromosomes
b
chr. X
chr. XVI
Mmyco
Sum of contacts with chr. X
chr. XVI
Mmyco
chrI
chrII
chrIII
chrIV
chrV
chrVI
chrVII
chrVIII
chrIX
chrX
chrXI
chrXII
chrXIII
chrXIV
chrXV
c
Mpneumo strain
Yeast chr.
Mpneumo
Slope of contact frequency
20kb
100kb
Separation
d
Mmyco strain
Yeast chr.
Mmymco
Slope of contact frequency
20kb
100kb
Separation
e
Mpneumo strain
Yeast chr.
Mpneumo
Median loop score
10kb 20kb 30kb 40kb 50kb 60kb 70kb
Size of detected loops
f
Mmyco: 500-1200kb (2kb)
Mpneumo: 100kb-800kb (2kb)
IV: 100kb-900kb (in Mpneumo strain)
Δwpl1/WT
WT
Δwpl1
Balanced scores (log10)
Contacts enriched in WT
Contacts enriched in Δwpl1
g
Mmyco strain
Mpneumo strain
Signal
Mask
FISH probe
DAPI
h
Mmyco FISH
Mpneumo FISH
Cumulative proportion of patches
Distance from patch center to nucleus border

Supplementary Figure 5. Folding of exogenic bacterial sequences within the yeast nucleus.

a, % of total trans-contacts made by endogenous yeast chromosomes, Mmyco (top) or Mpneumo (bottom) bacterial chromosomes, in G1 or G2/M.

b, Left: quantification of contacts between the entire Mmyco chromosome (blue) and endogenous yeast chr. X (grey). Right: similar analysis for the other 15 endogenous yeast chromosomes. Note the increase of Myco. contacts at yeast telomeres.

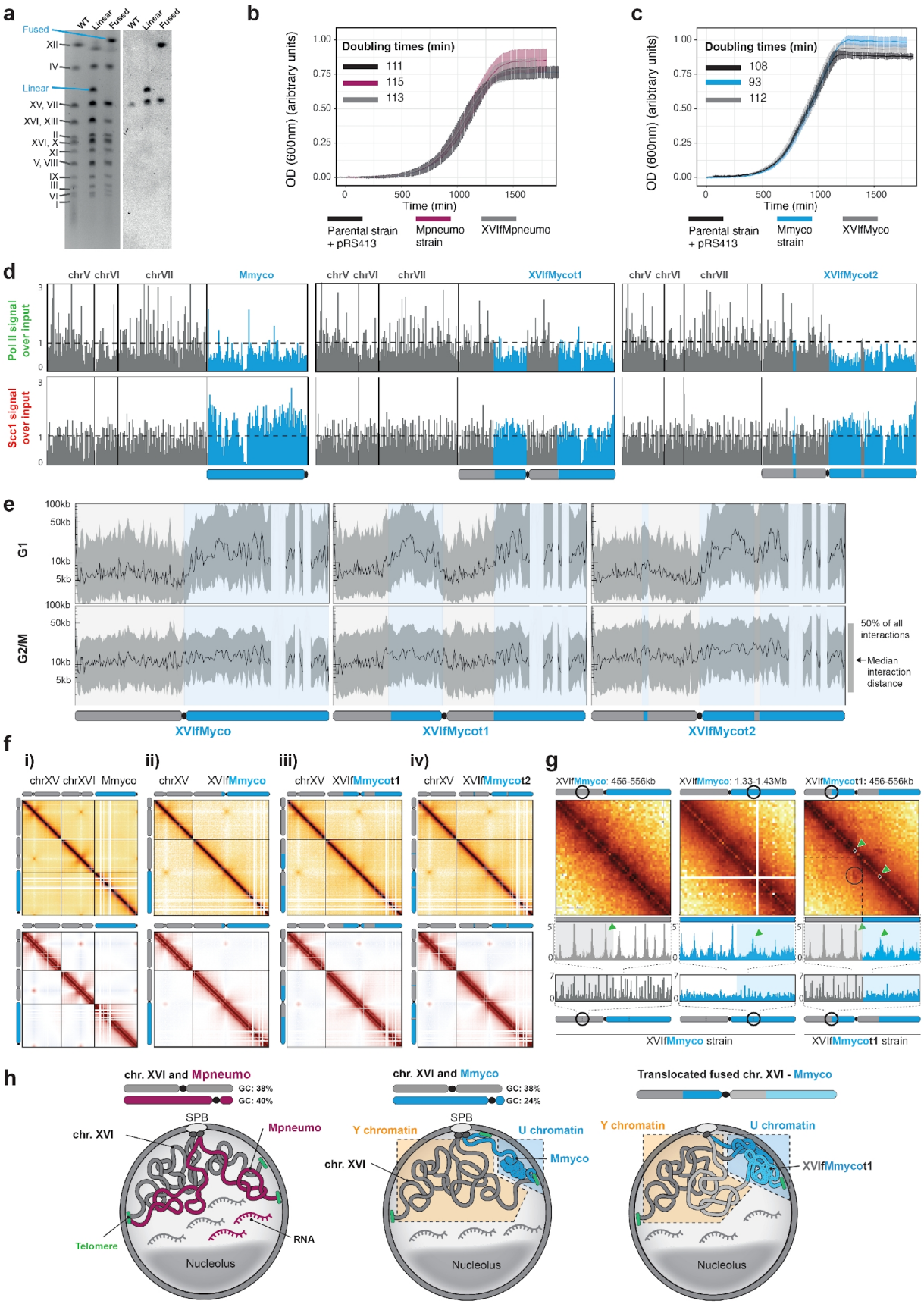
c, d, Slope of distance-dependent contact frequency of endogenous yeast chromosomes and bacterial chromosomes in G2/M Mpneumo (**c**) and Mmyco (**d**) strains.

e, Distance-dependent loop scores (computed using Chromosight [9]) for loops along either endogenous (grey) or Mpneumo (purple) chromosomes.

f, For Mmyco, Mpneumo and a regular endogenous chromosomes: Left, Hi-C contact maps of the endogenous yeast chromosome IV of the Mpneumo strain synchronized in G2/M in either WT and Wpl1 depleted cells (Δwpl1); Right, corresponding chr. IV ratio map (Δwpl1 over WT). Red (or blue) indicate enriched (or depleted) contacts in Δwpl1.

g, Series of nuclei from Mpneumo or Mmyco fixed cells labeled with DAPI and hybridized with a fluorescent probe generated from either purified Mmyco (left) or Mpneumo (right) chromosome (top row: probe signal; bottom row: DAPI signal).

h, Distance between the patch of fluorescent signal from either the Mmyco or Mpneumo chromosomes and the nucleus border. Note that the Mmyco patches are located closer to the nucleus border than Mpneumo ones.


a
WT
Linear
Fused
Fused
Linear
XII
IV
XV, VII
XVI, XIII
II
XVI, X
XI
V, VIII
IX
III
VI
I
b
Doubling times (min)
111
115
113
OD (600nm) (aribtrary units)
Time (min)
Parental strain + pRS413
Mpneumo strain
XVIfMpneumo
c
Doubling times (min)
108
93
112
OD (600nm) (aribtrary units)
Time (min)
Parental strain + pRS413
Mmyco strain
XVIfMyco
d
chrV chrVI chrVII
Mmyco
XVIfMycot1
XVIfMycot2
Pol II signal over input
Scc1 signal over input
e
G1
G2/M
100kb
50kb
10kb
5kb
50% of all interactions
Median interaction distance
XVIfMyco
XVIfMycot1
XVIfMycot2
f
i)
ii)
iii)
iv)
chrXV chrXVI Mmyco
chrXV XVIfMmyco
chrXV XVIfMmycot1
chrXV XVIfMmycot2
g
XVIfMmyco: 456-556kb
XVIfMmyco: 1.33-1.43Mb
XVIfMmycot1: 456-556kb
XVIfMmyco strain
XVIfMmycot1 strain
h
chr. XVI and Mpneumo
GC: 38%
GC: 40%
chr. XVI and Mmyco
GC: 38%
GC: 24%
Translocated fused chr. XVI - Mmyco
SPB
chr. XVI
Mpneumo
Telomere
RNA
Nucleolus
Y chromatin
U chromatin
Mmyco
XVIfMmycot1

Supplementary Figure 6. Compartmentalization of mosaic chromosomes composed of Y and U-type chromatin.

a, Left: pulsed-field gel electrophoresis of chromosomes from yeast strains containing the Mmyco chromosome, either linear or fused with endogenous yeast chr. XVI (XVIfMmyco strain). Right: Southern blot hybridization using a *his-3* probe present on the Mmyco chromosome sequence (note that *his-3* is also present on the endogenous chr. XV in the parental yeast strain).

b, Growth curves of WT, Mpneumo and XVIfMpneumo strains.

c, Growth curves of WT, Mmyco and XVIfMmyco strains.

d, PolII (top) and Scc1 (bottom) ChIP-seq deposition profiles along three representative yeast chromosomes and Mmyco chromosome (left) and mosaic chromosomes in XVIfMmycot1 (center) and XVIfMmycot2 (right) strains. Bin size: 10kb.

e, Average distance of interaction along the fused and mosaic Mmyco chromosomes: XVIfMmyco (left), XVIfMmycot1 (center) and XVIfMmycot2 (right). The shaded ribbon represents the interval between the 25% and 75% quantile of distance of interactions.

f, Top: G2/M Hi-C contact maps of chr. XV, XVI, and bacterial chromosomes in the Mmyco strain, and in its derivatives (i.e. the chr. XVI and Mmyco fusion, and the two strains with translocations resulting in alternating Y and U chromatin segments; Methods). Bottom: auto-correlation contact matrices in wt and mosaic chromosomes strains. The color scales are the same as in Fig. 5b.

g, Hi-C contact maps in G2/M of 100 kb window of either the XVIfMmyco (left and center) or XVIfMmycot1 (right) chromosomes, centered on the translocation position of the XVIfMmycot1 chromosome. In XVIfMmycot1, this window is effectively centered at the junction between the yeast chr. XVI segment (upstream) and the Myco chromosome segment (downstream). The Scc1 deposition profile measured by ChIP-seq in each strain is shown underneath each contact map. Green arrows: cohesin peaks flanking the junction between the yeast and the Myco segments in the XVIfMmycot1 chromosome.

h, Illustration of the behaviors and activity of Mpneumo (left panel) and Mmyco (middle and right) chromosome sequences integrated in the yeast genome. Note that yeast and Mpneumo chromosomes intermingle in a single nuclear compartment, whereas the Mmyco chromosome (independently or in a chimeric state within chr. XVI) is condensed and segregated at the nuclear periphery, thereby defining the U-type chromatin compartment.

Supplementary Results

Replication and pairing of artificial chromosomes

DNA sequence composition and chromatin factors, including nucleosomes, are essential drivers of the replication timing, prompting us to investigate how this timing is established in Y and U chromatin [1]. In yeast, replication initiates at the level of discrete, small autonomous replication sequences (ARS) positioned along all chromosomes. To investigate the replication timing profile of both bacterial chromosomes, which did not spontaneously evolve to contain these sequences, we used marker frequency analysis (MFA; Methods) [2]. Whereas Mpneumo MFA profile exhibits DNA copy number variation indicative of the early or mid-S phase firing of multiple replication origins, MFA along Mmyco unveil the early firing of only the centromere-proximal ARS which was artificially integrated during chromosome assembly[3] (**Supplementary Figure 2a**). The average copy number of the rest of the Mmyco chromosome appears flat and close to 1N in the MFA analysis. These observations confirm the role of chromatin composition in the replication profile as Y type chromatin found on the Mmyco chromosome is similar to the profile of endogenous chromosomes while the replication timing of U chromatin points at replication taking place late during S phase, most likely from cryptic ARS all along the AT-rich sequence [4]. This pattern is reminiscent of the random and late-replication pattern displayed by the human inactive chromosome X[5].

DNA replication and SCC are closely related as cohesion is established during S phase through entrapment of sister DNA molecules by cohesin rings as the replication fork progresses[6,7]. We therefore investigated SCC by performing image analysis of chromosome pairing (**Methods**; **Supplementary Figure 2b**). Despite the strong enrichment in Scc1 deposition, SCC in Mmyco appears significantly reduced, in agreement with the flat replication pattern suggesting a late and random initiation of the replication process. On the other hand, the endogenous yeast chromosomes in Mmyco strain did not display a significant decrease in SCC, despite the apparent loss of cohesin at centromeres (**Supplementary Figure 3e**).

Material and Methods

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 [10] and *Mycoplasma pneumoniae* strain M129 [3], respectively. Yeast strains were grown overnight at 30°C in 150ml of suitable media to attain $4{,}2x10^8$ cells. 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 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.

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 targeting remains relatively poorly efficient. Plasmids pML107 [11] or PAEF5 [12] carrying gRNA were co-transformed with 200bp DNA repair recombinant donor sequences carrying telomeric repeats seeds (Genscript Biotech Netherlands) (Figure Supp 1A).

We also applied a CrispR-Cas9 approach to concomitantly fuse chromosome XVI with the linear bacterial chromosomes and remove chromosome XVI centromere, as described in Luo et al. [13]. The resulting chromosome (XVIfMmyco) as one bacterial DNA arm carrying U chromatin and one yeast DNA arm made of Y chromatin (**Fig. 6A**). Strain XVIfMmyco's karyotype was verified using pulsed-field gel electrophoresis (**Fig. S6A**), and that it grows normally was verified by growth assay (**Fig. S6B, C**). Briefly, we used three gRNAs inserted in pGZ110 and pAEF5 [12] 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. 6A**). The same strategy was applied to generate reciprocal translocations between chromosome XVI and Mmyco sequences of the XVIfMmyco chromosome (**Fig. 6A**). We generated alternating domains of Y and U chromatin along the XVIfMmyco chromosome using CRISPR-induced reciprocal translocations between the two arms, resulting in strains XVIfMmycot1 and t2 carrying alternating regions of U and Y type chromatin along a chromosome (**Fig. 6A**). The chromatin composition of these chimeric chromosome consists of [TEL - 500kb Y - 700kb U - CEN - 400kb Y - 500kb U - TEL] for XVIfMmycot1 and of [TEL - 500kb Y - 50kb U - 350kb Y - CEN - 650kb U - 50kb Y - 500kb U - TEL] for XVIfMmycot2.

gRNA sites were chosen to optimize Crispr targeting specificity using chopchop.cbu.uib.no [14] or CRISPOR [15]. 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 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.

Liquid growth assays and segregational plasmid stability

Growth rate of parental, Mmyco and Mpneumo strains carrying linear or fusioned 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 100µl 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 [8]. 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 tures were diluted into non-selective media and grown for X generations (g). After approximately 12 generations, the percentage of cells containing the bacterial chromosome (P2) was alshe chromosome stability, culto 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 % 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 [16] 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 h].

DNA from the PFGE was transferred on a membrane. Southern blot was performed using digoxigenin-labeled DNA probes as described [17]. 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'.

Hi-C procedure and sequencing

Cell fixation with 3% formaldehyde (Sigma-Aldrich, Cat. F8775) was performed as described in Dauban et al. [18]. 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 120ul H2O and sonicated using Covaris (DNA 300bp) in Covaris microTUBE (Covaris, 520045). Biotinylated DNA was loaded on Dynabeads™ Streptavidin C1 (FISHER SCIENTIFIC, 10202333). 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 [19]. 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 4mL 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 2mL of 2.5M 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 2mL tubes containing glass beads (ozyme, P000913-LYSK0-A.0) before mechanical cells lysis. The soluble fraction was isolated by centrifugation at 2,000 rpm for 3min then transferred to sonication tubes (Covaris milliTUBE 1ml, 520135) and samples were sonicated to produce sheared chromatin with a size range of 200-1,000bp using a Covaris sonicator. After sonication the samples were centrifuged at 13,200 rpm at 4°C for 20min and the supernatant was transferred into 700µL of ChIP lysis buffer. 80µL (27µl of each sample) of the supernatant was removed (termed 'whole cell extract [WCE] sample') and store at -80°C. 5ug of antibody was added to the remaining supernatant which is then incubated overnight at 4°C (wheel cold room). 50µL of protein G Dynabeads or protein A Dynabeads was then added and incubated at 4°C for 2h. Beads were washed 2 times with ChIP lysis buffer, 3 times with high salt ChIP lysis buffer (50mMHEPES-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; 0.25MLiCl; 0.5% NP-40; 0.5% sodium deoxycholate; 1mM EDTA;1 mMPMSF) and 1 time with TE pH7.5. The immunoprecipitated chromatin was then eluted by incubation in 120µL TES buffer (50 mMTris-HCl, pH 8.0; 10 mM EDTA; 1% SDS) for 15min at 65°C and the supernatant is collected termed 'IP sample'. The WCE samples were mixed with 40µL of TES3 buffer (50 mM Tris-HCl, pH 8.0; 10 mM EDTA; 3% SDS). All (IP and WCE) samples were de-cross-linked by incubation at 65°C overnight. RNA was degraded by incubation with 2µL RNase A (10 mg/mL) for 1h at 37°C. Proteins were removed by incubation with 10µL of proteinase K (18 mg/mL) for 2h 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.

MNase-seq

Five 1.5ml Tubes were prepared with 1.5 x 10^8 asynchronous yeast cells: cells were centrifuged, pelleted, washed in 1x PBS and frozen at -80C. Cell pellets were thawed on ice and resuspended in 1mL 1xPBS supplemented with 27ul 37% formaldehyde. Cells were fixed for 20min at room temperature on rotation. 200ul of 2.5M Glycine was added for a further 5 min incubation. Cells were centrifuged at 2500g, 4C for 5 min. Pellets were gently washed with cold 1xPBS before resuspension in 500ul of lysis solution (Sorbitol 1M, EDTA 10mM, Zymo100T 250ug) and incubation at 37C with gently shaking for 30min/45min or until complete spheroplasting. Cells were gently washed with cold 1xPBS and resuspended in 50ul 1X Nuclease Digest buffer plus an increasing amount of Mnase (NEB, M0247S) enzyme starting with 0.2ul. Digestion was performed at 22C in a shaker set at 1250rpm for 15min. 5ul of 0.5M EGTA and 3ul of 20% SDS were added for incubation at 22C in a shaker set at 1250rpm for 5 min. 65ul of TES buffer (50 mMTris-HCl, pH 8.0; 10 mM EDTA; 1% SDS) was added and samples were de-cross-linked by incubation at 65°C overnight. RNA was degraded by incubation with 2µL RNase A (10 mg/mL) for 1h at 37°C. Proteins were removed by incubation with 10µL of proteinase K (18 mg/mL) for 2h at 65°C. DNA was purified using Zymo Research DNA Clean and Concentrator™-5 Kit. Mnase digestion was verified on a Bioanalyzer System using HS DNA kit. DNA was purified using 1.2x volume of AmpureXP beads and librairies prepared with Invitrogen TM Collibri TM PS DNA Library Prep Kit for Illumina following manufacturer instructions. Paired-end sequencing on an Illumina NextSeq500 (2x35 bp) was performed. Libraries were performed in two biological replicates and the duplicates were averaged for visualization. We calculated pairwise Pearson correlation scores between replicates and all showed high concordance.

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 50µl 1xTE. 100ng of soluble gDNA was transferred to sonication tubes and samples were sonicated to produce sheared chromatin with a size of about 300bp 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 [20] 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 [21]. PCR duplicates (defined as multiple pairs of reads positioned at the exact same position) were discarded. Contacts were binned at 1kb resolution and multiple-resolution balanced contact maps (in mcool format) were generated using cooler [22].

MNase-seq processing

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 and fragments smaller than 130 or longer than 200 bp were manually removed. Fragment coverage was normalized by library depth (CPM) and converted into a genomic track (bigwig) using deepTools.

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 1kb 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 stranded or unstranded genomic tracks (bigwig) using deepTools.

ChIP-seq processing

For standard ChIP-seq, bowtie2 was used to align paired-end RNA-seq data on the appropriate genome reference. Only fragments with an insert size shorter than 1kb 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 calibrated Scc1 ChIP-seq, CPM-normalized fragment coverages were also multiplied by the ORi factor for calibration ($WCE_{glabrata}$ x $IP_{cerevisiae}$ / $WCE_{cerevisiae}$ x $IP_{glabrata}$, 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. Calibrated Scc1 ChIP-seq tracks were also computed as in Hu et al 2015.

Analysis of genome-wide assays

All downstream analysis steps were performed in R 4.1.2 / Bioconductor 3.16 using in-house scripts, unless mentioned otherwise.

ChIP-seq analysis

Chromosome-wide ChIP-seq profiles were generated by averaging library-depth-normalized coverage tracks over non-overlapping windows of 10kb. Locus-specific ChIP-seq profiles were generated over 80kb windows (8kb windows) by averaging library-depth-normalized coverage tracks over non-overlapping windows of 100bp (10bp).

Aggregated ChIP-seq profiles (average +/- 95% CI and heatmap) were plotted over windows centered at Scc1 peaks or TSSs after averaging coverage over 1bp-moving 200bp-wide rolling windows.

Median Scc1 coverage was calculated over 20 kb windows centered at yeast centromeres or over 300bp non-overlapping 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 100bp 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 (REF), using ushuffle (REF) 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.

Transcription convergence and Scc1 binding

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

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

Genomic positions *j* for which $Dir_{j-101} > 0 \wedge Dir_{j-100} < 0$ 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:

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

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 $Conv_{\square}$ 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 [4].

Contact heatmaps

All the contact heatmaps and ratio maps were plotted with a log10 or linear scaling respectively, using “plotMatrix” function from HiContacts (https://github.com/koszullab/HiContacts).

Loop detection

Chromosight 1.3.1 was used to call loops de novo from contact maps binned at 1 kb and balanced with Cooler. 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) and its derivative were determined using pairs files (generated by hicstuff) and the “distanceLaw” function from HiContacts with default parameters, averaging the contact data by individual chromosomes.

Virtual 4C profiles

Virtual 4C profiles for 20kb-wide viewpoints were computed from 2kb-binned contact maps, using the “virtual4C” function from HiContacts. 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 Mnase, PolII and Scc1 profiles from the underlying genomic sequence. The signals were pretreated as follows. First, 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 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, IIV and XV for validation and XVI for test.

We implemented all the CNNs using the Keras library [23] and Tensorflow [24] as back-end. 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.

For the Mnase prediction task, our CNN architecture was similar to the one used in [25]. It consists of three convolutional layers with respectively 64, 16 and 8 kernels of shape (3x1x4), (8x1x 64) 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 PolII and Scc1 tasks, the architecture was modified as follows to take into account longer range influences [26]. 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 PolII, 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 (loss = MAE(x, y) + 1 - corr(x, y)). This loss function has been previously shown to enable both a faster and a more accurate convergence [25].

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

5 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 GCc content of the sequences on the overall profile heights, we generated 1000 random 10kb sequences and predicted each of the three profiles on these sequences. We then averaged these signals and computed the mean and standard deviation of these averages.

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 ul at 5mg/ml) was added every 30 min for 2 h. 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. [27], with some modifications [27]. The probes were obtained by direct labeling of the bacterial DNA (1.5 µg) 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 µl of water. The purified probe was then diluted in the probe mix buffer (50% formamide, 10% dextran sulfate, 2× SSC final). 20 OD of cells (1 OD corresponding to 107 cells) were grown to mid–logarithmic phase (1–2 × 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 minutes with Zymolyase (60 µg/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× SSC, 0.1% Tween, and 20 µg/ml RNase, the slides were washed in H20 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× 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× 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 (1µg/ml in 1× PBS) for 5 minutes and then washed twice in 1× 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×/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

1. Bodmer-Glavas, M., Edler, K. & Barberis, A. RNA polymerase II and III transcription factors can stimulate DNA replication by modifying origin chromatin structures. *Nucleic Acids Res* **29**, 4570–4580 (2001).
2. Müller, C. A. *et al.* The dynamics of genome replication using deep sequencing. *Nucleic Acids Res* **42**, e3 (2014).
3. Ruiz, E. *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).
4. Batrakou, D. G., Müller, C. A., Wilson, R. H. C. & Nieduszynski, C. A. DNA copy-number measurement of genome replication dynamics by high-throughput sequencing: the sort-seq, sync-seq and MFA-seq family. *Nat Protoc* **15**, 1255–1284 (2020).
5. Koren, A. & McCarroll, S. A. Random replication of the inactive X chromosome. *Genome Res.* **24**, 64–69 (2014).

6. Lengronne, A. *et al.* Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* **430**, 573–578 (2004).
7. Nasmyth, K. Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu Rev Genet* **35**, 673–745 (2001).
8. Christianson, T. W., Sikorski, R. S., Dante, M., Shero, J. H. & Hieter, P. Multifunctional yeast high-copy-number shuttle vectors. *Gene* **110**, 119–122 (1992).
9. Matthey-Doret, C. *et al.* Computer vision for pattern detection in chromosome contact maps. *Nat Commun* **11**, 5795 (2020).
10. Labroussaa, F. *et al.* Impact of donor–recipient phylogenetic distance on bacterial genome transplantation. *Nucleic Acids Research* **44**, 8501–8511 (2016).
11. Laughery, M. F. *et al.* New Vectors for Simple and Streamlined CRISPR-Cas9 Genome Editing in Saccharomyces cerevisiae. *Yeast* **32**, 711–720 (2015).
12. Agier, N., Fleiss, A., Delmas, S. & Fischer, G. A Versatile Protocol to Generate Translocations in Yeast Genomes Using CRISPR/Cas9. *Methods Mol Biol* **2196**, 181–198 (2021).
13. Luo, J., Sun, X., Cormack, B. P. & Boeke, J. D. Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast. *Nature* **560**, 392–396 (2018).
14. Labun, K. *et al.* CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res* **47**, W171–W174 (2019).
15. Concordet, J.-P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research* **46**, W242–W245 (2018).
16. Koszul, R., Caburet, S., Dujon, B. & Fischer, G. Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. *EMBO J* **23**, 234–243 (2004).
17. Viterbo, D. *et al.* A fast, sensitive and cost-effective method for nucleic acid detection using non-radioactive probes. *Biol Methods Protoc* **3**, bpy006 (2018).
18. Dauban, L. *et al.* Regulation of Cohesin-Mediated Chromosome Folding by Eco1 and Other Partners. *Mol Cell* **77**, 1279-1293.e4 (2020).
19. Hu, B. *et al.* Biological chromodynamics: a general method for measuring protein occupancy across the genome by calibrating ChIP-seq. *Nucleic Acids Res* **43**, e132 (2015).
20. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
21. Cournac, A., Marie-Nelly, H., Marbouty, M., Koszul, R. & Mozziconacci, J. Normalization of a chromosomal contact map. *BMC Genomics* **13**, 436 (2012).
22. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).
23. Chollet, F. *keras*. (2015).
24. Abadi, M. *et al. TensorFlow: Large-scale machine learning on heterogeneous systems*. (2015).
25. Routhier, E., Pierre, E., Khodabandelou, G. & Mozziconacci, J. Genome-wide prediction of DNA mutation effect on nucleosome positions for yeast synthetic genomics. *Genome Res.* **31**, 317–326 (2021).
26. Kelley, D. R. *et al.* Sequential regulatory activity prediction across chromosomes with convolutional neural networks. *Genome Res* **28**, 739–750 (2018).
27. Gotta, M., Laroche, T. & Gasser, S. M. Analysis of nuclear organization in Saccharomyces

cerevisiae. *Methods Enzymol* **304**, 663–672 (1999).
28. Henikoff, J. G., Belsky, J. A., Krassovsky, K., MacAlpine, D. M. & Henikoff, S. Epigenome characterization at single base-pair resolution. *Proceedings of the National Academy of Sciences* **108**, 18318–18323 (2011).
29. Rhee, H. S., Bataille, A. R., Zhang, L. & Pugh, B. F. Subnucleosomal structures and nucleosome asymmetry across a genome. *Cell* **159**, 1377–1388 (2014).
30. Wery, M. *et al.* Nonsense-Mediated Decay Restricts LncRNA Levels in Yeast Unless Blocked by Double-Stranded RNA Structure. *Mol Cell* **61**, 379–392 (2016).