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

Léa Meneu, Christophe Chapard, Jacques Serizay, Alex Westbrook, Etienne Routhier, Myriam Ruault, Manon Perrot, Alexandros Minakakis, Fabien Girard, Amaury Bignaud, et al.

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

Authors: Léa Meneu[1,2,Ψ], Christophe Chapard[1,Ψ,&], Jacques Serizay[1,Ψ,*], Alex Westbrook[2,3], Etienne Routhier[2,3,4], Myriam Ruault[5], Manon Perrot[1,2], Alexandros Minakakis[8], Fabien Girard[1], Amaury Bignaud[1,2], Antoine Even[5], Géraldine Gourgues[6], Domenico Libri[8], Carole Lartigue[6], Aurèle Piazza[1,#], Agnès Thierry[1], Angela Taddei[5], Frédéric Beckouët[7], Julien Mozziconacci[3,4,9*] and Romain Koszul[1,*]

Affiliations:

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

[2] Sorbonne Université, College Doctoral

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

[4] Laboratoire de Physique Théorique de la Matière Condensée, Sorbonne Université, CNRS, 75005 Paris, France

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

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

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

[8] Institut de Génétique Moléculaire de Montpellier (IGMM), 34090 Montpellier, France

[9] UAR 2700 2AD, Muséum National d'Histoire Naturelle, Paris 75005, France

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

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

Ψ These authors contributed equally

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

Abstract

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

Introduction

Genome sequence composition, broadly defined by its GC%, polynucleotide frequencies, DNA motifs and repeats, varies widely between species as well as within individual genomes (*1*, *2*). In eukaryotes, the sequence composition is known to correlate with: i) chromatin composition, which includes nucleosome formation and binding of structural and functional proteins to DNA (*3*), ii) chromatin activity, such as transcription and replication (*4*), and iii) functional 3D organization of the genome into loops and compartments (*5*, *6*). For instance in mammals, GC-rich regions are enriched in actively transcribed sequences, in chromatin loops mediated by the structural maintenance of chromosomes (SMC) cohesin, and coalesce into a specific compartment (*7*). These relationships between sequence and chromatin composition activities and folding reflect their continuous coevolution over millions of years.

Disruptive variations in sequence composition can emerge naturally during evolution, e.g. through transfer of genetic material across species by horizontal gene transfers or introgression, in viral infections (*8–11*), or even artificially, e.g. by introducing chromosome-long DNA molecules into chassis microbial strains or cell lines (*12–14*). Such transfer can lead to the long-term integration of foreign DNA whose sequence composition strongly diverges from their host's genome (e.g. the introgression in *Lachancea kluyveri* of a 1 Mb sequence with a GC content 12% higher that the rest of the genome (*15*)). Once integrated, these sequences are organized and processed by chromatin-associated proteins of the host genome, obeying new rules under which they have not coevolved. How a eukaryotic host packages, regulates the activity and folds long exogenous DNA sequences, and the importance of the sequence composition in this process, remain largely unknown.

Here, we investigate the behavior of natural, chromosome-sized bacterial and eukaryotic sequences, with different sequence composition, artificially introduced into the *S. cerevisiae* genome (*16*, *17*). We profile nucleosome, RNA polymerase and cohesin landscapes, transcriptional activity and 3D organization of these supernumerary chromosomes during the cell cycle. We show that highly divergent bacterial or eukaryotic chromosomes, with different GC content, present different chromatin composition and activities. This eventually leads to the spontaneous formation of two chromatin archetypes, one active and one inactive, each displaying different physical properties and segregating into distinct chromosomal compartments, similar to those observed in complex multicellular organisms such as mammals. This partitioning occurs

independently of heterochromatin formation but is driven by transcriptional activity. Sequence determinants, computationally learnt on yeast sequences are sufficient to predict the chromatin composition and activity of exogenous chromosomes integrated in the yeast, suggesting s that the fate of any DNA molecule introduced into a given cellular context, from nucleosome positioning up to its 3D folding and transcriptional activity, is governed by rules that are both deterministic and predictable.

Adaptation of supernumerary bacteria chromosomes integrated into yeast

To investigate 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* subspecies *mycoides* (referred to as "Mmyco") or *Mycoplasma pneumoniae* ("Mpneumo") genomes containing a yeast centromeric sequence and an autonomous replication sequence (ARS) (*18*) (**Methods**; **Table S1**). While the GC content of *S. cerevisiae* is 38%, the GC content of the Mmyco and Mpneumo chromosomes are 24% (GC-poor) and 40% (GC-neutral), respectively (**Fig. 1A**). Dinucleotide composition (*19*) also largely differs between yeast and bacterial chromosomes (**Fig. S1A**). We linearized these chromosomes, added yeast telomeres at the extremities (**Fig. 1A; Fig. S1B**; **Methods**), and investigated their replication and cohesion using split-dot assay and marker frequency analysis (**Fig. S1C,D** and **Supplementary Text**; **Methods**). Overall, these chromosomes do not impose a significant fitness cost on their eukaryotic host and have a segregation rate similar to that of a centromeric plasmid **(Fig. S1E**, **S1F**) making them powerful tools for studying chromatin assembly on foreign sequences.

Spontaneous chromatinization of bacterial chromosomes in a eukaryotic context

While Mmyco and Mpneumo chromosomes evolved in a non-eukaryotic context, both sequences lead to the formation of nucleosome arrays and local nucleosome-depleted regions (NDR) as assessed by time-course MNase-seq and H3 and H2A ChIP-seq experiments (**Fig. 1B, C; Fig. S2A-B**). We annotated nucleosome positions following an approach adapted from Brogaard et al. (*20*), and found that nucleosome arrays over the Mpneumo chromosome are similar to yeast's ones, with a linker length of ~14-18 bp and a nucleosome repeat length (NRL) of 160 bp (**Fig. 1D,E)**. In contrast, nucleosome arrays in the Mmyco chromosome have a longer ~ 25 bp linker and an NRL of 174 bp (**Fig. 1D,E, S2D)**. Sequence features favoring nucleosome

organization (short poly(dA) or poly(dT) tracks and strong WW 10 bp periodicity) are enriched in Mpneumo, while sequence features limiting nucleosome assembly (long poly(dA) or poly(dT) tracks and weak WW 10 bp periodicity) are enriched in Mmyco chromosome (**Fig. S2C,D**).

Chromatin accessibility profiling by ATAC-seq confirmed that the Mpneumo chromosome has a similar density and breadth of accessible loci to endogenous yeast chromosomes and that these accessible sites are devoid of nucleosomes, effectively forming nucleosome-depleted regions (NDRs) (**Fig. 1B, Fig. S3A,B, S3D-E**). In comparison, we only detected 16 weaker ATAC peaks over the Mmyco chromosome, which were weakly covered but not completely depleted in nucleosomes (**Fig. 1C, Fig. S3C-E**).

The cohesin (Scc1) and RNA polymerase II (Pol. II) profiles along the Mpneumo chromosome also appear similar to wild-type (WT) yeast chromosomes, with discrete Scc1 peaks preferentially located in nucleosome-depleted and Pol. II-enriched regions (**Fig. 1F, Fig. S2F-H**). In contrast, Scc1 and Pol. II binding profiles in the Mmyco strain show significant differences. On the one hand, Scc1 is strongly enriched over the whole Mmyco chromosome, but does not form the distinct peaks observed on endogenous chromosomes (**Fig. 1F, Fig. S2F,G**). This broad enrichment is correlated with a strongly reduced Scc1 occupancy at centromeres of endogenous *S. cerevisiae* chromosomes (**Fig. S2E**), suggesting that the Mmyco chromosome titrates the dynamic cohesin pool enriched over yeast centromeres (*21*, *22*). The Pol. II occupancy is on the other hand greatly reduced along the Mmyco chromosome compared to yeast chromosomes (**Fig. 1F**).

These results show that large exogenous bacterial chromosomes placed in a eukaryotic context spontaneously adopt features of eukaryotic chromatin, as histones, Pol II and cohesins can all bind bacterial DNA. However, the chromatin landscapes over the two bacterial chromosomes delineate two different chromatin archetypes: (1) “Y” (yeast-like) chromatin landscape, found over Mpneumo (whose 40% GC content is close to the native *S. cerevisiae* GC content), and (2) “U” (for Unconventional) chromatin, found over Mmyco (with a low 24% GC content), featuring less packed nucleosomes, a reduced Pol II coverage, and a broad binding of cohesins across the entire chromosome.

Transcriptional activity of bacterial genomes in a yeast context

Consistent with Pol. II ChIP-seq profiles (**Fig. 1F**), we find that the Y chromatin type on Mpneumo is transcribed to levels similar to those of endogenous yeast chromosomes (**Fig. 2A**).

Mpneumo transcription tracks are significantly longer than yeast genes (4.9 kb vs 3.4 kb, p-value $< 2e^{-4}$, two-sided Student's t-test) and do not systematically display clear boundaries, with more loci where both strands are transcribed compared to yeast (**Fig. 2B**, black triangle). They do not preferentially occur over bacterial gene bodies nor seem to initiate at bacterial promoters, consistent with an absence of Pol. II enrichment at these loci (**Fig. S4A-C**). In sharp contrast, the Mmyco U chromatin type is only sparsely and lowly transcribed (**Fig. 2A,B**), again in good agreement with the reduced levels of Pol. II deposition (**Fig. 1F**).

Remarkably, for both bacterial genomes, the steady-state orientation of transcription tracks preferentially follow those of the bacterial genes annotated along these sequences, despite the fact that the transcription machinery evolved independently for billions of years (**Fig. 2C**, see the orientation of genes and stranded tracks in **Fig. 2A,B**). This preference correlates with the over-representation of A vs. T and G vs. C on the coding strand of both yeast and bacterial genes (**Fig. S4D**, see **Discussion**). To test whether the steady-state unidirectional orientation of transcription is the outcome of active degradation of overlapping antisense transcripts, we profiled nascent transcripts by CRAC-seq, a technique that quantifies RNA molecules still bound to RNA Pol II (**Fig. S5A**). We found a strong correlation between RNA-seq and CRAC-seq (**Fig. S5B**) which confirmed that transcription can initiate at the hundreds of accessible NDRs along the Mpneumo chromosome (**Fig. 2D, E**) and that transcription is extremely sparse over the Mmyco chromosome (**Fig. S5A**). We further validated these observations by performing RNA-seq in *Δupf1*, *Δrrp6* or *Δupf1/rrp6* Mpneumo strains, respectively incompetent for nonsense-mediated RNA decay (NMD), nuclear RNA surveillance by the exosome, or both. This showed that although some overlapping antisense transcripts sporadically emerge in the mutant strains, eukaryotic transcription over a bacterial chromosome preferentially follows the direction of the bacterial genes (**Fig. S5C**).

Inactive U chromatin forms a distinct compartment in the interphase nucleus

Chromatin conformation capture (Hi-C) contact maps show that, as expected in a Rabl configuration (*23*, *24*), yeast centromeres added to both Mmyco and Mpneumo chromosomes cluster with native ones. The two bacterial chromosomes exhibit nonetheless very distinct structural characteristics. The Y-type Mpneumo chromosome behaves similarly to an endogenous *S. cerevisiae* chromosome (**Fig. 3A**), with comparable trans contacts (**Fig. S6A**) and a slope of the

contact probability (p) as a function of the genomic distance (*s)* (or genomic contact decay curve) close to -1.5, a value corresponding to a typical random coil structure observed in simulations (**Fig. 3C**) (*25*). DNA-FISH labeling of Mpneumo DNA also revealed an extended and contorted structure within the intranuclear space, confirming the intermixing of the chromosome with the actively transcribed native chromosomes (**Fig. 3E, Fig. S6C, D**).

In contrast, few contacts were detected between the U-type Mmyco and the endogenous chromosomes (**Fig. 3A, Fig. S6A**), predominantly with the 32 yeast subtelomeric regions (**Fig. 3A**, dotted rectangles; **Fig. S6B**). The slope of the p(s) curve also differs dramatically with a value of -1 at shorter distances, corresponding to a crumpled globule (**Fig. 3C**)(*26*, *27*). DNA-FISH labeling of Mmyco confirmed the globular conformation of the chromosome and revealed its preferential position at the nuclear periphery, reflecting its proximity with sub/telomeres (**Fig. 3E; Fig. S6C, D**). These results show that inactive, U chromatin spontaneously forms a distinct compartment in the nuclear space segregated from the active endogenous chromosomal set.

Cohesin compacts both exogenous bacterial chromosomes in G2/M

In yeast G2/M-arrested cells, chromatin is folded into arrays of cohesin-anchored loops (*28*, *29*). This compaction results in a decrease of inter-chromosomal contacts and in a shifted p(s) curve, with an inflection point around the average loop length (*23*, *30*). These structural features are observed for both bacterial chromosomes, indicating a similar mitotic compaction of both Y and U chromatin types (**Fig. 3B, 3D, Fig. S6A,E,F**). We identified 59 loops across the Mpneumo chromosome in G2/M, spanning over slightly longer (28 kb) genomic distances than the loops found along yeast sequences (22 kb) (**Fig. S6G**; **Methods**). Strong Scc1 peaks were positioned close (within 500bp) to these loop anchors (**Fig. 3F,H**), at sites of convergent transcription (**Fig. 2B**, green dotted lines, **Fig. S6H**), with strong cohesin peaks associated with sharper transcriptional convergence (**Fig. 2B**, arrows; **Fig. S6I,J**). The dotted grid pattern in Mpneumo is reminiscent of multiple DNA loops observed along mammalian interphase chromosomes (*31*), suggesting that some loop anchors can be involved in loops of various sizes, a phenomenon not observed across native *S. cerevisiae* chromosomes (**Fig. 3F**). On the other hand, no visible discrete loops were observed along the Mmyco chromosome where cohesins uniformly cover inactive U chromatin regions without clear peaks, probably reflecting the absence of (convergent) transcription (**Fig. 3G**, and also **Fig. 1F** and **Fig. 2B**). Cohesin-mediated loops along yeast chromosomes are formed

through loop extrusion, a process by which cohesin complexes organize DNA by capturing and gradually enlarging small loops (*32*, *33*). In absence of Wpl1, which impairs cohesin removal and results in longer loops (*28*, *33*), we observe an increase of long-range contacts for all chromosomes, including Mmyco and Mpneumo (**Fig. 3I; Fig. S6K**), consistent with active cohesin-mediated loop extrusion proceeding on both Y and U chromatin.

Mosaic 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 compartments described along metazoan chromosomes (*34*). To test whether and how they can coexist on a single chromosome, we fused the Mmyco chromosome with the native yeast chromosome XVI (XVIfMmyco chromosome) and induced translocations to generate two additional strains with chromosomes harboring alternating U and Y chromatin regions over hundreds of kb (XVIfMmycot1), 50 kb (XVIfMmycot2) or as small as 15 kb (XVIfMmycot2') (**Fig. 4A**; **Fig. S7A-B** Methods). The overall Hi-C contact maps of XVIfMmyco strain synchronized in G1 showed little differences compared to the parental Mmyco strain (but for the deleted and fused regions, **Fig. 4B**, compare panel **i** with **ii**). The Y-type chr. XVI arm intermixes with the other 15 yeast chromosomes while the U-type Mmyco arm remains isolated from the rest of the genome, contacting subtelomeric regions (**Fig. 4B,** panel **ii**). In the translocated strains, the alternating chromatin type regions resulted in striking checkerboard contact patterns within XVIfMmycot1 and XVIfMmycot2 chromosomes (**Fig. 4B**, panels **iii** and **iv**). U-type regions can thus make specific contacts over long distances, bypassing the Y-type regions found in between (**Fig. 4C, D**; **Fig. S7E,** G1). Similarly, Y-type regions of chromosome XVI are also involved in specific Y-Y contacts over longer distances (**Fig. 4C, D**). Intra-U and Y regions contact decay is different in the two chromatin types, U chromatin being prone to longer range contacts in *cis* (**Fig. S7E**). Strikingly, translocation of a U-chromatin segment at the end of chromosome XIII also led to a strong increase of inter-chromosomal contacts with U-chromatin segments in *trans* (**Fig. 4E, F**), illustrating that inactive U chromatin can span multiple chromosomes, a finding that broadens our understanding of chromatin compartmentalization and that further relates Y and U-chromatin with eu- and heterochromatin compartments found in higher eukaryotes (*34*).

Pol. II ChIP-seq in XVIfMmycot2 strain revealed that 50 kb-long translocated Y or U chromatin segments did not exhibit any change in Pol. II occupancy profiles, demonstrating that Pol. II binding is independent of the broader chromatin context (**Fig. S7C**). We further performed RNA-seq profiling in the XVIfMycot2' strain, which harbors a 15 kb segment of Mmyco translocated within the euchromatic yeast chromosome XVI and vice versa. Compared to the XVIfMmycot2 strain, the right end of this 15kb Mmyco segment is directly adjacent to yeast chromatin (**Fig. S7D**). In this context, we observed that transcription immediately at the junction with yeast chromatin can progress over ~1 kb into the Mmyco segment (**Fig. S7D**, black arrow). In contrast, transcription of yeast genes translocated close to Mmyco chromatin is unaffected. This suggests that although active transcription can spread from Y chromatin to Mmyco U chromatin, the U-type chromatin locally exerts a strong inhibitory effect on the transcription initiation machinery.

In larger genomes, euchromatin and heterochromatin compartments are abolished following loop extrusion-mediated metaphase chromosome compaction (*32*, *35*, *36*). Similarly, in our chimeric strains, Hi-C maps of the mosaic chromosomes in G2/M reveal that intra- and inter-chromosomal U-type compartments all disappear upon cohesin-mediated compaction (**Fig. 4E, Fig. S7F**), with cohesin being over-enriched along inactive Mmyco regions independently of the proximity with the centromere (**Fig. S7C**). At this stage, all the chimeric chromosome regions have the same distance-dependent contact frequency (**Fig. S7E**, G2/M), and loops can span over the Y/U chromatin junction, bridging the nearest cohesin enrichment sites on each side (**Fig. S7G**, green arrows) (*37*, *38*).

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

We investigated the molecular mechanisms that might be responsible for the formation of this U-type heterochromatin-like chromatin compartment. In *S. cerevisiae*, heterochromatin formation does not rely on the canonical H3K9me3 modification found in most eukaryotes (*39*). Instead, heterochromatin is formed and maintained at telomeres, mating-type and rDNA loci by the SIR (Silent Information Regulator) complex, which consists of Sir2 histone deacetylase (HDAC class III), Sir3, a structural chromatin-binding protein stabilizing deacetylated histones and Sir4, another structural protein bridging Sir2 and Sir3 and interacting with nuclear envelope associated proteins

(doi: 10.1534/genetics.112.145243). We found that 1) H4K16ac levels are reduced overall on Mmyco chromosome (**Fig. S8A**), 2) that Sir2 inhibitor nicotinamide (NAM) only increases H4 acetylation levels locally at telomeric regions of both yeast and Mmyco chromosomes (**Fig. S8B**) and 3) that Sir3 is not enriched along the Mmyco chromosome (**Fig. S8C**). These observations reveal that U chromatin is hypo-acetylated independently of SIR-driven telomere heterochromatinization. We further investigated whether widespread deacetylation by other histone deacetylases could lead to heterochromatinization of Mmyco. Following treatment with Trichostatin A (TSA, an HDAC I/II inhibitor), we found that H4 acetylation levels increase globally over yeast segments but not over Mmyco segments (**Fig. S8D**) and that the Mmyco U compartment and the p(s) contact decay curves were not affected (**Fig. S8F-H**). This indicates that neither hypo-acetylation of U-chromatin nor its compartmentalization rely on histone deacetylation activity.

The unusual 10bp-longer NRL measured in the Mmyco chromosome (**Fig. 1D,E**) suggests that Histone 1 (H1) could bind the DNA linkers between consecutive nucleosomes leading to the formation of U-type chromatin. We tested this hypothesis in Hho1 (yeast H1) deleted mutant and found that the longer NRL, the distance-dependent contact frequency and the Mmyco compartment remained unchanged (**Fig. S8E, S8I-K**), showing that H1 is not required for U-type chromatin formation.

Finally, we assessed whether active transcription of yeast chromatin could be responsible for the segregation of the inactive Mmyco chromosome into a separate compartment. We treated cells with thiolutin, a RNA polymerase inhibitor leading to its rapid dissociation from chromatin. Pol. II ChIP-seq profiling confirmed its unloading from yeast chromatin upon treatment (**Fig. S8L**) and calibrated RNA-seq revealed that transcription of yeast genes was reduced overall (**Fig. S8M,** 23-34% reduction in steady-state transcription). Concomitantly, Hi-C revealed that long-range interactions within the yeast genome increased, and that Y/U compartments were strongly affected (**Fig. 4G-I**), suggesting that chromatin compartmentalization is dependent on yeast transcriptional activity. To test this under physiological conditions, we performed Hi-C in quiescent yeast cells, where transcription is largely silenced (*40*). Hi-C revealed that the Mmyco U-type compartment disappears upon entry into quiescence (**Fig. 4J-L**), and that the Mmyco chromatin segregates away from the yeast telomere hypercluster appearing in quiescence (*41*) (**Fig. S8N**), supporting the independence between U-type chromatin and the telomeric compartment.

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

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

We explored the extent to which intrinsic DNA sequence composition can be predictive of specific chromatin features of foreign DNA integrated in yeast. We trained convolutional neural networks (CNNs) using yeast DNA sequences from chromosomes I to XV as independent variables to predict nucleosome, cohesin and Pol II coverage tracks along these sequences (see **Methods, Fig. S9A**). These models were then validated on the held-out sequence of yeast chromosome XVI (correlation with experimental signals of 0.63, 0.82 and 0.68 for nucleosome, Scc1 ChIP and Pol. II ChIP predictions respectively, **Fig. 5A,** see **Methods**), confirming that the genomic sequence is predictive of chromatin composition and activity in yeast (*42–44*).

Using our CNRR models, we then predicted coverages over Mpneumo and Mmyco chromosomes, and observed that the rules learned from yeast sequences were sufficient to accurately predict many of the features experimentally characterized on these chromosomes. The predictions recapitulate the features of Y chromatin on Mpneumo, but also the U chromatin features on Mmyco sequences, including the increased linker length (**Fig. 5A insert i, 5B-C**), (2) Scc1 over-enrichment with no discrete peaks (**Fig. 5A insert ii**), and (3) the lower unstructured Pol. II coverage, comparable to the background signal over yeast chromosomes (**Fig. 5A insert iii**).

To broaden the spectrum of exogenous DNA origin, we next experimentally characterized the nucleosome coverage and spatial organization of two other YACs integrated in yeast, containing sequences from the *Plasmodium falciparum* eukaryotic genome (91 kb and 58 kb, 18% GC content; Methods) and the other containing a 284 kb-long, 21% GC YAC from *Candidatus Phytoplasma vitis*, a non-mycoides bacterial species (**Fig. S10**). Both bacterial and eukaryotic AT-rich exogenous chromosomes display features of U-type chromatin, including (1) a NRL longer than that of endogenous yeast chromosomes (~182 bp) and (2) segregation into a distinct compartment, with reduced trans-chromosomal contacts and increased contacts with yeast telomeres (**Fig. S10A-E**). Of note, trans-chromosomal contacts between the two *P. falciparum*

YACs were frequent (**Fig. S10A**), reminiscent of trans-chromosomal contacts between Mmyco translocated segments (**Fig. 4E**). We found that the increased NRL in these two foreign chromosomes is correctly predicted by the CNN model (**Fig. S10F, G**). The CNN model also accurately predicts nucleosome coverage over an artificial 18kb-long GC-rich sequence with 50% GC for which MNase-seq data in yeast was already generated (*45*).

To further validate the accuracy of the CNN model, we compared Pol. II coverage predictions over several natural and artificial sequences that have been integrated and characterized in yeast (**Fig. S10H**). In particular, the human HPRT1 sequence or its reverse sequence (HPRT1r) (~ 41% GC), and a 18kb-long artificial sequence (~ 50% GC) have been shown to be transcriptionally more active than yeast endogenous chromosomes, when integrated in yeast (*46*). Over these sequences, we predict a higher Pol II occupancy compared to yeast chromosomes, consistent with published experimental results (*45*, *46*).

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

To better characterize the relationship between GC content and chromatin composition, we predicted the average nucleosome, Scc1 and Pol. II coverage over thousands of 1kb artificial random sequences with variable GC content (**Fig. 5D**; **Methods**). We observed different behaviors as a function of GC content: (1) the nucleosome signal was higher for intermediate GC % (i.e. between 30% and 50%) and was strongly decreased outside of this range; (2) the cohesin signal continuously decreased with increasing GC%; (3) the Pol. II signal was minimal up to 25% GC, constant up to 45% GC and then increased up to 85% GC. Importantly, these predictions on random sequences accurately recapitulate experimental measurements over 1kb segments from yeast, Mmyco and Mpneumo sequences (**Fig. 5D)**, suggesting that these predicted variations indeed reflect intrinsic properties of the chromatin assembled on DNA with different GC content.

When comparing CNNs performances with two simpler linear regression models, either based on the GC% or on the dinucleotide composition as predictor variables, we found that all models accurately capture dinucleotide signatures for nucleosome, cohesin or Pol. II coverage (e.g. AG, AC, CT, CA, GT, TG in nucleosome) (**Fig. S9B**) but also that the CNN approach performs better in predicting the actual tracks (**Fig. S9C**). A motif analysis based on the saliency computed from the trained networks also allowed us to identify some of the DNA motifs involved in the

sequence-dependent chromatin composition (**Fig. S9D**). These results show that besides GC% and dinucleotide signatures, more complex sequence features influence chromatin composition, regardless of their evolutionary origin.

Discussion

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

Several studies have shown that exogenous or random DNA segments in the yeast nucleus with a GC% relatively similar to that of yeast chromosomes are actively transcribed (*45–48*). Here we show that this is not always the case, and that DNA sequences with different GC content adopt one of two archetypes of chromatin that we called Y and U for yeast-like and unconventional. Y chromatin displays nucleosome arrays with a canonical 160bp NRL, recruits Pol. II and is transcriptionally active, whereas U chromatin is characterized by a longer 174 bp NRL 174bp, and a lower Pol. II occupancy and transcriptional activity (**Fig. 5E**).

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

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

Transcriptomics signal on bacterial chromosomes in yeast follows, on average, bacteria genes orientation (**Fig. 2B, C**). Sequence determinants, including GC and AT skews, predate the divergence of eucaryotes and procaryotes, and are notably found in both Mpneumo and Mmyco chromosomes. Such determinants are known to influence polymerase directionality (*49*, *50*). Eventually, the resulting conserved orientation could facilitate the domestication of exogenous sequences during horizontal transfer/introgression events between distant species. Whether these RNA molecules are translated, and peptides of bacterial origins exist in the yeast cell, remains to be determined. If so, they could provide a source of diversity and adaptation.

Cohesin-mediated chromatin folding along non-transcribed DNA templates

Cohesins have been proposed to actively extrude DNA loops until they encounter an obstacle and/or a release signal (*6*, *32*). In yeast and other species, these anchors are determined by a combination of convergent transcription, replication fork progression during S phase and/or the presence at these positions of stably bound cohesin promoting sister chromatid cohesion (*51–53*). Here, we show that the barely transcribed Mmyco chromosome is compacted by cohesin at G2/M, without focal loop anchoring, suggesting that transcription is neither necessary for loading nor a primary driver for translocation. We propose that cohesins can move freely along this template without encountering significant obstacles, making it a suitable model for studying potentially blocking sequences or molecules.

Compartmentalization of transcriptionally inactive foreign DNA in host's nucleus

We show that the introduction of a foreign DNA with a lower GC content spontaneously promotes the formation of transcriptionally inactive chromatin which folds into an isolated crumpled globule compartment at the periphery of the budding yeast nucleus (**Fig. 3E; Fig. S6H**), a behavior that mirrors the metazoan compartment "B" formed by inactive, H3K9me3/HP1-mediated heterochromatin, which is absent in yeast. In contrast, we find that chromatin assembled on sequences with a composition similar to that of the host and transcriptionally active adopts a random coil structure and intermixes with transcribed yeast chromosomes (**Fig. 5E**).

We further show that (1) high or low Pol. II occupancy can be accurately predicted in both chromatin archetypes based on their sequence alone, and that (2) thiolutin treatment, a drug that inhibits transcription by dissociating Pol. II from chromatin, reduces the segregation of U and Y chromatin. These findings suggest that the recruitment of the transcriptional machinery, which correlates to the underlying DNA sequence, directly contributes to the physical segregation of adjacent Pol. II-depleted chromatin into a distinct, globular, inactive compartment. This behavior may be explained by the different chromatin composition (e.g., acetylated histone tail residues or intrinsically disordered proteins composing the transcription machinery) in each compartment

These different fates for two foreign chromosomal sequences raise interesting evolutionary considerations in the context of the invasion of the genome by exogenous mobile elements. This could lead either to the spontaneous isolation of inactive foreign DNA or, on the contrary, to the

co-option of a set of active sequences that could represent a reservoir of genetic innovations. We show that an active chromatin region as small as 15 kb is not sensitive to the surrounding silenced chromatin environment, suggesting that the harnessing of small genome-attuned regions could readily occur during introgressions or HGT events. This sequence-dependent mechanism may have contributed to the heterochromatinization of AT-rich transposable elements integrated in mammalian genomes.

Acknowledgements: We are very grateful to Michael Lanzer (Heidelberg University) and Cecilia Sanchez for providing us with *P. falciparum* YACs, and to Sebastian Baumgarten for long-read sequencing of these YACs. We also thank Bernard Dujon, Micheline Fromont-Racine, Alain Jacquier, Gianni Liti, Bertrand Llorente, Marcelo Nollmann, Cosmin Saveanu, Benoit le Tallec and all members of the laboratory Régulation Spatiale des Génomes for fruitful comments on the work and the manuscript. We thank Cyril Matthey-Doret and Guillaume Mercy for help during the earlier steps of the project, the Biomics the PICT-IBiSA@Pasteur Imaging Facility of the Institut Curie, a member of the France Bioimaging National Infrastructure (ANR-10-INBS-04), and particularly Mickael Garnier for his help on FISH quantification.

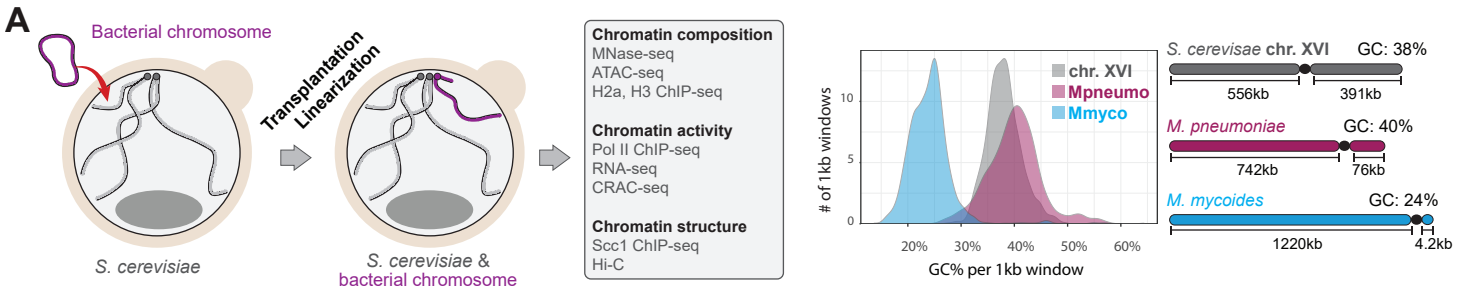
Funding: This work was supported by the European Research Council under the Horizon 2020 Program (ERC grant agreement 771813) and Agence Nationale pour la Recherche (ANR-19-CE13-0027-02) to RK. RK, FB, JM and AnT also received support from Agence Nationale pour la Recherche (ANR-22-CE12-0013-01). CC was supported by a Pasteur-Roux-Cantarini fellowship. JS was supported by an ARC fellowship. 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.

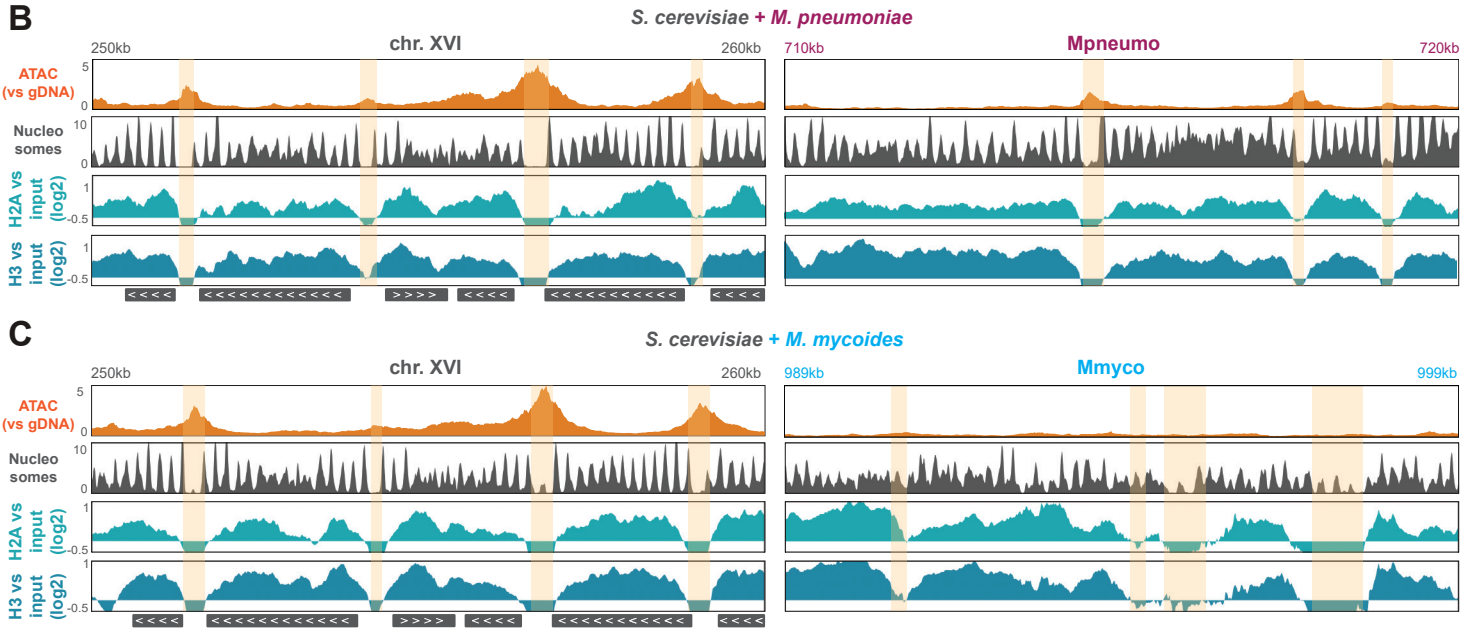
Author contributions: Conceptualization: CC, LM, JS, JM and RK. Methodology: LM, CC, JS, CL, JM, RK. Software: JS. Validation: CC, JS, LM. Investigation: LM, CC, with contributions from MP, FG, AP, AB, AE, AgT, MR and FB. Formal analysis: JS (all data processing and integration), AW and ER (CNN models), with contributions from CC, LM, FB, MR, AnT. Data Curation: JS, with contributions from CC, LM. Resources: GG, CL. Visualization: JS. Writing - original draft preparation: CC, JS, LM, JM and RK. Writing – Editing: all authors. Writing –

revisions: JS, LM, AW, JM, RK. Supervision: AT, DL, JM, RK. CC co-supervised a student. Funding acquisition: AT, DL, FB, JM, RK. Project Administration: RK.

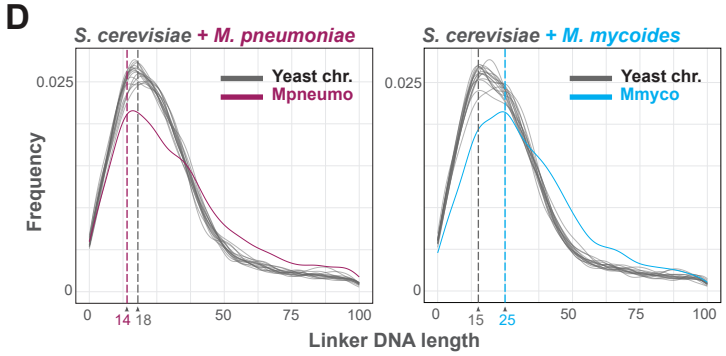
Competing interests: The authors declare no competing interests.

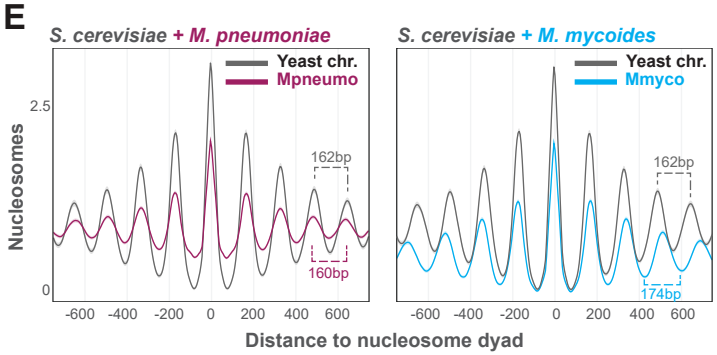
Data and materials availability: Sample description and raw sequences for all figures are accessible on GEO database through the following accession number: GSE217022 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE217022, token *arepcamenvafhuz*). All custom-made code of the analysis of sequencing data is available online https://github.com/koszullab/. Open-access versions of the programs and pipeline used (Hicstuff) are available online on the github account of the Koszul lab: Hicstuff (https://github.com/koszullab/hicstuff) version 3.1.2, Chromosight (version 1.4.1 available online at https://github.com/koszullab/chromosight/), Bowtie2 (version 2.4.5 available online at http://bowtie-bio.sourceforge.net/bowtie2/), SAMtools (version 1.9 available online at http://www.htslib.org/), Bedtools86 (version 2.29.1 available online at https://bedtools.readthedocs.io/en/latest/content/installation.html) and Cooler (versions 0.8.7–0.8.11 available online at https://cooler.readthedocs.io/en/latest/).


A
Bacterial chromosome
Transplantation
Linearization
S. cerevisiae
S. cerevisiae &
bacterial chromosome
Chromatin composition
MNase-seq
ATAC-seq
H2a, H3 ChIP-seq
Chromatin activity
Pol II ChIP-seq
RNA-seq
CRAC-seq
Chromatin structure
Scc1 ChIP-seq
Hi-C
# of 1kb windows
chr. XVI
Mpneumo
Mmyco
GC% per 1kb window
S. cerevisae chr. XVI GC: 38%
556kb
391kb
M. pneumoniae GC: 40%
742kb
76kb
M. mycoides GC: 24%
1220kb
4.2kb


B
S. cerevisiae + M. pneumoniae
chr. XVI
Mpneumo
ATAC (vs gDNA)
Nucleo somes
H2A vs input (log2)
H3 vs input (log2)

C
S. cerevisiae + M. mycoides
chr. XVI
Mmyco
ATAC (vs gDNA)
Nucleo somes
H2A vs input (log2)
H3 vs input (log2)


D
S. cerevisiae + M. pneumoniae
S. cerevisiae + M. mycoides
Yeast chr.
Mpneumo
Mmyco
Frequency
Linker DNA length


E
S. cerevisiae + M. pneumoniae
S. cerevisiae + M. mycoides
Yeast chr.
Mpneumo
Mmyco
Nucleosomes
162bp
160bp
174bp
Distance to nucleosome dyad

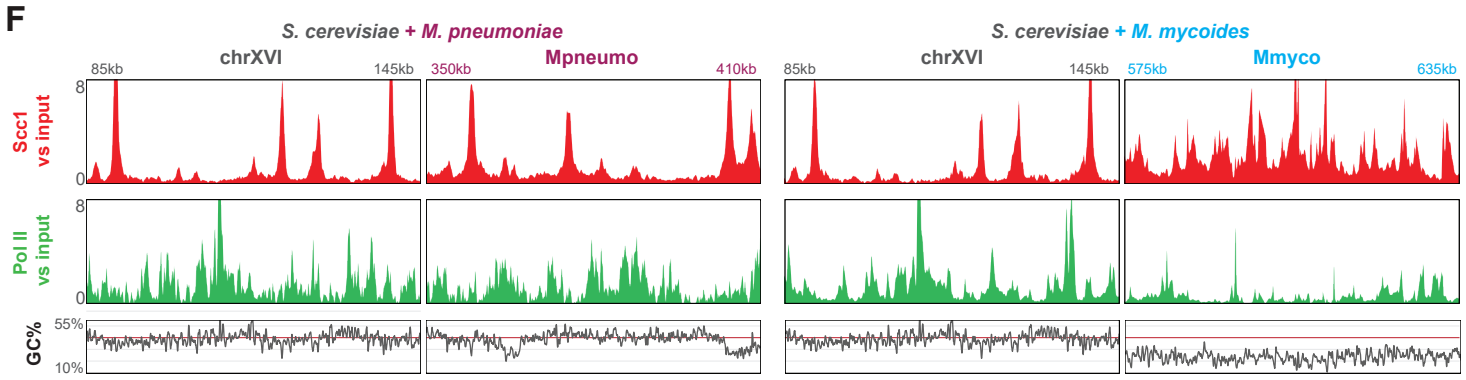

F
S. cerevisiae + M. pneumoniae
S. cerevisiae + M. mycoides
chrXVI
Mpneumo
Mmyco
Scc1 vs input
Pol II vs input
GC%

Fig. 1

Figure legends

Fig. 1. Chromatin composition of bacterial chromosomes integrated in yeast.

A, Schematic 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. Right: distribution of GC% for 1kb windows over yeast chromosome XVI, *M. pneumoniae* and *M. mycoides* chromosomes.

B, ATAC-seq (orange, CPM), nucleosomal track (gray, see **Methods**) and H2A and H3 ChIP-seq (shades of blue, IP vs input, log2) profiles obtained in the Mpneumo strain (*S. cerevisiae* + *M. pneumoniae*). 10kb-long genomic windows from the chromosome XVI (left: 250-260kb) and the Mpneumo chromosome (right: 710-720kb) are shown at the same scale. Nucleosome-depleted regions are highlighted in yellow.

C, Same as **B** in the Mmyco strain (*S. cerevisiae* + *M. mycoides*). 10kb-long genomic windows from chromosome XVI (left: 250-260kb) and the Mmyco chromosome (right: 989-999kb) are shown at the same scale.

D, Frequency of nucleosome linker DNA length in Mpneumo and Mmyco strains. For each strain, the distribution is calculated for each chromosome separately. The dashed line indicates 14 bp and the dotted line indicates 25 bp. The x axis only shows linker DNA lengths in the 0-100 bp range.

E, Nucleosomal track centered on nucleosome peaks for *S. cerevisiae*, Mpneumo and Mmyco. The Y axis represents the average nucleosomal track (130-165 bp MNase-seq fragments, resized to 40bp, piled-up and normalized to sequencing depth).

F, Scc1 (red) and RNA Pol II (green) ChIP-seq profiles (IP vs input, log2) obtained in the Mpneumo (left) or in the Mmyco (right) strains. For each strain, 60kb-long genomic windows from the bacterial chromosome and yeast chromosome XVI are shown at the same scale. GC% in sliding 1kb windows is shown below the ChIP-seq profiles.

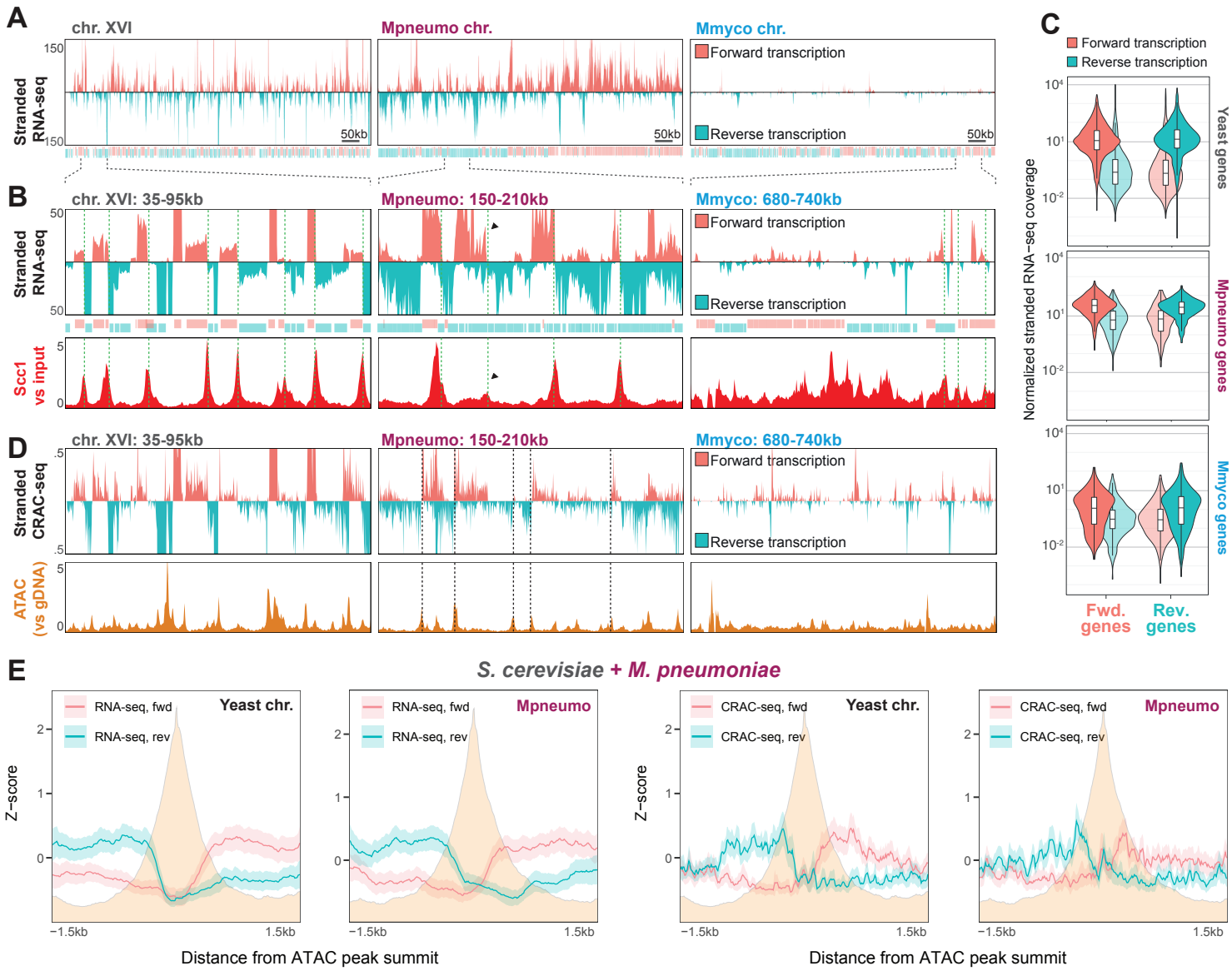

A
chr. XVI
Mpneumo chr.
Mmyco chr.
Stranded RNA-seq
Forward transcription
Reverse transcription
50kb
B
chr. XVI: 35-95kb
Mpneumo: 150-210kb
Mmyco: 680-740kb
Scc1 vs input
C
Normalized stranded RNA−seq coverage
Yeast genes
Mpneumo genes
Mmyco genes
Fwd. genes
Rev. genes
D
Stranded CRAC-seq
ATAC (vs gDNA)
E
S. cerevisiae + M. pneumoniae
RNA-seq, fwd
RNA-seq, rev
Yeast chr.
Mpneumo
CRAC-seq, fwd
CRAC-seq, rev
Z−score
−1.5kb
1.5kb
Distance from ATAC peak summit

Fig. 2

Fig. 2. Transcriptional activity of exogenic bacterial sequences in budding yeast.

A, Stranded RNA-seq profiles along either yeast chromosome XVI or bacterial chromosomes Mpneumo and Mmyco. Forward (pink) and reverse (turquoise) genes along yeast or bacterial sequences are indicated as transparent segments under the tracks. Pink and turquoise represent forward and reverse transcription, respectively.

B, Top: stranded RNA-seq profiles along a 60kb window along yeast chromosome XVI, Mpneumo or Mmyco. Bottom: Scc1 (cohesin) deposition profiles of the corresponding loci. Green dotted lines indicate identified loci of convergent transcription (see Methods). Black arrowheads: overlapping bidirectional transcription.

C, Forward and reverse RNA-seq coverage of forward- and reverse-oriented yeast or bacterial genes. Scores are normalized by each genomic feature length.

D, Top: stranded CRAC-seq profiles along the region shown in **B**. Bottom: corresponding ATAC-seq profile. Dotted lines point at sites of bidirectional initiation along Mpneumo.

E, Metaplot of 3kb regions centered on ATAC peaks over yeast or Mpneumo chromosomes, of stranded RNA-seq (left panels) or CRAC-seq (right panels) profiles.

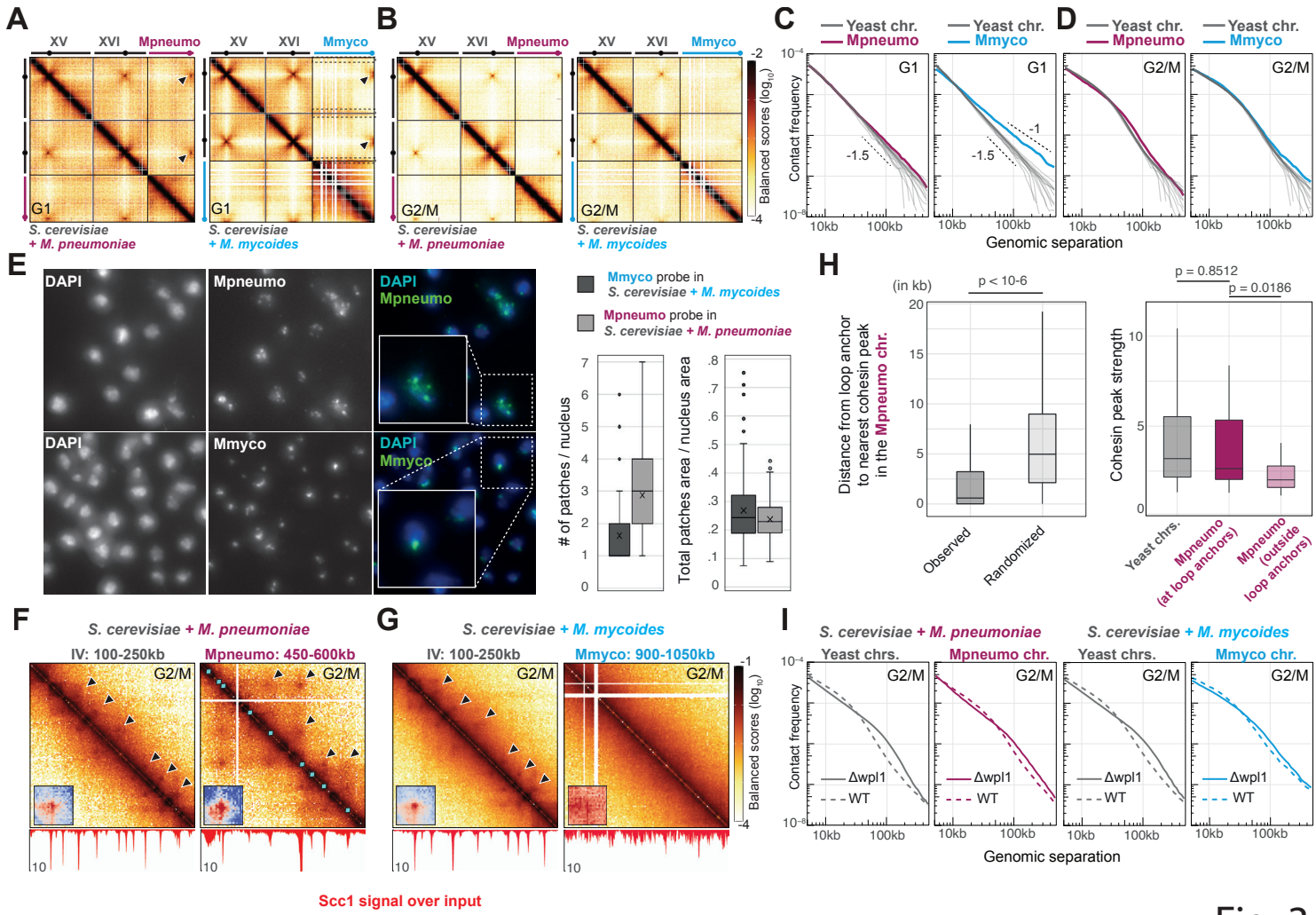

A
XV
XVI
Mpneumo
Mmyco
G1
G2/M
S. cerevisiae + M. pneumoniae
S. cerevisiae + M. mycoides
Balanced scores (log10)
B
C
Yeast chr.
Mpneumo
Mmyco
Contact frequency
-1.5
-1
10kb
100kb
D
Genomic separation
E
DAPI
Mpneumo
Mmyco
Mmyco probe in S. cerevisiae + M. mycoides
Mpneumo probe in S. cerevisiae + M. pneumoniae
# of patches / nucleus
Total patches area / nucleus area
H
(in kb)
p < 10-6
Distance from loop anchor to nearest cohesin peak in the Mpneumo chr.
Observed
Randomized
p = 0.8512
p = 0.0186
Cohesin peak strength
Yeast chrs.
Mpneumo (at loop anchors)
Mpneumo (outside loop anchors)
F
S. cerevisiae + M. pneumoniae
IV: 100-250kb
Mpneumo: 450-600kb
Scc1 signal over input
G
S. cerevisiae + M. mycoides
IV: 100-250kb
Mmyco: 900-1050kb
I
Yeast chrs.
Mpneumo chr.
Mmyco chr.
Δwpl1
WT

Fig. 3

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

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

C, D, Contact frequency (p) as a function of genomic distance (s) plots of endogenous yeast chromosomes (long arms) and of Mmyco and Mpneumo bacterial chromosomes in G1 (**C**) and G2/M (**D**).

E, Left: FISH imaging. Representative field of either (top) Mpneumo or (bottom) Mmyco fixed cells labeled with DAPI (left panel) and hybridized with a fluorescent probe generated from either the Mpneumo or Mmyco chromosome, respectively. Right: For each probe, number of patches detected per nucleus and surface occupied by these patches relative to the whole nucleus surface (Methods).

F, G, Top: magnification of 150kb windows from Hi-C contact maps in G2/M from either an endogenous or the bacterial chromosome in Mpneumo (**F**) and Mmyco (**G**) strains (1kb resolution). Bottom: Scc1 ChIP-seq deposition profile. Black arrowheads: loops. Cyan diamonds in Mpneumo Hi-C contact map: Scc1 peaks positions reported on the contact map diagonal. Inset in each map: *Chromosight* pileup of contacts between cohesin enrichment peaks along either S. cerevisiae or bacterial chromosomes (see **Methods**).

H, Left: distance between chromatin loop anchors and the nearest Scc1 peak in Mpneumo (with and without a random shuffle of peak positions). Right: Scc1 peak strengths in yeast or Mpneumo chromosome, near (< 1kb) or outside loop anchors (p-values from two-sided Student's t-test).

I, Distance-dependent contact frequency in endogenous yeast and in bacterial chromosomes (left, Mpneumo; right, Mmyco), in WT (dashed) and in Δwpl1 mutants.

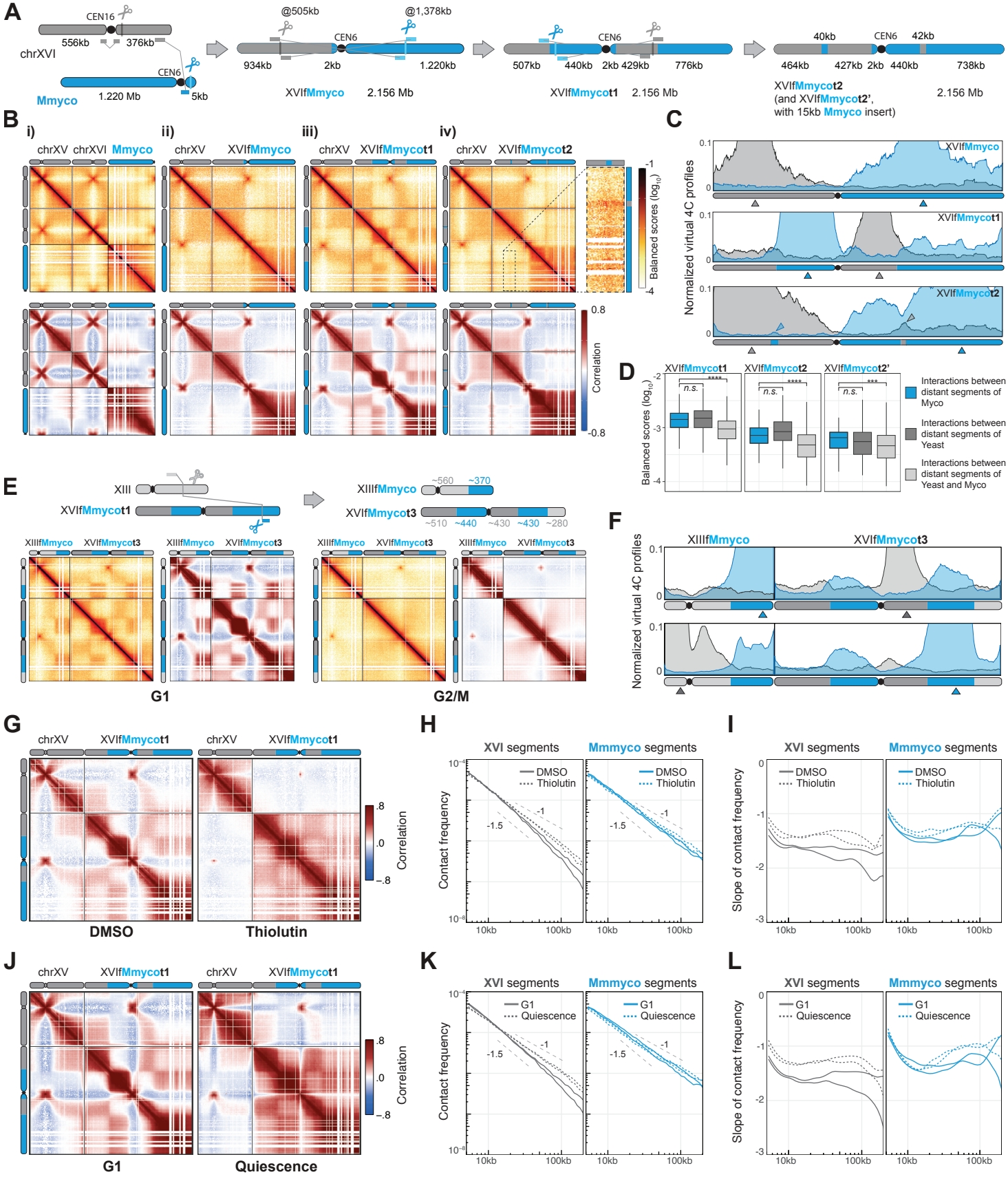

A
CEN16
556kb
376kb
chrXVI
CEN6
1.220 Mb
5kb
Mmyco
@505kb
@1,378kb
CEN6
934kb
2kb
1.220kb
XVIfMmyco 2.156 Mb
CEN6
507kb
440kb
2kb 429kb
776kb
XVIfMmycot1 2.156 Mb
40kb
CEN6
42kb
464kb
427kb 2kb 440kb
738kb
XVIfMmycot2
(and XVIfMmycot2',
with 15kb Mmyco insert)
2.156 Mb
B
i)
chrXV chrXVI Mmyco
ii)
chrXV XVIfMmyco
iii)
chrXV XVIfMmycot1
iv)
chrXV XVIfMmycot2
Balanced scores (log10)
Correlation
C
Normalized virtual 4C profiles
XVIfMmyco
XVIfMmycot1
XVIfMmycot2
D
XVIfMmycot1
XVIfMmycot2
XVIfMmycot2'
Balanced scores (log10)
n.s.
****
***
Interactions between distant segments of Myco
Interactions between distant segments of Yeast
Interactions between distant segments of Yeast and Myco
E
XIII
XVIfMmycot1
XIIIfMmyco
~560
~370
XVIfMmycot3
~510 ~440 ~430 ~430 ~280
XIIIfMmyco XVIfMmycot3
G1
G2/M
F
Normalized virtual 4C profiles
XIIIfMmyco
XVIfMmycot3
G
chrXV XVIfMmycot1
Correlation
DMSO
Thiolutin
H
XVI segments
Mmmyco segments
DMSO
Thiolutin
Contact frequency
-1
-1.5
10kb
100kb
I
XVI segments
Mmmyco segments
Slope of contact frequency
J
chrXV XVIfMmycot1
Correlation
G1
Quiescence
K
XVI segments
Mmmyco segments
G1
Quiescence
Contact frequency
L
XVI segments
Mmmyco segments
G1
Quiescence
Slope of contact frequency

Fig. 4

Fig. 4. Compartmentalization of mosaic chromosomes composed of Y and U-type chromatin.

A, Schematic representation of the CRISPR strategy used to generate the Mmyco mosaic chromosomes with alternating Y and U chromatin regions.

B, Top: G1 Hi-C contact maps of chr. XV, XVI, and bacterial chromosomes in the Mmyco strain, and in translocated derivatives from panel **A**) (4kb resolution). Bottom: correlation matrices of the corresponding contact maps.

C, Virtual 4C profiles of viewpoints (indicated as gray/blue arrowheads) located within yeast segments (in gray) or Mmyco chromosome segments (in blue) of the chimeric chromosome XVI, in XVIfMmyco (top), XVIfMmycot1 (middle) and XVIfMmycot2 (bottom) strains.

D, Quantification of long-range interactions (800kb-1.5Mb) between pairs of distant Mmyco segments (blue), pairs of distant yeast segments (gray) and interactions between a Mmyco segment and a yeast segment (light gray), in double-translocation XVIfMmyco (top), XVIfMmycot1 (middle) and XVIfMmycot2 (bottom) strains.

E, Left: Schematic representation of the CRISPR strategy used to generate the mosaic XVIfMmycot3 strain with alternating yeast and Mmyco segments in chromosomes XIII and XVI. Middle: G1 Hi-C maps of mosaic chromosomes XIII and XVI in XVIfMmycot3 strain. Same color scales as in **B**.

F, Virtual 4C profiles of viewpoints located within yeast segments (gray arrowheads) or Mmyco segments (blue arrowheads) of the chimeric chromosomes XIII and XVI, in the XVIfMmycot3 strain.

G, Correlation matrices of the contacts in chr. XV and XVIfMmycot1 in G1, after the addition of DMSO (left) or thiolutin (right). Same color scale as in **B**.

H, Contact frequency (p) as a function of genomic distance (s), for contacts in yeast segments (gray) or Mmmyco segments (blue) of the chimeric chromosome XVIfMmycot1, in G1 after the addition of DMSO (solid) or thiolutin (dotted).

I, Derivatives of curves from **H**.

J, Chr. XV and XVIfMmycot1 correlation maps in G1 or in quiescence. Same color scale as in **B.**

K, Same p(s) as in H, but in G1 and quiescence.

L, Derivatives of curves from **K**.

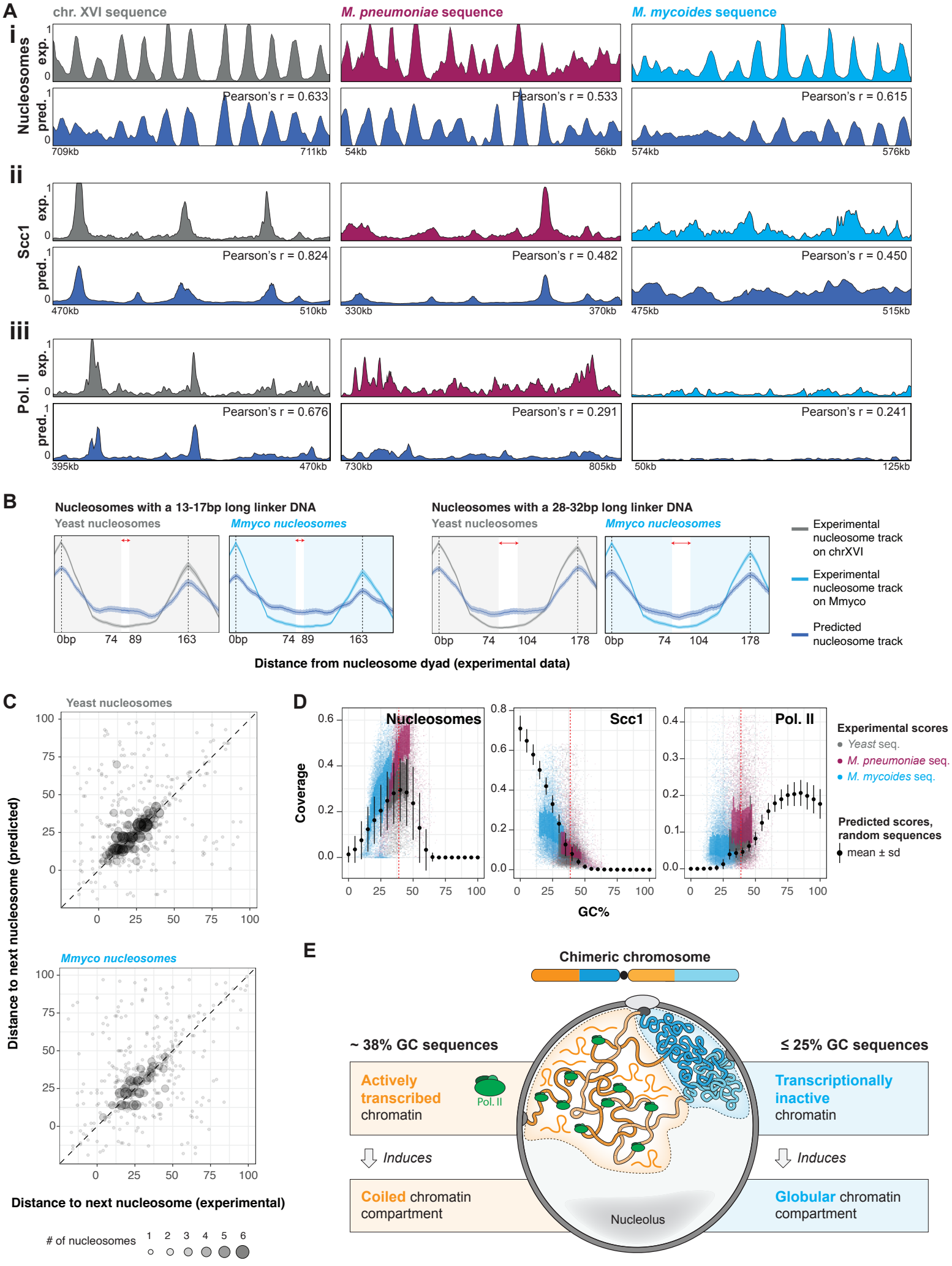

A
chr. XVI sequence
M. pneumoniae sequence
M. mycoides sequence
i
Nucleosomes
exp.
pred.
Pearson's r = 0.633
Pearson's r = 0.533
Pearson's r = 0.615
709kb
711kb
54kb
56kb
574kb
576kb
ii
Scc1
Pearson's r = 0.824
Pearson's r = 0.482
Pearson's r = 0.450
470kb
510kb
330kb
370kb
475kb
515kb
iii
Pol. II
Pearson's r = 0.676
Pearson's r = 0.291
Pearson's r = 0.241
395kb
470kb
730kb
805kb
50kb
125kb
B
Nucleosomes with a 13-17bp long linker DNA
Nucleosomes with a 28-32bp long linker DNA
Yeast nucleosomes
Mmyco nucleosomes
0bp
74
89
104
163
178
Experimental nucleosome track on chrXVI
Experimental nucleosome track on Mmyco
Predicted nucleosome track
Distance from nucleosome dyad (experimental data)
C
Yeast nucleosomes
Mmyco nucleosomes
Distance to next nucleosome (predicted)
Distance to next nucleosome (experimental)
# of nucleosomes
1
2
3
4
5
6
D
Nucleosomes
Scc1
Pol. II
Coverage
GC%
Experimental scores
Yeast seq.
M. pneumoniae seq.
M. mycoides seq.
Predicted scores, random sequences
mean ± sd
E
Chimeric chromosome
~ 38% GC sequences
≤ 25% GC sequences
Actively transcribed chromatin
Pol. II
Transcriptionally inactive chromatin
Induces
Induces
Coiled chromatin compartment
Globular chromatin compartment
Nucleolus

Fig. 5

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

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

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

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

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

E, Schematic of a chimeric chromosome composed of alternating Y- and U-type chromatin.

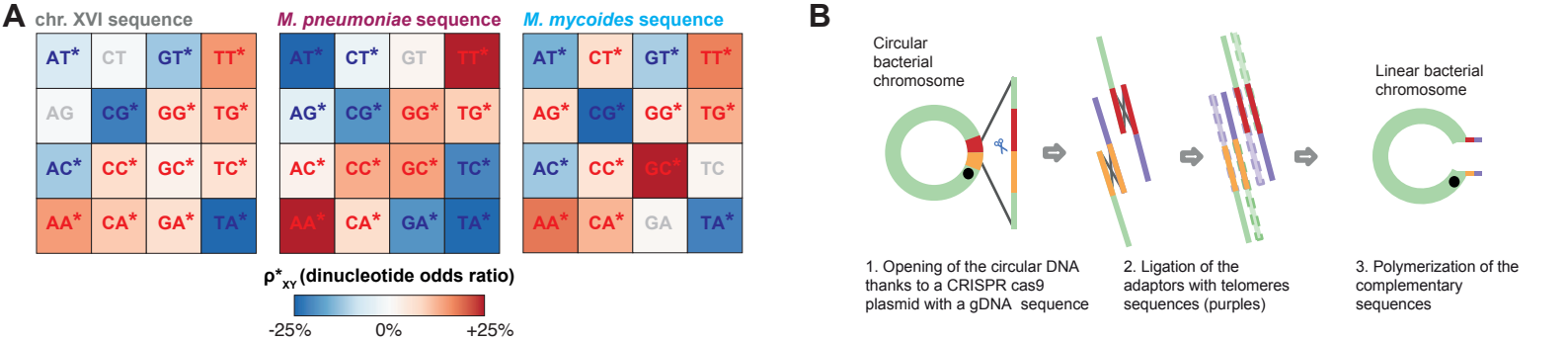
Supplementary Text

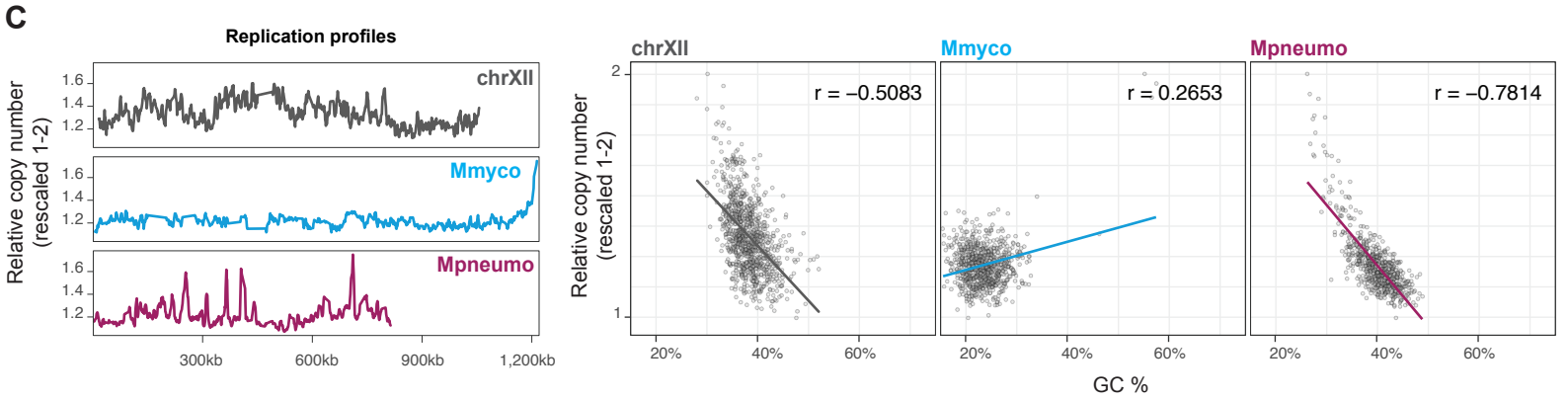
Replication and pairing of artificial chromosomes

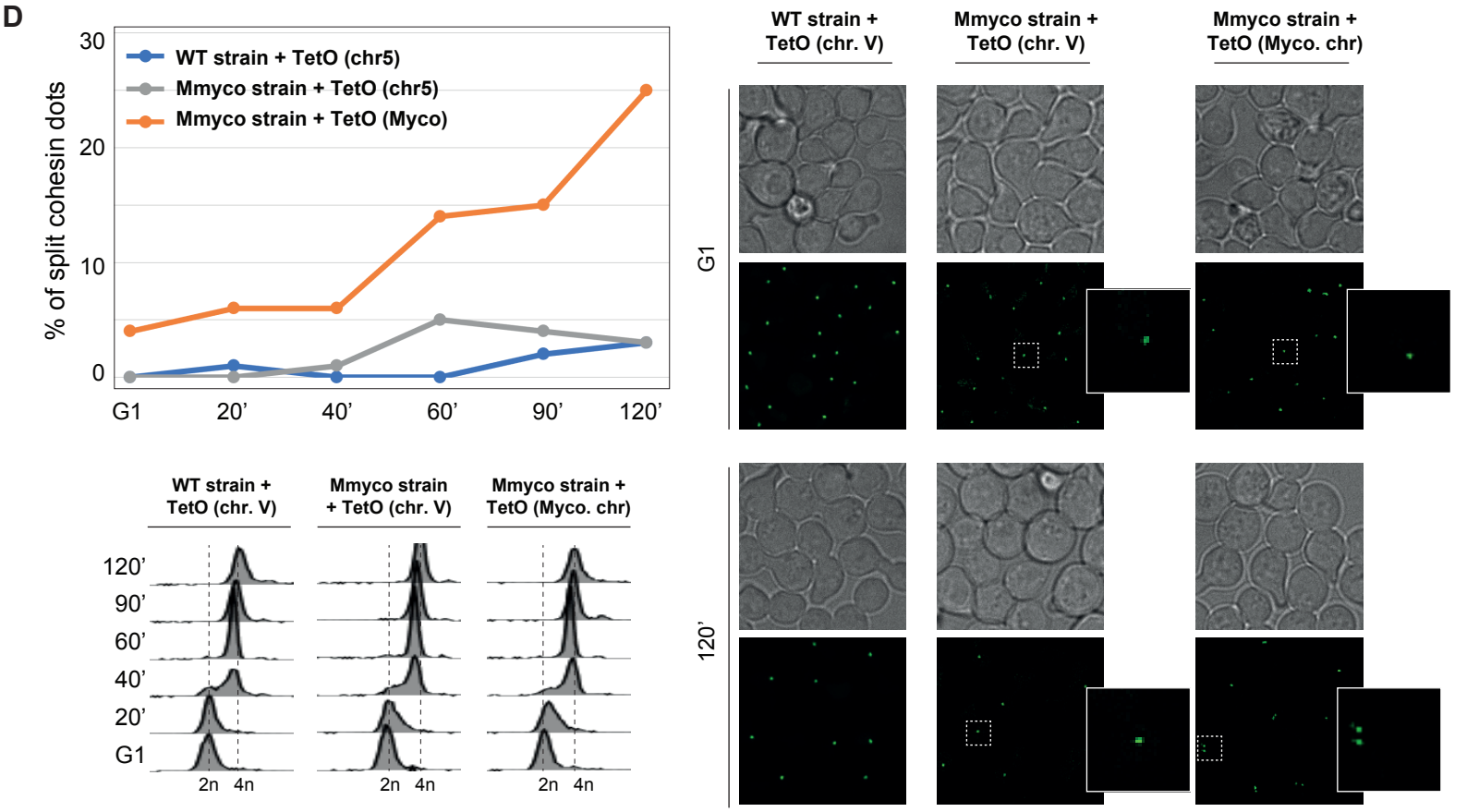
DNA sequence composition and chromatin factors, including nucleosomes, are essential drivers of the replication timing, prompting us to investigate how this timing is established in Y and U chromatin (96).

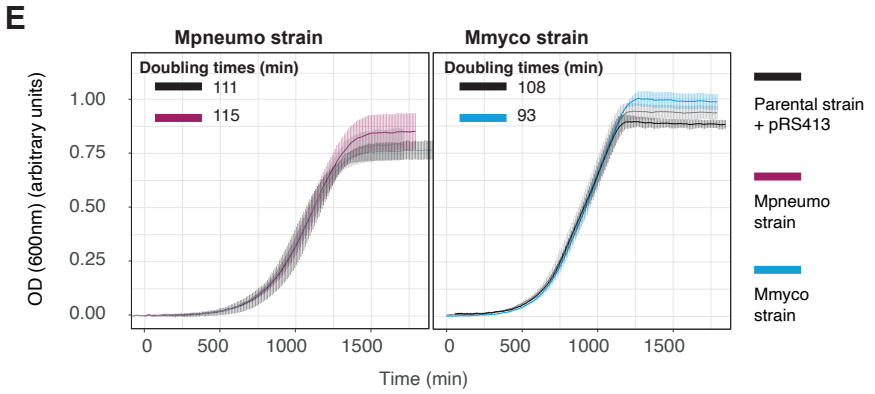
In yeast, replication initiates at the level of discrete, small autonomous replication sequences (ARS) positioned along all chromosomes. To investigate the replication timing profile of both bacterial chromosomes, which did not spontaneously evolve to contain these sequences, we used Repliscore (Methods)(83). Mpneumo Repliscore profile exhibits DNA copy number variation comparable to that observed along chrXII, indicative of early or mid-S phase firing of multiple replication origins (Fig. S1C). This copy number is anti-correlated with GC%, consistent with AT-rich regions replicating earlier. In contrast, DNA copy number variation along Mmyco unveils the early firing of only the centromere-proximal ARS which was artificially integrated during chromosome assembly (Fig. S1C). The average copy number of the rest of the Mmyco chromosome appears flat and is not anti-correlated with the GC%. Thus, the replication of U-type chromatin occurs later during the S phase despite its AT-rich sequence composition. These observations confirm the predominant role of chromatin composition in the replication timing, and indicate a heterochromatin-like effect exerted by the U-type chromatin. This pattern is reminiscent of the random and late-replication pattern displayed by the human inactive chromosome X (93).

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 (20). We therefore investigated SCC by performing image analysis of chromosome pairing (Methods; Fig. S1D). 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 (Fig. S1D), despite the apparent loss of cohesin at centromeres.









F

	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

Fig. S1

Fig. S1. Characterization of Mpneumo and Mmyco strains metabolism.

A, Dinucleotide over-representation (ρ*(XY)) in yeast chromosome XVI, Mpneumo and Mmyco chromosomes.

B, Schematic representation of the CRISPR-Cas9 strategy applied to linearize and telomerize the circular bacterial genomes.

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

D, 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.

E, 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.

F, The chromosome stability and the segregation rate were measured as described in (66) (**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.

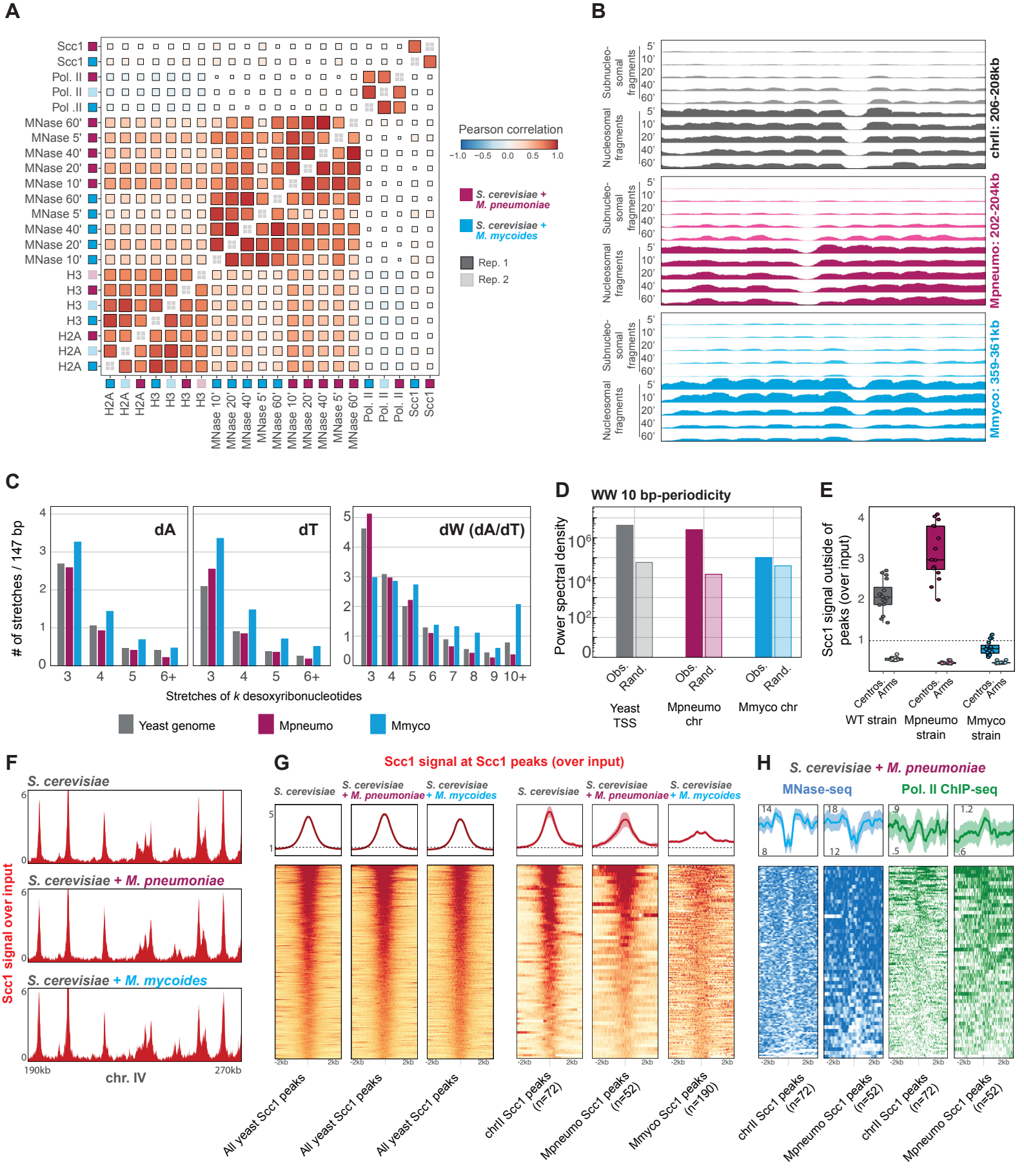

A
Scc1
Scc1
Pol. II
Pol. II
Pol .II
MNase 60'
MNase 5'
MNase 40'
MNase 20'
MNase 10'
MNase 60'
MNase 5'
MNase 40'
MNase 20'
MNase 10'
H3
H3
H3
H3
H2A
H2A
H2A
H2A H2A H2A H3 H3 H3 MNase 10' MNase 20' MNase 40' MNase 5' MNase 60' MNase 10' MNase 20' MNase 40' MNase 5' MNase 60' Pol. II Pol. II Pol. II Scc1 Scc1
Pearson correlation
−1.0 −0.5 0.0 0.5 1.0
S. cerevisiae + M. pneumoniae
S. cerevisiae + M. mycoides
Rep. 1
Rep. 2
B
Subnucleosomal fragments
Nucleosomal fragments
5' 10' 20' 40' 60'
chrII: 206-208kb
Mpneumo: 202-204kb
Mmyco: 359-361kb
C
# of stretches / 147 bp
dA
dT
dW (dA/dT)
Stretches of k desoxyribonucleotides
Yeast genome
Mpneumo
Mmyco
D
WW 10 bp-periodicity
Power spectral density
Obs. Rand.
Yeast TSS
Mpneumo chr
Mmyco chr
E
Scc1 signal outside of peaks (over input)
Centros. Arms
WT strain
Mpneumo strain
Mmyco strain
F
S. cerevisiae
S. cerevisiae + M. pneumoniae
S. cerevisiae + M. mycoides
Scc1 signal over input
190kb
chr. IV
270kb
G
Scc1 signal at Scc1 peaks (over input)
S. cerevisiae
S. cerevisiae + M. pneumoniae
S. cerevisiae + M. mycoides
-2kb 2kb
All yeast Scc1 peaks
chrII Scc1 peaks (n=72)
Mpneumo Scc1 peaks (n=52)
Mmyco Scc1 peaks (n=190)
H
S. cerevisiae + M. pneumoniae
MNase-seq
Pol. II ChIP-seq
chrII Scc1 peaks (n=72)
Mpneumo Scc1 peaks (n=52)

Fig. S2

Fig. S2. Replication of bacterial genomes in yeast.

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

B, Coverage of sub-nucleosomal (smaller than 130bp) or nucleosomal (between 130 and 165bp) fragments over 2kb-long genomic loci from chr. II, Mpneumo and Mmyco chromosomes, over an MNase digestion timecourse. All the tracks are displayed at the same scale (0-20 CPM).

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

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

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

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

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

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

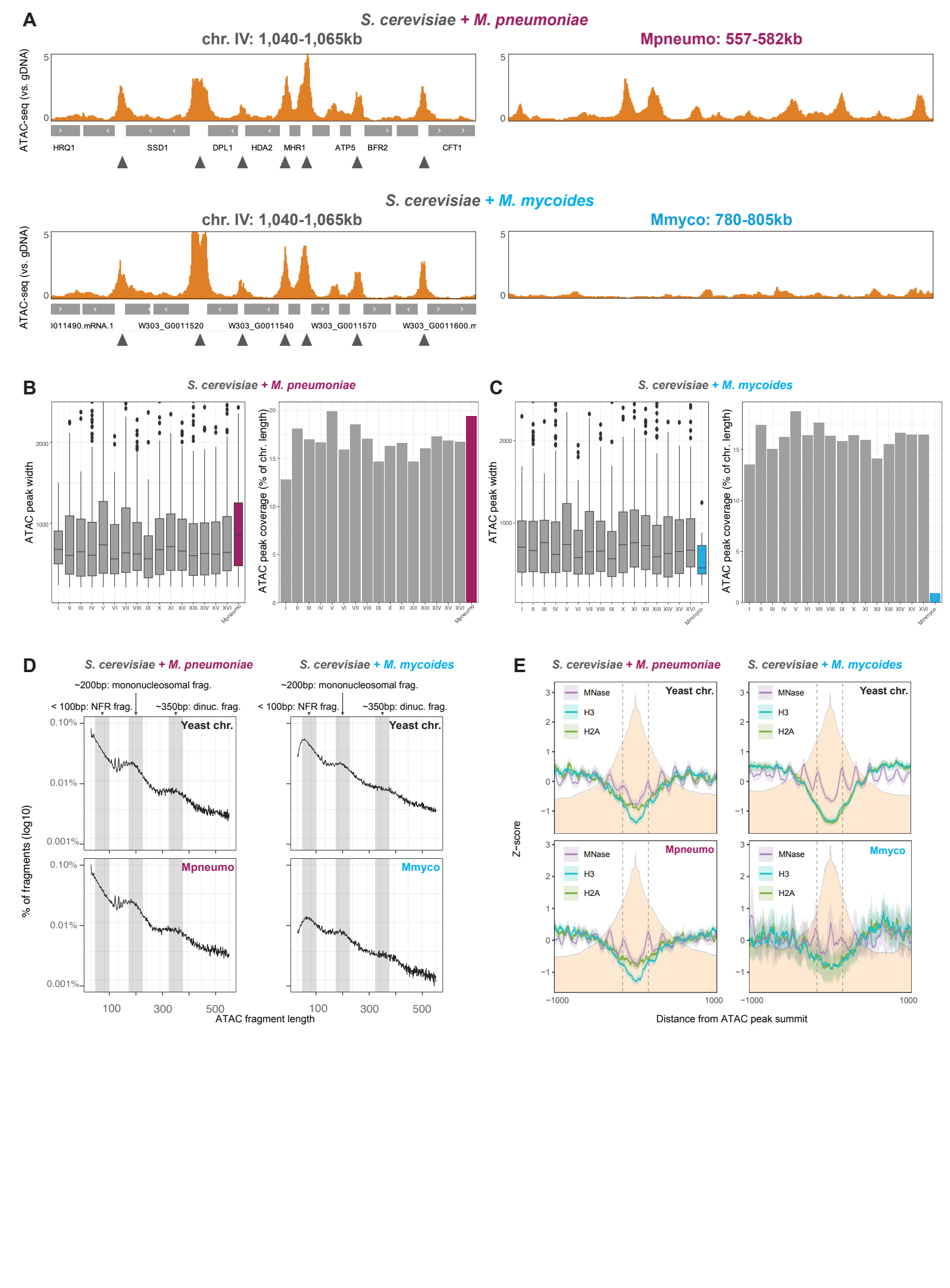

A
S. cerevisiae + M. pneumoniae
chr. IV: 1,040-1,065kb
Mpneumo: 557-582kb
ATAC-seq (vs. gDNA)
HRQ1
SSD1
DPL1
HDA2
MHR1
ATP5
BFR2
CFT1
S. cerevisiae + M. mycoides
chr. IV: 1,040-1,065kb
Mmyco: 780-805kb
W303_G0011520
W303_G0011540
W303_G0011570
B
S. cerevisiae + M. pneumoniae
ATAC peak width
ATAC peak coverage (% of chr. length)
C
S. cerevisiae + M. mycoides
D
S. cerevisiae + M. pneumoniae
S. cerevisiae + M. mycoides
~200bp: mononucleosomal frag.
< 100bp: NFR frag.
~350bp: dinuc. frag.
Yeast chr.
Mpneumo
Mmyco
% of fragments (log10)
ATAC fragment length
E
S. cerevisiae + M. pneumoniae
S. cerevisiae + M. mycoides
MNase
H3
H2A
Z-score
Distance from ATAC peak summit

Fig. S3

Fig. S3. Chromatin accessibility of bacterial chromosomes in yeast.

A, ATAC-seq profiles (chromatin/gDNA) obtained in the Mpneumo strain (S. cerevisiae + M. pneumoniae) and Mmyco strain (S. cerevisiae + M. mycoides) (10kb-long genomic windows over the chromosome IV, Mpneumo and Mmyco chromosomes).

B-C, ATAC-seq analysis comparing S. cerevisiae with M. pneumoniae (**B**) or with M. mycoides (**C**). Boxplots display ATAC peak widths and barplots summarize the total % of each chromosome covered by ATAC-seq peaks.

D, ATAC-seq fragment length distribution for S. cerevisiae combined with M. pneumoniae (left) and M. mycoides (right). ATAC-seq fragments obtained from nucleosome-free regions (<100 bp), mononucleosomal (~200 bp), and dinucleosomal (~350 bp) fragments are highlighted.

E, Average signal of nucleosome and histone ChIP-seq tracks over ATAC-seq peaks.

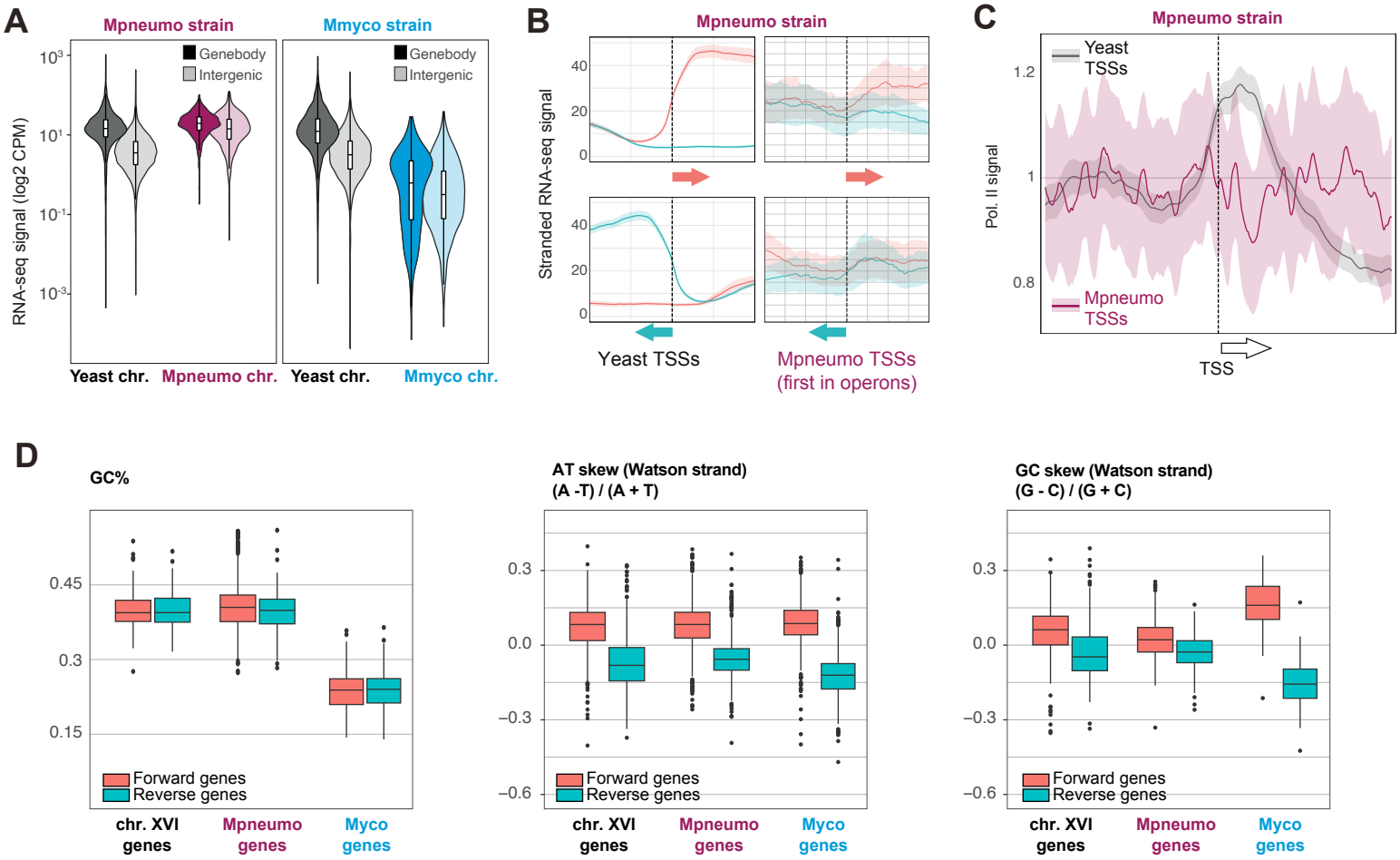

A
Mpneumo strain
Mmyco strain
Genebody
Intergenic
RNA-seq signal (log2 CPM)
Yeast chr.
Mpneumo chr.
Yeast chr.
Mmyco chr.
B
Mpneumo strain
Stranded RNA-seq signal
Yeast TSSs
Mpneumo TSSs
(first in operons)
C
Mpneumo strain
Yeast
TSSs
Pol. II signal
Mpneumo
TSSs
TSS
D
GC%
AT skew (Watson strand)
(A -T) / (A + T)
GC skew (Watson strand)
(G - C) / (G + C)
Forward genes
Reverse genes
chr. XVI
genes
Mpneumo
genes
Myco
genes

Fig. S4

Fig. S4. Transcription orientation of bacterial chromosomes in yeast.

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

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

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

D, GC content (left), average AT (middle) and GC (right) skews (computed on Watson strand) over yeast, Myco and Mpneumo forward or reverse genes.

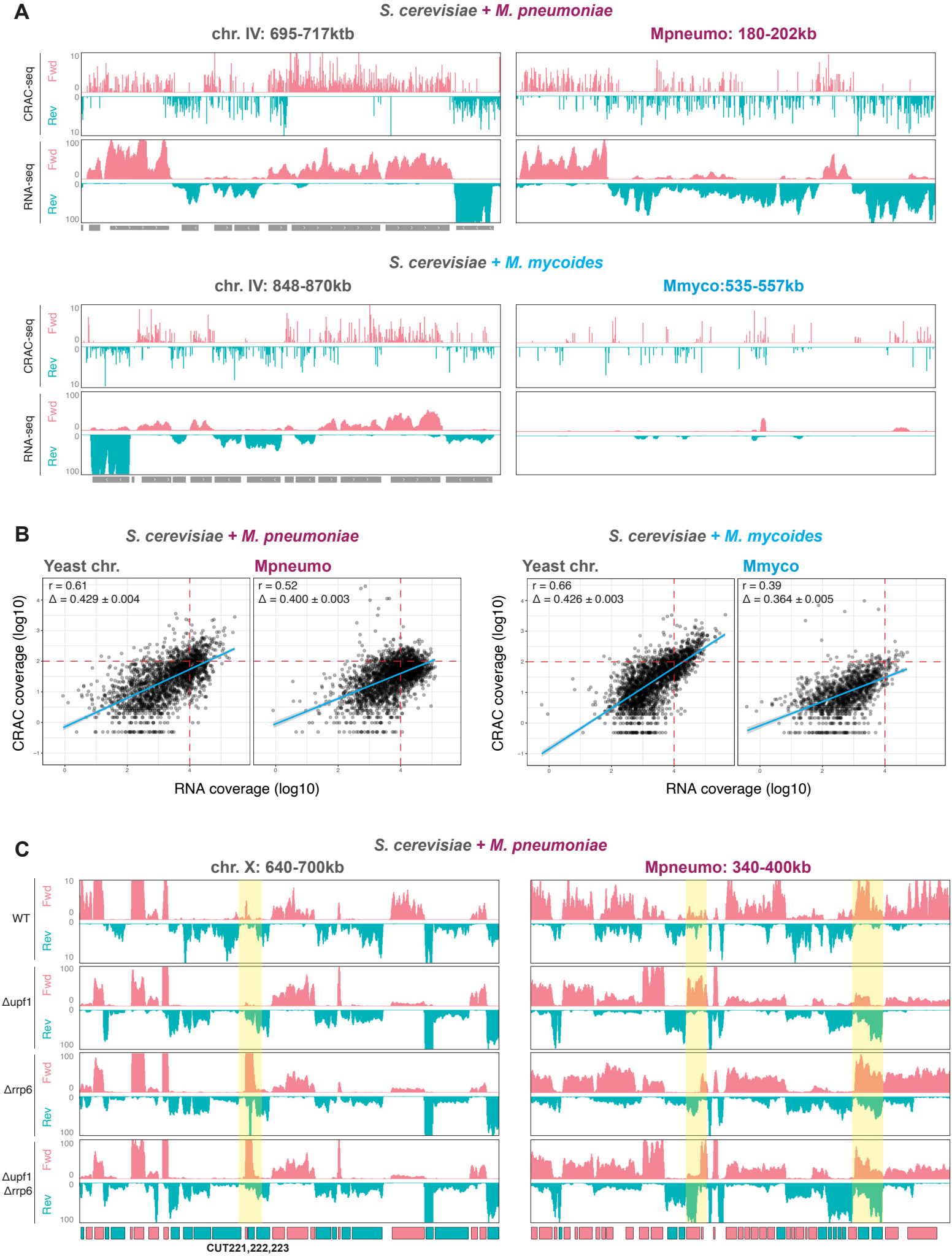

A
S. cerevisiae + M. pneumoniae
chr. IV: 695-717ktb
Mpneumo: 180-202kb
CRAC-seq
RNA-seq
Fwd
Rev
S. cerevisiae + M. mycoides
chr. IV: 848-870kb
Mmyco:535-557kb
B
S. cerevisiae + M. pneumoniae
Yeast chr.
Mpneumo
r = 0.61
Δ = 0.429 ± 0.004
r = 0.52
Δ = 0.400 ± 0.003
CRAC coverage (log10)
RNA coverage (log10)
S. cerevisiae + M. mycoides
Yeast chr.
Mmyco
r = 0.66
Δ = 0.426 ± 0.003
r = 0.39
Δ = 0.364 ± 0.005
C
S. cerevisiae + M. pneumoniae
chr. X: 640-700kb
Mpneumo: 340-400kb
WT
Δupf1
Δrrp6
Δupf1 Δrrp6
CUT221,222,223

Fig. S5

Fig. S5. Nascent transcription of bacterial chromosomes in yeast.

A. Stranded CRAC-seq and RNA-seq profiles along chr. IV, Mpneumo and Mmyco chromosomes (CPM). Forward transcription profiles are shown in pink, and reverse transcription profiles are shown in turquoise.

B, Relationship between CRAC and RNA coverage, in S. cerevisiae strains with M. pneumoniae (left) or with M. mycoides chromosome (right).

C, Stranded RNA-seq profiles of Mpneumo strains in WT or Δupf1, Δrrp6 and Δupf1/Δrrp6 mutants, along chr. X or Mpneumo chromosomes. Forward transcription profiles are shown in pink, and reverse transcription profiles are shown in turquoise.

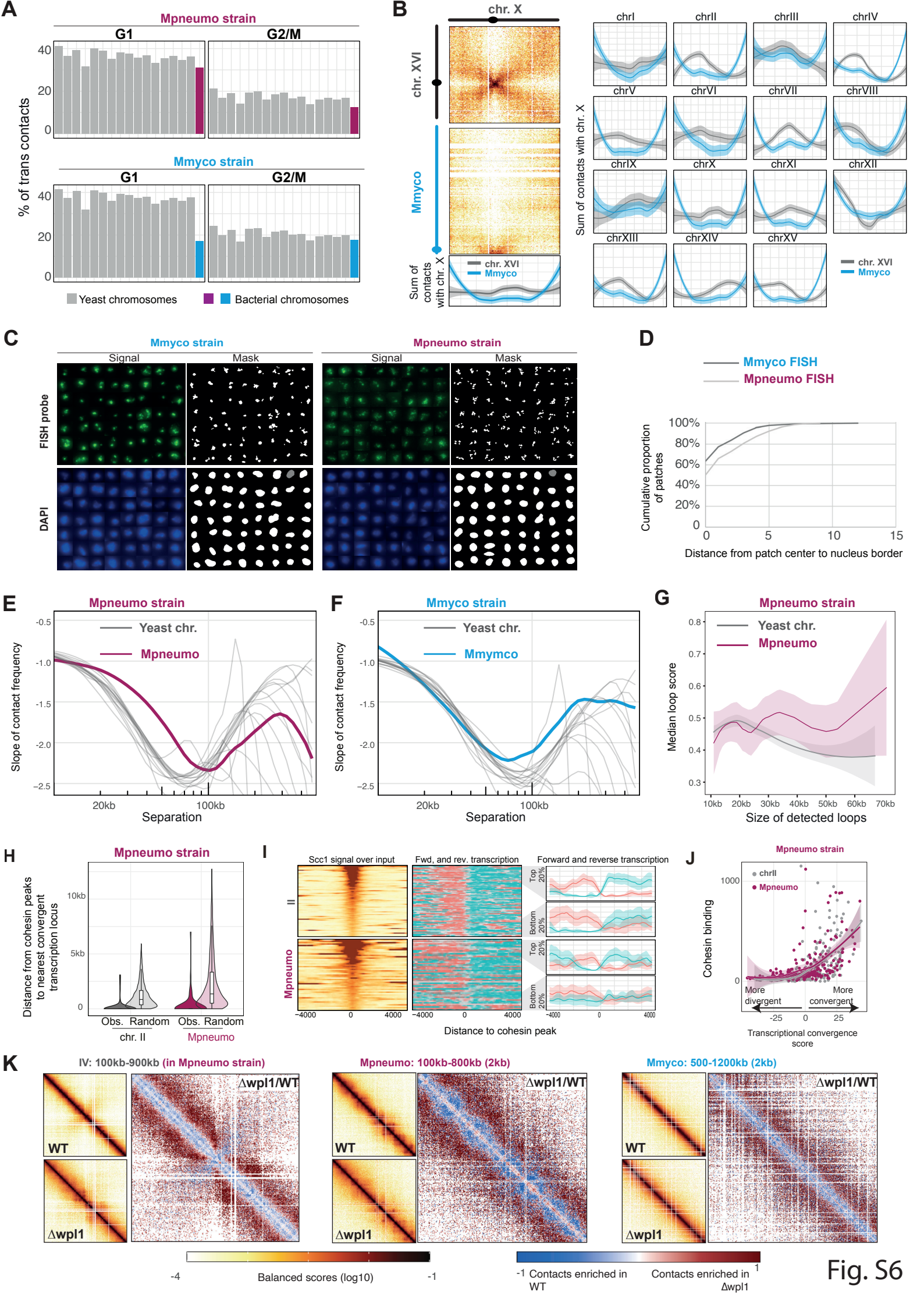

A
Mpneumo strain
G1
G2/M
% of trans contacts
Mmyco strain
Yeast chromosomes
Bacterial chromosomes
B
chr. X
chr. XVI
Mmyco
Sum of contacts with chr. X
chr. XVI
Mmyco
chrI
chrII
chrIII
chrIV
chrV
chrVI
chrVII
chrVIII
chrIX
chrX
chrXI
chrXII
chrXIII
chrXIV
chrXV
C
Mmyco strain
Mpneumo strain
Signal
Mask
FISH probe
DAPI
D
Mmyco FISH
Mpneumo FISH
Cumulative proportion of patches
Distance from patch center to nucleus border
E
Mpneumo strain
Yeast chr.
Mpneumo
Slope of contact frequency
Separation
F
Mmyco strain
Yeast chr.
Mmymco
Slope of contact frequency
Separation
G
Mpneumo strain
Yeast chr.
Mpneumo
Median loop score
Size of detected loops
H
Mpneumo strain
Distance from cohesin peaks to nearest convergent transcription locus
Obs.
Random
chr. II
Mpneumo
I
Scc1 signal over input
Fwd, and rev. transcription
Forward and reverse transcription
Mpneumo
Distance to cohesin peak
J
Mpneumo strain
chrII
Mpneumo
Cohesin binding
More divergent
More convergent
Transcriptional convergence score
K
IV: 100kb-900kb (in Mpneumo strain)
Mpneumo: 100kb-800kb (2kb)
Mmyco: 500-1200kb (2kb)
WT
Δwpl1
Δwpl1/WT
Balanced scores (log10)
Contacts enriched in WT
Contacts enriched in Δwpl1

Fig. S6

Fig. S6. Folding of exogenic bacterial sequences within the yeast nucleus.

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

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

C, 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).

D, 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.

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

G, Distance-dependent loop scores (computed using Chromosight (*84*)) for loops along either endogenous (grey) or Mpneumo (purple) chromosomes.

H, 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.

I, 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.

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

K, For yeast chromosome IV, Mpneumo and Mmyco: 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.

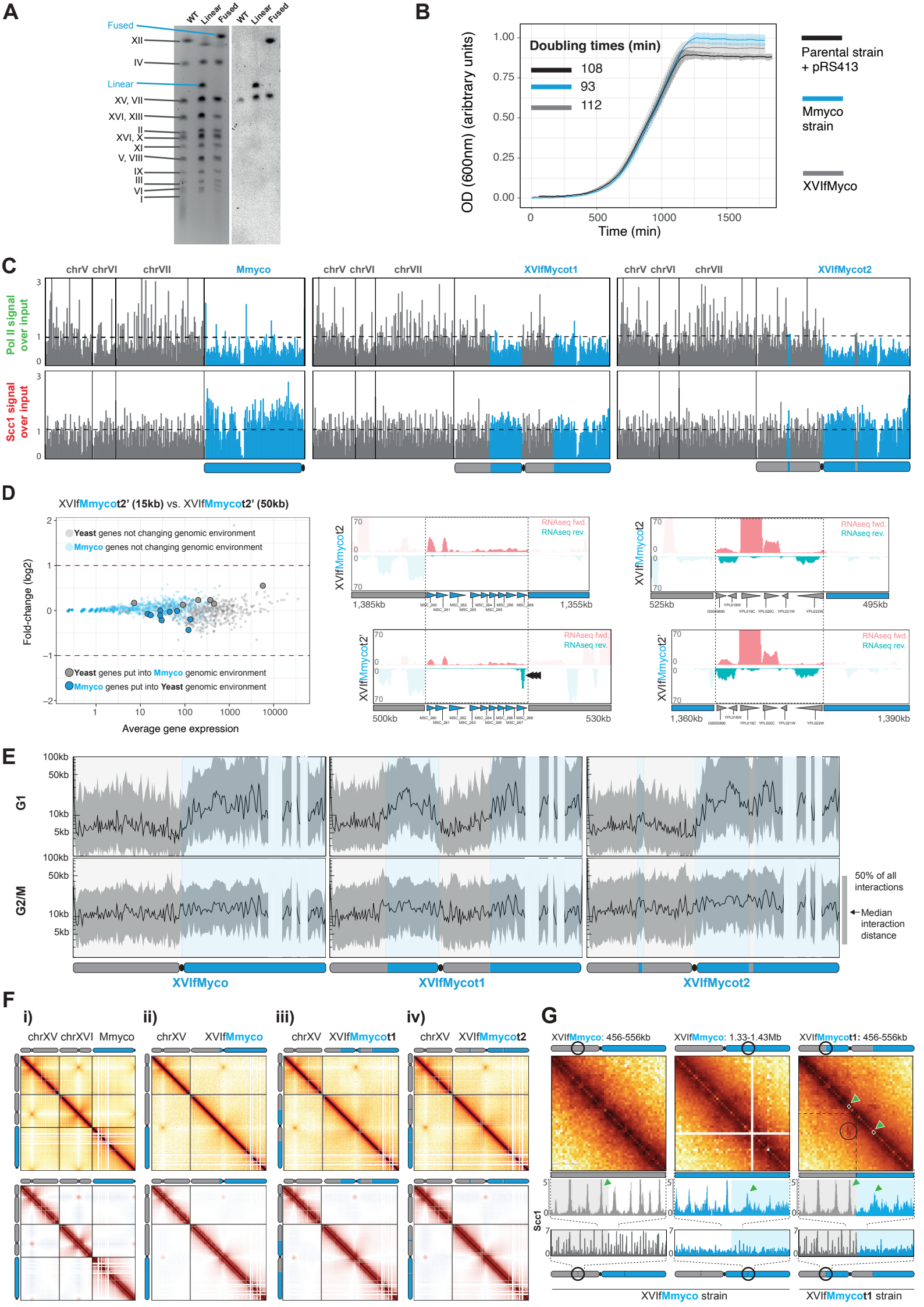

A
WT
Linear
Fused
WT
Linear
Fused
Fused
XII
IV
Linear
XV, VII
XVI, XIII
II
XVI, X
XI
V, VIII
IX
III
VI
I
B
Doubling times (min)
108
93
112
OD (600nm) (aribtrary units)
Time (min)
Parental strain + pRS413
Mmyco strain
XVIfMyco
C
chrV chrVI chrVII Mmyco
chrV chrVI chrVII XVIfMycot1
chrV chrVI chrVII XVIfMycot2
Pol II signal over input
Scc1 signal over input
D
XVIfMmycot2' (15kb) vs. XVIfMmycot2' (50kb)
Yeast genes not changing genomic environment
Mmyco genes not changing genomic environment
Yeast genes put into Mmyco genomic environment
Mmyco genes put into Yeast genomic environment
Fold-change (log2)
Average gene expression
XVIfMmycot2
XVIfMmycot2'
RNAseq fwd.
RNAseq rev.
1,385kb
1,355kb
525kb
495kb
500kb
530kb
1,360kb
1,390kb
E
G1
G2/M
100kb
50kb
10kb
5kb
50% of all interactions
Median interaction distance
XVIfMyco
XVIfMycot1
XVIfMycot2
F
i)
ii)
iii)
iv)
chrXV chrXVI Mmyco
chrXV XVIfMmyco
chrXV XVIfMmycot1
chrXV XVIfMmycot2
G
XVIfMmyco: 456-556kb
XVIfMmyco: 1.33-1.43Mb
XVIfMmycot1: 456-556kb
Scc1
XVIfMmyco strain
XVIfMmycot1 strain

Fig. S7. 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, Mmyco and XVIfMmyco strains.

C, Pol. II (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.

D, Expression fold-change (log2) against average expression, for genes located in yeast segments or Mmmcyo segments, between the XVIfMmycot2 and the XVIfMmycot2' strains. Genomic tracks illustrating stranded RNA-seq coverage over yeast (gray) and Mmmcyo (blue) genes, and their surrounding sequences (yeast: gray; Mmmcyo: blue).

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

F, Top: G2/M Hi-C contact maps of chr. XV, XVI, and bacterial chromosomes in the Mmyco strain, and in its derivatives (i.e. the chr. XVI and Mmyco fusion, and the two strains with translocations resulting in alternating Y and U chromatin segments; Methods). Bottom: correlation contact matrices in wt and mosaic chromosomes strains. The color scales are the same as in Fig. 4B.

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

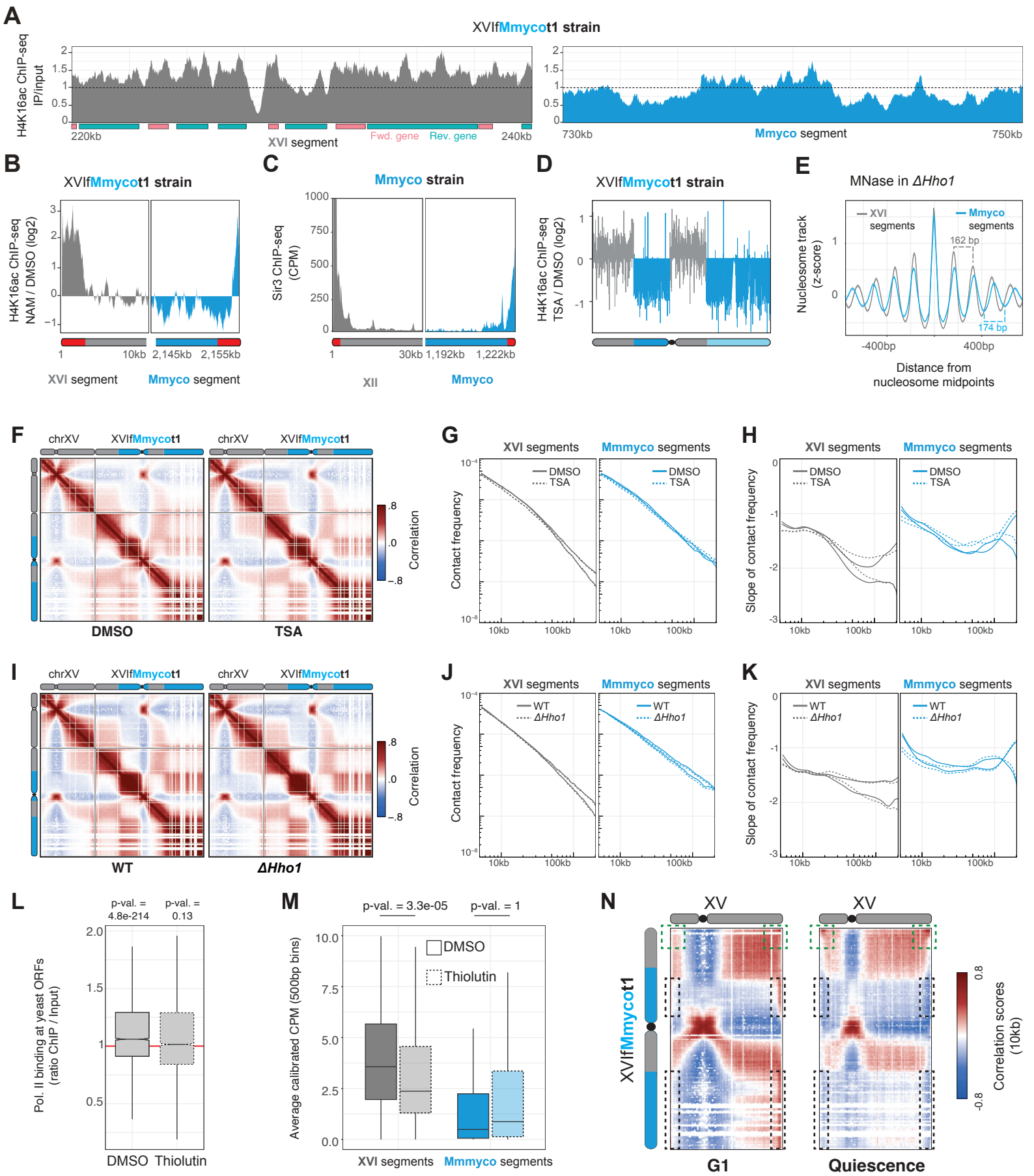

A
XVIfMmycot1 strain
H4K16ac ChIP-seq IP/input
220kb
XVI segment
Fwd. gene
Rev. gene
240kb
730kb
Mmyco segment
750kb
B
XVIfMmycot1 strain
H4K16ac ChIP-seq NAM / DMSO (log2)
1
10kb
2,145kb
2,155kb
XVI segment
Mmyco segment
C
Mmyco strain
Sir3 ChIP-seq (CPM)
1
30kb
1,192kb
1,222kb
XII
Mmyco
D
XVIfMmycot1 strain
H4K16ac ChIP-seq TSA / DMSO (log2)
E
MNase in ΔHho1
Nucleosome track (z-score)
XVI segments
Mmyco segments
162 bp
174 bp
-400bp
400bp
Distance from nucleosome midpoints
F
chrXV
XVIfMmycot1
chrXV
XVIfMmycot1
Correlation
DMSO
TSA
G
XVI segments
Mmmyco segments
DMSO
TSA
Contact frequency
10kb
100kb
H
XVI segments
Mmmyco segments
DMSO
TSA
Slope of contact frequency
I
chrXV
XVIfMmycot1
chrXV
XVIfMmycot1
Correlation
WT
ΔHho1
J
XVI segments
Mmmyco segments
WT
ΔHho1
Contact frequency
K
XVI segments
Mmmyco segments
WT
ΔHho1
Slope of contact frequency
L
p-val. = 4.8e-214
p-val. = 0.13
Pol. II binding at yeast ORFs (ratio ChIP / Input)
DMSO
Thiolutin
M
p-val. = 3.3e-05
p-val. = 1
DMSO
Thiolutin
Average calibrated CPM (500bp bins)
XVI segments
Mmmyco segments
N
XV
XV
XVIfMmycot1
Correlation scores (10kb)
G1
Quiescence

Fig. S8

Fig. S8. Functional investigation of heterochromatinization mechanisms of U-type compartment.

A, H4K16ac ChIP-seq profile (ratio IP/input) in the XVIfMmycot1 strain, over a yeast chromosome segment (left) or a Myco chromosome segment (right).

B, Comparison of H4K16ac ChIP-seq profiles in the XVIfMmycot1 strain, in cells treated by Nicotinamide (NAM) vs. DMSO (log2 scale). Telomeres and subtelomeric domains (2.5kb) are shown in red.

C, ChIP-seq profiles of Sir3 in the XVIfMmycot1 strain. Telomeres and subtelomeric domains (2.5kb) are shown in red.

D, Comparison of H4K16ac ChIP-seq profiles in the XVIfMmycot1 strain, in cells treated by Trichostatin A (TSA) vs. DMSO (log2 scale).

E, Nucleosome track over yeast (gray) or Mmyco nucleosomes (blue), in ΔHho1 yeast strain.

F, Correlation matrices of the contacts in chr. XV and XVIfMmycot1 in G1, after the addition of DMSO (left) or TSA (right).

G, Contact frequency (p) as a function of genomic distance (s), for contacts in yeast segments (gray) or Mmmyco segments (blue) of the chimeric chromosome XVIfMmycot1, in G1 after the addition of DMSO (solid) or TSA (dotted).

H, Derivatives of curves from **F**.

I, Correlation matrices of the contacts in chr. XV and XVIfMmycot1 in G1, after the addition of DMSO (left) or TSA (right).

J, Contact frequency (p) as a function of genomic distance (s), for contacts in yeast segments (gray) or Mmmyco segments (blue) of the chimeric chromosome XVIfMmycot1, in WT (solid) or ΔHho1 mutant (dotted).

K, Derivatives of curves from **I**.

L, Average Pol. II ChIP-seq coverage (ratio IP/input) at every yeast ORF, in DMSO (solid) or after thiolutin treatment (dotted). P-values from one-sample two-tailed Student t-test.

M, Average RNA-seq coverage over 500bp bins from XVI (gray) or Mmmcyo (blue) segments, in DMSO (solid) or after thiolutin treatment (dotted). P-values from one-tailed Student's t-test.

N, Correlation matrices of the trans-chromosomal contacts between chr. XV and XVIfMmycot1 in G1, in G1 (left) or quiescence (right).

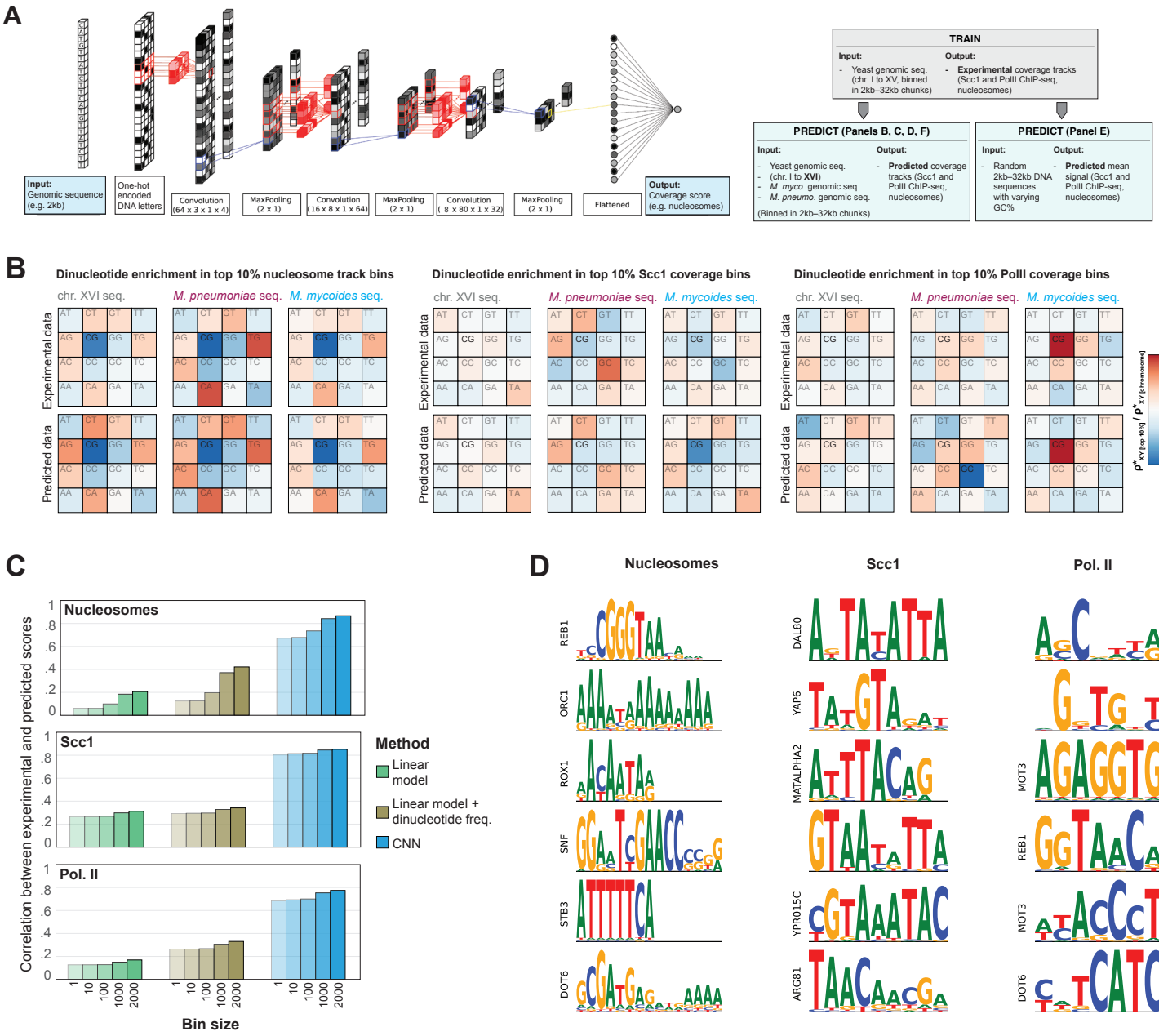

A
Input: Genomic sequence (e.g. 2kb)
One-hot encoded DNA letters
Convolution (64 x 3 x 1 x 4)
MaxPooling (2 x 1)
Convolution (16 x 8 x 1 x 64)
MaxPooling (2 x 1)
Convolution (8 x 80 x 1 x 32)
MaxPooling (2 x 1)
Flattened
Output: Coverage score (e.g. nucleosomes)
TRAIN
Input:
- Yeast genomic seq. (chr. I to XV, binned in 2kb–32kb chunks)
Output:
- Experimental coverage tracks (Scc1 and PolII ChIP-seq, nucleosomes)
PREDICT (Panels B, C, D, F)
Input:
- Yeast genomic seq. (chr. I to XVI)
- M. myco. genomic seq.
- M. pneumo. genomic seq.
(Binned in 2kb–32kb chunks)
Output:
- Predicted coverage tracks (Scc1 and PolII ChIP-seq, nucleosomes)
PREDICT (Panel E)
Input:
- Random 2kb–32kb DNA sequences with varying GC%
Output:
- Predicted mean signal (Scc1 and PolII ChIP-seq, nucleosomes)
B
Dinucleotide enrichment in top 10% nucleosome track bins
Dinucleotide enrichment in top 10% Scc1 coverage bins
Dinucleotide enrichment in top 10% PolII coverage bins
chr. XVI seq.
M. pneumoniae seq.
M. mycoides seq.
Experimental data
Predicted data
+15%
0%
-15%
C
Nucleosomes
Scc1
Pol. II
Correlation between experimental and predicted scores
Bin size
Method
Linear model
Linear model + dinucleotide freq.
CNN
D
Nucleosomes
Scc1
Pol. II
REB1
ORC1
ROX1
SNF
STB3
DOT6
DAL80
YAP6
MATALPHA2
YPR015C
ARG81
MOT3
REB1
MOT3
DOT6

Fig. S9

Fig. S9. Sequence-based prediction of chromatin composition of exogenous bacterial chromosomes in yeast.

A, Left: Schematic representation of the convolution neural network used in this study. The input of the network is a 2 to 30kb DNA sequence. The output is the corresponding value of the nucleosome, Scc1 or Pol. II signal. Details about the size of the input/output and the architecture used are discussed in the Material and Methods. Right: Overall training/prediction strategy. We trained a CNN on sequences of chr. I-XV to predict (i) genome-wide nucleosome, Scc1 or Pol. II ChIP-seq coverage tracks over chrXVI and bacterial chromosomes, and (ii) the average nucleosome, Scc1 or Pol. II ChIP-seq signal over 10kb random sequences with varying GC%.
B, Dinucleotide enrichment in genomic loci with 10% greatest nucleosome (left), Scc1 (middle) or Pol. II (right) ChIP-seq coverage (100 bp bins), extracted from experimental or predicted tracks over chromosome XVI, Mpneumo or Mmyco. The dinucleotide composition in these loci is compared to the chromosome-wide dinucleotide composition ($\rho^*(XY)_{[chromosome]}$ / $\rho^*(XY)_{[chromosome]}$).
C, Correlation scores between experimental and predicted nucleosome, Scc1 or Pol. II ChIP-seq data, using a linear model accounting for GC% only (green) or dinucleotide composition (kaki), or using CNNs (blue).
D, Motifs identified de novo as relevant for predicting nucleosome, Scc1 or Pol. II profiles. Matches to similar motifs in yeast databases are shown on the left.

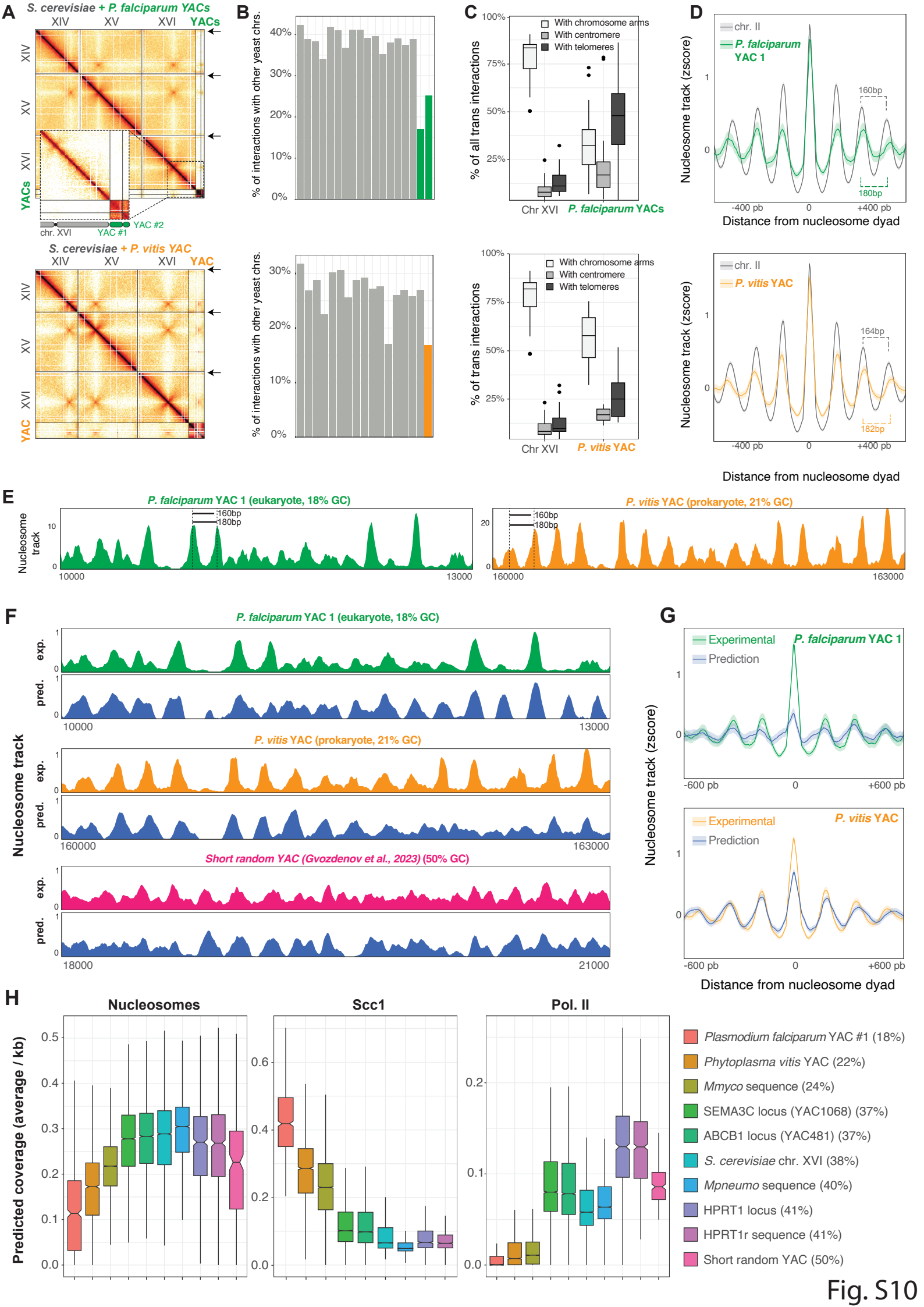

A
S. cerevisiae + P. falciparum YACs
XIV
XV
XVI
YACs
chr. XVI
YAC #1
YAC #2
S. cerevisiae + P. vitis YAC
YAC
B
% of interactions with other yeast chrs.
C
% of all trans interactions
With chromosome arms
With centromere
With telomeres
Chr XVI
P. falciparum YACs
P. vitis YAC
D
chr. II
P. falciparum YAC 1
P. vitis YAC
Nucleosome track (zscore)
160bp
164bp
180bp
182bp
-400 pb
+400 pb
Distance from nucleosome dyad
E
P. falciparum YAC 1 (eukaryote, 18% GC)
P. vitis YAC (prokaryote, 21% GC)
Nucleosome track
F
exp.
pred.
Short random YAC (Gvozdenov et al., 2023) (50% GC)
G
Experimental
Prediction
H
Nucleosomes
Scc1
Pol. II
Predicted coverage (average / kb)
Plasmodium falciparum YAC #1 (18%)
Phytoplasma vitis YAC (22%)
Mmyco sequence (24%)
SEMA3C locus (YAC1068) (37%)
ABCB1 locus (YAC481) (37%)
S. cerevisiae chr. XVI (38%)
Mpneumo sequence (40%)
HPRT1 locus (41%)
HPRT1r sequence (41%)
Short random YAC (50%)

Fig. S10

Fig. S10. U-type chromatin compartmentalization of AT-rich prokaryotic and eukaryotic YACs.

A, Hi-C maps showing chromosomal interactions between S. cerevisiae chromosomes (XIV, XV, XVI) and artificial yeast chromosomes (YACs) from P. falciparum (top) and P. vitis (bottom). The inset for P. falciparum highlights the increased trans-chromosomal contacts between the two P. falciparum YACs in comparison with the decreased trans-chromosomal contacts between each of the YACs and the yeast chromosome XVI.

B, Percentage of interactions that each YAC has with yeast chromosomes.

C, Percentage of trans interactions between YACs and specific chromosomal regions (chromosome arms, centromere and telomeres), for P. falciparum (top) and P. vitis (bottom).

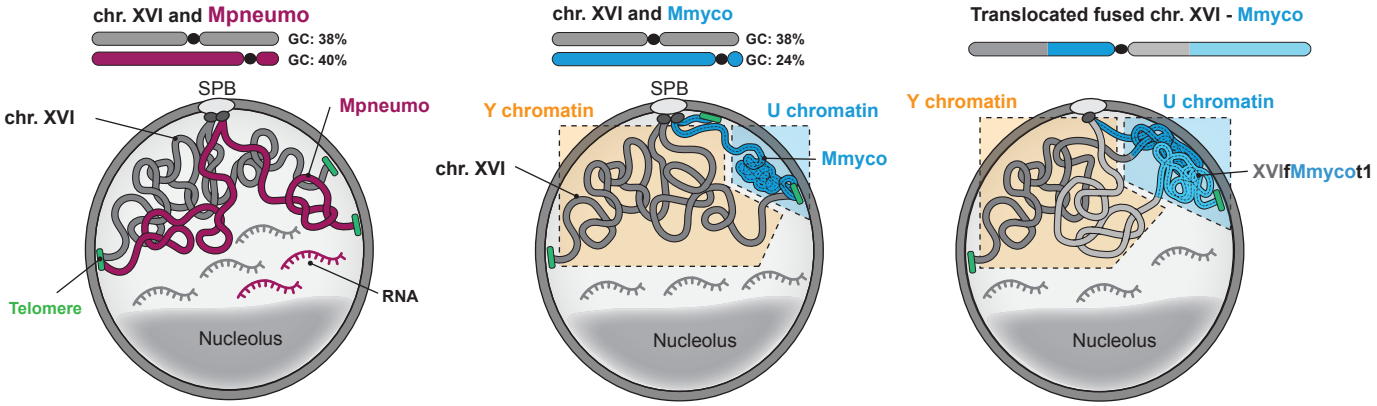
D, Average nucleosome signal centered at nucleosomes annotated over P. falciparum YAC (top, green) and P. vitis YAC (bottom, orange), compared to yeast nucleosomes (gray).

E, Nucleosome track over 3kb-wide segments of the P. falciparum YAC (left, green) or P. vitis YAC (right, orange).

F, Comparison of experimental (exp.) and predicted (pred.) nucleosome tracks for P. falciparum YAC (green), P. vitis YAC (orange), and a short random YAC (pink)(45).

G, Comparison of experimental (green or orange) and predicted nucleosome signals (dark blue), centered at nucleosomes annotated over P. falciparum YAC (top) and P. vitis YAC (bottom).

H, Predicted nucleosome, Scc1 and Pol. II coverages in chromosome sequences from different genomes. Coverage values are averaged using 1kb bins tiling each chromosome sequence.


chr. XVI and Mpneumo
GC: 38%
GC: 40%
SPB
chr. XVI
Mpneumo
Telomere
RNA
Nucleolus
chr. XVI and Mmyco
GC: 38%
GC: 24%
SPB
Y chromatin
U chromatin
chr. XVI
Mmyco
Nucleolus
Translocated fused chr. XVI - Mmyco
Y chromatin
U chromatin
XVIfMmycot1
Nucleolus

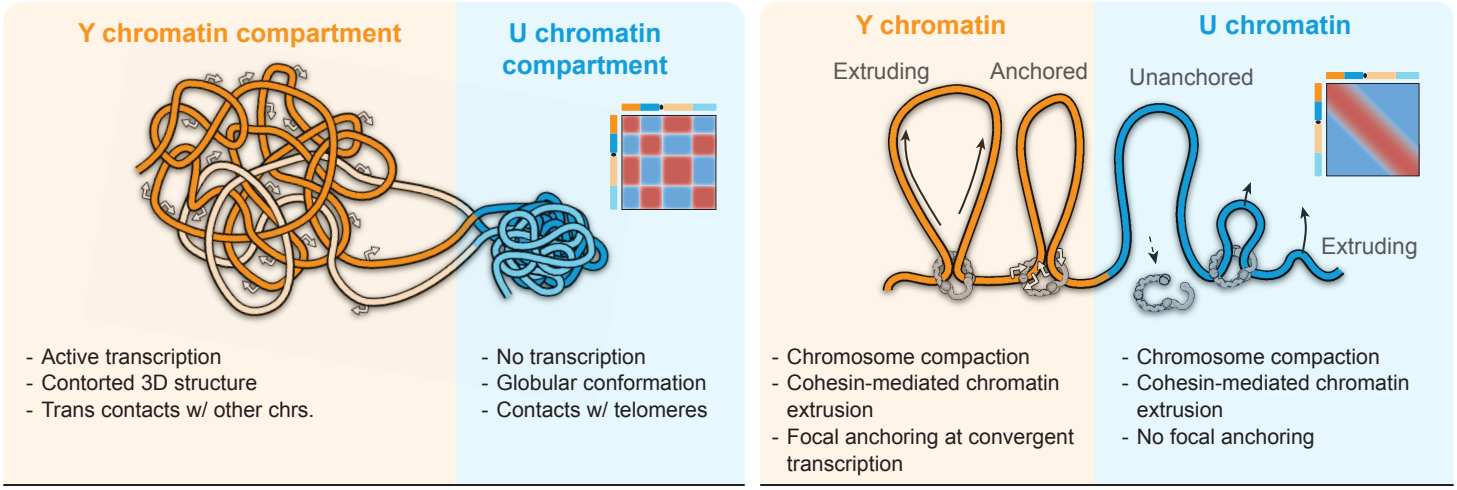

Y chromatin compartment
U chromatin compartment
- Active transcription
- Contorted 3D structure
- Trans contacts w/ other chrs.
- No transcription
- Globular conformation
- Contacts w/ telomeres
G1
Y chromatin
U chromatin
Extruding
Anchored
Unanchored
Extruding
- Chromosome compaction
- Cohesin-mediated chromatin extrusion
- Focal anchoring at convergent transcription
- Chromosome compaction
- Cohesin-mediated chromatin extrusion
- No focal anchoring
G2/M

Fig. S11

Fig. S11. Behavior and activity of bacterial genomes integrated in the yeast genome.

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

Table S1. Summary table of strains

	Strains name	Description	Parental strain	Genetic background	MAT	Genotype	Reference
Mmmyco	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	(*16*)
	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, 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, ΔWAPL	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_Y001375	"Translocation n°1" strain, Δhho1	RSG_Y1136	W303	a	ade2-1, leu2-3,112, his3-11,15, trp1-1, can1-100, RAD5, Δhho1, 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
	RSG_Y001237	XVIfMmycot2' : "Translocation n°2" strain (15 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_Y001268	XVIfMmycot2' : "Translocation n°2" strain (15 kb block) strain with SCC1-PK9	RSG_Y001237	W303	a	ade2-1, leu2-3,112, his3-11,15, trp1-1, can1-100, RAD5, SCC1-pk9::KanMX, XVIfMmycot2'(CEN6/ARS-HIS3)	this study
	RSG_Y1274	XVIfMmycot3 : "Triple Translocation n" strain	RSG_Y1136	W303	a	ade2-1, leu2-3,112, his3-11,15, trp1-1, can1-100, RAD5, XVIfMmycot3(CEN6/ARS-HIS3)	this study
	yLD126-36c	Control split dot assay	NA	W303	a	Met3p-CDC20, tetR-GFP, ura3-1::URA3tetO	(*94*)
	FB176	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	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	(*17*)
	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_Y1037	XVIfMpneumo : "Fused "strain, Mpneumo chromosome merged with the XVI	RSG_Y960	VL6-48N	a	trp1-Δ1, ura3-Δ1, ade2-101, his3-Δ200, lys2, met14 cir° + XVIfMpneumo(CEN6/ARS-HIS3)	this study
	RSG_Y001349	XVIfMpneumo, RNH1::HTP	RSG_Y1037	VL6-48N	a	trp1-Δ1, ura3-Δ1, ade2-101, his3-Δ200, lys2, met14 cir°, RNH1::HTP + XVIfMpneumo(CEN6/ARS-HIS3)	this study
	RSG_Y001273	XVIfMpneumo ; Δupf1	RSG_Y1037	VL6-48N	a	trp1-Δ1, ura3-Δ1, ade2-101, his3-Δ200, lys2, met14 cir° Δupf1::KanMX + XVIfMpneumo(CEN6/ARS-HIS3),	this study
	RSG_Y001308	XVIfMpneumo ; Δrrp6	RSG_Y1037	VL6-48N	a	trp1-Δ1, ura3-Δ1, ade2-101, his3-Δ200, lys2, met14 cir°, Δrrp6 + XVIfMpneumo(CEN6/ARS-HIS3),	this study
	RSG_Y001298	XVIfMpneumo ; Δupf1 and Δrrp6	RSG_Y001273	VL6-48N	a	trp1-Δ1, ura3-Δ1, ade2-101, his3-Δ200, lys2, met14 cir°, Δupf1::KanMX, Δrrp6 + XVIfMpneumo(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	(*69*)
	W303 strain + scc1-pk	W303 strain + scc1-pk	W303	W303	a	Scc1-pk9::KanMX	(*69*)
YACs	RSG_Y001392	W303a + FDp92, regions of ~300kb - 22%GC	W303	W303	a	leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15	this study
	RSG_Y001388	W303a + 2 regions of *Plasmodium falciparum*	AB1380	AB1380	a	tiV ura3 trp1 ade2-1canl-100lys2-1his5	*(93)*

Table S2. gRNA sequence (CRISPR targeting).

gRNA sequences					
	Name	**Transformation expected**	**Target**	**gRNA 5' 3'**	**Plasmid used**
Mmyco strains	RSG_B700	Chromosome linearisation	DSB on the circular Mmyco (@ 739,680 bp)	TCTGCTAATCCTGTTACCAGTGG	pML107
	RSG_B1158	Fusion between the end of the chromosome XVI to the chromosome Mmyco (generation of XVIfMmyco strain)	DSB next to the centromere of chr XVI (@ 555,995 bp)	TTAGAATTACGACAACATAACGG	pAEF5
			DSB next to the right telomere of chrXVI (@ 936,139 bp)	TGGTGTTATATAGTGGCACCGGG	
	RSG_B1150		DSB next to the right telomere of Mmyco (@ 1,219,826 bp)	GCGTGGACAAAGGTACAACGAGG	pGZ110
	RSG_B1200	Translocation 1 (generation of XVIfMmycot1 strain)	DSB in the middle of the left arm of XVIfMmyco (XVI DNA, @ 506,748 bp)	GAGAGCACAGGTGTACTGGAGGG	pAEF5
			DSB in the middle of the right arm of XVIfMmyco (Mmyco DNA, @1,379,095 bp)	ACTTATTAGAAAATGAGCAGTGG	
	RSG_B1201	Translocation 2 50 kb (generation of XVIfMmycot2 strain)	DSB ~ 50kb from translocation 1 breakpoint on left arm of XVIfMmycot1 (in XVI sequence; @ 1,691,364 bp)	CTAATCATGTCTCACTCCCGCGG	pAEF5
			DSB ~ 50kb from translocation 1 breakpoint on right arm of XVIfMmycot1 (in Mmyco sequence, @ 737,819 bp)	CGTTCATGTTATGTGAAAAGAGG	
Mpneumo strains	**Name**	**Transformation expected**	**Target**	**gRNA 5' 3'**	**Plasmid used**
		Chromosome linearisation	DSB on the circular Mpneumo (@ xxx,xxx bp)	GTCAACGGTCAAAAAAACCGAGG	pAEF5

Table S3. DNA donor sequences.

DNA donor				
	Transformation expected	**Purpose**	**Sequence length**	**Sequence 5' 3'**
Mmyco strains	Linearisation	Myco Telomeres repeats TGG - left part	450	ggcgcgccCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAA GGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTG TAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACggattgctactgaccgctctGTGGTGTGGGTGTGGTGTGGGTGTGGTGTGGGT GTGGTGTGTGTGGGTGTGGGGTGTGGGTGTGGTGTGGGTGTGGTGTGGTGTGGGTGTGGTGTGTGGGTGTGGTGTGGGTGTGG TGTGGGTGTGGGTGTGGGTGGTGTGGGTGTGTGGGGGTGTGGTGTGTGGGTGTGGTGTGGGTGTGGTGTGGGTGTGGGTGTGG GTGGTGTGGGTGTGTGGcgcgcc
		Myco Telomeres repeats TGG - right part	450	ggcgcgccCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGC TCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCG GTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGgcctttagtacctcgcaacgGTGGTGTGGGTGTGGTGTGGGTGTGGTGTGGGTG TGGTGTGTGTGGGTGTGGGGTGTGGGTGTGGTGTGGGTGTGGTGTGTGGGTGTGGTGTGGGTGTGGT GTGGGTGTGGGTGTGGGTGGTGTGGGTGTGTGGGGGTGTGGTGTGTGGGTGTGGTGTGGGTGTGGGTGTGGG TGGTGTGGGTGTGTGGcgcgcc
	XVIfMyco	Loss of the centromere XVI	420	ttcatttatttTAGTCAAAGATGATTATAGGCGTATTATAGAGTCGATAAATAAATCTCAAGGTTAACTGtttattaaatatataataacatatttatttgtttaA CAAGCCTGAAATAAGAAATTCTTTTATAAATACATCATCGAGATGATCATCAAAGTGTCTCTATAGTTAATATAAGCTTATGCA AAGGTTGAAGgattgctactgaccgctctctatGAGCAAACAATTTgaacagaaaaataaaattggGGAAGTGACACACCATGGTAGCGGTTCTAAAGC GAAATCGGCAAAGCGGCTAAATAGCAGTTTTGATGACTTACTCCACACTGAAAATGGATGACCTTAAATAGGAGATAAAGCTT TTTCATCCCTATGTATTTAAGATGACTGGCTTGTCAAGCATTCTaat
		Fusion between Mmyco tel and XVI tel	420	TATATATTCAATGTAATTTTAATTACATTTGCTTTTAATAAAAACACTACTTAATAGAGAAAGGAAATATAAGATCCTTTGAAT GGAGGAAAATCACATGAAAATTATTAATATTGGAGTTTTAGCTCATGTTGATGCAGGAAAAACTACCTTAACAGAAAGCTTAT TATATAACAGTGGAGCGATTACAGAATTAGGAAgcctttagtacctcgcaacgCTATGCTATACTATAATAGCAATACACATTAACGCAGA TGCTATAATATCTTTACTTCATCTCACCAACTCAACTGATAATCATACAGACAGTACTATACAATACAGTCATTTACTAAAACG AATGCCGCTCCATGATATCCCAAACTTATACTGGCACCATTTTGAAACATTATATGTCACCTCTTCATCCTAATGACT
	XVIfMycot1	Translocation n°1 - left part	412	GAATTCAGTCCCTCTTCCAATGTAACCAACAACCTGAAGTGTTCTCTGTTTGCCTTTATTTTGCGACTTTTCTTCATCATTGAAG TTAGAAAAAGCTTTACTCGCTTATCTCCTGGAGTAAGCTATAGCCACAGTCAATATCGCGTCTAATGAAAAGAAAAAGTCGGG ACGAGTAAACTTTTGTCTGCGTGGCGCATAGGAGAGGACAATCTGTAAAATAGGTTCAAAGTATACCTTTACAATTAGGGGTA TACTTTTTTGTTATATAATTAAATTTATGTTAATGTGTTTTTTAGAAAGAGGTTTATATTTAATGATAAAAAAAATTGGAATTTT GACTTCTGGAGGGGATGCTCCTGGAATGAATAATGCCATTGCTGGAGTTGTAAAAACAGCAGCTTCAAAAGAATTC
		Translocation n°1 - right part	412	GAATTCTGATTGGAGTCGTCGGgttcatcttcatttttaattttagaACTTTGAGATGCTTGAGTTGTCTGCCTCGTAAGTTCCTCCTGGAGCTCA TCATCAGAGTCGCCCTGATAAAATAACTGAGACATCTGGAAAGATATTTTCAACAGGATAGTgcaatattatttttacaCATTTAGCAAAT GCTCTACCAACGTCCTGAAGAAAATAAACAAGTAATTTAGACTAATAAAATATAAAATATTAGTATAGAAGAAAGGAGATAA CTAAATAACTGTTCAAATTTTATACTTACTTAAAAAAACAAAATGACTTACTTTTTTTATAGACTTAAATAATAGTAAACAAAT ATAAATTTTTTTTCAAGTAATTCTTATAAAATTTTGAATTTTTGAAAAAAGATGAATTC
	XVIfMycot2	Translocation n°2 "50 kb" - left part	413	GAATTCAAAGAACAACCGTATCAAGTGGCGAAGCAGCTGCTTATAAGTAAAGCCCAAGGCCCACCAACGCATACATGCAAGG CCTGTGCAAGCCGACTGCTGCTTGTTTCATCTTCCGCTCGCGGTGACCCACTCGGGGAATTTCTCACGCTTGCTGCTGCAGGTC GTTGGGCGTTGATTCCTGTTTTTTTCTTTCGTTTTTTCAATCTAAGTTATTTACAAGTAATGGTATTTTATATTCATTATTTCTATC TATTTTATATGCAATTGTAGCTGGGATGCTATGCGTTTTAATTGGATATCCGATTGCTTTAATGATGGCTCAAATGAAATCAAA AATTTTAGCTAGAAATATGTGAGTCATTGTAACAATGCCTATGTGAATTTCTATGCTTTTAAAAGTATTAGAATTC
		Translocation n°2 "50 kb" - right part	412	GAATTCTAAAGTACCTTGAGACATGGCGTGCAAAAACTGTATCTGATCTTGAGTTTCTATGTTCTGTATGCTCTAATAACAGAA AATACAATAACAAGGAGGGAAGACCAATTTGCGGGCTGCCAAGCAACACAACCAAACCCTGCTGATAAAcgaaaacaagaaaactaataa aaaaataaataaggAACTCTACAAGGACTACATTGGCTGGTTTTCACTTAATTGAAACTTGTTGATCTACATCATATTCATTGATTGTAT

				GAATAATTCATTGTCTTTTTTTACTAGTTTCAACAACTATTTCTCAGTGTATTCCTTTAAAAGTTGTTTTAATAACTTTTGCATTA AAAAATCCATTATTTGGGTTTTTAATAATAATATCTTCTGGTCTAATGATAATATCGAATTC
Mpneumo strains	**Transformation expected**	**Purpose**	**Length**	**Sequence 5' 3'**
	Linearisation	Pneumo Telomeres repeats TGG	409	ggcgcgccAGTACACCAAGTGTCAGTTGTTCACCATTAATAAGCAACTAAGTAAGACAAAAAGCCCCAGAAAGACTATCCCACCC CAAAGGTTGTACTTTTGGAGATAGTGCGTTTTATTAGGATTGTCCAACTGGTAACCCTTGGCTTGTAGCGCAAACGCGACTGgga ttgctactgaccgctctGTGGTGTGGGTGTGGTGTGGGTGTGGTGTGGGTGTGGTGTGTGTGGGTGTGGGGTGTGGGTGTGGTGTGGGTG TGGTGTGGTGTGGGTGTGGTGTGTGGGTGTGGTGTGGGTGTGGTGTGGGTGTGGGTGTGGGTGGTGTGGGTGTGTGGGGGTGT GGTGTGTGGGTGTGGTGTGGGTGTGGTGTGGGTGTGGGTGTGGGTGGTGTGGGTGTGTGGcgcgcc
		Pneumo Telomeres repeats TGG	409	ggcgcgccTTAAGCATCTTATTTGCACTGGTGTTTAAACTGTTACCCATTGTTAAGGGTGAATTAAAACGGATTAAACAAGCGCAA GCAATCCGTGGTTTTAAGTACGGTAAGCTAGCCTTTTTAAATCCAGTCAAACTCAAAACGCTCTTTATTCCGGTGTTATTGgcctttt agtacctcgcaacgGTGGTGTGGGTGTGGTGTGGGTGTGGTGTGGGTGTGGTGTGTGTGGGTGTGGGGTGTGGGTGTGGTGTGGGTGT GGTGTGGTGTGGGTGTGGTGTGTGGGTGTGGTGTGGGTGTGGTGTGGGTGTGGGTGTGGGTGGTGTGGGTGTGTGGGGGTGTG GTGTGTGGGTGTGGTGTGGGTGTGGTGTGGGTGTGGGTGTGGGTGGTGTGGGTGTGTGGcgcgcc

Table S4. Primer sequences.

Checking primers					
	Transformation expected	**Purpose**	**Primers sequence forward 5' 3'**	**Primers sequence reverse 5' 3'**	**Amplicons Length**
Mmyco	Linearisation	Circular loss	TTCCACTGAGCGTCAGACC	TGCCGCTTACCGGATACC	-
		Check telomere addition R	CGTTGCGAGGTACTAAAGGC	TGCCGCTTACCGGATACC	242
		Check telomere addition L	TTCCACTGAGCGTCAGACC	agagcggtcagtagcaatcc	294
	XVIfMyco	Check loss centromere XVI	GCCCTAGATCAAACTGATCCAG	CCGGTAGAAGCCTTTGTACC	555
		Check fusion	CCCGCATAAGTACGTGTAGCT	CAGTGAAGCACCAGTTTCTG	688
	XVIfMycot1	Check translocation R	CGATGCATACGTTCCATC	CTGGTAATCTTGCAGATCC	729
		Check translocation L	GAGGGTATAACACTGCATCG	CCATATCCAGCCATATTTGC	1127
	XVIfMycot2	Check translocation R	CTTGTCGTCCTGTGAAAG	CCAGAAAATGCTTGAGTAGC	760
		Check translocation L	GCAGTCGTTCTCAAACGG	GCATAGAAATTCACATAGGC	903
	Transformation expected	**Purpose**	**Primers sequence forward 5' 3'**	**Primers sequence reverse 5' 3'**	**Amplicons Length**
Mpneumo	Linearisation	Circular loss	GGTGGGTAAAGGTAAAGC	cgaatTAGCATGCTGTTTGC	-
		Check telomere addition R	ggattgctactgaccgctct	GGTGGGTAAAGGTAAAGC	215
		Check telomere addition L	gcctttagtacctcgcaacg	cgaatTAGCATGCTGTTTGC	300

Table S5. Hi-C statistics.

	HiC library	background	strain	stage	Total sequenced fragments	Aligned fragments	Fragments retained in map
CH196	Hi-C of Mmyco strain (G2/M)	W303_Mmmyco	RSG_Y712	G2/M	302589048	261015801	66875527
CH02	Hi-C of Mmyco strain (G1)	W303_Mmmyco	RSG_Y712	G1	109056382	81014812	20087984
CH198	Hi-C of Mmyco strain (Wpl1 deletion, G2/M)	W303_Mmmyco	RSG_Y955	G2/M	41113718	35092403	12373773
CH195	Hi-C of Mpneumo strain (G2/M)	S288C_Mpneumo	RSG_Y960	G2/M	76019938	65125087	19632515
CH210	Hi-C of Mpneumo strain (G1)	S288C_Mpneumo	RSG_Y960	G1	71379048	60466145	12456326
CH197	Hi-C of Mpneumo strain (Wpl1 deletion, G2/M)	S288C_Mpneumo	RSG_Y954	G2/M	39684936	35230228	11676521
LM34	Hi-C of XVIfMmyco strain (G2/M)	XVIfMmyco	RSG_Y1033	G2/M	56889456	42820436	9754819
LM33	Hi-C of XVIfMmyco strain (G1)	XVIfMmyco	RSG_Y1033	G1	43712974	32935796	6683250
LM58	Hi-C of XVIfMmycot1 strain (G2/M)	XVIfMmycot1	RSG_Y1136	G2/M	53729240	42557126	13123109
LM63	Hi-C of XVIfMmycot1 strain (G1)	XVIfMmycot1	RSG_Y1136	G1	43841832	36113087	9330391
LM59	Hi-C of XVIfMmycot2 strain (G2/M)	XVIfMmycot2	RSG_Y1137	G2/M	49569888	40547069	13600786
LM62	Hi-C of XVIfMmycot2 strain (G1)	XVIfMmycot2	RSG_Y1137	G1	50005472	41366045	12902417
LM119	Hi-C of XVIfMmycot3 strain (G2/M)	XVIfMmycot3	RSG_Y1235	G2/M			
LM122	Hi-C of XVIfMmycot3 strain (G1)	XVIfMmycot3	RSG_Y1235	G1			