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 orderly folding of long DNA molecules into functional chromosomes. In eukaryotes, this multiscale folding involves several molecular complexes and structures, ranging from nucleosomes to large cohesin-mediated DNA loops. To directly explore the causal relationships between the DNA sequence composition and the spontaneous loading and activity of these complexes in the absence of co-evolution, we used and characterized yeast strains carrying exogenous bacterial chromosomes that diverged from eukaryotic sequences over 1.5 billion years ago. By combining this synthetic approach with deep learning-based in silico analysis, we show that sequence composition drives chromatin assembly, transcriptional activity, folding, and compartmentalization in this cellular context. These results are also a step forward in understanding the molecular events at play following natural horizontal gene transfers, and could also be considered in synthetic genomic engineering projects.

Introduction

In all domains of life, genomic sequences have coevolved with DNA-binding proteins and their complexes to regulate processes such as gene expression, DNA replication and the functional organization of the genome in 3D. These interdependence between sequences and molecular regulators raise intriguing questions in light of the transfer of genetic material across species. These include horizontal gene transfers (HGT) or introgression events that occur between species sometimes very divergent evolutionarily, and invasion by DNA viruses or retrotransposons (Crisp et al. 2015; Edelman et Mallet 2021; Peter et al. 2018; Van Etten et Bhattacharya 2020). These elements often show base compositions that are divergent from the nucleotide content of their host's genome, such as the ~1Mb introgression event in yeast *Lachancea kluyveri* that involved a DNA segment with 12% higher GC% than the rest of the genome (Payen et al. 2009). In the emerging field of synthetic genomics, whole genomes or long DNA molecules carrying entire metabolic and genetic pathways by combining genes from other species can be introduced into chassis microbial strains or cell lines (Baby et al. 2018; Coradini, Hull, et Ehrenreich 2020; Currin et al. 2021; Galanie et al. 2015; Meng et Ellis 2020; Trolle et al. 2022; Pinglay et al. 2022). How the host genome copes with such exogenous sequence remains unclear, with some of these studies reporting broadly variable levels of transcription by the host cell of these sequences (e.g. Baby et al. 2018; Camellato et al. 2022). All these observations and events outline the need to better understand how an exogenous DNA molecule is processed by a host' molecular machinery. Answering this question is not only of fundamental interest for genome evolution and synthetic genomics approaches, but also holds important promises regarding our understanding of the metabolism of the host genome since being able to predict the behavior of exogenic sequences necessarily implies a better characterization of the molecular processes at play.

Genomes show a great diversity in sequence composition, with GC content ranging from 20 to 65% (D 2018). Variation in sequence composition within a given genome can also behave differently, being either homogeneous (e.g. budding yeast Saccharomyces cerevisiae;(Goffeau et al. 1996)) or heterogeneous (e.g. the human genome; Bernardi et al. 1985; Lander et al. 2001). The local GC composition is linked to the activity, with regions of higher GC being on average enriched in coding sequences, thus more likely to be actively transcribed and replicated early (Holmquist 1989). This difference in gene density and activity is reflected by structural

differences at different scales. In mammals, the spacing between consecutive nucleosomes is influenced by the GC content, with AT-rich, inactive regions displaying a longer spacing (Valouev et al. 2011). The formation of higher order structural features mediated by structural maintenance of chromosome (SMCs) ring complexes, such as chromatin loops and topologically associating domains (TADs), are also influenced by the GC content (Davidson et Peters 2021; Mirny et Dekker 2022; Rao et al. 2017). In GC-rich, gene-rich regions, CTCF binding to chromatin leads to the formation of loops and TADs. In contrast, in AT-rich, inactive regions, CTCF binding sites are often methylated leading to the absence of these structures (Spracklin et al. 2021). In addition, variations in sequence composition also correlate with the genome 3D organization at larger scales, as chromatin organized on sequences of similar composition tends to coalesce in different regions of the nucleus, forming distinct compartments in large genomes, such as those of metazoans and plants (Lieberman-Aiden et al. 2009; Cournac et al. 2016). In contrast, the shorter, compact genomes of eukaryotic microorganisms such as budding yeast do not exhibit large GC variations, nor compartments.

Here we present a synthetic approach to explore the causal relationships between sequence-dependent chromatinization principles and the structure and activity of a eukaryotic genome, including transcription, replication, and multiscale 3D organization. To this end, we study the behavior of very large bacterial chromosomal sequences, several Mb in length, and with highly divergent GC contents, introduced in the GC homogeneous *S. cerevisiae* genome. We characterize chromatin assembly, activity and folding of these sequences in this nuclear environment within which they have not evolved by mapping nucleosome assembly, polymerase activity and cohesin binding. In parallel, we trained neural networks (NN) on the profiles along native yeast chromosomes to learn their sequence composition determinants. We then confronted the experimental results obtained on the bacterial chromosomes with the predictions made by the trained NN. Both approaches converge and reveal that bacterial sequences with a GC content similar to those of wild-type yeast spontaneously lead to regular nucleosome positioning, active oriented transcription, and cohesin-mediated loop formation. In contrast, bacterial sequences with a low GC content lead to aberrant protein occupancy profiles that remain nevertheless predicted by the trained NN. These DNA sequences do not lead to the formation of yeast-like chromosomes: they are hardly transcribed and present unusual levels of cohesin deposition in metaphase. Cohesin complexes are still able to compact the chromosome, presumably through active loop extrusion. Using targeted chromosome engineering, we further show that alternating regions with different GC contents within the same chromosome still lead to different regional chromatinization profiles. We found that these

alternating regions self-associate in 3D during interphase, leading to the spontaneous formation of inactive, chromosomal compartments similar to the ones observed in complex multicellular organisms.

These experimental and computational approaches therefore reveal 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” strain or chromosome) or *Mycoplasma pneumoniae* (“Mpneumo”) genomes (**Methods**; **Table S1**). These bacterial sequences contain a yeast centromeric sequence along an autonomous replication sequence (ARS) allowing them to self-replicate and segregate (Lartigue et al. 2009). Each circular bacterial chromosome was linearized using CRISPR-Cas9, and the resulting double-strand breaks extremities repaired by homologous recombination and primers carrying yeast telomere sequences (Lustig 1992) (**Fig. 1A and S1A**; **Methods**). The resulting linear Mmyco presents a 1.22 Mb long left arm and a 4.2 kb long right arm with an overall GC content of 24%, whereas the Mpneumo linear chromosome harbors a 742 kb long left arm and a 76 kb long right arm with an overall GC content of 40% (**Fig. 1A**). For comparison, the GC content of *S. cerevisiae* is 38%.

These chromosomes do not impose a fitness cost on their eukaryotic host, resulting in similar growth rates compared to their parental strains in selective media (**Fig. S1B**) with an estimated segregation rate similar to a centromeric plasmid (Christianson et al. 1992) (**Fig. S1C**). We then proceeded with analyzing the behavior of these two exogenous prokaryotic chromosomes in a eukaryotic environment.

Spontaneous chromatinization of bacterial chromosomes in a eukaryotic context

We first investigated the ability of these exogenous chromosomes to chromatinize. 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; **Fig S2A**).

Strikingly, the chromatin composition of bacterial sequences varies greatly depending on which chromosome is inserted. Genome-wide profiles along the Mpneumo chromosome, the endogenous yeast chromosomes of the same strain, and the parental strain chromosomes appear highly similar in average and amplitude (compare **Fig. 1B**, panel **i** with **ii,**). In sharp contrast, these profiles on both the bacterial and the endogenous yeast chromosomes in the Mmyco strain showed several differences from the profiles found in the parental wild type strain (compare **Fig. 1B**, panel **iii** with **i**).

Firstly, while ChIP-seq profiles of histones H2A and H3 show that these two histones are present along the Mmyco chromosome at levels even slightly higher than along endogenous chromosomes (**Fig S2B**), the MNase-seq signal was strongly reduced compared to the signal on endogenous and parental strain chromosomes (**Fig 1B, Fig S2B**). These results suggest that histones bind DNA but cannot form well-assembled nucleosomes, leaving DNA unprotected. Physical properties of DNA, such as molecular stiffness and bendability, depend on the local sequence features and impact the ability of DNA to wrap around histone cores (Anselmi et al. 1999; De Santis, Morosetti, et Scipioni 2010; Miele et al. 2008; Widom 2001). We found that poly(dA) and poly(dT) tracks, which increase DNA stiffness and reduce bendability, are over-represented in Mmyco DNA, while WW (where W is A or T) (dA/dT) 10-bp periodicity, which facilitates DNA bending around histone cores, is reduced (**Fig S2C-D**). Of note, MNase has more affinity for AT-rich sequences (Chereji, Ocampo, et Clark 2017) and this property may also lead to a decrease of MNase-seq coverage along Mmyco DNA.

Secondly, the PolII signal along the Mmyco chromosome is on average reduced and lower than the control input signal (**Fig. 1B**), whereas Scc1 is enriched across the entire chromosome, with no distinctive enrichment at specific positions, as observed on the endogenous chromosomes. At the same time, Scc1 levels appear respectively strongly and moderately reduced at centromeres and along the arms of the sixteen endogenous *S. cerevisiae* chromosomes (**Fig. 1B**, panel **i** and panel **iii, Fig 1F**). 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.

At a finer scale, we noticed the expected presence of regular nucleosomal arrays on gene bodies as well as peaks of both PolII and Scc1 on endogenous chromosomes in all the strains (**Fig. 1C, Fig. S2B**). We called de novo these peaks using MACS2 (Methods) and plotted the

pile-up distribution of the Scc1 signal within 4 kb windows centered at Scc1 peak summits (**Fig. 1D**). Scc1 peaks across the Mpneumo chromosome appear sharp and with a strong signal-to-noise ratio (SNR) (**Fig. 1D**). Furthermore, these peaks are preferentially located in nucleosome-depleted regions (NDR) and overlap with a moderate local increase of PolII, as observed for Scc1 peaks located in yeast chromosomes of the same strain (**Fig. 1E**). In contrast, Scc1 peaks across the Mmyco chromosome appear poorly defined, with a broader contact signal and a weaker SNR compared to those located in the endogenous chromosomes or in the wild-type strain (**Fig. 1D, Fig. S2E-F**).

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 the Mpneumo chromosome, 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 the Mmyco chromosome, 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 or Mpneumo chromosomes 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 (Eraslan et al. 2019; Routhier et Mozziconacci 2022). We trained a CNN to learn MNase-seq, and Scc1 or PolII ChIP-seq profiles using yeast chromosomes I to XV sequences (**Fig. 2A**; Methods) (Routhier et al. 2021). For each strain, we first checked that the trained CNN could predict these three different 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; **Fig. 2B**, **2C** and **2D**), confirming that genomic sequence is a primary determinant of the chromatin landscape (Chereji et Clark 2018; Levo et al. 2015; Routhier et al. 2021; Sahu et al. 2022; Struhl et Segal 2013). While we could attain a high level of correlation between the

predicted and experimental profiles at the scale of chromosomes, individual cohesin peaks present at the centromeres were not correctly predicted by the trained NN (**Fig. 2C**).

Next, we applied the trained CNNs to Mmyco or the 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 (**Fig. 2B, 2C** and **2D)**, 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 (**Fig. 2B**, **2C** and **2D**).

The GC % being a major difference between the two exogenous chromosomes, we postulated that this GC skew could explain the differences in chromatinization patterns observed between Mmyco and Mpneumo. To test this hypothesis, we predicted using the trained CNNs the MNase-seq signal, and PolII and Scc1 ChIP-seq profiles on artificial, randomly generated 30 kb sequences with variable GC contents ranging from 0 to 100 % (Methods). The MNase-seq, Scc1 and PolII ChIP-seq averaged predicted signals display a strong variability with respect to the GC content. The Mnase signal observed along endogenous chromosomes was only recovered for 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 (**Fig 2E**, **2F** and **2G**), 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.

Replication and pairing of exogenous 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 (Bodmer-Glavas, Edler, et Barberis 2001). 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)(Müller et al. 2014). 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 integrated during chromosome assembly (Ruiz et al. 2019) (**Fig. S3A**). 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 (Batrakou et al. 2020). This pattern is reminiscent of the random and late-replication pattern displayed by the human inactive chromosome X (Koren et McCarroll 2014).

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 (Lengronne et al. 2004; Nasmyth 2001). We therefore investigated SCC by performing image analysis of chromosome pairing (Methods; **Fig. S3B**). 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 (**Fig. S1D**).

Transcriptional activity of bacterial genomes in a yeast context

Chromatin and transcription are closely related to each other (Rando et Winston 2012). We explored whether transcription occurs on the bacterial chromosomes harboring either Y or U chromatin types by performing total, stranded RNA-seq analysis (**Fig. 3A, 3B,** method). Consistent with PolII ChIP-seq profiles (**Fig. 1B**), the Mpneumo chromosome is transcribed all along its length, to levels similar to those of endogenous yeast chromosomes while the Mmyco chromosome is barely transcribed (**Fig. 3A ii** and **iii**). 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 (**Fig. 3B**). Bacterial genes across the Mpneumo chromosome are not transcribed specifically over their coding regions (**Fig. 3C**), and transcription does not initiate at bacterial TSSs, consistent with an absence of PolII enrichment at these loci (**Fig. S4A, S4B**). While the Y chromatin type found on the Mpneumo chromosome is transcribed, the U chromatin type found on the Mmyco chromosome is sparsely and lowly transcribed (**Fig. 3Aiii**), again in good agreement with the poor levels of PolII deposition (**Fig. 1B**). The low transcription of the Mmyco genome is not explained by the recruitment of the SIR

Sir2/3/4 silencing complex, as evidenced by the ChIP-seq profile of Sir3 deposition, which reveals only significant enrichment near the Mmyco and yeast endogenous (sub)telomeres (**Fig. S4C**). A possibility is that the low transcription rate rather results from the impairment of a proper nucleosomal organization with clear nucleosome depleted regions, known to be required at TSS (Tirosh et Barkai 2008)

The stranded RNA-seq profiles also allow characterization of the orientation of transcription tracks. 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 (**Fig. 3D**, see the orientation of genes and stranded tracks in **Fig. 3B**). The results suggest either a slight preference of the yeast RNA polymerase for transcribing one of the two strands, or a differential instability of the RNA molecules from one or the other strand (see discussion).

Interplay between transcription and cohesin localization

In yeast, cohesin accumulates at sites of convergent transcription (Lengronne et al. 2004; Glynn et al. 2004). In agreement, we found that cohesin peaks along endogenous yeast chromosomes are closer to convergent transcription sites than expected by chance (**Fig. 3B**, green dotted lines, **Fig. S4D**). This behavior extends to the Mpneumo chromosome (**Fig. 3B, Fig. S4D**). Comparison of the Scc1 ChIP-seq and stranded RNA-seq profiles over the Mpneumo chromosome suggest that sharper cohesin peaks are associated with sharper transcriptional convergence (**Fig. 3B**, compare the shapes of the peaks marked by arrows in the Mpneumo chromosome). To quantify this behavior, we aggregated the strand specific RNA-seq signal in a 8 kb window centered at Scc1 peaks found either in yeast chromosome II or in Mpneumo chromosome (**Fig. 3E**). The strongest Scc1 peaks are correlated with a stronger transcriptional convergence both in yeast and over the Mpneumo chromosome (**Fig. 3E-F**, Methods), underlying the similarity of mechanisms at work within the Y chromatin regions. In sharp contrast, high cohesin levels span the entire transcriptionally silent U chromatin regions, with no clear peaks, most likely reflecting the lack of convergent transcription sites (**Fig 1B** and **Fig 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 (Dekker et al. 2002; Lieberman-Aiden et al. 2009) on the Mmyco or Mpneumo strains arrested either in G1 or G2/M (**Fig. 4A,B**).

In G1-arrested cells, yeast chromosomes exhibit a Rabl configuration, with their centromeres clustered at the microtubule organizing center and subtelomeric regions colocalizing at the nuclear periphery, resulting in a dotted pattern in Hi-C contact maps (Bystricky et al. 2004; Duan et al. 2010; Lazar-Stefanita et al. 2017). For both Mmyco and Mpneumo chromosomes, the added yeast centromeric sequence clusters with the native yeast centromeres, as expected. However, the rest of the bacterial chromosome sequences, being either covered with Y or U chromatin, exhibited very distinct structural characteristics (**Fig. 4A**).

In G1-arrested cells HiC maps, the Y-chromatin type Mpneumo chromosome displays the same structural characteristics as endogenous *S. cerevisiae* chromosomes (**Fig. 4A**). The proportion of trans-contacts with all other chromosomes (**Fig. S5B**), and of long-range cis-contacts (**Fig. 4C**), are comparable to those found for yeast chromosomes. The slope of the p(s) average contact probability as a function of the genomic distance (s) is close to -1.5 both for yeast and the bacterial chromosomes, a value corresponding to the typical polymer structure observed in simulations (**Fig. S5C**) (Wong et al. 2012). DNA-FISH labeling of Mpneumo DNA reveals an extended and contorted structure within the intranuclear space, confirming the intermixing and the positioning of the chromosome in the vicinity of other active regions of native *S. cerevisiae* chromosomes (**Fig. 4E, S5H, S5I**) (Methods).

In contrast, trans-contacts between the Mmyco and endogenous yeast chromosomes were much weaker than those of yeast chromosomes with each other (**Fig. S5B**). The entire Mmyco chromosome appears isolated from the 16 native yeast chromosomes, interacting predominantly with their 32 subtelomeric regions (**Fig. 4A**, dotted rectangles; **Fig. S5A**). The slope of the p(s) curve also changes dramatically towards a value of -1 at shorter distances, corresponding to a polymer globule (**Fig. 4C, Fig. S5C**) (Lieberman-Aiden et al. 2009). DNA-FISH labeling of the Mmyco chromosome confirmed the globular conformation of the chromosome and revealed its preferential position at the nuclear periphery that could account for its preferential interaction with the telomeres of the yeast chromosomes (**Fig. 4E; Fig. S5H**, **S5I**).

Both imaging and Hi-C data showed that the physical nature of Y and U chromatin are such that they result in very different structures: Mmyco AT-rich sequence folds into a compact, self-interacting globular structure at the nuclear periphery while the Mpneumo GC rich sequence unfolds as a WT yeast chromosome.

We then generated Hi-C contact maps and p(s) decay curves 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 (Garcia-Luis et al. 2019; Lazar-Stefanita et al. 2017; Schalbetter et al. 2017). The decrease of interchromosomal contacts (**Fig 4B, Fig S5B)** and the change in p(s) (**Fig 4D, S5D**) observed for both native and bacterial chromosomes indicate that mitotic compaction is happening in a similar way for Y and U chromatin. Although the two Y and U chromatin types displayed different contact patterns at short scales. On the one hand, endogenous yeast and Mpneumo chromosomes exhibit focal enrichments of medium-range contacts corresponding to anchored loops (black triangles on **Fig. 4F**)(Costantino et al. 2020; Dauban et al. 2020; Garcia-Luis et al. 2019). On the other hand, the Mmyco chromosome, although covered by cohesin, displayed a strikingly different pattern with no visible loops (**Fig. 4G**), suggesting that cohesins are able to promote compaction but that the loop anchors are not preferentially positioned along the sequence (**Fig. S5D**). Using the loop caller chromosight (Matthey-Doret et al. 2020), 59 loops were found across the Mpneumo chromosome. These loops span over slightly longer genomic distances (28 kb on average) than the ones found along endogenous yeast sequences (22 kb on average on the endogenous chromosomes) (**Fig S5E**). The dotted grid pattern delimiting regions of enriched contacts in Mpneumo is reminiscent of DNA loops and TADs observed along mammalian interphase chromosomes (Rao et al. 2014). This suggests that, on the Mpneumo chromosome, some genomic positions have the ability to be involved in long-range loops of various sizes, a phenomenon not observed across native *S. cerevisiae* wild-type chromosomes (**Fig. 4F**). Whether these loops are formed simultaneously within the same cells, with the anchors involved in multiple loop boundaries, or whether different cells present different loops, remain unknown (Costantino et al. 2020; Dauban et al. 2020).

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 mitotic DNA loops in G2/M.

Cohesin compacts exogenous bacterial chromosomes in yeast

As shown above, Mmyco and Mpneumo chromosomes display different Scc1 binding landscapes. The Mmyco chromosome shows a large Scc1 recruitment but no defined peaks, whereas Mpneumo shows a Scc1 profile similar to that over native yeast chromosomes. In yeast, we and others previously showed that these cohesin peaks correspond to the loop basis observed in metaphase Hi-C maps (Garcia-Luis et al. 2019; Dauban et al. 2020; Costantino et al. 2020). We therefore investigated the relationship between cohesin deposition and contact maps patterns for the bacterial sequences. As for WT yeast chromosomes, strong Scc1 peaks are close to anchors of focal DNA loops in the Mpneumo chromosome (median distance of 566 bp from nearest loop anchor) (**Fig. 5A, B**). The colocalization of the loop basis and Scc1 peaks suggest 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 (Goloborodko et al. 2016; Haarhuis et al. 2017). To test this hypothesis, we performed Hi-C in absence of the cohesin-holocomplex member Wpl1. In absence of Wpl1 (human Wapl), cohesin removal from DNA is impaired: as a result, extended loop expansion occurs and results in longer loops readily visible by the inflection of the p(s) curve towards longer distances in Hi-C contact maps of human and yeast cells (Dauban et al. 2020; Haarhuis et al. 2017). For both strains, we observed in absence of Wpl1 decreased mid-range and increased long-range contacts in all chromosomes, including Mmyco and Mpneumo (**Fig. 5C, 5D, Fig. S5F**). Consistently, the strength of short-range loops identified at discrete positions in yeast chromosomes and Mpneumo decreased upon Wpl1 depletion (**Fig. S5G**). These results show that Y and U chromatin are prone to dynamic cohesin-dependent DNA compaction from the S to G2/M phase (**Fig. 5E**).

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 organisms' chromosomes (Lieberman-Aiden et al. 2009). We therefore wondered whether they can coexist on a single chromosome and designed three strains for this purpose. 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; see schematic representation on **Fig. 6A**; Methods; see also **Fig. S6B**, **S6C** for validation). We then investigated the behavior of chromatin along these chromosomes using ChIP-seq and Hi-C.

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 (**Fig. S6D**). Also, besides the expected changes induced by the loss of chr. XVI centromere and the fusion of the two chromosomes (e.g. (Lazar-Stefanita et al. 2022; Shao et al. 2018)) the overall Hi-C contact maps of XVIfMmyco strain synchronized in G1 showed little differences compared to the parental Mmyco strain (**Fig. 6A**, 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, but for its contacts with subtelomeric regions (**Fig. 6A**, **6B**). In the translocated strains, however, the shuffling of chromatin type resulted in strikingly contrasted Hi-C contact maps, with a chessboard pattern appearing within XVIfMmycot1 and XVIfMmycot2 chromosomes (**Fig. 6A**, panels **iii** and **iv**, see correlation maps in **Fig. S6E**). This chessboard pattern reflects the fact that U-type regions make specific contacts over long distances, bypassing the Y-type regions found in between (**Fig. 6B**; see also **Fig. S6F** G1). Similarly, Y-type regions of chromosome XVI also are also involved in specific Y-Y contacts over longer distances (**Fig. 6B**). 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 (**Fig. S6F**). These results unambiguously show that alternate Y and U regions along chromosomes are spontaneously segregated within the nuclear space, a phenomena reminiscent of eu- and hetero chromatin compartments found in higher eukaryotes (Lieberman-Aiden et al. 2009).

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 (**Fig. S6D**). In addition, cohesin coating remained enriched over 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 were observed in G2/M Hi-C contact maps at the junction between Mmyco and the endogenous yeast sequences. There, a DNA loop appears over the boundary (**Fig. 6C**). The basis 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. This suggests that loop extrusion can proceed through the

Y/U chromatin junction to form a loop bridging the two nearest cohesin enrichment sites on each side (Nora et al. 2017; Spracklin et al. 2021).

At a broader scale, Hi-C maps of the mosaic chromosomes in G2/M show that U-type compartments disappear upon cohesin-mediated compaction (**Fig. 6A**). The analysis of distance-dependent contact frequency shows that all chromosome regions behave similarly at this stage (**Fig. S6F**). This observation shows that, as observed in multicellular organisms, LE mediated metaphase chromosome compaction abolishes chromatin compartments to prepare chromosomes for segregation (Gibcus et al. 2018; Goloborodko et al. 2016; Naumova et al. 2013).

Discussion

In this work, we explore how Mb-long DNA sequences behave in a cellular context into which they have not evolved. We show that exogenous DNA sequences behave in a way that is largely predictable by neural networks trained on endogenous sequences, suggesting that the behavior of the cellular machinery is extremely robust, and indiscriminately processes any sequence. The GC-content appears as one of the main drivers of chromatin composition and activity. Chromosomes with two different values of the GC-content 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. The Y and U compartments disappear during chromosome compaction in mitosis preparing for chromosome segregation (Gibcus et al. 2018).

Chromatin structure and activity is driven by the sequence composition

Our results suggest that the chromatin assembled in a given nuclear context on a DNA molecule depends on first approximation on its GC content. DNA is always covered with histones, as this association only depends on the attraction between positively charged proteins and the negatively charged double helix, but a nucleosome's proper assembly depends on the GC content of the sequence. Extremely low levels of GC are not compatible with the wrapping of DNA into nucleosomes and we hypothesize that this could be due to the frequent occurrence of poly(A) motifs that are known to prevent the formation of nucleosomes (Routhier et al. 2021). This disordered coating of the DNA molecule with histones may result in the presence of intra

nucleosomal patches along the molecule that are prone to sticky interactions. The resulting self-attracting potential would lead to the collapse of the whole chromosome into a globule (Barbieri et al. 2012), in agreement with both the observed p(s) and FiSH microscopy images. When such domains of U chromatin are fragmented within a Y type chromatin chromosome, the appearance of long range contacts within chromatin of similar type could also be explained by the attractive potential resulting from the non-physiological nucleosome folding. We also observed a correlation between this compact, globular state and a suppression of transcription, although we cannot conclude here about causal relationship. The formation of arrays of well-assembled nucleosomes on an exogenous sequence within the GC range of the evolved host DNA is also associated with the binding of polII and the initiation of transcriptional tracks. The resulting RNA is stable enough to be sequenced, suggesting that all transcriptional and cell quality control mechanisms are unable to repress this mechanism.

In metaphase, the SMC cohesin complex binds to all exogenous sequences, regardless of their transcriptional status and GC content, and appears to start expanding, presumably through extrusion, DNA loops through an active ATPase dependent process (Bastié et al. 2022; Davidson et Peters 2021). Along Y chromatin, these complexes relocate to convergent transcription tracks, leading to an accumulation of cohesins in these regions (Dauban et al. 2020; Lengronne et al. 2004; Ocampo-Hafalla et al. 2016). 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.

Part of these results are in line with previous works that characterized the behavior of exogenous DNA within the yeast nucleus. For instance, DeBoer et al. showed using short (13kb) random DNA sequences that TSS could be formed and that those correspond most likely to nucleosome depleted regions that exist along the chromosome (de Boer et al. 2014). In later works, the same group showed that random DNA could be used to generate yeast promoters and used this property to decipher the regulatory logic of transcription activation (de Boer et al. 2020; Vaishnav et al. 2022).

Prokaryotic DNA contains determinants that favor transcription or RNA orientation in eukaryote

A large bacterial DNA molecule that never evolved in a yeast environment will be spontaneously processed by the yeast nuclear machinery. Active transcription of both normal GC% and the AT-rich bacterial chromosome follow, on average, the orientation of the bacteria genes (**Fig. 3D**). In yeast, the Nonsense-Mediated Decay (NMD) complex destabilizes

eukaryotic transcripts with long 3' UTR. One possibility is that NMD will be more frequently activated for bacteria antisense RNA compared to sense RNA, given the former will contain more random, short CDS than the latter that are more likely to contain long CDS. Alternatively, binding sites of the Nrd1-Nab3-Sen1 (NNS) complex that lead to transcription termination are enriched in yeast along ncRNA and absent along mRNA, a bias influenced by the codon usage. It is also possible that the bacteria codon usage bias would result in a similar enrichment in NNS sites in the ncRNA strand (Schulz et al. 2013). Another, more speculative possibility is that sequence determinants predating the divergence of eucaryotes and procaryotes could 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. From the standpoint of synthetic biology and genomics, our results show that the GC content 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.

Cohesin-mediated loop formation along non-transcribed DNA templates

Investigations of chromosome folding have focused in recent years on the roles of SMC complexes in promoting formation of DNA loops in all domains of life (reviewed in Davidson et Peters 2021; Mirny et Dekker 2022). The regulatory mechanisms of the loop extrusion process are highly conserved between human and yeast (Bastié et al. 2022; Dauban et al. 2020; Haarhuis et al. 2017; Wutz et al. 2020; 2017; van Ruiten et al. 2022). Our results on long exogenous sequences provide new insights on the mechanisms behind cohesin-mediated loops. They show that (1) cohesin is loaded on long lowly transcribed DNA sequences, independently of centromere vicinity or active promoters, and (2) that cohesin compacts DNA along these regions, without forming visible DNA loops with discrete basis. The positions of loop basis 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 (Dauban et al. 2020; Jeppsson et al. 2022; Costantino et al. 2020). Loop positioning, stabilization and maintenance at discrete loci also involves Eco1 dependent acetylation of cohesin Smc3 (Bastié et al. 2022; Dauban et al. 2020; Wutz et al. 2020; van Ruiten et al. 2022).

We propose that late replication and low transcription account 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 residual transcription may still serve this purpose and/or interfere with loop extrusion process (Jeppsson et al. 2022). In agreement with *in vitro* data (Davidson et Peters 2021), 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 (Lengronne et al. 2004; Glynn et al. 2004). 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. In addition, it remains unclear why some positions become involved in multiple loops while others are restricted to the same partner.

Compartmentalization of the genome

The nucleus of WT yeast (and other fungi) is typically separated into two compartments: the larger one corresponding to 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 and expressed Y chromatin occupies the central space of the nucleus. 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 located 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, CC, LM, ER, 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. 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 to RK (ERC grant agreement 771813). 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

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.

D) Aggregated profile of Scc1 deposition centered at Scc1 peaks (+/- 2kb) called over the yeast chromosome II (left), the Mpneumo (middle) and the Mmyco (right) chromosomes.

E) 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)

F) 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.

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

A) 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%.

B-D) 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.

E-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.

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) 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).

D) 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.

E) 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.

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

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**).

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) 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).

Figure 5. Role of cohesin in the folding of exogenic bacterial sequences.

A) Top: Magnification of the Hi-C contact map generated in G2 over a 150kb window within the Mpneumo bacterial chromosome. Circles represent chromatin loops mapped with Chromosight and black triangles indicate loop anchors. Bottom: Scc1 ChIP-seq deposition profile. Red triangles indicate Scc1 peaks identified with MACS2.

B) 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.

C) Hi-C contact maps of Mmyco and Mpneumo chromosomes in “wild-type” and upon Wpl1 depletion (Δwpl1) (top) and ratio map of Δwpl1 over WT (bottom). Red (or blue) indicate enriched (or depleted) contacts in Δwpl1.

D) Distance-dependent contact frequency in endogenous yeast chromosomes and in bacterial chromosomes in Mmyco (top) and Mpneumo (bottom) strains, in “wild-type” (dashed) and in Δwpl1 mutants.

E) Model of chromatin folding upon cohesin-dependent compaction of *M. mycoides* U-type chromatin (top) or of *M. pneumoniae* Y-type chromatin (bottom) sequences. Along Mmyco U-type chromatin, cohesin binding over the bacterial chromosome is increased without specific enrichment at convergent transcription sites. We propose that cohesin compacts DNA independently of transcription. On Mpneumo Y-type chromatin, cohesin dependent DNA loops are stabilized at convergent transcription sites.

Figure 6. 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) G1 and G2/M 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).

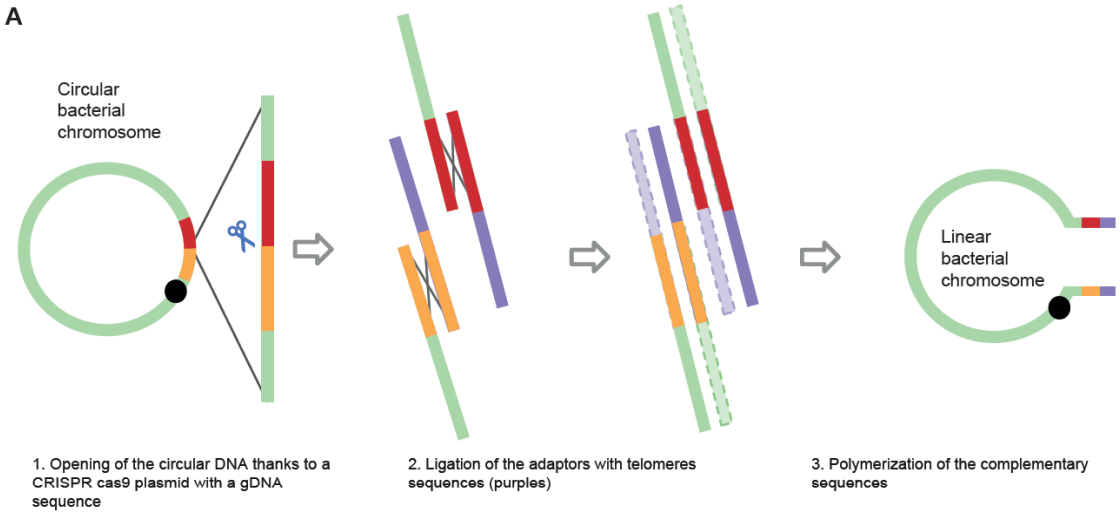
C) Virtual 4C profiles of viewpoints located within yeast chr. XVI (in grey) or Mmyco chromosome segments (in blue), in XVIfMmyco, XVIfMmycot1 and XVIfMmycot2 strains. The large grey/blue arrows indicate the location of each viewpoint. Note the increase in contacts of chr. XVI Y chromatin segments with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

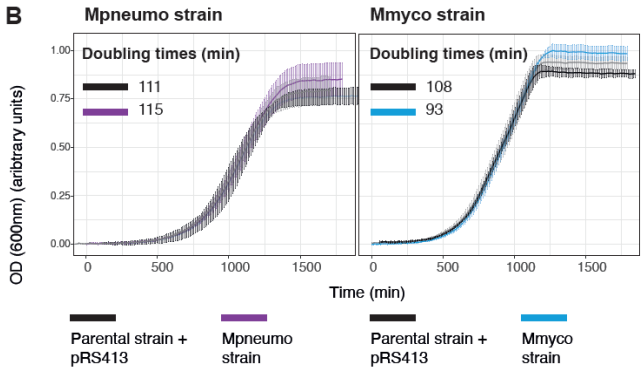
D) Hi-C contact maps in G2/M of 500 kb window of either the XVIfMmyco (top) or XVIfMmycot1 (bottom) chromosomes centered on the translocation position. The Scc1 deposition profile measured by ChIP-seq in each strain is shown below underneath the contact maps. In XVIfMmycot1, this window is effectively centered at the junction between the yeast chr. XVI segment (upstream) and the Myco chromosome segment (downstream).

E) Same as **D**) but with 100 kb window of either the XVIfMmyco (left) or XVIfMmycot1 (right) strains centered on the sequences involved in a junction of the XVIfMmycot1 The Scc1 deposition profile measured by ChIP-seq in each strain is shown atop the contact maps. Green arrows: cohesin peaks flanking the breakpoint.

F) Illustration of the behaviors and activity of Mpneumo (left panel) and Mmyco (middle and right) chromosome sequences in the yeast nucleus.

Supplementary figure legends





C

	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

Figure S1. 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 (Christianson et al. 1992) (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.

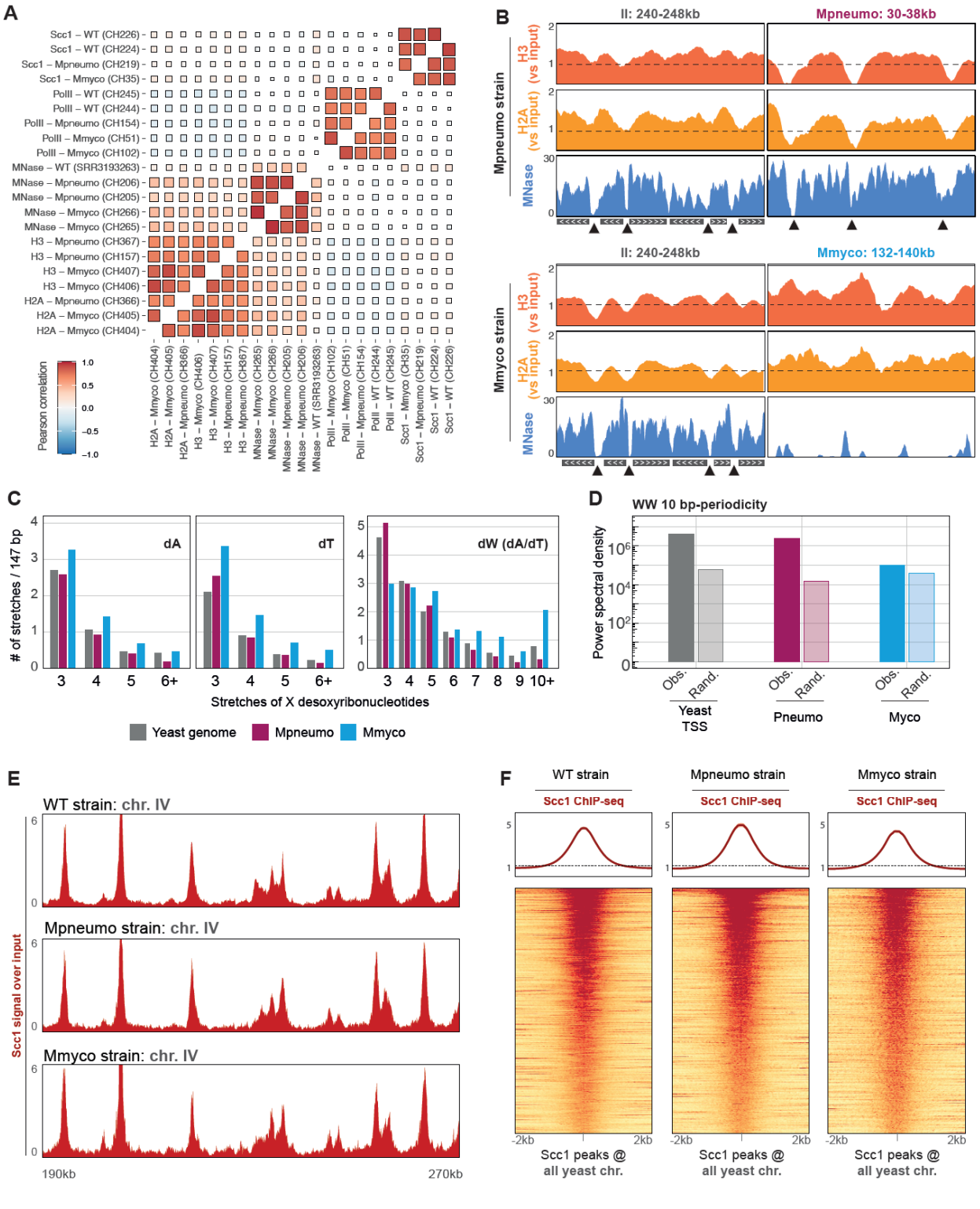


Figure S2. Chromatin structuration of the bacterial chromosomes in yeast.

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

F) Aggregated profile of Scc1 deposition centered at Scc1 peaks (+/- 4kb) called over all endogenous yeast chromosomes in the WT, Mpneumo and Mmyco strains.

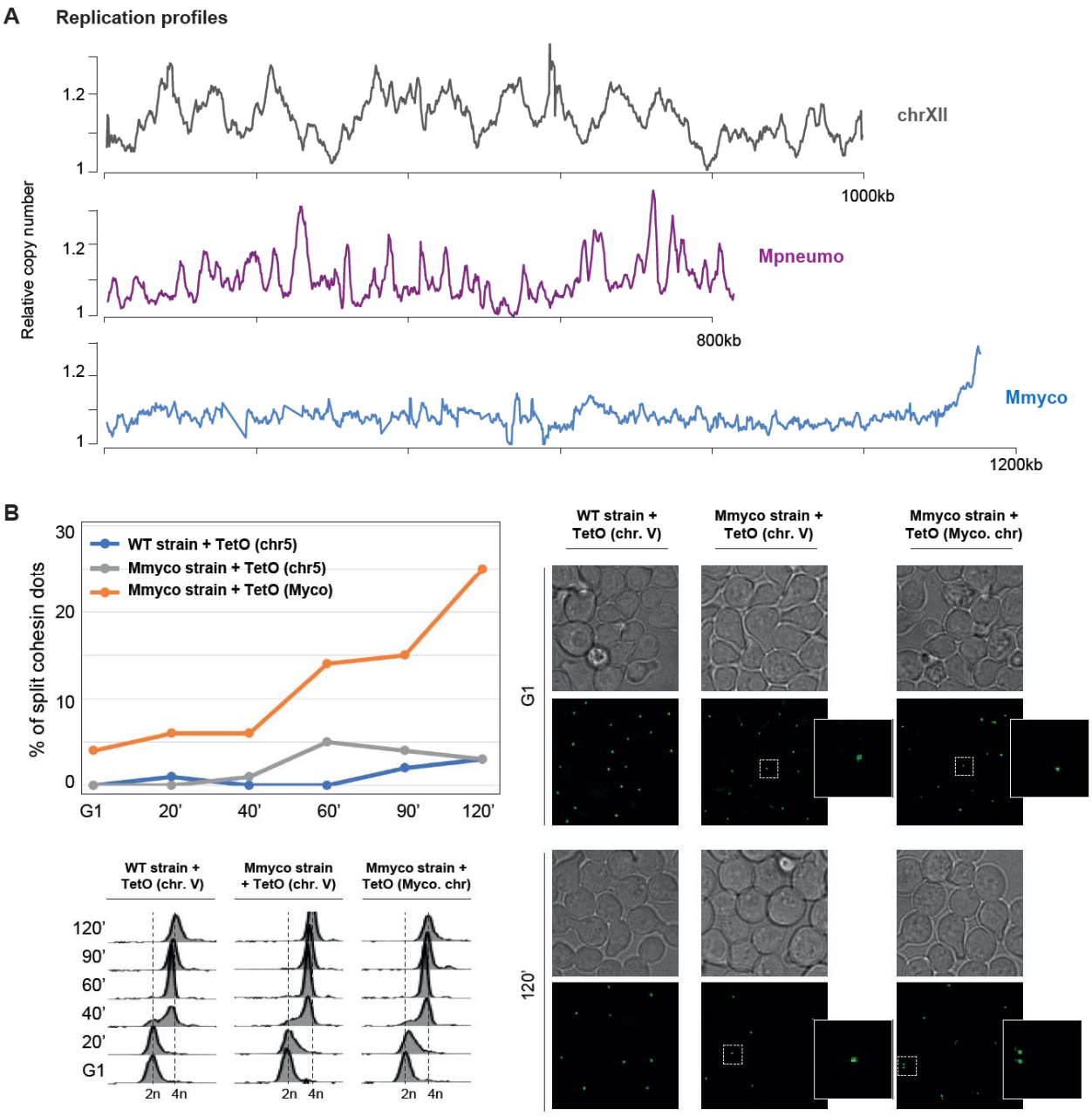


Figure S3. Replication of bacterial genomes in yeast.

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.

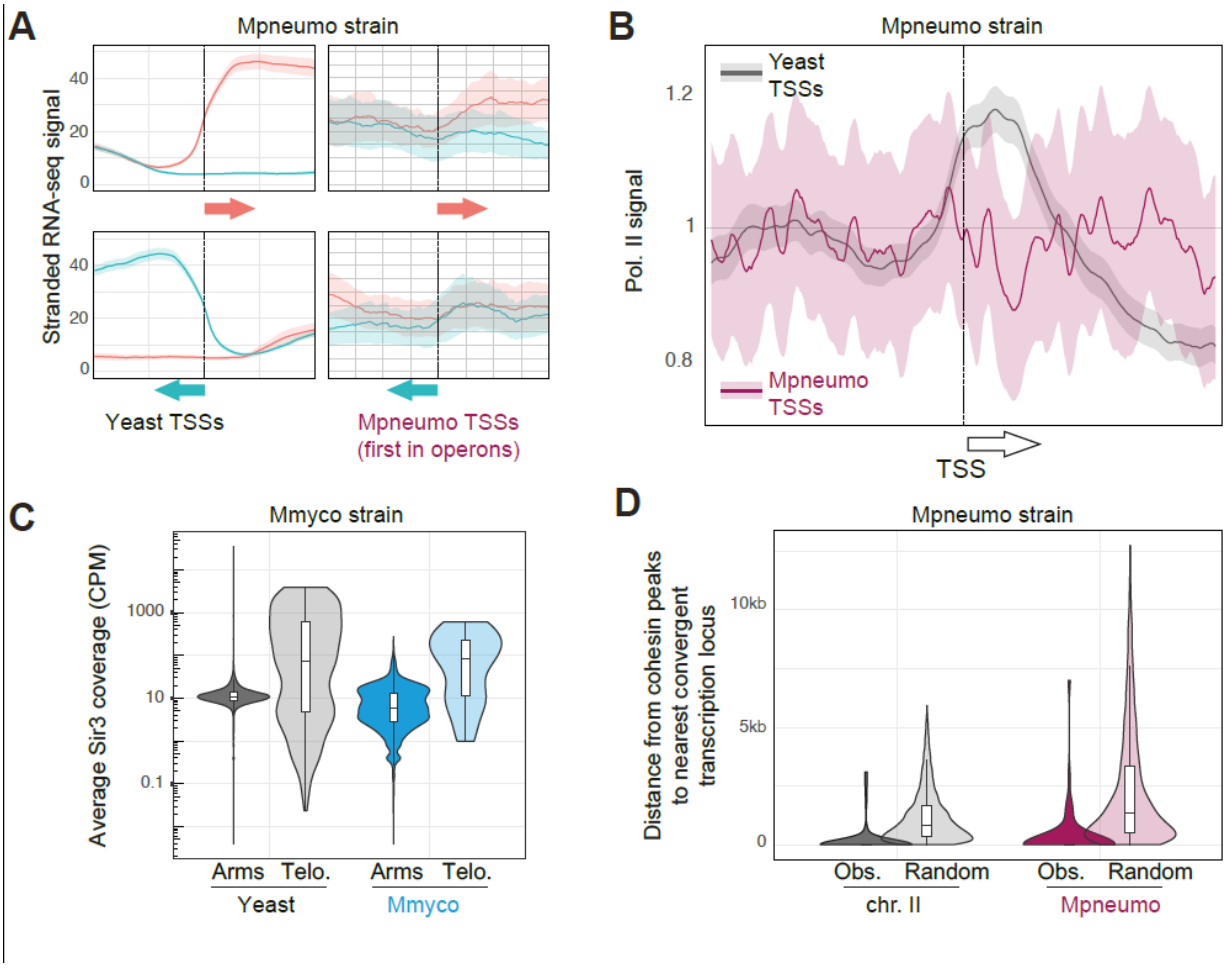


Figure S4. Transcription orientation of bacterial genomes in yeast.

A) 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.

B) 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.

C) 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.

D) 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.

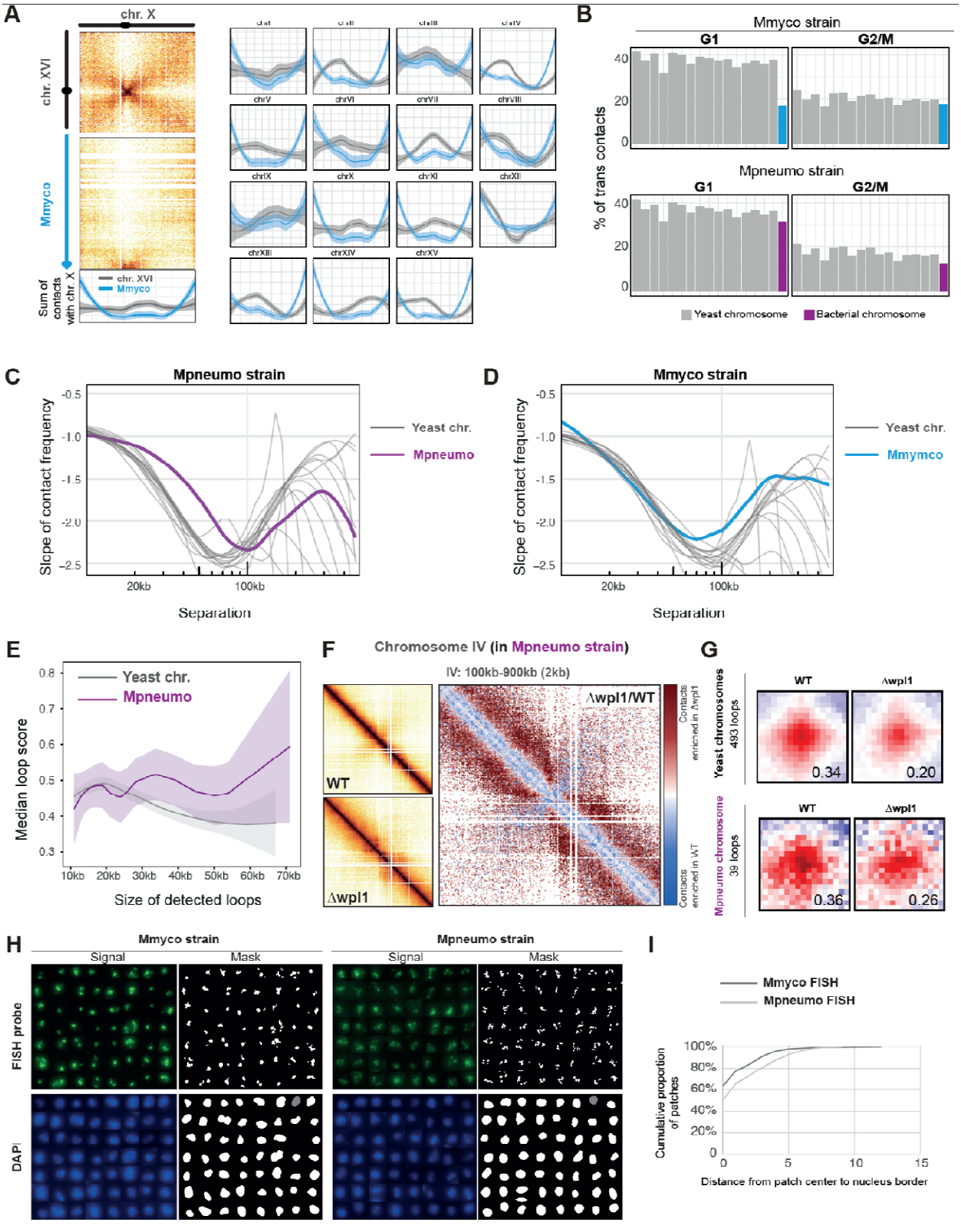


Figure S5. Folding of exogenic bacterial sequences within the yeast nucleus.

A) 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.

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

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 (Matthey-Doret et al. 2020)) for loops along either endogenous (grey) or Mpneumo (purple) chromosomes.

F) 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) Average loop score and aggregated Hi-C signal centered at loops over endogenous yeast (top) and Mpneumo (bottom) chromosomes, in “wild-type” and upon Wpl1 depletion (Δwpl1).

H) 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).

I) 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.

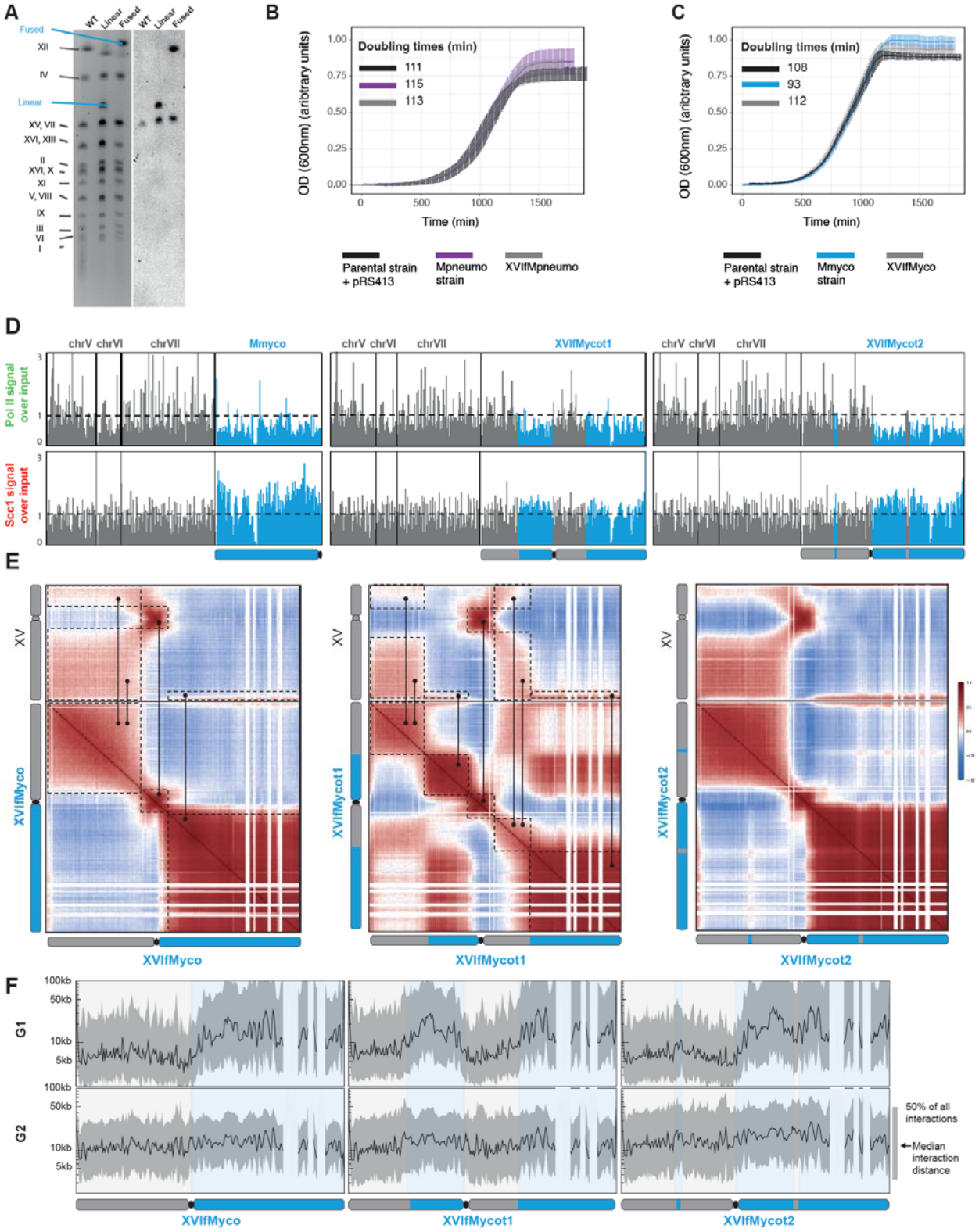


Figure S6. Compartmentalization of mosaic chromosomes composed of Y and U-type chromatin.

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) Auto-correlation contact matrices in mosaic chromosomes from XVIfMmyco (left), XVIfMmycot1 (center) and XVIfMmycot2 (right) strains.

F) 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.

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. m*ycoides* strain PG1 (Labroussaa et al. 2016) and *Mycoplasma pneumoniae* strain M129 (Ruiz et al. 2019), 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 (Laughery et al. 2015) or PAEF5 (Agier et al. 2021) 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. (2018). 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 (Agier et al. 2021) 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 (Labun et al. 2019) or CRISPOR (Concordet et Haeussler 2018). 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 (Christianson et al. 1992). 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 (Koszul et al. 2004) 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 (Viterbo et al. 2018). 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. (2020). 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 (Hu et al. 2015). 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.

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 (Langmead et Salzberg 2012) 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 (Cournac et al. 2012). 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 (Abdennur et Mirny 2020)**.**

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:

$$\square\square\square_{\square} = \frac{\sum_{\square=\square}^{\square+200} (\square\square\square_{\square\square\square,\square} - \square\square\square_{\square\square\square,\square})}{\sum_{\square=\square-200}^{\square} (\square\square\square_{\square\square\square,\square} - \square\square\square_{\square\square\square,\square})}$$

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

$$\square\square\square\square_{\square} = \left[\sum_{\square=\square-2000}^{\square} (\square\square\square_{\square\square\square,\square} > 20) - \sum_{\square=\square-2000}^{\square} (\square\square\square_{\square\square\square,\square} > 20) + \sum_{\square=\square}^{\square+2000} (\square\square\square_{\square\square\square,\square} > 20) - \sum_{\square=\square}^{\square+2000} (\square\square\square_{\square\square\square,\square} > 20) \right] \times 100 / (4000 \times 4)$$

Thus, $\square\square\square\square_{\square}$ scores range between -50 and 50 and a positive (negative) $\square\square\square\square_{\square}$ corresponds to a local genomic position $\square$ of convergent (divergent) transcription. The relationship between $\square\square\square\square$ 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 (Batrakou et al. 2020).

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 (Chollet 2015) and Tensorflow (Abadi et al. 2015) 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 (Routhier et al. 2021). 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 setl 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 (Kelley et al. 2018). 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 (Routhier et al. 2021).

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. (1999), with some modifications (Gotta, Laroche, et Gasser 1999). 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

Abadi, Martín, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, et al. 2015. *TensorFlow: Large-scale machine learning on heterogeneous systems*. tensorflow.org.

Abdennur, Nezar, et Leonid A. Mirny. 2020. « Cooler: Scalable Storage for Hi-C Data and Other Genomically Labeled Arrays ». *Bioinformatics (Oxford, England)* 36 (1): 311□16. https://doi.org/10.1093/bioinformatics/btz540.

Agier, Nicolas, Aubin Fleiss, Stéphane Delmas, et Gilles Fischer. 2021. « A Versatile Protocol to Generate Translocations in Yeast Genomes Using CRISPR/Cas9 ». *Methods in Molecular Biology (Clifton, N.J.)* 2196: 181□98. https://doi.org/10.1007/978-1-0716-0868-5_14.

Anselmi, C, G Bocchinfuso, P De Santis, M Savino, et A Scipioni. 1999. « Dual Role of DNA Intrinsic Curvature and Flexibility in Determining Nucleosome Stability11Edited by T. Richmond ». *Journal of Molecular Biology* 286 (5): 1293□1301. https://doi.org/10.1006/jmbi.1998.2575.

Baby, Vincent, Fabien Labroussaa, Joëlle Brodeur, Dominick Matteau, Géraldine Gourgues, Carole Lartigue, et Sébastien Rodrigue. 2018. « Cloning and Transplantation of the Mesoplasma florum Genome ». *ACS Synthetic Biology* 7 (1): 209□17. https://doi.org/10.1021/acssynbio.7b00279.

Barbieri, Mariano, Mita Chotalia, James Fraser, Liron-Mark Lavitas, Josée Dostie, Ana Pombo, et Mario Nicodemi. 2012. « Complexity of Chromatin Folding Is Captured by the Strings and Binders Switch Model ». *Proceedings of the National Academy of Sciences of the United States of America* 109 (40): 16173□78. https://doi.org/10.1073/pnas.1204799109.

Bastié, Nathalie, Christophe Chapard, Lise Dauban, Olivier Gadal, Frédéric Beckouët, et Romain Koszul. 2022. « Smc3 Acetylation, Pds5 and Scc2 Control the Translocase Activity That Establishes Cohesin-Dependent Chromatin Loops ». *Nature Structural & Molecular Biology*, juin, 1□11. https://doi.org/10.1038/s41594-022-00780-0.

Batrakou, Dzmitry G., Carolin A. Müller, Rosemary H. C. Wilson, et Conrad A. Nieduszynski. 2020. « DNA Copy-Number Measurement of Genome Replication Dynamics by High-Throughput Sequencing: The Sort-Seq, Sync-Seq and MFA-Seq Family ». *Nature Protocols* 15 (3): 1255□84. https://doi.org/10.1038/s41596-019-0287-7.

Bernardi, Giorgio, Birgitta Olofsson, Jan Filipski, Marino Zerial, Julio Salinas, Gerard Cuny, Michele Meunier-Rotival, et Francis Rodier. 1985. « The Mosaic Genome of Warm-Blooded Vertebrates ». *Science* 228 (4702): 953□58. https://doi.org/10.1126/science.4001930.

Bodmer-Glavas, Morana, Karin Edler, et Alcide Barberis. 2001. « RNA polymerase II and III transcription factors can stimulate DNA replication by modifying origin chromatin structures ». *Nucleic Acids*

Research 29 (22): 4570□80.

Boer, Carl G. de, Harm van Bakel, Kyle Tsui, Joyce Li, Quaid D. Morris, Corey Nislow, Jack F. Greenblatt, et Timothy R. Hughes. 2014. « A Unified Model for Yeast Transcript Definition ». *Genome Research* 24 (1): 154□66. https://doi.org/10.1101/gr.164327.113.

Boer, Carl G. de, Eeshit Dhaval Vaishnav, Ronen Sadeh, Esteban Luis Abeyta, Nir Friedman, et Aviv Regev. 2020. « Deciphering Eukaryotic Gene-Regulatory Logic with 100 Million Random Promoters ». *Nature Biotechnology* 38 (1): 56□65. https://doi.org/10.1038/s41587-019-0315-8.

Bystricky, Kerstin, Patrick Heun, Lutz Gehlen, Jörg Langowski, et Susan M. Gasser. 2004. « Long-Range Compaction and Flexibility of Interphase Chromatin in Budding Yeast Analyzed by High-Resolution Imaging Techniques ». *Proceedings of the National Academy of Sciences of the United States of America* 101 (47): 16495□500. https://doi.org/10.1073/pnas.0402766101.

Camellato, ProfileBrendan, Ran Brosh, Matthew Maurano, et Jef Boeke. 2022. « Genomic analysis of a synthetic reversed sequence reveals default chromatin states in yeast and mammalian cells ». *bioRxiv*, juin. https://doi.org/10.1101/2022.06.22.496726.

Chereji, Răzvan V., et David J. Clark. 2018. « Major Determinants of Nucleosome Positioning ». *Biophysical Journal* 114 (10): 2279□89. https://doi.org/10.1016/j.bpj.2018.03.015.

Chereji, Răzvan V., Josefina Ocampo, et David J. Clark. 2017. « MNase-Sensitive Complexes in Yeast: Nucleosomes and Non-Histone Barriers ». *Molecular Cell* 65 (3): 565-577.e3. https://doi.org/10.1016/j.molcel.2016.12.009.

Chollet, Francois. 2015. *keras*. https://github.com/keras-team/keras.

Christianson, T. W., R. S. Sikorski, M. Dante, J. H. Shero, et P. Hieter. 1992. « Multifunctional Yeast High-Copy-Number Shuttle Vectors ». *Gene* 110 (1): 119□22. https://doi.org/10.1016/0378-1119(92)90454-w.

Concordet, Jean-Paul, et Maximilian Haeussler. 2018. « CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens ». *Nucleic Acids Research* 46 (W1): W242□45. https://doi.org/10.1093/nar/gky354.

Coradini, Alessandro L. V., Cara B. Hull, et Ian M. Ehrenreich. 2020. « Building Genomes to Understand Biology ». *Nature Communications* 11 (1): 6177. https://doi.org/10.1038/s41467-020-19753-2.

Costantino, Lorenzo, Tsung-Han S. Hsieh, Rebecca Lamothe, Xavier Darzacq, et Douglas Koshland. 2020. « Cohesin Residency Determines Chromatin Loop Patterns ». *ELife* 9 (novembre): e59889. https://doi.org/10.7554/eLife.59889.

Cournac, Axel, Romain Koszul, et Julien Mozziconacci. 2016. « The 3D Folding of Metazoan Genomes Correlates with the Association of Similar Repetitive Elements ». *Nucleic Acids Research* 44 (1): 245□55. https://doi.org/10.1093/nar/gkv1292.

Cournac, Axel, Hervé Marie-Nelly, Martial Marbouty, Romain Koszul, et Julien Mozziconacci. 2012. « Normalization of a Chromosomal Contact Map ». *BMC Genomics* 13 (août): 436. https://doi.org/10.1186/1471-2164-13-436.

Currin, Andrew, Steven Parker, Christopher J. Robinson, Eriko Takano, Nigel S. Scrutton, et Rainer Breitling. 2021. « The Evolving Art of Creating Genetic Diversity: From Directed Evolution to Synthetic Biology ». *Biotechnology Advances* 50 (octobre): 107762. https://doi.org/10.1016/j.biotechadv.2021.107762.

D, Wang. 2018. « GCevobase: An Evolution-Based Database for GC Content in Eukaryotic Genomes ». *Bioinformatics (Oxford, England)* 34 (12). https://doi.org/10.1093/bioinformatics/bty068.

Dauban, Lise, Rémi Montagne, Agnès Thierry, Luciana Lazar-Stefanita, Nathalie Bastié, Olivier Gadal, Axel Cournac, Romain Koszul, et Frédéric Beckouët. 2020. « Regulation of Cohesin-Mediated Chromosome Folding by Eco1 and Other Partners ». *Molecular Cell* 77 (6): 1279-1293.e4. https://doi.org/10.1016/j.molcel.2020.01.019.

Davidson, Iain F., et Jan-Michael Peters. 2021. « Genome Folding through Loop Extrusion by SMC Complexes ». *Nature Reviews. Molecular Cell Biology* 22 (7): 445□64.

https://doi.org/10.1038/s41580-021-00349-7.

De Santis, Pasquale, Stefano Morosetti, et Anita Scipioni. 2010. « Prediction of Nucleosome Positioning in Genomes: Limits and Perspectives of Physical and Bioinformatic Approaches ». *Journal of Biomolecular Structure and Dynamics* 27 (6): 747‑64. https://doi.org/10.1080/07391102.2010.10508583.

Dekker, Job, Karsten Rippe, Martijn Dekker, et Nancy Kleckner. 2002. « Capturing Chromosome Conformation ». *Science (New York, N.Y.)* 295 (5558): 1306‑11. https://doi.org/10.1126/science.1067799.

Duan, Zhijun, Mirela Andronescu, Kevin Schutz, Sean McIlwain, Yoo Jung Kim, Choli Lee, Jay Shendure, Stanley Fields, C. Anthony Blau, et William S. Noble. 2010. « A Three-Dimensional Model of the Yeast Genome ». *Nature* 465 (7296): 363‑67. https://doi.org/10.1038/nature08973.

Eraslan, Gökcen, Žiga Avsec, Julien Gagneur, et Fabian J. Theis. 2019. « Deep Learning: New Computational Modelling Techniques for Genomics ». *Nature Reviews Genetics* 20 (7): 389‑403. https://doi.org/10.1038/s41576-019-0122-6.

Galanie, Stephanie, Kate Thodey, Isis J. Trenchard, Maria Filsinger Interrante, et Christina D. Smolke. 2015. « Complete biosynthesis of opioids in yeast ». *Science* 349 (6252): 1095‑1100. https://doi.org/10.1126/science.aac9373.

Garcia-Luis, Jonay, Luciana Lazar-Stefanita, Pilar Gutierrez-Escribano, Agnes Thierry, Axel Cournac, Alicia García, Sara González, et al. 2019. « FACT Mediates Cohesin Function on Chromatin ». *Nature Structural & Molecular Biology* 26 (10): 970‑79. https://doi.org/10.1038/s41594-019-0307-x.

Gibcus, Johan H., Kumiko Samejima, Anton Goloborodko, Itaru Samejima, Natalia Naumova, Johannes Nuebler, Masato T. Kanemaki, et al. 2018. « A pathway for mitotic chromosome formation ». *Science* 359 (6376): eaao6135. https://doi.org/10.1126/science.aao6135.

Glynn, Earl F., Paul C. Megee, Hong-Guo Yu, Cathy Mistrot, Elcin Unal, Douglas E. Koshland, Joseph L. DeRisi, et Jennifer L. Gerton. 2004. « Genome-Wide Mapping of the Cohesin Complex in the Yeast Saccharomyces Cerevisiae ». *PLoS Biology* 2 (9): E259. https://doi.org/10.1371/journal.pbio.0020259.

Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, et al. 1996. « Life with 6000 Genes ». *Science (New York, N.Y.)* 274 (5287): 546, 563‑67. https://doi.org/10.1126/science.274.5287.546.

Goloborodko, Anton, Maxim V. Imakaev, John F. Marko, et Leonid Mirny. 2016. « Compaction and Segregation of Sister Chromatids via Active Loop Extrusion ». *ELife* 5 (mai): e14864. https://doi.org/10.7554/eLife.14864.

Gotta, M., T. Laroche, et S. M. Gasser. 1999. « Analysis of Nuclear Organization in Saccharomyces Cerevisiae ». *Methods in Enzymology* 304: 663‑72. https://doi.org/10.1016/s0076-6879(99)04040-9.

Haarhuis, Judith H. I., Robin H. van der Weide, Vincent A. Blomen, J. Omar Yáñez-Cuna, Mario Amendola, Marjon S. van Ruiten, Peter H. L. Krijger, et al. 2017. « The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension ». *Cell* 169 (4): 693-707.e14. https://doi.org/10.1016/j.cell.2017.04.013.

Holmquist, Gerald P. 1989. « Evolution of Chromosome Bands: Molecular Ecology of Noncoding DNA ». *Journal of Molecular Evolution* 28 (6): 469‑86. https://doi.org/10.1007/BF02602928.

Hu, Bin, Naomi Petela, Alexander Kurze, Kok-Lung Chan, Christophe Chapard, et Kim Nasmyth. 2015. « Biological Chromodynamics: A General Method for Measuring Protein Occupancy across the Genome by Calibrating ChIP-Seq ». *Nucleic Acids Research* 43 (20): e132. https://doi.org/10.1093/nar/gkv670.

Jeppsson, Kristian, Toyonori Sakata, Ryuichiro Nakato, Stefina Milanova, Katsuhiko Shirahige, et Camilla Björkegren. 2022. « Cohesin-Dependent Chromosome Loop Extrusion Is Limited by Transcription

and Stalled Replication Forks ». *Science Advances* 8 (23): eabn7063. https://doi.org/10.1126/sciadv.abn7063.

Kelley, David R., Yakir A. Reshef, Maxwell Bileschi, David Belanger, Cory Y. McLean, et Jasper Snoek. 2018. « Sequential Regulatory Activity Prediction across Chromosomes with Convolutional Neural Networks ». *Genome Research* 28 (5): 739‑50. https://doi.org/10.1101/gr.227819.117.

Koren, Amnon, et Steven A. McCarroll. 2014. « Random Replication of the Inactive X Chromosome ». *Genome Research* 24 (1): 64‑69. https://doi.org/10.1101/gr.161828.113.

Koszul, Romain, Sandrine Caburet, Bernard Dujon, et Gilles Fischer. 2004. « Eucaryotic Genome Evolution through the Spontaneous Duplication of Large Chromosomal Segments ». *The EMBO Journal* 23 (1): 234‑43. https://doi.org/10.1038/sj.emboj.7600024.

Labroussaa, Fabien, Anne Lebaudy, Vincent Baby, Géraldine Gourgues, Dominick Matteau, Sanjay Vashee, Pascal Sirand-Pugnet, Sébastien Rodrigue, et Carole Lartigue. 2016. « Impact of donor–recipient phylogenetic distance on bacterial genome transplantation ». *Nucleic Acids Research* 44 (17): 8501‑11. https://doi.org/10.1093/nar/gkw688.

Labun, Kornel, Tessa G. Montague, Maximilian Krause, Yamila N. Torres Cleuren, Håkon Tjeldnes, et Eivind Valen. 2019. « CHOPCHOP v3: Expanding the CRISPR Web Toolbox beyond Genome Editing ». *Nucleic Acids Research* 47 (W1): W171‑74. https://doi.org/10.1093/nar/gkz365.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, et al. 2001. « Initial Sequencing and Analysis of the Human Genome ». *Nature* 409 (6822): 860‑921. https://doi.org/10.1038/35057062.

Langmead, Ben, et Steven L. Salzberg. 2012. « Fast Gapped-Read Alignment with Bowtie 2 ». *Nature Methods* 9 (4): 357‑59. https://doi.org/10.1038/nmeth.1923.

Lartigue, Carole, Sanjay Vashee, Mikkel A. Algire, Ray-Yuan Chuang, Gwynedd A. Benders, Li Ma, Vladimir N. Noskov, et al. 2009. « Creating Bacterial Strains from Genomes That Have Been Cloned and Engineered in Yeast ». *Science (New York, N.Y.)* 325 (5948): 1693‑96. https://doi.org/10.1126/science.1173759.

Laughery, Marian F., Tierra Hunter, Alexander Brown, James Hoopes, Travis Ostbye, Taven Shumaker, et John J. Wyrick. 2015. « New Vectors for Simple and Streamlined CRISPR-Cas9 Genome Editing in Saccharomyces cerevisiae ». *Yeast (Chichester, England)* 32 (12): 711‑20. https://doi.org/10.1002/yea.3098.

Lazar-Stefanita, Luciana, Jingchuan Luo, Remi Montagne, Agnes Thierry, Xiaoji Sun, Guillaume Mercy, Julien Mozziconacci, Romain Koszul, et Jef D. Boeke. 2022. « Karyotype Engineering Reveals Spatio-Temporal Control of Replication Firing and Gene Contacts ». *Cell Genomics*, juillet, 100163. https://doi.org/10.1016/j.xgen.2022.100163.

Lazar-Stefanita, Luciana, Vittore F. Scolari, Guillaume Mercy, Héloise Muller, Thomas M. Guérin, Agnès Thierry, Julien Mozziconacci, et Romain Koszul. 2017. « Cohesins and Condensins Orchestrate the 4D Dynamics of Yeast Chromosomes during the Cell Cycle ». *The EMBO Journal* 36 (18): 2684‑97. https://doi.org/10.15252/embj.201797342.

Lengronne, Armelle, Yuki Katou, Saori Mori, Shihori Yokobayashi, Gavin P. Kelly, Takehiko Itoh, Yoshinori Watanabe, Katsuhiko Shirahige, et Frank Uhlmann. 2004. « Cohesin Relocation from Sites of Chromosomal Loading to Places of Convergent Transcription ». *Nature* 430 (6999): 573‑78. https://doi.org/10.1038/nature02742.

Levo, Michal, Einat Zalckvar, Eilon Sharon, Ana Carolina Dantas Machado, Yael Kalma, Maya Lotam-Pompan, Adina Weinberger, Zohar Yakhini, Remo Rohs, et Eran Segal. 2015. « Unraveling determinants of transcription factor binding outside the core binding site ». *Genome Research* 25 (7): 1018‑29. https://doi.org/10.1101/gr.185033.114.

Lieberman-Aiden, Erez, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, et al. 2009. « Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome ». *Science (New York, N.Y.)* 326 (5950): 289‑93.

https://doi.org/10.1126/science.1181369.

Luo, Jingchuan, Xiaoji Sun, Brendan P. Cormack, et Jef D. Boeke. 2018. « Karyotype Engineering by Chromosome Fusion Leads to Reproductive Isolation in Yeast ». *Nature* 560 (7718): 392□96. https://doi.org/10.1038/s41586-018-0374-x.

Lustig, A J. 1992. « Hoogsteen G-G base pairing is dispensable for telomere healing in yeast. » *Nucleic Acids Research* 20 (12): 3021□28.

Matthey-Doret, Cyril, Lyam Baudry, Axel Breuer, Rémi Montagne, Nadège Guiglielmoni, Vittore Scolari, Etienne Jean, et al. 2020. « Computer Vision for Pattern Detection in Chromosome Contact Maps ». *Nature Communications* 11 (1): 5795. https://doi.org/10.1038/s41467-020-19562-7.

Meng, Fankang, et Tom Ellis. 2020. « The Second Decade of Synthetic Biology: 2010-2020 ». *Nature Communications* 11 (1): 5174. https://doi.org/10.1038/s41467-020-19092-2.

Miele, Vincent, Cédric Vaillant, Yves d'Aubenton-Carafa, Claude Thermes, et Thierry Grange. 2008. « DNA physical properties determine nucleosome occupancy from yeast to fly ». *Nucleic Acids Research* 36 (11): 3746□56. https://doi.org/10.1093/nar/gkn262.

Mirny, Leonid, et Job Dekker. 2022. « Mechanisms of Chromosome Folding and Nuclear Organization: Their Interplay and Open Questions ». *Cold Spring Harbor Perspectives in Biology* 14 (7): a040147. https://doi.org/10.1101/cshperspect.a040147.

Müller, Carolin A., Michelle Hawkins, Renata Retkute, Sunir Malla, Ray Wilson, Martin J. Blythe, Ryuichiro Nakato, et al. 2014. « The dynamics of genome replication using deep sequencing ». *Nucleic Acids Research* 42 (1): e3. https://doi.org/10.1093/nar/gkt878.

Nasmyth, K. 2001. « Disseminating the Genome: Joining, Resolving, and Separating Sister Chromatids during Mitosis and Meiosis ». *Annual Review of Genetics* 35: 673□745. https://doi.org/10.1146/annurev.genet.35.102401.091334.

Naumova, Natalia, Maxim Imakaev, Geoffrey Fudenberg, Ye Zhan, Bryan R. Lajoie, Leonid A. Mirny, et Job Dekker. 2013. « Organization of the Mitotic Chromosome ». *Science (New York, N.Y.)* 342 (6161): 948□53. https://doi.org/10.1126/science.1236083.

Nora, Elphège P., Anton Goloborodko, Anne-Laure Valton, Johan H. Gibcus, Alec Uebersohn, Nezar Abdennur, Job Dekker, Leonid A. Mirny, et Benoit G. Bruneau. 2017. « Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization ». *Cell* 169 (5): 930-944.e22. https://doi.org/10.1016/j.cell.2017.05.004.

Ocampo-Hafalla, Maria, Sofía Muñoz, Catarina P. Samora, et Frank Uhlmann. 2016. « Evidence for Cohesin Sliding along Budding Yeast Chromosomes ». *Open Biology* 6 (6): 150178. https://doi.org/10.1098/rsob.150178.

Payen, Célia, Gilles Fischer, Christian Marck, Caroline Proux, David James Sherman, Jean-Yves Coppée, Mark Johnston, Bernard Dujon, et Cécile Neuvéglise. 2009. « Unusual composition of a yeast chromosome arm is associated with its delayed replication ». *Genome Research* 19 (10): 1710□21. https://doi.org/10.1101/gr.090605.108.

Pinglay, Sudarshan, Milica Bulajić, Dylan P. Rahe, Emily Huang, Ran Brosh, Nicholas E. Mamrak, Benjamin R. King, et al. 2022. « Synthetic Regulatory Reconstitution Reveals Principles of Mammalian Hox Cluster Regulation ». *Science (New York, N.Y.)* 377 (6601): eabk2820. https://doi.org/10.1126/science.abk2820.

Rando, Oliver J., et Fred Winston. 2012. « Chromatin and Transcription in Yeast ». *Genetics* 190 (2): 351□87. https://doi.org/10.1534/genetics.111.132266.

Rao, Suhas S. P., Su-Chen Huang, Brian Glenn St Hilaire, Jesse M. Engreitz, Elizabeth M. Perez, Kyong-Rim Kieffer-Kwon, Adrian L. Sanborn, et al. 2017. « Cohesin Loss Eliminates All Loop Domains ». *Cell* 171 (2): 305-320.e24. https://doi.org/10.1016/j.cell.2017.09.026.

Rao, Suhas S. P., Miriam H. Huntley, Neva C. Durand, Elena K. Stamenova, Ivan D. Bochkov, James T. Robinson, Adrian L. Sanborn, et al. 2014. « A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping ». *Cell* 159 (7): 1665□80.

https://doi.org/10.1016/j.cell.2014.11.021.

Routhier, Etienne, et Julien Mozziconacci. 2022. « Genomics Enters the Deep Learning Era ». *PeerJ* 10: e13613. https://doi.org/10.7717/peerj.13613.

Routhier, Etienne, Edgard Pierre, Ghazaleh Khodabandelou, et Julien Mozziconacci. 2021. « Genome-Wide Prediction of DNA Mutation Effect on Nucleosome Positions for Yeast Synthetic Genomics ». *Genome Research* 31 (2): 317‑26. https://doi.org/10.1101/gr.264416.120.

Ruiten, Marjon S. van, Démi van Gent, Ángela Sedeño Cacciatore, Astrid Fauster, Laureen Willems, Maarten L. Hekkelman, Liesbeth Hoekman, et al. 2022. « The Cohesin Acetylation Cycle Controls Chromatin Loop Length through a PDS5A Brake Mechanism ». *Nature Structural & Molecular Biology* 29 (6): 586‑91. https://doi.org/10.1038/s41594-022-00773-z.

Ruiz, Estelle, Vincent Talenton, Marie-Pierre Dubrana, Gabrielle Guesdon, Maria Lluch-Senar, Franck Salin, Pascal Sirand-Pugnet, Yonathan Arfi, et Carole Lartigue. 2019. « CReasPy-Cloning: A Method for Simultaneous Cloning and Engineering of Megabase-Sized Genomes in Yeast Using the CRISPR-Cas9 System ». *ACS Synthetic Biology* 8 (11): 2547‑57. https://doi.org/10.1021/acssynbio.9b00224.

Sahu, Biswajyoti, Tuomo Hartonen, Päivi Pihlajamaa, Bei Wei, Kashyap Dave, Fangjie Zhu, Eevi Kaasinen, et al. 2022. « Sequence Determinants of Human Gene Regulatory Elements ». *Nature Genetics* 54 (3): 283‑94. https://doi.org/10.1038/s41588-021-01009-4.

Schalbetter, Stephanie Andrea, Anton Goloborodko, Geoffrey Fudenberg, Jon-Matthew Belton, Catrina Miles, Miao Yu, Job Dekker, Leonid Mirny, et Jonathan Baxter. 2017. « SMC Complexes Differentially Compact Mitotic Chromosomes According to Genomic Context ». *Nature Cell Biology* 19 (9): 1071‑80. https://doi.org/10.1038/ncb3594.

Schulz, Daniel, Bjoern Schwalb, Anja Kiesel, Carlo Baejen, Phillipp Torkler, Julien Gagneur, Johannes Soeding, et Patrick Cramer. 2013. « Transcriptome Surveillance by Selective Termination of Noncoding RNA Synthesis ». *Cell* 155 (5): 1075‑87. https://doi.org/10.1016/j.cell.2013.10.024.

Shao, Yangyang, Ning Lu, Zhenfang Wu, Chen Cai, Shanshan Wang, Ling-Li Zhang, Fan Zhou, et al. 2018. « Creating a Functional Single-Chromosome Yeast ». *Nature* 560 (7718): 331‑35. https://doi.org/10.1038/s41586-018-0382-x.

Spracklin, George, Nezar Abdennur, Maxim Imakaev, Neil Chowdhury, Sriharsa Pradhan, Leonid Mirny, et Job Dekker. 2021. « Heterochromatin diversity modulates genome compartmentalization and loop extrusion barriers ». *bioRxiv*, janvier, 2021.08.05.455340. https://doi.org/10.1101/2021.08.05.455340.

Struhl, Kevin, et Eran Segal. 2013. « Determinants of Nucleosome Positioning ». *Nature Structural & Molecular Biology* 20 (3): 267‑73. https://doi.org/10.1038/nsmb.2506.

Tirosh, Itay, et Naama Barkai. 2008. « Two Strategies for Gene Regulation by Promoter Nucleosomes ». *Genome Research* 18 (7): 1084‑91. https://doi.org/10.1101/gr.076059.108.

Trolle, Julie, Ross M. McBee, Andrew Kaufman, Sudarshan Pinglay, Henri Berger, Sergei German, Liyuan Liu, et al. 2022. « Resurrecting Essential Amino Acid Biosynthesis in Mammalian Cells ». *ELife* 11 (septembre): e72847. https://doi.org/10.7554/eLife.72847.

Vaishnav, Eeshit Dhaval, Carl G. de Boer, Jennifer Molinet, Moran Yassour, Lin Fan, Xian Adiconis, Dawn A. Thompson, Joshua Z. Levin, Francisco A. Cubillos, et Aviv Regev. 2022. « The Evolution, Evolvability and Engineering of Gene Regulatory DNA ». *Nature* 603 (7901): 455‑63. https://doi.org/10.1038/s41586-022-04506-6.

Valouev, Anton, Steven M. Johnson, Scott D. Boyd, Cheryl L. Smith, Andrew Z. Fire, et Arend Sidow. 2011. « Determinants of Nucleosome Organization in Primary Human Cells ». *Nature* 474 (7352): 516‑20. https://doi.org/10.1038/nature10002.

Viterbo, David, Astrid Marchal, Valentine Mosbach, Lucie Poggi, Wilhelm Vaysse-Zinkhöfer, et Guy-Franck Richard. 2018. « A Fast, Sensitive and Cost-Effective Method for Nucleic Acid Detection Using Non-Radioactive Probes ». *Biology Methods & Protocols* 3 (1): bpy006.

https://doi.org/10.1093/biomethods/bpy006.

Widom, J. 2001. « Role of DNA Sequence in Nucleosome Stability and Dynamics ». *Quarterly Reviews of Biophysics* 34 (3): 269□324. https://doi.org/10.1017/S0033583501003699.

Wong, Hua, Hervé Marie-Nelly, Sébastien Herbert, Pascal Carrivain, Hervé Blanc, Romain Koszul, Emmanuelle Fabre, et Christophe Zimmer. 2012. « A Predictive Computational Model of the Dynamic 3D Interphase Yeast Nucleus ». *Current Biology: CB* 22 (20): 1881□90. https://doi.org/10.1016/j.cub.2012.07.069.

Wutz, Gordana, Rene Ladurner, Brian Glenn St Hilaire, Roman R. Stocsits, Kota Nagasaka, Benoit Pignard, Adrian Sanborn, et al. 2020. « ESCO1 and CTCF Enable Formation of Long Chromatin Loops by Protecting CohesinSTAG1 from WAPL ». *ELife* 9 (février): e52091. https://doi.org/10.7554/eLife.52091.

Wutz, Gordana, Csilla Várnai, Kota Nagasaka, David A. Cisneros, Roman R. Stocsits, Wen Tang, Stefan Schoenfelder, et al. 2017. « Topologically Associating Domains and Chromatin Loops Depend on Cohesin and Are Regulated by CTCF, WAPL, and PDS5 Proteins ». *The EMBO Journal* 36 (24): 3573□99. https://doi.org/10.15252/embj.201798004.

Table S1 | Summary table of strains

	Strains name	Description	Parental strain	Genetic background	MAT	Genotype	Reference
Mmyco	RSG_Y711	Strain carrying the circular chromosome Mycoplasma mycoides subsp. mycoides strain PG1	W303	W303	a	ade2-1, leu2-3,112, his3-11,15, trp1-1, can1-100, RAD5 + M. mycoides circular genome with CEN6/ARS-HIS3	PMID: 27488189 DOI: 10.1093/nar/gkw688
	RSG_Y712	Mmyco chromosome linearized	RSG_Y711	W303	a	ade2-1, leu2-3,112, his3-11,15, trp1-1, can1-100, RAD5 + M. mycoides linear genome with CEN6/ARS-HIS3	this study
	RSG_Y951	Mmyco chromosome linearized with SCC1-PK9	RSG_Y712	W303	a	ade2-1, leu2-3,112, his3-11,15, trp1-1, can1-100, RAD5 + M. mycoides linear genome with CEN6/ARS-HIS3, SCC1-pk9::KanMX	this study
	RSG_Y955	Linearized Mmyco with delta WAPL	RSG_Y712	W303	a	ade2-1, leu2-3,112, his3-11,15, trp1-1, can1-100, RAD5 + M. mycoides linear genome with CEN6/ARS-HIS3, Rad61::KanMX	this study
	RSG_Y1033	XVIfMmyco : "Fused "strain, Mmyco chromosome merged with the XVI	RSG_Y712	W303	a	ade2-1, leu2-3,112, his3-11,15, trp1-1, can1-100, RAD5, XVIfMmyco(CEN6/ARS-HIS3)	this study
	RSG_Y1053	XVIfMmyco : "Fused" strain with SCC1-PK9	RSG_Y1033	W303	a	ade2-1, leu2-3,112, his3-11,15, trp1-1, can1-100, RAD5 + XVI-M. mycoides linear genome with CEN6/ARS-HIS3 SCC1-pk9::KanMX	this study
	RSG_Y1136	XVIfMmycot1 : "Translocation n°1" strain	RSG_Y1033	W303	a	ade2-1, leu2-3,112, his3-11,15, trp1-1, can1-100, SCC1-pk9::KanMX, RAD5, XVIfMmyco(CEN6/ARS-HIS3)	this study
	RSG_Y1146	"Translocation n°1" strain with SCC1-PK9	RSG_Y1136	W303	a	ade2-1, leu2-3,112, his3-11,15, trp1-1, can1-100, RAD5, XVIfMmycot1(CEN6/ARS-HIS3)	this study
	RSG_Y1137	XVIfMmycot2 : "Translocation n°2" strain (50 kb block)	RSG_Y1136	W303	a	ade2-1, leu2-3,112, his3-11,15, trp1-1, can1-100, RAD5, XVIfMmycot2(CEN6/ARS-HIS3)	this study
	RSG_Y1147	"Translocation n°2" strain (50 kb block) with SCC1-PK9	RSG_Y1137	W303	a	ade2-1, leu2-3,112, his3-11,15, trp1-1, can1-100, RAD5, SCC1-pk9::KanMX, XVIfMmycot2(CEN6/ARS-HIS3)	this study
	yLD126-36c	WT control for split dot assay (Michaelis et al., 1997)	NA	W303	a	Met3p-CDC20, tetR-GFP, ura3-1::URA3tetO	Michaelis et al., 1997
	FB176	Strain with fluorescent spot on linear Mmyco	RSG_Y712	W303	a	ade2-1, leu2-3,112::LEU2::tetR-GFP, his3-11,15, trp1-1, can1-100, ura3-1, Met3p-CDC20, M. mycoides linear genome with CEN6/ARS-HIS3 carrying 40000-42000::URA3::NATtetO (40Kb from centro)	this study
	FB200	Strain with linear Mmyco, and fluorescent spot on endogenous yeast chromosome	RSG_Y712	W303	a	ade2-1, leu2-3,112::LEU2::tetR-GFP, his3-11,15, trp1-1, can1-100, ura3-1::URA3tetO, Met3p-CDC20, M. mycoides linear genome with CEN6/ARS-HIS3	this study
Mpneumo	RSG_Y681	Strain carrying the *Mycoplasma pneumoniae strain M129*	VL6-48N	VL6-48N	alpha	trp1-Δ1, ura3-Δ1, ade2-101, his3-Δ200, lys2, met14 cir° + Mp circular genome with CEN6/ARS-HIS3	PMID: 31663334 DOI: 10.1021/acssynbio.9b00224
	RSG_Y960	Pneumo chromosome linearized	RSG_Y681	VL6-48N	a	trp1-Δ1, ura3-Δ1, ade2-101, his3-Δ200, lys2, met14 cir° + Mp linear genome with CEN6/ARS-HIS3	this study
	RSG_Y1080	Pneumo chromosome linearized with SCC1-PK9	RSG_Y960	VL6-48N	a	trp1-Δ1, ura3-Δ1, ade2-101, his3-Δ200, lys2, met14 cir° + Mp linear genome with CEN6/ARS-HIS3, SCC1-pk9::KanMX	this study
	RSG_Y954	Pneumo chromosome linearized with delta WAPL	RSG_Y960	VL6-48N	a	trp1-Δ1, ura3-Δ1, ade2-101, his3-Δ200, lys2, met14 cir° + Mp linear genome with CEN6/ARS-HIS3, Rad61::KanMX	this study
"Control" strains	RSG_Y784	W303 strain + pRS313	W303	W303	a	his3-11, 15 trp11 leu2-3,112 ura3-1 ade2-1 can1-100 [HIS3]	this study
	RSG_Y1158	VL6-48N strain + pRS413	VL6-48N	VL6-48N	a	trp1-Δ1 ura3-Δ1 ade2-101 his3-Δ200 lys2 met14 cir [HIS3]	this study
	RSG_Y973	Glabrata strain with SCC1-PK9	*C.glabrata*	*C.glabrata*	a	Scc1PK9::NatMX	Hu et al, 2015

	W303 strain + scc1-pk	W303 strain + scc1-pk	W303	W303	a	Scc1-pk9::KanMX	Hu et al, 2015