



HAL
open science

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

Christophe Chapard, Léa Meneu, Jacques Serizay, Alex Westbrook, Etienne Routhier, Myriam Ruault, Amaury Bignaud, Agnès Thierry, Géraldine Gourgues, Carole Lartigue, et al.

► To cite this version:

Christophe Chapard, Léa Meneu, Jacques Serizay, Alex Westbrook, Etienne Routhier, et al.. Exogenous chromosomes reveal how sequence composition drives chromatin assembly, activity, folding and compartmentalization. 2023. hal-04278801

HAL Id: hal-04278801

<https://hal.science/hal-04278801v1>

Preprint submitted on 10 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

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

Christophe Chopard^{1,Ψ}, Léa Meneu^{1,2,Ψ}, Jacques Serizay^{1,Ψ,*}, Alex Westbrook^{2,3}, Etienne Routhier^{2,3,4}, Myriam Ruault⁵, Amaury Bignaud^{1,2}, Agnès Thierry¹, Géraldine Gourgues⁶, Carole Lartigue⁶, Aurèle Piazza^{1,#}, Angela Taddei⁵, Frédéric Beckouët⁷, Julien Mozziconacci^{3,4,*} and Romain Koszul^{1,*}

Affiliation

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

² Sorbonne Université, Collège Doctoral

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

⁴ Laboratoire de Physique Théorique des Liquides, 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

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

Ψ These authors contributed equally

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

1 Abstract

2 Genomic sequences co-evolve with DNA-associated proteins to ensure the multiscale folding of
3 long DNA molecules into functional chromosomes. In eukaryotes, different molecular complexes
4 organize the chromosome's hierarchical structure, ranging from nucleosomes and cohesin-
5 mediated DNA loops to large scale chromatin compartments. To explore the relationships
6 between the DNA sequence composition and the spontaneous loading and activity of these
7 DNA-associated complexes in the absence of co-evolution, we characterized chromatin
8 assembly and activity in yeast strains carrying exogenous bacterial chromosomes that diverged
9 from eukaryotic sequences over 1.5 billion years ago. We show that nucleosome assembly,
10 transcriptional activity, cohesin-mediated looping, and chromatin compartmentalization can
11 occur in a bacterial chromosome with a largely divergent sequence integrated in a eukaryotic
12 host, and that the chromatinization of bacterial chromosomes is highly correlated with their
13 sequence composition. These results are a step forward in understanding how foreign
14 sequences are interpreted by a host nuclear machinery during natural horizontal gene transfers,
15 as well as in synthetic genomics projects.

16 **Introduction**

17 Genome sequence composition, broadly defined by its GC%, polynucleotide frequencies,
18 DNA motifs and repeats, varies widely between species as well as within individual genomes^{1,2}.
19 In eukaryotes, the sequence composition is known to correlate with: 1) chromatin composition,
20 which includes nucleosome formation and binding of structural and functional proteins to DNA³,
21 2) chromatin activity, such as transcription and replication⁴, and 3) functional 3D organization of
22 the genome into loops and compartments^{5,6}. For instance in mammals GC-rich regions are
23 enriched in actively transcribed sequences and in chromatin loops mediated by the structural
24 maintenance of chromosomes (SMC) cohesin, and genomic loci with different sequence
25 composition tend to coalesce in distinct compartments⁷. These relationships between sequence
26 and chromatin composition activities and folding reflect their continuous coevolution over
27 millions of years.

28 Disruptive variations in sequence composition can also emerge naturally during evolution,
29 e.g. through transfer of genetic material across species by horizontal gene transfers or
30 introgression, in viral infections⁸⁻¹¹, or even artificially, e.g. by introducing chromosome-long
31 DNA molecules into chassis microbial strains or cell lines¹²⁻¹⁴. Such transfer can lead to the
32 long-term integration of foreign DNA whose sequence composition strongly diverges from their
33 host's genome (e.g. the introgression of a 12% GC-divergent 1Mb sequence in *Lachancea*
34 *kluyveri*¹⁵). Once integrated, these sequences are organized and processed by chromatin-
35 associated proteins of the host genome, obeying new rules under which they have not
36 coevolved. How a eukaryotic host can successfully package, fold and regulate the activity of
37 chromosome-long exogenous bacterial DNA sequences, and the importance of the sequence
38 composition, remain largely unknown. Addressing this question would improve our
39 understanding of the evolutionary processes and sequence determinants involved in chromatin
40 folding and activity.

41 Here we explored whether and how foreign bacterial full-length chromosomes acquire
42 chromatin features when integrated in a eukaryotic nucleus, and the relationship between their
43 sequence composition and their chromatinization. To this end, we investigated the behavior of
44 two individual bacterial chromosomes with different sequence composition, artificially introduced
45 into the *S. cerevisiae* genome^{16,17}. We profiled nucleosome assembly, polymerase and cohesin
46 binding and activities, as well as the 3D chromosome organization during the cell cycle. We
47 show that bacterial chromosomes with different sequence composition present different profiles
48 of chromatin composition and activities, eventually leading to the spontaneous formation of
49 active / inactive chromosomal compartments, similar to the ones observed in complex
50 multicellular organisms. Using neural networks (NNs), we further show that eukaryotic sequence
51 determinants are sufficient to predict the chromatin composition and activity of exogenous
52 chromosomes integrated in the yeast genome. Consequently, this study suggests that the fate
53 of any DNA molecule introduced into a given cellular context, including its organization from
54 nucleosome positioning up to 3D folding and transcriptional activity, is governed by rules that
55 are both deterministic and predictable.

56

57 **Results**

58 **Adaptation of supernumerary bacteria chromosomes to yeast**

59 To investigate large sequences which have not evolved in a eukaryotic context, we
60 exploited *S. cerevisiae* strains carrying an extra 17th circular chromosome made either from the
61 *Mycoplasma mycoides* subsp. *mycoides* (referred to as “Mmyco”) or *Mycoplasma pneumoniae*
62 (“Mpneumo”) genomes containing a yeast centromeric sequence and an autonomous
63 replication sequence (ARS)¹⁸ (**Methods; Table S1**). While the GC content of *S. cerevisiae* is
64 38%, the GC content of the Mmyco and Mpneumo chromosomes 24% (GC-poor) and 40% (GC-
65 neutral), respectively (**Fig. 1a**). Dinucleotide composition, estimated by the dinucleotide odds

66 ratio $\rho^*(XY)$ ¹⁹, also largely differs between yeast and bacterial chromosomes (**Extended Data**
67 **Fig. 1a**).

68 We linearized these chromosomes, added yeast telomeres at the extremities (**Fig. 1a**;
69 **Extended Data Fig. 1b**; **Methods**), and investigated their replication and cohesion using split-
70 dot assay and marker frequency analysis (**Extended Data Fig. 1c,d** and **Supplementary**
71 **Results**; **Methods**). Overall, these chromosomes do not impose a significant fitness cost on
72 their eukaryotic host, with each strain displaying similar growth rates compared to their parental
73 strains in selective media (**Extended Data Fig. 1e**). They also have an estimated segregation
74 rate similar to a centromeric plasmid²⁰ (**Extended Data Fig. 1f**).

75

76 **Spontaneous chromatinization of bacterial chromosomes in a eukaryotic context**

77 We first assessed whether integrated Mmyco and Mpneumo chromosomes were able to
78 chromatinize by performing MNase-seq and H3 and H2A ChIP-seq experiments (Methods; **Fig.**
79 **1b, c**; **Extended Data Fig. 2a**). Strikingly, both bacterial chromosomes were able to form
80 nucleosomes and local nucleosome-depleted regions (NDR; **Fig. 1b, c**; **Extended Data Fig.**
81 **2b-d**). Nucleosomes over the Mpneumo chromosome have structural features comparable to
82 yeast nucleosomes, with a linker DNA of ~14 bp and a nucleosomal repeat length (NRL) of 160
83 bp (**Fig. 1d,e**). In contrast, nucleosomes in the Mmyco chromosome have a longer linker DNA
84 and thus are more spaced than in yeast or Mpneumo chromosomes, with an NRL of 177 bp
85 (**Fig. 1d,e**). Two sequence features may explain this: first, poly(dA) and poly(dT) tracks, known
86 to promote NDR formation, are over-represented along the Mmyco genome. Secondly, the
87 AA/TT 10-bp periodicity, which facilitates DNA bending around histone cores, is reduced
88 (**Extended Data Fig. 2e,f**). The analysis of MNase-seq time course experiments also revealed
89 that Mmyco nucleosomal fragments are more rapidly degraded by MNase than yeast or
90 Mpneumo nucleosomal fragments (**Extended Data Fig. 2c,d**), reflecting either the “fragile”
91 nature of AT-rich nucleosomes or the MNase bias for these sequences^{21,22}.

92 We further profiled the chromatin composition of Mmyco and Mpneumo chromosomes by
93 performing RNA polymerase II (PolII) and cohesin (Scc1) ChIP-seq experiments (**Fig. 1f,**
94 **Extended Data Fig. 2g-j**). We found that the Scc1 and PolII binding profiles along the
95 Mpneumo chromosome appear similar to wild-type (WT) yeast chromosomes, with discrete
96 Scc1 peaks (**Fig. 1f, Extended Data Fig. 2h,i**) preferentially located in nucleosome-depleted
97 and PolII-enriched regions (**Extended Data Fig. 2j**). In contrast, Scc1 and PolII binding profiles
98 in the Mmyco strain show significant differences. On the one hand, Scc1 is strongly enriched
99 over the whole Mmyco chromosome, but does not form the distinct peaks observed on
100 endogenous chromosomes (**Fig. 1f, Extended Data Fig. 2h,i**). Interestingly, Scc1 levels also
101 appear strongly reduced at centromeres of endogenous *S. cerevisiae* chromosomes (**Extended**
102 **Data Fig. 2g**), suggesting that the Mmyco chromosome titrates cohesins enriched over yeast
103 centromeres. On the other hand, the PolII profile is greatly reduced along the Mmyco
104 chromosome, compared with yeast chromosomes.

105 These results show that large exogenous bacterial chromosomes placed in a eukaryotic
106 context spontaneously adopt eukaryotic chromatin composition features: histones, Pol II and
107 cohesins all bind bacterial DNA irrespectively of their sequence composition. However, striking
108 differences in the chromatin profiles led us to define two chromatin types: (1) “Y” (yeast-like)
109 chromatin landscape, found over Mpneumo (whose 40% GC content is close to the native *S.*
110 *cerevisiae* GC content), and (2) “U” (for Unconventional) chromatin, found over Mmyco (with a
111 low 24% GC content), featuring less packed nucleosomes, a reduced Pol II enrichment, and a
112 broad binding of cohesins across the entire chromosome.

113 Finally, we find that a convolutional neural network trained on yeast chromosome
114 sequences was able to predict nucleosome, Scc1 and PolII binding profiles along integrated
115 bacterial chromosomes (**Supplementary Results and Extended Data Fig. 3; Methods**). This
116 indicates that the complex, underlying eukaryotic sequence determinants are successfully

117 reused to generate these different chromatin landscapes along integrated bacterial
118 chromosomes.

119

120 **Transcriptional activity of bacterial genomes in a yeast context**

121 We next explored Y and U chromatin activities by first performing total, stranded RNA-seq
122 (**Fig. 2a,b; Method**). Consistent with PolIII ChIP-seq profiles (**Fig. 1f**), we find that the Y
123 chromatin type on Mpneumo is transcribed to levels similar to those of endogenous yeast
124 chromosomes (**Fig. 2a**). However, Mpneumo transcription tracks are significantly longer than
125 yeast genes (4.9kb vs 3.4kb, p-value < 2e-4, two-sided Student's t-test) and do not
126 systematically display clear boundaries (**Fig. 2b**). They also do not preferentially occur over
127 bacterial gene bodies nor seem to initiate at bacterial promoters, also consistent with an
128 absence of PolIII enrichment at these loci (**Extended Data Fig. 4a-c**).

129 In sharp contrast, the Mmyco U chromatin type is only sparsely and lowly transcribed (**Fig.**
130 **2a,b**), again in good agreement with the reduced levels of PolIII deposition (**Fig. 1f**). This
131 transcriptional inactivity is not associated with the recruitment of the yeast silencing complex
132 SIR Sir2/3/4²³ (**Extended Data Fig. 4d**).

133 Intriguingly, the transcription tracks orientation along bacterial chromosomes follows, on
134 average, the orientation of the bacterial genes annotated along these sequences, although the
135 transcription machinery has diverged billions years ago (**Fig. 2c**, see the orientation of genes
136 and stranded tracks in **Fig. 2a,b**). This correlates with the over-representation of A (G)
137 compared to T (C) on the coding strand, both detected in yeast genes and bacterial genes
138 (**Extended Data Fig. 4e**, see **Discussion**).

139

140 **Inactive U chromatin forms a compartment into the nuclear space in G1**

141 To assess the 3D organization of Y and U chromatin types, we performed capture of
142 chromosome conformation (Hi-C) ²⁴ experiments on the Mmyco or Mpneumo strains arrested
143 either in G1 or G2/M (**Fig. 3a, b**).

144 In G1-arrested cells, yeast chromosomes exhibit the expected Rab1 configuration, with
145 clustered centromeres and subtelomeric regions, resulting in a dotted pattern in Hi-C contact
146 maps ^{25,26}. For both Mmyco and Mpneumo chromosomes, the added yeast centromeric
147 sequence clusters with the native yeast centromeres. However, the two bacterial chromosomes
148 exhibit very distinct structural characteristics (**Fig. 3a**). The Y-type Mpneumo chromosome
149 behaves as an endogenous *S. cerevisiae* chromosome (**Fig. 3a**), with similar trans-contacts
150 (**Extended Data Fig. 5a**) and long-range cis-contacts (**Fig. 3c**), and a slope of the averaged
151 contact probability (p) as a function of the genomic distance (s) close to -1.5, a value
152 corresponding to the typical polymer structure observed in simulations (**Fig. 3c**) ²⁷. In addition,
153 DNA-FISH labeling of Mpneumo DNA reveals an extended and contorted structure within the
154 intranuclear space, confirming the intermixing of the chromosome with the actively transcribed
155 native chromosomes (**Fig. 3e, Extended Data Fig. 5c,d**).

156 In contrast, in the U-type Mmyco chromosome, (1) only few contacts were detected with the
157 endogenous chromosomes (**Extended Data Fig. 5a**), and predominantly with the 32 yeast
158 subtelomeric regions (**Fig. 3a**, dotted rectangles; **Extended Data Fig. 5b**); (2) the slope of the
159 $p(s)$ curve also changes dramatically towards a value of -1 at shorter distances, corresponding
160 to a polymer globule (**Fig. 3c**) ²⁴; (3) DNA-FISH labeling of the Mmyco chromosome confirmed
161 the globular conformation of the chromosome and revealed its preferential position at the
162 nuclear periphery, reflecting its proximity with sub/telomeres (**Fig. 3e; Extended Data Fig.**
163 **5c,d**). These results show that inactive, U chromatin forms a compartment in the nuclear space
164 segregated from the active endogenous chromosomal set.

165

166 **Cohesin compacts similarly both exogenous bacterial chromosomes during G2/M**

167 We then generated Hi-C contact maps of both strains synchronized in G2/M. In yeast,
168 chromosome compaction in G2/M-arrested cells reflect cohesin-mediated chromatin folding into
169 arrays of loops, with cohesin peaks corresponding to loop anchors^{28,29}. This compaction results
170 in a typical p(s) curve that presents an inflection point around the average loop length²⁵. The
171 decrease of interchromosomal contacts (**Fig. 3b, Extended Data Fig. 5a**) and of p(s) (**Fig. 3d,**
172 **Extended Data Fig. 5e,f**) observed for both native and bacterial chromosomes indicate a
173 similar mitotic compaction of both Y and U chromatin types.

174 Yeast and Mpneumo chromosomes exhibit focal contact Scc1 enrichments corresponding
175 to cohesin-anchored loops (black triangles on **Fig. 3f**)²⁸. We found that cohesin accumulates at
176 sites of convergent transcription not only, as expected, on yeast chromosomes^{30,31}, but also on
177 Mpneumo (**Fig. 2b**, green dotted lines, **Extended Data Fig. 5g**), with strong cohesin peaks
178 associated with sharper transcriptional convergence (**Fig. 2b**, arrows; **Extended Data Fig.**
179 **5h,i**). Using Chromosight, we identified 59 loops across the Mpneumo chromosome, spanning
180 over slightly longer genomic distances (28 kb on average) than the loops found along yeast
181 sequences (22 kb) (**Extended Data Fig. 5j; Methods**), and with strong Scc1 peaks positioned
182 close to anchors, with a median distance of 566 bp (**Fig. 3f, h**). The dotted grid pattern in
183 Mpneumo is reminiscent of multiple DNA loops observed along mammalian interphase
184 chromosomes³², suggesting that some loop anchors have the ability to be involved in loops of
185 various sizes, a phenomenon not observed across native *S. cerevisiae* chromosomes (**Fig. 3f**).

186 On the other hand, no visible discrete loops were observed along the Mmyco chromosome
187 where cohesins cover inactive U chromatin regions without clear peaks, probably reflecting the
188 absence of (convergent) transcription (**Fig. 3g**) (**Fig. 1f** and **Fig. 2b**). These results suggest that
189 cohesin can bind DNA and promote compaction in absence of active transcription, but does
190 not accumulate at discrete positions.

191 Cohesin-mediated loops along yeast chromosomes are formed through loop extrusion (LE),
192 a process by which cohesin complexes organize DNA by capturing and gradually enlarging

193 small loops^{33,34}. To test whether LE compacts bacterial genomes, we performed Hi-C in
194 absence of Wpl1 (human Wapl), which impairs cohesin removal and results in longer loops
195 readily visible by a shift of the inflection point in the p(s) curve^{28,34}. In absence of Wpl1, we
196 observed increased long-range contacts in all chromosomes, including Mmyco and Mpneumo
197 **(Fig. 3i; Extended Data Fig. 5k)**, consistent with active cohesin-mediated LE proceeding on
198 both Y and U chromatin.

199 Overall, these observations highlight two different spatial configurations of Mmyco and
200 Mpneumo exogenous chromosomes. On the one hand, the inactive Mmyco chromosome in G1
201 is compacted and positioned at the periphery of the nucleus, away from the active chromosome
202 set. Upon S-phase and G2 arrest, the DNA is nevertheless compacted in a cohesin-dependent
203 manner with no visible loops at discrete positions. On the other hand, the active Mpneumo
204 chromosome behaves as endogenous yeast chromosomes, displaying discrete mitotic DNA
205 loops in G2/M.

206

207 **Mosaic chimeric chromosomes display spontaneous, chromatin type-dependent** 208 **DNA compartmentalization**

209 The structural and functional features of the Y- and U-type chromatin are reminiscent of the
210 euchromatin and heterochromatin compartments described along metazoan chromosomes²⁴.
211 We wondered whether they can coexist on a single chromosome. Using CRISPR-Cas9, we first
212 fused the Mmyco chromosome with the native yeast chromosome XVI (XVIfMmyco
213 chromosome). We then induced translocations to generate two other strains with chromosomes
214 harboring alternating U and Y chromatin regions as small as 50 kb (XVIfMmycot1 and
215 XVIfMmycot2; **Fig. 4a; Extended Data Fig. 6a-c** Methods). ChIP-seq profiles of RNA PolIII on
216 Mmyco regions were similar before and after translocation, demonstrating that PolIII binding is
217 independent of the broader regional context (**Extended Data Fig. 6d**). The overall Hi-C contact
218 maps of XVIfMmyco strain synchronized in G1 showed little differences compared to the

219 parental Mmyco strain (but for the deleted and fused regions) (**Fig. 4b**, compare panel **i** with **ii**).
220 The Y-type chr. XVI arm intermixes with the other 15 yeast chromosomes while the U-type
221 Mmyco arm remains isolated from the rest of the genome, contacting subtelomeric regions (**Fig.**
222 **4b**). In the translocated strains, however, the alternating chromatin type regions resulted in
223 striking checkerboard contact patterns within XVIIfMmycot1 and XVIIfMmycot2 chromosomes
224 (**Fig. 4b**, panels **iii** and **iv**). This pattern reflects the fact that U-type regions make specific
225 contacts over long distances, bypassing the Y-type regions found in between (**Fig. 4c**;
226 **Extended Data Fig. 6e** G1). Similarly, Y-type regions of chromosome XVI also are also
227 involved in specific Y-Y contacts over longer distances (**Fig. 4c**). Intra U and Y regions contact
228 decay is also different in the two chromatin types, U chromatin being prone to longer range
229 contacts in cis (**Extended Data Fig. 6e**). We then induced another translocation to integrate a
230 segment of U chromatin in a *trans* position at the end of chromosome XIII and generated the
231 corresponding Hi-C map (**Fig. 4d**). The correlation and 4C-like plot revealed an increase in
232 inter-chromosomal contacts between Y chromatin segments with distant segments of itself, and
233 in alternance with decreased contacts with Mmyco U chromatin segments (and vice versa),
234 showing that the inactive U chromatin compartment can contain segments positioned on
235 different chromosomes (**Fig. 4d**). These results unambiguously show that in yeast alternated or
236 inter-chromosomal segments of active and inactive chromatin are spontaneously segregated
237 within the nuclear space in G1 (**Fig. 4e**, **Extended Data Fig. 6h**), a phenomenon reminiscent of
238 eu- and heterochromatin compartments found in higher eukaryotes ²⁴.

239 In G2/M, the Scc1 deposition pattern, including centromere depletion, remained conserved
240 along mosaic chromosomes (**Extended Data Fig. 6d**). Cohesin coating remained enriched over
241 U-type Mmyco sequences, including over the 50 kb encompassed within 850 Mb of yeast DNA
242 on XVIIfMmycot2 chromosome. These results show that cohesin is loaded along inactive Mmyco
243 regions independently of the proximity with the centromere (as along WT chromosomes) and

244 independently of transcription, suggesting that the larger chromatin context plays little role in
245 cohesin-mediated loading. In addition, Hi-C maps of the mosaic chromosomes in G2/M show
246 that U-type compartments disappear upon cohesin-mediated compaction (**Extended Data Fig.**
247 **6f**). The analysis of distance-dependent contact frequency shows that all chromosome regions
248 behave similarly at this stage (**Extended Data Fig. 6e**, G2/M). The only noticeable difference
249 resulting from the fusion/translocation events in G2/M Hi-C contact maps consists in a DNA loop
250 appearing over the boundary between Mmyco and the endogenous yeast sequence (**Extended**
251 **Data Fig. 6f,g**). The two anchors of the loop are a strongly cohesin-enriched region already
252 present on chr. XVI sequence, and another region slightly enriched in cohesin already present
253 on the Mmyco sequence (**Extended Data Fig. 6g**, green arrows).

254 Together with observations from the previous section, these results show that Y and U
255 chromatin types are both prone to dynamic cohesin-dependent DNA compaction and loop
256 formation in G2/M phase (**Fig. 4e**) and further suggest that loop extrusion can proceed through
257 the Y/U chromatin junction to form a loop bridging the nearest cohesin enrichment sites on each
258 side ^{35,36}. Therefore, as observed in multicellular organisms, LE-mediated metaphase
259 chromosome compaction abolishes chromatin compartments to prepare chromosomes for
260 segregation (**Fig. 4e**) ^{33,37}.

261

262 **Discussion**

263 *Chromatin composition and activity of bacterial DNA follows yeast sequence determinants*

264 Several studies have shown that exogenous or random DNA segments in the yeast nucleus
265 with a GC% relatively similar to that of yeast chromosome are actively transcribed ³⁸⁻⁴¹. Here we
266 explore the relationships between sequence composition and chromatin composition and
267 activity by characterizing large, divergent bacterial DNA sequences in a eukaryotic cellular
268 context into which they have not evolved. We find that *M. pneumoniae* and *M. mycoides*

269 chromosomes integrated in yeast adopt one of two archetypes of chromatin that we called Y
270 and U for yeast-like and unconventional. In both types, nucleosomes can assemble albeit with a
271 longer NRL in the U chromatin (177 vs 160 bp), which also correlates with a low transcription. In
272 G1, transcriptionally inactive U chromatin segments spontaneously collapse into a globule
273 leading to the formation of a new nuclear compartment, whereas Y chromatin intermixes with
274 yeast endogenous DNA. In G2/M, SMC cohesin complexes bind to both chromatin types and
275 compact DNA through the extrusion of DNA loops ⁶. Along Y chromatin, these complexes
276 accumulate at convergent transcription loci ^{28,30}, whereas along U chromatin, the absence of
277 active transcription leads to a homogeneous enrichment pattern.

278 The chromatin composition and activity experimentally measured over these exogenous
279 chromosomes can be predicted using convolutional neural networks solely trained on yeast
280 sequences, showing that exogenous DNA composition obeys a set of host rules based on
281 sequence determinants that include GC content, dinucleotide composition but also complex
282 sequence features. Sequence composition is therefore an important driver of the chromatin
283 composition of exogenous sequences.

284

285 *Prokaryotic DNA contains some determinants of transcription orientation in eukaryote*

286 Active transcription on bacterial chromosomes in yeast follows, on average, bacteria genes
287 orientation (**Fig. 2**). Several mechanisms could account for this observation. Sequence
288 determinants predating the divergence of eucaryotes and procaryotes, including GC and AT
289 skews, are known to influence polymerase directionality ^{47,48}. Alternatively, bacteria codon
290 usage bias could result in enrichment in Nrd1-Nab3-Sen1 (NNS) binding sites along the ncRNA
291 strand, leading to transcription termination ^{39,49}. The Nonsense-Mediated Decay (NMD)
292 complex, which destabilizes in the cytoplasm eukaryotic transcripts with long 3' UTR, could also
293 be more frequently activated by bacteria RNA transcribed in an antisense orientation.
294 Eventually, the resulting conserved orientation could facilitate the domestication of exogenous

295 sequences during horizontal transfer/introgression events between distant species. Whether
296 these RNA molecules are translated, and peptides of bacterial origins exist in the yeast cell,
297 remains to be determined. If so, they could provide a source of diversity and adaptation.

298

299 *Cohesin-mediated chromatin folding along non-transcribed DNA templates*

300 It is proposed that cohesins expand DNA loops by an active process of loop extrusion until
301 they encounter an obstacle and/or a release signal ^{6,33}. For instance, the transcriptional
302 repressor CTCF constitutes a cohesin roadblock to LE in mammals, but SMC-mediated loops
303 have been identified in genomes lacking CTCF across domains of life. In yeast and other
304 species, anchors are thought to be determined by a combination of convergent transcription,
305 replication fork progression during S phase and/or the presence at these positions of stably
306 bound cohesin promoting sister chromatid cohesion ⁵⁰⁻⁵². The barely transcribed Mmyco
307 chromosome is compacted by cohesin at G2/M, without discrete loop anchoring, suggesting that
308 transcription is neither necessary for loading nor a primary driver for translocation. We propose
309 that cohesins can move freely along this template without encountering significant obstacles,
310 making it a suitable model for studying potentially blocking sequences or molecules.

311

312 *Compartmentalization of foreign DNA in host's nucleus*

313 We show that the introduction a foreign DNA with a lower GC content is sufficient to
314 spontaneously promote the formation of a distinct inactive chromatin type, with increased NRL
315 and a compact globular structure, all features reminiscent of mammalian heterochromatin. This
316 chromatin forms a new compartment at the periphery of the budding yeast nucleus (**Fig. 5e;**
317 **Extended Data Fig. 6h**), like the metazoan compartment "B" formed by inactive,
318 H3K9me3/HP1-mediated heterochromatin, which is absent in yeast. The mechanisms behind
319 compartmentalization remain unknown, but we hypothesize that the unmixing between U and Y
320 chromatin could be due either to NRL change, protein composition and/or activity. Importantly,

321 longer NRLs are found in heterochromatin in mammals, and are associated with histone H1⁵³.

322 In contrast, Y chromatin intermixes with active yeast DNA.

323 Depending on its sequence composition, a DNA molecule will therefore either
324 compartmentalize or intermix with the host DNA. These different fates of the two chromatin
325 types raise interesting evolutionary considerations. In the context of the invasion of the genome
326 by exogenous mobile elements, this could lead either to the spontaneous isolation of inactive
327 foreign DNA or, on the contrary, to the co-option of a new set of active sequences that could
328 represent a reservoir of genetic innovations. This sequence-dependent mechanism may have
329 contributed to the heterochromatinization of AT-rich transposable elements integrated in
330 mammalian genomes.

331

332 **Competing interests**

333 The authors declare no competing interests.

334

335 **Authors Contributions**

336 Conceptualization: CC, LM, JS, JM and RK. Methodology: LM, CC, JS, CL, JM, RK. Software:
337 JS. Validation: CC, JS, LM. Investigation: LM, CC, with contributions from AP, AB, AgT MR and
338 FB. Formal analysis: JS, AW, ER, CC, LM, FB, MR, AnT. Data Curation: JS, with contributions
339 from CC, LM. Resources: GG, CL. Visualization: JS. Writing - original draft preparation: CC, JS,
340 LM, JM and RK. Writing – Editing: all authors. Writing – Revisions: JS, LM, AW, JM, RK.
341 Supervision: JM, RK. CC co-supervised a student. Funding acquisition: RK. Project
342 Administration: RK.

343 **Acknowledgements**

344 We are grateful to Bernard Dujon, Micheline Fromont-Racine, Alain Jacquier, Gianni Liti,
345 Bertrand Llorente, Marcelo Nollmann, Cosmin Saveanu and Benoit le Tallec for fruitful
346 comments on the work and the manuscript. This research was supported by the European
347 Research Council under the Horizon 2020 Program (ERC grant agreement 771813) and
348 Agence Nationale pour la Recherche (ANR-19-CE13-0027-02) to RK. RK, FB, JM and AnT also
349 received support from Agence Nationale pour la Recherche (ANR-22-CE12-0013-01). CC was
350 supported by a Pasteur-Roux-Cantarini fellowship. JS was supported by a postdoctoral ARC
351 fellowship. We thank all our colleagues from the laboratory régulation spatiale des génomes for
352 fruitful discussions. We especially thank Cyril Matthey-Doret for support during the earlier steps
353 of the project. E. Turc and L. Lemée at Biomix Platform, C2RT, Institut Pasteur, Paris, France,
354 are supported by France Génomique (ANR-10-INBS-09-09) and IBISA for processing and
355 sequencing RNA samples.

356 **Figure legends**

357

358 **Figure 1. Chromatin composition of bacterial chromosomes in yeast.**

359 **a**, Schematic representation of the conversion from circular to linear chromosomes integrated in
360 yeast. The purple and blue colors represent the *M. pneumoniae* (Mpneumo) and *M. mycoides*
361 (Mmyco) bacterial sequence in all figures, respectively. Right: distribution of GC% for 1kb
362 windows over yeast chromosome XVI, *M. pneumoniae* and *M. mycoides* chromosomes.

363 **b**, Nucleosomal track (grey, see **Methods**) and H2A and H3 ChIP-seq (shades of blue, IP vs
364 input, log₂) profiles obtained in the Mpneumo strain (*S. cerevisiae* + *M. pneumoniae*). 10kb-long
365 genomic windows from the chromosome XVI (left: 250kb-260kb) and the Mpneumo
366 chromosome (right: 710kb-720kb) are shown at the same scale. Genomic loci highlighted in
367 yellow correspond to nucleosome-depleted regions.

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

371 **d**, Frequency of nucleosome linker DNA length in Mpneumo and Mmyco strains. For each
372 strain, the distribution is calculated for each chromosome separately. The dashed line indicates
373 14 bp and the dotted line indicates 25 bp.

374 **e**, Auto-correlation of the nucleosome track in Mpneumo and Mmyco strains, computed for each
375 chromosome separately (lag ≤ 1000). Nucleosome repeat length (NRL) is indicated for yeast,
376 Mpneumo and Mmyco chromosomes.

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

381

382 **Figure 2. Expression of exogenic bacterial sequences in budding yeast.**

383 **a**, Stranded RNA-seq profiles along yeast chromosome XVI or along bacterial chromosomes
384 Mpneumo and Mmyco in the corresponding strains. Forward (pink) and reverse (turquoise)
385 genes along yeast or bacterial sequences are indicated as transparent segments under the
386 tracks. Pink and turquoise represent forward and reverse transcription, respectively.

387 **b**, Top: stranded RNA-seq profiles along a 60kb window along yeast chromosome XVI,
388 Mpneumo or Mmyco. Bottom: Scc1 (cohesin) deposition profiles of the corresponding loci.
389 Green dotted lines indicate identified loci of convergent transcription (see Methods).

390 **c**, Forward and reverse RNA-seq coverage of forward- and reverse-oriented yeast or bacterial
391 genes. Scores were normalized by the length of each genomic feature.

392

393 **Figure 3. Folding of exogenic bacterial sequences within the yeast nucleus.**

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

396 **c, d**, Contact frequency (ρ) as a function of genomic distance (s) plots of endogenous yeast
397 chromosomes (long arms) and of Mmyco and Mpneumo bacterial chromosomes in G1 (**c**) and
398 G2/M (**d**).

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

404 **f, g**, Top: magnification of 150kb windows from Hi-C contact maps in G2/M from either an
405 endogenous or the bacterial chromosome in Mpneumo (**f**) and Mmyco (**g**) strains (1kb

406 resolution). Bottom: Scc1 ChIP-seq deposition profile. Black arrowheads: loops. Cyan
407 diamonds: Scc1 peaks positions.

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

412 **i**, Distance-dependent contact frequency in endogenous yeast and in bacterial chromosomes
413 (left, Mpneumo; right, Mmyco), in WT (dashed) and in $\Delta wpl1$ mutants.

414

415 **Figure 4. Compartmentalization of mosaic chromosomes composed of Y and U-type**
416 **chromatin.**

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

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

422 **c**, Virtual 4C profiles of viewpoints (grey/blue arrows) located within yeast chr. XVI (in grey) or
423 Mmyco chromosome segments (in blue), in XVIfMmyco (top), XVIfMmycot1 (middle) and
424 XVIfMmycot2 (bottom) strains.

425 **d**, Left: Schematic representation of the CRISPR strategy used to generate the Mmyco mosaic
426 XVIfMmycot3 strain, with alternating yeast and Mmyco segments in chromosomes XIII and XVI.
427 Middle: G1 Hi-C maps of mosaic chromosomes XIII and XVI in XVIfMmycot3 strain. The color
428 scales for the normalized contact map and the correlation matrix are the same as in **b**. Right:
429 Virtual 4C profiles of viewpoints located within yeast chr. XIII (in grey) or within yeast chr. XVI (in
430 grey) or Mmyco chromosome (in blue) segments in XVIfMmycot3 strain. Grey/blue arrows:Y
431 and U chromatin viewpoints position.

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

435 **References**

- 436 1. Romiguier, J., Ranwez, V., Douzery, E. J. P. & Galtier, N. Contrasting GC-content dynamics
437 across 33 mammalian genomes: relationship with life-history traits and chromosome sizes.
438 *Genome Res* **20**, 1001–1009 (2010).
- 439 2. Tajbakhsh, J. *et al.* Spatial Distribution of GC- and AT-Rich DNA Sequences within Human
440 Chromosome Territories. *Experimental Cell Research* **255**, 229–237 (2000).
- 441 3. Valouev, A. Determinants of nucleosome organization in primary human cells. *Nature* **474**,
442 516–520 (2011).
- 443 4. Holmquist, G. P. Evolution of chromosome bands: Molecular ecology of noncoding DNA. *J*
444 *Mol Evol* **28**, 469–486 (1989).
- 445 5. Mirny, L. & Dekker, J. Mechanisms of Chromosome Folding and Nuclear Organization:
446 Their Interplay and Open Questions. *Cold Spring Harb Perspect Biol* **14**, a040147 (2022).
- 447 6. Davidson, I. F. & Peters, J.-M. Genome folding through loop extrusion by SMC complexes.
448 *Nat Rev Mol Cell Biol* (2021) doi:10.1038/s41580-021-00349-7.
- 449 7. Belmont, A. S. Nuclear Compartments: An Incomplete Primer to Nuclear Compartments,
450 Bodies, and Genome Organization Relative to Nuclear Architecture. *Cold Spring Harb*
451 *Perspect Biol* **14**, a041268 (2022).
- 452 8. Crisp, A., Boschetti, C., Perry, M., Tunnacliffe, A. & Micklem, G. Expression of multiple
453 horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes.
454 *Genome Biol* **16**, 50 (2015).
- 455 9. Edelman, N. B. & Mallet, J. Prevalence and Adaptive Impact of Introgression. *Annu Rev*
456 *Genet* **55**, 265–283 (2021).
- 457 10. Van Etten, J. & Bhattacharya, D. Horizontal Gene Transfer in Eukaryotes: Not if, but How
458 Much? *Trends Genet* **36**, 915–925 (2020).

- 459 11. Peter, J. *et al.* Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*
460 **556**, 339–344 (2018).
- 461 12. Baby, V. *et al.* Cloning and Transplantation of the *Mesoplasma florum* Genome. *ACS Synth*
462 *Biol* **7**, 209–217 (2018).
- 463 13. Coradini, A. L. V., Hull, C. B. & Ehrenreich, I. M. Building genomes to understand biology.
464 *Nat Commun* **11**, 6177 (2020).
- 465 14. Currin, A. *et al.* The evolving art of creating genetic diversity: From directed evolution to
466 synthetic biology. *Biotechnol Adv* **50**, 107762 (2021).
- 467 15. Payen, C. *et al.* Unusual composition of a yeast chromosome arm is associated with its
468 delayed replication. *Genome Res.* **19**, 1710–1721 (2009).
- 469 16. Labroussaa, F. *et al.* Impact of donor-recipient phylogenetic distance on bacterial genome
470 transplantation. *Nucleic Acids Res* **44**, 8501–8511 (2016).
- 471 17. Ruiz, E. *et al.* CReasPy-Cloning: A Method for Simultaneous Cloning and Engineering of
472 Megabase-Sized Genomes in Yeast Using the CRISPR-Cas9 System. *ACS Synth Biol* **8**,
473 2547–2557 (2019).
- 474 18. Lartigue, C. Creating bacterial strains from genomes that have been cloned and engineered
475 in yeast. *Science* **325**, 1693–1696 (2009).
- 476 19. Karlin, S. & Mrázek, J. Compositional differences within and between eukaryotic genomes.
477 *Proceedings of the National Academy of Sciences* **94**, 10227–10232 (1997).
- 478 20. Christianson, T. W., Sikorski, R. S., Dante, M., Shero, J. H. & Hieter, P. Multifunctional yeast
479 high-copy-number shuttle vectors. *Gene* **110**, 119–122 (1992).
- 480 21. Chereji, R. V., Ocampo, J. & Clark, D. J. MNase-Sensitive Complexes in Yeast:
481 Nucleosomes and Non-histone Barriers. *Mol Cell* **65**, 565-577.e3 (2017).
- 482 22. Xi, Y., Yao, J., Chen, R., Li, W. & He, X. Nucleosome fragility reveals novel functional states
483 of chromatin and poises genes for activation. *Genome Res.* **21**, 718–724 (2011).

- 484 23. Kueng, S., Oppikofer, M. & Gasser, S. M. SIR proteins and the assembly of silent chromatin
485 in budding yeast. *Annu Rev Genet* **47**, 275–306 (2013).
- 486 24. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals
487 Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).
- 488 25. Lazar-Stefanita, L. *et al.* Cohesins and condensins orchestrate the 4D dynamics of yeast
489 chromosomes during the cell cycle. *EMBO J* **36**, 2684–2697 (2017).
- 490 26. Duan, Z. *et al.* A three-dimensional model of the yeast genome. *Nature* **465**, 363–367
491 (2010).
- 492 27. Wong, H. A predictive computational model of the dynamic 3D interphase yeast nucleus.
493 *Curr Biol* **22**, 1881–1890 (2012).
- 494 28. Dauban, L. *et al.* Regulation of Cohesin-Mediated Chromosome Folding by Eco1 and Other
495 Partners. *Mol Cell* **77**, 1279-1293.e4 (2020).
- 496 29. Costantino, L., Hsieh, T.-H. S., Lamothe, R., Darzacq, X. & Koshland, D. Cohesin residency
497 determines chromatin loop patterns. *Elife* **9**, e59889 (2020).
- 498 30. Lengronne, A. *et al.* Cohesin relocation from sites of chromosomal loading to places of
499 convergent transcription. *Nature* **430**, 573–578 (2004).
- 500 31. Glynn, E. F. Genome-wide mapping of the cohesin complex in the yeast *Saccharomyces*
501 *cerevisiae*. *PLoS Biol* **2**, E259, (2004).
- 502 32. Rao, S. S. P. A 3D map of the human genome at kilobase resolution reveals principles of
503 chromatin looping. *Cell* **159**, 1665–1680 (2014).
- 504 33. Goloborodko, A., Imakaev, M. V., Marko, J. F. & Mirny, L. Compaction and segregation of
505 sister chromatids via active loop extrusion. *Elife* **5**, e14864 (2016).
- 506 34. Haarhuis, J. H. I. *et al.* The Cohesin Release Factor WAPL Restricts Chromatin Loop
507 Extension. *Cell* **169**, 693-707.e14 (2017).
- 508 35. Spracklin, G. *et al.* Diverse silent chromatin states modulate genome compartmentalization
509 and loop extrusion barriers. *Nat Struct Mol Biol* **30**, 38–51 (2023).

- 510 36. Nora, E. P. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome
511 Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 (2017).
- 512 37. Naumova, N. Organization of the mitotic chromosome. *Science* **342**, 948–953 (2013).
- 513 38. Struhl, K. & Segal, E. Determinants of nucleosome positioning. *Nat Struct Mol Biol* **20**, 267–
514 273 (2013).
- 515 39. Gvozdenov, Z., Barcutean, Z. & Struhl, K. Functional analysis of a random-sequence
516 chromosome reveals a high level and the molecular nature of transcriptional noise in yeast
517 cells. *Molecular Cell* **0**, (2023).
- 518 40. Luthra, I. *et al.* Biochemical activity is the default DNA state in eukaryotes.
519 2022.12.16.520785 Preprint at <https://doi.org/10.1101/2022.12.16.520785> (2022).
- 520 41. Camellato, P., Brosh, R., Maurano, M. & Boeke, J. Genomic analysis of a synthetic reversed
521 sequence reveals default chromatin states in yeast and mammalian cells. bioRxiv. (2022).
- 522 42. Ocampo-Hafalla, M., Muñoz, S., Samora, C. P. & Uhlmann, F. Evidence for cohesin sliding
523 along budding yeast chromosomes. *Open Biol* **6**, 150178 (2016).
- 524 43. de Boer, C. G. *et al.* Deciphering eukaryotic gene-regulatory logic with 100 million random
525 promoters. *Nat Biotechnol* **38**, 56–65 (2020).
- 526 44. Vaishnav, E. D. The evolution, evolvability and engineering of gene regulatory DNA. *Nature*
527 **603**, 455–463 (2022).
- 528 45. Koster, C. C., Postma, E. D., Knibbe, E., Cleij, C. & Daran-Lapujade, P. Synthetic Genomics
529 From a Yeast Perspective. *Front. Bioeng. Biotechnol.* **10**, 869486 (2022).
- 530 46. Schindler, D. *et al.* Design, Construction, and Functional Characterization of a tRNA
531 Neochromosome in Yeast. 2022.10.03.510608 Preprint at
532 <https://doi.org/10.1101/2022.10.03.510608> (2022).
- 533 47. Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chédin, F. R-loop formation is a
534 distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell* **45**, 814–
535 825 (2012).

- 536 48. Lobry, J. R. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol*
537 *Evol* **13**, 660–665 (1996).
- 538 49. Schulz, D. Transcriptome Surveillance by Selective Termination of Noncoding RNA
539 Synthesis. *Cell* **155**, 1075–1087 (2013).
- 540 50. Bastie, N. *et al.* Sister chromatid cohesion halts DNA loop expansion. 2023.07.31.551217
541 Preprint at <https://doi.org/10.1101/2023.07.31.551217> (2023).
- 542 51. Banigan, E. J. *et al.* Transcription shapes 3D chromatin organization by interacting with loop
543 extrusion. *Proceedings of the National Academy of Sciences* **120**, e2210480120 (2023).
- 544 52. Dequeker, B. J. H. *et al.* MCM complexes are barriers that restrict cohesin-mediated loop
545 extrusion. *Nature* **606**, 197–203 (2022).
- 546 53. Hergeth, S. P. & Schneider, R. The H1 linker histones: multifunctional proteins beyond the
547 nucleosomal core particle. *EMBO Rep* **16**, 1439–1453 (2015).
- 548

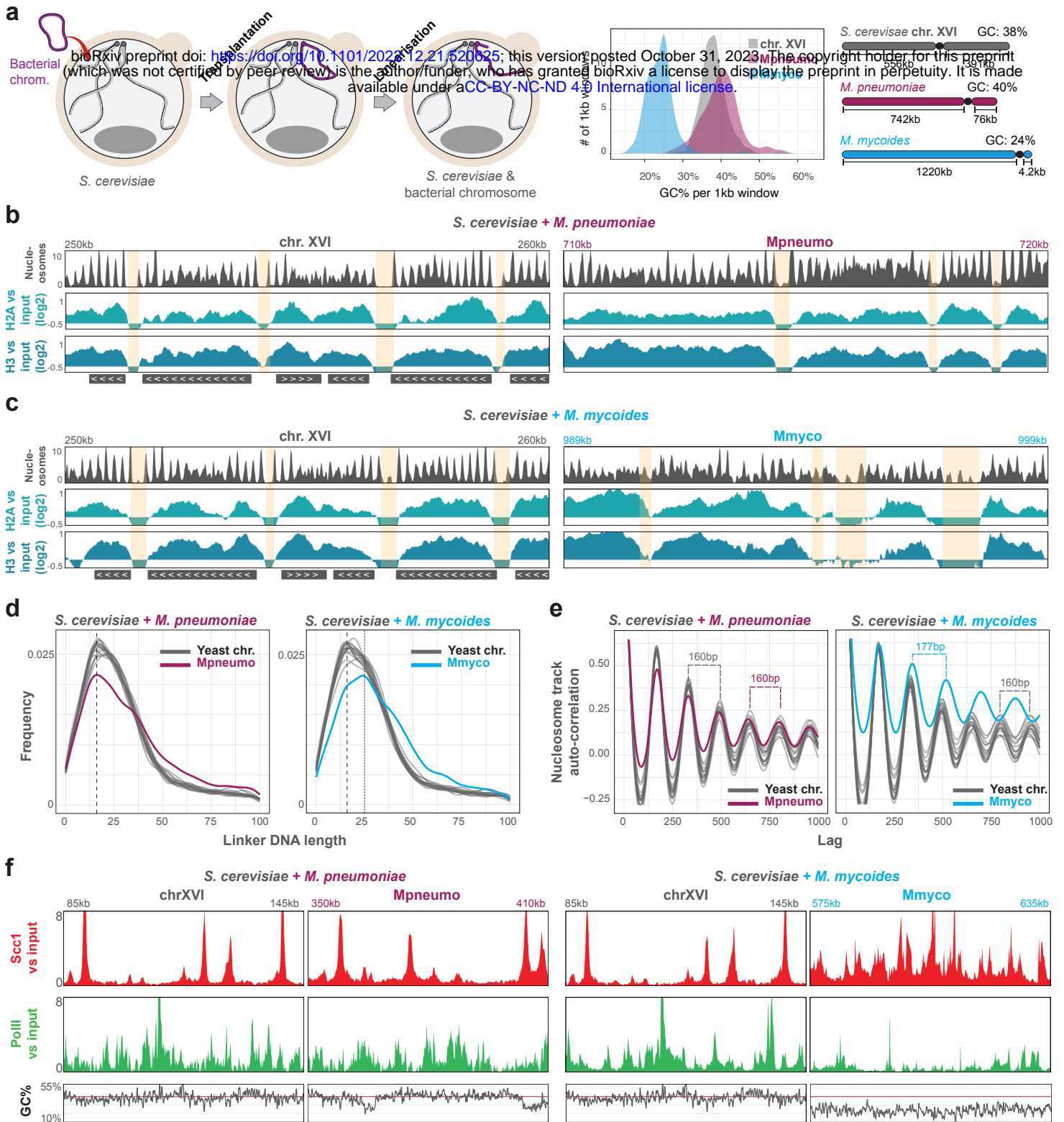


Fig. 1

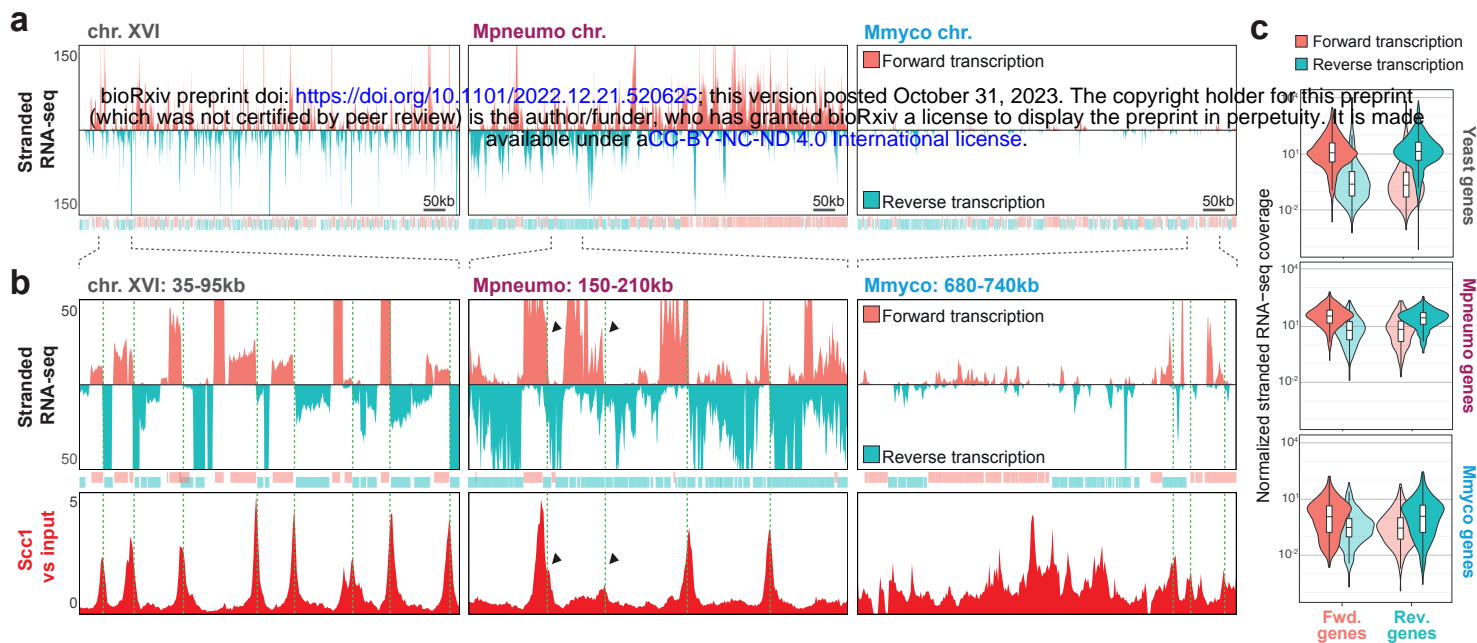


Fig. 2

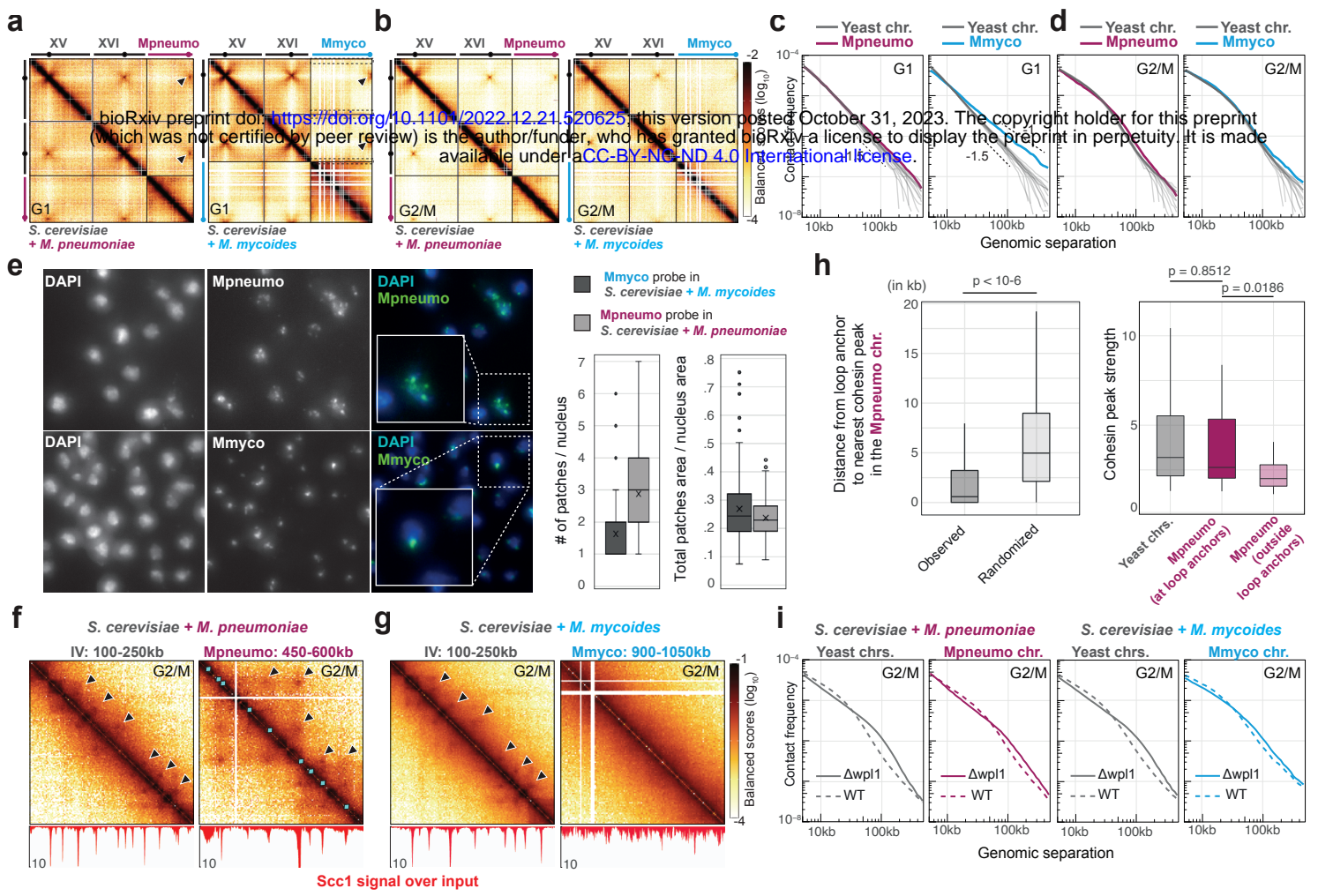


Fig. 3

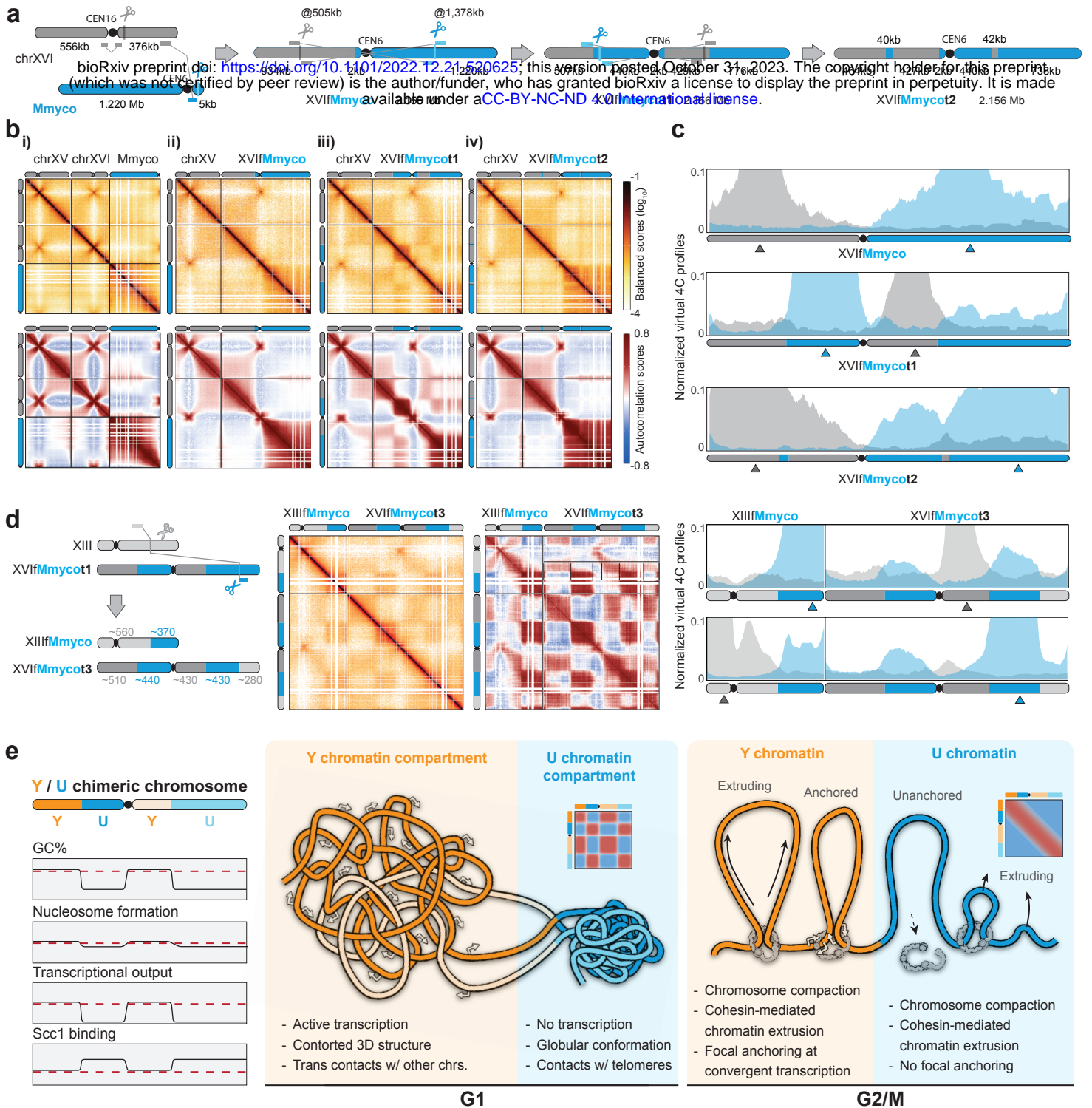


Fig. 4