



HAL
open science

Dynamics of the compartmentalized *Streptomyces* chromosome during metabolic differentiation

Virginia S Lioy, Jean-Noel Lorenzi, Soumaya Najah, Thibault Poinsignon, Herve Leh, Corinne Saulnier, Bertrand Aigle, Sylvie Lautru, Annabelle Thibessard, Olivier Lespinet, et al.

► **To cite this version:**

Virginia S Lioy, Jean-Noel Lorenzi, Soumaya Najah, Thibault Poinsignon, Herve Leh, et al.. Dynamics of the compartmentalized *Streptomyces* chromosome during metabolic differentiation. 2021. hal-03128776v1

HAL Id: hal-03128776

<https://hal.science/hal-03128776v1>

Preprint submitted on 28 Apr 2021 (v1), last revised 6 Sep 2021 (v2)

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.

23 **Abstract**

24 *Streptomyces* are among the most prolific bacterial producers of specialized
25 metabolites, including antibiotics. The linear chromosome is partitioned into a central
26 region harboring core genes and two extremities enriched in specialized metabolite
27 biosynthetic gene clusters (SMBGCs). The molecular mechanisms governing
28 structure and function of these compartmentalized genomes remain mostly unknown.
29 Here we show that in exponential phase, chromosome structure correlates with
30 genetic compartmentalization: conserved, large and highly transcribed genes form
31 boundaries that segment the central part of the chromosome into domains, whereas
32 the terminal ends are transcriptionally, largely quiescent compartments with different
33 structural features. Onset of metabolic differentiation is accompanied by remodeling
34 of chromosome architecture from an 'open' to a rather 'closed' conformation, in which
35 the SMBGCs are expressed forming new boundaries. Altogether, our results reveal
36 that *S. ambofaciens*' linear chromosome is partitioned into structurally distinct
37 entities, indicating a link between chromosome folding, gene expression and genome
38 evolution.

39

40 **Introductory paragraph**

41 Bacteria of the genus *Streptomyces* are amongst the most prolific producers of
42 specialized metabolites with applications in medicine, agriculture and the food
43 industry¹. The biosynthesis of these specialized metabolites (e.g. antibiotics and
44 pigments) generally occurs after 'metabolic differentiation', a physiological transition
45 from primary to specialized metabolism^{2,3}. This transition coincides with
46 morphological differentiation⁴ (e.g. formation of the secondary multinucleated
47 mycelium, sporulation) that occurs late in the growth phase. These metabolic and

48 morphological changes are controlled by highly interconnected regulatory
49 mechanisms that have been studied mostly at the transcriptional, translational and
50 post-translational levels⁴⁻⁷. However, a clear link between global chromosome
51 organization and metabolic differentiation has yet to be established.

52 *Streptomyces* possess an unusual linear chromosome. Moreover, its
53 extremities, or ‘terminal arms’, contain terminal inverted repeats (TIRs) capped by
54 telomere-like sequences. In addition, the chromosome is large (6-15 Mb) with an
55 extreme GC content (circa 72%). Finally, the *Streptomyces* chromosome presents a
56 remarkable genetic compartmentalization, with a distinguishable central region and
57 the terminal arms. The central region primarily contains core genes, common to all
58 *Streptomyces* species, these being often essential⁸⁻¹³. The terminal arms are mainly
59 composed of conditionally adaptive genes, in particular, enriched in specialized
60 metabolite biosynthetic gene clusters (SMBGCs)¹⁴. In addition, they are prone to
61 large DNA rearrangements and frequent recombination^{12,15,16,17}. Importantly, many
62 SMBGCs appear silent or poorly expressed under laboratory conditions, giving rise to
63 the concept of ‘cryptic’ SMBGCs. The basis of the regulation of these clusters in
64 relation to the overall genome dynamics remains to be explored.

65 The development of chromosome conformation capture (3C) methods coupled
66 to deep sequencing provided novel concepts in bacterial chromosome biology^{18,19}.
67 Notably pioneer studies revealed that bacterial chromosomes present a high degree
68 of 3D-organization in macrodomains and/or ‘chromosome interacting domains’
69 (CIDs), mediated by multiple structural factors, including transcription and
70 replication²⁰⁻²⁷. Here, we explore the dynamics of the *Streptomyces* chromosome
71 during metabolic differentiation. By combining multi-omic approaches, we show that
72 the dynamics of gene expression correlate with the folding of the linear chromosome

73 of *Streptomyces ambofaciens* ATCC 23877 into transcriptionally active and silent
74 compartments. Moreover, metabolic differentiation is accompanied by a huge
75 remodeling of chromosome architecture from an ‘open’ to a rather ‘closed’
76 conformation, SMBGCs forming new boundaries. Altogether, our results highlight a
77 link between chromosome folding, gene expression and genome evolution.

78

79 **Results**

80 **Compartmentalization of the *Streptomyces ambofaciens* genome highlighted** 81 **by comparative genomics and functional annotation**

82 The genetic compartmentalization of the *S. ambofaciens* ATCC 23877 genome
83 was previously reported¹². However, here we took advantage of the numerous
84 *Streptomyces* available sequences to update the cartography of the genome. We
85 also used gene persistence²⁸ as a new indicator to describe the organization of
86 *Streptomyces* genome. Gene persistence reflects the tendency for genes to be
87 conserved in a large number of genomes. As shown in other bacteria, gene
88 persistence is associated with gene essentiality and similar expression levels²⁸⁻³⁰.
89 Here, we calculated a gene persistence index by determining the frequency of a
90 given gene in a panel of 125 complete *Streptomyces* genomes (**Supplementary**
91 **Table 1**)³¹. The highest level of persistence associated with the best reciprocal
92 matches between genes of all these genomes defines the core-genome³¹. In
93 addition, we mapped the genes encoding the ‘actinobacterial signature’ previously
94 identified³² as those coding sequences (CDSs) that are nearly universal among
95 actinobacteria. In contrast, we searched the *S. ambofaciens* ATCC 23877 genome
96 for variable regions, potentially acquired by horizontal gene transfer. For this
97 purpose, we identified unique genes as well as genomic islands (GIs,

98 **Supplementary Table 2, Supplementary Fig. 1**). These are defined as DNA
99 sequences that are inserted in regions of synteny (i.e. same orthologous gene order
100 in closely related genomes, see Methods). Thus genes from the core, the
101 actinobacterial signature and/or presenting a high level of persistence are enriched in
102 the central region, whereas GIs and unique CDSs are enriched in chromosome
103 extremities. The results clearly highlight the compartmentalization of the *S.*
104 *ambofaciens* genome (**Fig.1, Supplementary Table 3**). Synteny gradually
105 disappears in terminal arms (**Fig.1**), as previously reported¹². This makes it difficult to
106 have an operational delineation of the limits of these arms. However, since the first
107 and last ribosomal operons approximately mark the limits of a central region beyond
108 which the synteny level falls (**Fig.1**), these operons were used as limits for definition
109 of the left and right extremities in this study.

110 Functional annotation also highlighted a bias of gene distribution along the
111 chromosome (**Fig.1, Supplementary Table 3**). As a proxy for major metabolic
112 processes, we used the functional RNA genes (encoding rRNA, tRNA, tmRNA,
113 RNase-P RNA, SRP-RNA) and genes encoding functions related to translation
114 and/or RNA stability. We also mapped the genes encoding nucleoid-associated
115 proteins (NAPs) and chromatin structural factors, further shorten 'NAPSFs', that play
116 a central role in the dynamic organization of the bacterial chromosome and are
117 enriched in the central region (**Fig.1, Supplementary Table 3**). Finally, we used the
118 antiSMASH secondary metabolite genome mining pipeline (version 5.1.0)³³ to identify
119 putative SMBGCs (**Supplementary Table 4**). These genes are preferentially located
120 in the terminal arms and almost half of them are present within GIs (**Fig.1,**
121 **Supplementary Table 3**). Accordingly, GIs are 3.5-fold enriched in SMBGC genes (p
122 value $< 2.2 \cdot 10^{-6}$, Fisher's exact test for count data). These results illustrate that the

123 high variability of *Streptomyces* extremities relates in particular to functions involved
124 in metabolic differentiation.

125 Together, these analyses confirm the strong genetic compartmentalization of
126 the *S. ambofaciens* chromosome. In this context, we then explored the extent to
127 which this genetic organization correlates with gene expression and chromosome
128 architecture.

129

130 **Transcription of the *S. ambofaciens* genome is strongly compartmentalized**

131 To focus specifically on transcriptome dynamics during metabolic
132 differentiation, while limiting cellular physiological heterogeneity in the colony, we
133 took advantage of the fact that *S. ambofaciens*, like many *Streptomyces*², does not
134 sporulate in liquid medium. We chose MP5 and YEME liquid media (**Table 1**) in
135 which *S. ambofaciens* grows in a rather dispersed manner that makes the cells also
136 more accessible to further 3C-treatments. Moreover, for *S. ambofaciens*, MP5
137 medium was previously reported to be suited for the production of the antibiotics
138 spiramycin³⁴ and congocidine³⁵ by *S. ambofaciens* while there is only limited
139 antibiotic production in YEME medium (**Extended Data Fig.1.A&B**). The results
140 showed that the transcriptomes were rather similar in exponential phase in both
141 media, while commitment to specialized metabolism (C4, C5 and C7 conditions) was
142 accompanied by major transcriptional changes that form a distinct cluster
143 (**Fig.2.A&B, Extended Data Fig.1.C&D**). Notably, at the earliest time points in the
144 growth period in both media, the terminal regions were rather poorly expressed,
145 whereas transcription gradually increased toward terminal ends over the growth
146 period (**Fig.2.B**), as previously reported in *S. coelicolor*^{10,36}. We observed that the *S.*
147 *ambofaciens* genome has a transcriptional landscape with approximately 90 % of the

148 genes significantly expressed ('CAT_1' level or more) and/or regulated (DESeq2
149 statistical analysis) in at least one condition (**Extended Data Fig.2.A&B,**
150 **Supplementary Table 5**).

151 As shown by scanning the genome with respect to gene expression levels, we
152 found a marked bias in the localization of poorly *versus* highly expressed genes
153 along the genome as well as a positive correlation between gene persistence and
154 expression (**Fig. 2C, Extended Data Fig.2.C & D**). The 332 chromosomal genes
155 expressed at very high level in all conditions are mainly enriched in the central region
156 of the genome (**Supplementary Table 3, Fig.3D**), whereas most (> 60 %) of the
157 SMBGCs, GIs and mobile genetic elements (pSAM1, pSAM2 and a prophage) are
158 silent or poorly expressed in at least one condition. Up to 19 and 23 % of SMBGC
159 and GI genes, respectively, were found in this category in all tested conditions
160 (**Extended Data Fig.2.B**). However, compared to other poorly conserved genes,
161 SMBGCs have a higher proportion of highly induced genes, from 'CAT_0' to the
162 highest ('CAT_3' or 'CAT_4') categories (**Extended data Fig.2.C**).

163 Interestingly, as previously described in *S. coelicolor*³⁶, we observed
164 significant antisense-transcription throughout the genome. Considering the absolute
165 number of transcripts, this antisense-transcription was positively correlated to the
166 level of sense-transcription (**Extended Data Fig.2.E**), suggesting that spurious
167 antisense-transcription can arise from highly transcribed regions. To investigate the
168 relative importance of this antisense-transcription, we defined the 'antisense index' as
169 the level of antisense-transcription over the total (sense plus antisense) transcription.
170 Interestingly, in exponential phase, the antisense index was particularly high for GIs,
171 mobile genetic elements and SMBGCs and very low in conserved regions enriched in
172 the central region (except for pSAM2) (**Extended Data Fig.2.F**). During metabolic

173 differentiation the antisense indices of GIs, pSAM1, prophage and especially
174 SMBGCs tended to decrease (**Extended Data Fig.2.F**). This observation suggests
175 that antisense-transcription could be a consequence of or directly involved in the
176 regulation of gene expression concerning metabolic differentiation.

177 Together, these results indicate that the genetic compartmentalization of the
178 *Streptomyces* genome clearly correlates with a compartmentalization of transcription,
179 both sense and antisense.

180

181 **Compartmentalized architecture of the *Streptomyces ambofaciens*** 182 **chromosome in exponential phase**

183 We next asked the question, is a compartmentalized transcriptome correlated
184 with a specific chromosome folding. We first performed 3C-seq on cells harvested in
185 the exponential growth phase (**Fig.3, Fig.4.A**). The contact map displayed a main
186 diagonal reflecting the frequency of contacts, which extends up to 1 Mb (**Extended**
187 **Fig.3.A**). This diagonal contains loci acting as boundaries delimiting segments
188 (visualized as squares, **Fig.3.A, Fig.4.A**) reminiscent of CIDs in other bacteria²⁰⁻²².
189 To define these domains precisely in *S. ambofaciens*, we computed the ‘frontier
190 index’, a domain boundary indicator built from a multiscale analysis of the contact
191 map³⁷. Briefly, two indices are computed, reflecting the intensity of the loss of contact
192 frequencies when going downwards or upwards, respectively, to each genome
193 position (**Extended Fig.3.B**). In this context, a boundary is defined by a significant
194 change in both the downstream (green peaks, **Fig.3.B**) and upstream (orange peaks,
195 **Fig.3.B**) directions (see method section for details). In exponential phase in YEME
196 medium (non-optimal for antibiotic production, condition C6), we found 10 boundaries
197 that defined the central region, delimited by the first and last rDNA operons, 9

198 domains ranging in size from 240 kbp to 700 kbp (**Fig.3.B**). These central domains
199 resembled regular CIDs^{20,23} both in size and in the presumably nature of its formation
200 (see below). The central boundaries are also conserved in exponential phase in MP5
201 medium, which is optimized for antibiotic production (**Fig.4.A**). Only four additional
202 boundaries are observed in this medium (**Fig.4.A**). All rDNA operons coincide with
203 the sharpest boundaries (**Fig.3.B, Fig.4.A**). The genes surrounding these operons,
204 as well as those of other boundaries, are enriched in consecutive genes expressed at
205 very high level (**Fig.4.B, Supplementary Table 5**), as previously reported in other
206 bacterial models^{20,23}. Interestingly, in exponential phase, we report a clear correlation
207 between the level of gene conservation and the formation of boundaries (**Fig.4.C**).
208 Moreover, the transcription of the genes present within boundaries tends to be
209 oriented in the direction of continuous replication (odds ratio 1.8, p value $1.4 \cdot 10^{-8}$,
210 Fisher's exact test for count data), with a very low antisense index (**Extended Data**
211 **Fig.3.C**).

212 Genome extremities form two large terminal compartments of 1.46 and 2.46
213 Mbp, respectively, with different structural features (**Fig.3, Fig.4.A and Extended**
214 **Data Fig.3B**). Interestingly, within the terminal arms, the low transcriptional activity in
215 exponential phase in YEME medium correlates well with the absence of boundaries.
216 Of note, the boundaries present in each terminal domain in MP5 medium (one of
217 which is composed of an active prophage) do not contain highly expressed genes
218 and do not divide the terminal compartment into two parts (**Fig.4.A**). To compare the
219 dynamics of contacts within regions, we thus calculated a dispersion index, reflecting
220 the variability of the 3C-seq signal in each region (**Fig.4.D, Extended Data Fig.3.E**).
221 Despite that the plots for the probability of contacts as a function of genomic distance
222 are similar for both the terminal and central regions (**Extended Data Fig.3.D**), the

223 terminal ends present a higher dispersion index than the central compartment in both
224 media for contacts longer than 100 kbp (**Fig.4.D, Extended Data Fig.3.E**). This
225 indicates that contacts between loci in the terminal regions are more variable than in
226 the central compartment.

227 Thus, at the conformational level, the chromosome of exponentially growing
228 cells of *S. ambofaciens* is partitioned into three compartments: a central
229 compartment, actively transcribed and structured in multiple domains formed by the
230 high-level expression of persistent genes and two large terminal compartments rich in
231 GIs and SMBCGs and rather transcriptionally silent in which large-scale contacts are
232 highly variable. Since in exponential phase, the boundaries are very similar in YEME
233 and MP5 media and correspond mostly to persistent CDSs (**Fig.4.C**) or rDNA, their
234 orthologous sequences may also structure the spatial organization of the
235 chromosome in other *Streptomyces* species at a similar growth phase.

236

237 **Chromosome remodeling during metabolic differentiation**

238 As shown above, during metabolic differentiation we observed that the level of
239 transcription gradually increased within the terminal ends (**Fig.2**). To determine
240 whether transcription dynamics influence the 3D-organization of the chromosome, we
241 performed 3C-seq on *Streptomyces* cells during metabolic differentiation in MP5
242 medium (48 h, condition C4) (**Fig.4.E**). The frontier index revealed changes in the
243 boundaries along the primary diagonal; the boundaries identified in exponential
244 phase were not detected anymore after 48 h in the growth medium (**Fig.4.E**). By
245 contrast, the appearance of a very sharp boundary within the right terminal region
246 that divides this arm into two domains of 1,350 kpb and 1,000 kpb, correlates very
247 well with the increased level of transcription of the congocidine biosynthetic gene

248 cluster (BGC, **Fig.4.B & E**). In a replicate experiment, we observed that the
249 expression of another SMBGC, encoding the biosynthesis of a siderophore, could
250 generate the formation of an additional boundary (**Supplementary Fig.3.C**). This
251 illustrates some variability in the expression of SMBGCs during metabolic
252 differentiation. Together, these results highlight the correlation between SMBGC
253 expression and the formation of boundaries after 48 h. Accordingly, the level of
254 conservation of genes present within the boundaries switches from highly persistent
255 in exponential phase to poorly conserved during metabolic differentiation (**Fig.4.C**). In
256 addition, a second boundary detected within the right compartment is located in the
257 stambomycin BGC that is not or is poorly expressed in this condition. Interestingly,
258 accompanying the formation of boundaries and the increase of transcription in the
259 terminal compartments, the variability in the long-range contacts within the terminal
260 compartments is comparable to that of the central compartment (**Fig 4.D**).

261 The contact map also revealed the appearance of a secondary diagonal after
262 48 h of growth, indicating an increase in the frequency of contacts between regions
263 along the entire arms of the chromosome. This suggests that late in the cell cycle and
264 in the absence of a central compartment segmented into multiple domains, the two
265 chromosome arms are closer to each other (**Fig.4.E & F**). Remarkably, the second
266 diagonal is slightly tilted when it moves away from the origin, representing an
267 asymmetry in the contacts between loci within the terminal compartments.

268 In summary, our results indicate that the central and transcriptionally active
269 compartments in exponential phase present no boundaries after 48 h of growth,
270 whereas the terminal domains are locally remodeled concomitantly with the
271 expression of SMBGCs and with a decrease in the variability of the long-range
272 contacts. In addition, inter-arm contacts become more frequent. These results

273 indicate that metabolic differentiation is accompanied by major remodeling of the 3D-
274 architecture of the chromosome, both at the local and global levels (**Fig.4.F**).

275

276 **Discussion**

277 In this study, we demonstrated that in *Streptomyces*, compartmentalization of
278 gene organization, transcription and architecture are correlated. We explore the
279 transcriptional landscape of *S. ambofaciens* ATCC 23877 during metabolic
280 differentiation: its large genome is highly dynamic, most of the genes ($\approx 90\%$) being
281 significantly expressed in at least one condition. This situation is very similar to
282 *Bacillus subtilis*, another Gram-positive bacteria from soil³⁸. Interestingly, the
283 SMBGCs (e.g. antibiotic clusters) are generally considered as ‘cryptic’ under most
284 growth conditions^{10,39}. We observed that they are poorly expressed in exponential
285 phase. However their expression is characterized by up regulation after 48 h in the
286 growth medium, compared to the rest of the chromosome including other genes
287 putatively acquired by horizontal gene transfer (e.g. GIs). This result highlights that
288 SMBGCs in *S. ambofaciens* have evolved regulatory mechanisms, such as the
289 induction of cluster-situated transcriptional factor genes (**Supplementary Fig.2**) that
290 efficiently and specifically regulate gene expression during metabolic differentiation,
291 as reported in other *Streptomyces* species⁴⁻⁷.

292 Moreover, we show that changes in gene expression dynamics are correlated
293 with metabolic differentiation. Indeed, the terminal regions become transcriptionally
294 active after 48 h of growth (see also references^{10,36}) (**Fig.2**). Interestingly, the ratio,
295 antisense- over sense-transcription is high within the terminal compartments but
296 decreases over the growth period, especially in the SMBGCs (**Extended Data**
297 **Fig.2.F**). This suggests that this antisense-transcription may reflect regulatory

298 processes that remain to be explored. This is particularly interesting since some
299 NAPs suppress antisense-transcription in *S. venezuelae*⁴⁰.

300 In addition, 3C-seq analysis revealed that in exponential phase the central
301 compartment forms a multiple domain structure, delineated by boundaries, whereas
302 the terminal regions form two large compartments in which contacts are more
303 variable at larger distances. As previously shown for bacteria with circular
304 genomes^{20-23,26}, long and highly expressed genes (LHEGs encoding rRNA, ribosomal
305 proteins, or respiratory chain components) are found at the boundaries in the
306 *Streptomyces* chromosome (**Supplementary Table 5, Fig.4**). The positions of these
307 boundaries are both conserved (under certain circumstances) and dynamic, since
308 they change over the growth phase (²², this work). Notably, most of the boundaries
309 observed at early growth times were correlated with the presence of persistent genes
310 (**Fig.4C**). Within the central compartment, domains are a direct consequence of
311 boundary formation. Additionally, no trivial role in gene expression or in chromosome
312 conformation have been assigned to chromosome interacting domains (CIDs) in
313 bacteria⁴¹. Here we propose that boundaries are likely to act as transcriptional hubs
314 for clustered and persistent genes⁴², associated with very high levels of transcription
315 in exponential phase. Moreover, in exponential phase, the first and last rDNA
316 operons form the sharpest boundaries, recapitulating the borders that were arbitrarily
317 used to define the central region beyond which genome synteny falls (**Figure 1**). We
318 thus propose that they also constitute an evolutionary barrier that limits the
319 occurrence of single recombination events within the central region. Interestingly,
320 some boundaries in *Streptomyces* do not display high levels of gene expression
321 (**Supplementary Table 5**). These are generally located within the terminal arms,

322 suggesting the existence of other mechanisms, yet to be discovered that impose
323 chromosome constraints.

324 Furthermore, we show for the first time that the central and terminal
325 compartments present different organizational features, an observation
326 independently demonstrated for the *S. venezuelae* chromosome (see accompanying
327 paper, Szafran *et al.*). We favor the idea that the different organization of the terminal
328 compartments could be a consequence of multiple factors: i) the lack of constraints
329 imposed by active and sense-transcription; ii) the high level of anti-sense
330 transcription; iii) the enrichment of horizontally acquired NAPs and other
331 transcriptional regulators that, by increasing the dynamics of DNA intramolecular
332 interactions, ensure appropriate SMBGC repression in exponential phase.

333 Interestingly, it was shown in *S. venezuelae* that during sporulation the HupS NAP is
334 involved in promoting optimal contacts along the whole chromosome, but seems to
335 be particularly important for the organization of the terminal regions (see
336 accompanying paper, Szafran *et al.*).

337 During metabolic differentiation, the changes in transcriptional dynamics are
338 accompanied by a huge remodeling of chromosome folding, switching from an 'open'
339 to a 'closed' conformation, in which highly expressed SMBGC genes form new
340 boundaries (**Fig.4.E & F**). This may illustrate a relocation of transcriptional hubs from
341 clustered and persistent genes (in exponential phase) to SMBGCs at the onset of
342 metabolic differentiation. Remarkably, the central region is no longer structured into
343 multiple domains. The lack of boundaries in the central compartment after 48 h
344 growth is correlated with a slight decrease in transcripts from genes that were
345 present in boundaries in exponential phase. Indeed most of these genes remain
346 expressed at high level ('CAT-3' or 'CAT_4') but below a threshold classically

347 associated with boundaries (a string of genes with above 20 000 normalized reads
348 per kb, **Supplementary Table 5**). Moreover, we cannot exclude the possibility that
349 some of these transcripts are stable RNAs produced in exponential rather than
350 stationary phase⁴³. This loss of boundaries in the central compartment is reminiscent
351 of the phenomena observed in eukaryotes during G1-to-mitosis transition⁴⁴ and could
352 reflect a more general phenomenon of genome compaction during the cell cycle.

353 At larger scales, there is an increase in the frequency of inter-arm DNA
354 contacts. The most likely candidate to be involved in this reorganization is the Smc-
355 ScpAB condensin complex, which is recruited to the origin region by ParB bound to
356 *parS*^{45,46}. From the *parS* sites, Smc-ScpAB promotes inter-arm contacts by
357 translocating to the terminal region of the chromosome, likely by loop
358 extrusion^{20,21,24,27,47-52}. *Streptomyces* spp. encode both a ParABS system⁵³⁻⁵⁵ and a
359 Smc-ScpAB complex⁵⁶. Interestingly, during sporulation *Streptomyces* chromosomes
360 need to be compacted and segregated to ensure correct spore formation^{57 56}. Indeed,
361 HiC studies during sporogenic development in *S. venezuelae* showed that inter-arm
362 contacts are dependent on Smc-ScpAB (Szafran *et al.*, see accompanying paper).
363 The global remodeling of the *Streptomyces* chromosome going from an 'open' to a
364 'closed' conformation seems to occur similarly during metabolic (this work) and
365 sporogenic development (Szafran *et al.*, see accompanying paper). Such a global
366 rearrangement has not been previously described in wild-type bacteria. Interestingly,
367 we observed a slight tilt in the secondary diagonal in contact analysis when it
368 approaches to the terminal compartments (**Fig.4.E, Supplementary Fig.3.C**). This is
369 consistent with a slow-down of Smc-ScpAB activity when interacting with the
370 transcription machinery^{50,58}, in agreement with the fact that transcription in the

371 terminal compartment is higher than in the central compartment during metabolic
372 differentiation (**Fig.4.E, Supplementary Fig.3.C**).

373 Collectively, these results indicate a link between evolutionary processes,
374 including genome-compartmentalization and the molecular mechanisms (e.g.
375 transcription, 3D-folding) that shape the structure and function of genes and
376 genomes in *Streptomyces*. We therefore hypothesize that the efficiency of the
377 regulatory processes controlling conditional expression of SMBGCs may be an
378 emerging property of spatial compartmentalization. We believe that this study will
379 open new insights into setting the rules governing chromosome spatial organization,
380 expression and stability, and the optimal design of *Streptomyces* genomes for
381 SMBCG expression.

382

383 **Methods**

384 ***Streptomyces* strains and growth conditions**

385 *S. ambofaciens* ATCC 23877 was grown on solid soy flour-mannitol (SFM)
386 medium⁵⁹ at 28°C unless otherwise indicated. The strain was grown in the following
387 media: MP5 medium (7 g/l yeast extract, 20.9 g/l MOPS, 5 g/l NaCl, 1 g/l NaNO₃, 36
388 ml/l glycerol – pH 7.5)³⁴, or YEME medium 10.3 % sucrose (3 g/l yeast extract, 5 g/l
389 bactotryptone, 3 g/l malt extract, 10 g/l glucose, 103 g/l sucrose; pH 7.0-7.2 ; adapted
390 from⁵⁹). Twenty million spores of *S. ambofaciens* ATCC 23877 were inoculated in
391 100 ml of media before growth at 30°C in a shaking agitator (220 rpm, INFORS HT
392 multitron standard).

393

394 **Bioassays**

395 For bioassays from liquid cultures, 50 µl of filtered supernatants were spotted on
396 to agar medium using cylinder (diameter 0.5 cm) and allowed to dry until complete
397 penetration into a plate containing 50 ml of DIFCO antibiotic medium 5 (DIFCO
398 227710). Thereafter 7 ml of SNA medium (2 g/l agar, 8 g/l nutrient broth MP
399 Biomedicals Cat#1007917) containing *Micrococcus luteus* (final OD_{600nm} 0.04) were
400 overlaid on the plate and incubated at 37°C. The growth inhibition area was
401 measured 24 h later.

402 **Genomic island (GI) identification**

403 We designed the Synteruptor program ([http://bim.i2bc.paris-](http://bim.i2bc.paris-saclay.fr/synteruptor/)
404 [saclay.fr/synteruptor/](http://bim.i2bc.paris-saclay.fr/synteruptor/))⁶⁰ to compare the sequences of chromosomes of species close
405 enough to have synteny blocks, and to identify the genomic islands existing in each
406 respective chromosome. We define synteny breaks as genomic regions between two
407 consecutive synteny blocks when comparing two genomes, with the two blocks
408 having the same direction. The genome of *S. ambofaciens* ATCC 23877 was
409 compared to the chromosome of 7 closely related strains: *S. ambofaciens*
410 DSM40697, *S. coelicolor* A3(2), *S. lividans* 1326, *S. lividans* TK24, *S. mutabilis*,
411 *Streptomyces* sp. FXJ7.023, *Streptomyces* spp. M1013. To define a region
412 harbouring GIs, we used a threshold of 15 CDS as the minimal number of CDSs
413 within the synteny break, in at least one of the strains used for the pairwise
414 comparison. Of note, a GI can therefore correspond to 15 CDS in one strain but less
415 CDS in the other one. However, when the corresponding position within the *S.*
416 *ambofaciens* genome contained less than 2 CDS or only tRNAs, compared to at least
417 15 CDSs in the chromosome of the compared species, it was considered as an
418 insertion point (but not a GI) within *S. ambofaciens*. When the GIs were identified
419 during several pairwise comparisons of *S. ambofaciens* ATCC 23877 and/or

420 overlapping, they were fused and considered thereafter as a single GI. The complete
421 list of GIs identified in the *S. ambofaciens* genome is presented in **Supplementary**
422 **Fig. 1** and **Supplementary Table 2**.

423

424

425 **SMBGCs identification**

426 We used antiSMASH5.1.0³³ to identify putative SMBGCs. Thirty SMBGCs have
427 thus been identified in the *S. ambofaciens* genome, the one encoding kinamycin
428 biosynthesis being duplicated owing to its location within the TIRs. The definition of
429 cluster boundaries has been manually refined on the basis of literature data for
430 characterized SMBGCs (**Supplementary Table 4**). Of note, SMBGC genes are ≈ 1.5
431 times larger than the average, and only 13 of them (from 4 SMBGCs, namely
432 'CL4_Indole', 'CL10_Furan', 'CL11_NRPS', 'CL20_Hopenoid') belong to the core
433 genome.

434 **NAPSFs, identification**

435 In this study, we considered NAPs and chromosome structural factors *sensu lato*
436 by including orthologues of classical NAPs and structural factors (HU, sIHF, Lsr2,
437 Lrp/AsnC, SMC, Dps, CbpA, DnaA or IciA family proteins)⁶¹ and/or proteins
438 associated with *S. coelicolor* chromatin⁶². This list is available in **Supplementary**
439 **Table 5**.

440 **Definition of indexes for genome conservation analyses**

441 We selected 125 *Streptomyces* genomes from the NCBI database representative
442 of the *Streptomyces* genus by keeping only complete genomes of distinct species.
443 When genomes share an average nucleotide identity (ANIb) value greater than or
444 equal to 96 %, they are considered as members of the same species³¹

445 **(Supplementary Table 1)**. We made one exception by keeping two strains of *S.*
446 *ambofaciens* (ATCC 23877 and DSM 40697). Orthologous genes were identified by
447 BLASTp best hits⁶³ with at least 40 % of identity, 70 % coverage and an E-value
448 lower than $1e^{-10}$ ³¹. The gene persistence index was calculated as N_{orth}/N , where N_{orth}
449 is the number of genomes carrying a given orthologue and N the number of genomes
450 searched²⁸. Pairwise comparisons are achieved using a sliding window and
451 comparing a reference strain (species A) to another (species B). Here is considered a
452 window containing 8 genes noted from 1 to 8 in the reference species. To calculate
453 the GOC index (Gene Order Conservation), which is the number of contiguous
454 syntenic orthologs between the two chromosomes over the number of orthologs
455 between the two chromosomes, calculated in a sliding window, the pair of contiguous
456 genes present in the window in the reference species are searched and counted as
457 contiguous pairs of orthologues in the whole genome of species B related to the
458 number of orthologs is defined by the window.

459 **Transcriptome analysis**

460 For RNA-seq analysis performed in liquid cultures, $\approx 2 \cdot 10^7$ spores of *S.*
461 *ambofaciens* ATCC 23877 were inoculated in 100 ml of liquid medium. Thereafter, 25
462 ml (for samples harvested after 24 h growth) or 10 ml of cultures (for other time
463 points) were added to an equal volume of cold ethanol.

464 Cells were then harvested by centrifugation for 15 min at 4,000 *g* at 4°C, and
465 stored at -20°C. Pellets were washed with 1 ml of DEPC water, centrifuged for 5 min
466 at 16,000 *g* at 4°C and homogenized with glass beads ($\leq 106 \mu\text{m}$; G4649, SIGMA) in
467 350 μl of lysis buffer (RNeasy Plus Mini Kit, QIAGEN) supplemented with 10 $\mu\text{l/ml}$ β -
468 mercaptoethanol. Samples were processed 3 times for 45 sec each in FastPrep-24TM
469 5G instrument (MP Biomedicals) at setting 6 with 1 min cooling between the stages.

470 After centrifugation for 10 min at 16,000 *g* at 4°C, total RNAs were isolated from the
471 supernatants using an RNeasy Plus Mini Kit (QIAGEN) and gDNA Eliminator
472 columns, following the manufacturer's recommendations. To remove genomic DNA,
473 RNA samples were incubated for 30 min at 37°C with 20 U of RNase-free DNase I
474 (Roche) in a final reaction volume of 30 µl. RNAs were then cleaned up using the
475 RNeasy Mini Kit (QIAGEN), following the manufacturer's recommendations. The
476 absence of DNA in the preparations was checked by PCR on an aliquot. RNA
477 samples were quantified using Qubit™ RNA HS Assay kit (ThermoFischer Scientific),
478 following manufacturer's recommendations.

479 Total RNA quality was assessed in an Agilent Bioanalyzer 2100, using RNA 6000
480 pico kit (Agilent Technologies). 500 ng of total RNA were treated with DNase
481 (Baseline Zero DNase, Epicentre) prior to ribosomal RNA depletion using the
482 RiboZero bacteria magnetic Kit from Illumina according to the manufacturer's
483 recommendations. After the Ribo-Zero step, the samples were checked in the Agilent
484 Bioanalyzer for complete rRNA depletion. Directional RNA-seq libraries were
485 constructed using the Illumina ScriptSeq kit V2 (discontinued) for samples
486 corresponding to C5 conditions and Illumina Stranded library preparation kit for all
487 other samples, according to the manufacturer's recommendations.

488 Libraries were pooled in equimolar proportions and sequenced (Paired-end 2x43
489 bp) with an Illumina NextSeq500 instrument, using a NextSeq 500 High Output 75
490 cycles kit. Demultiplexing was done (bcl2fastq2 V2.2.18.12) and adapters
491 (adapter_3p_R1: AGATCGGAAGAGCACACGTCTGAACT; adapter_3p_R2:
492 AGATCGGAAGAGCGTCGTGTAGGGA) were trimmed with Cutadapt1.15, only
493 reads longer than 10 bp were kept.

494 STAR software⁶⁴ was used for mapping RNA-seq to the reference genome
495 (genome-build-accession NCBI_Assembly: GCF_001267885.1) containing only one
496 terminal inverted repeat (TIR). This avoids any biases with multiple mapping within
497 the duplicated extremities of the genome (since the two TIR sequences are
498 indistinguishable). We used the *featureCounts* program⁶⁵ to quantify reads (in sense-
499 and antisense- orientation) generated from RNA-sequencing technology in terms of
500 “Gene” feature characteristics of *S. ambofaciens* ATCC 23877 most recent
501 annotation (GCF_001267885.1_ASM126788v1_genomic.gff – released on the
502 06/15/2020).

503 **Bioinformatic analysis of RNA-seq count data**

504 SARTools (Statistical Analysis of RNA-Seq data Tools) DESeq2-based R
505 pipeline⁶⁶ was used for systematic quality controls and the detection of differentially
506 expressed genes. PCA and sample clustering used homoscedastic data transformed
507 with Variance Stabilizing Transformation (VST). For the differential analysis, the
508 Benjamini and Hochberg's method was used with a threshold of statistical
509 significance adjusted to 0.05. The reference condition was C1 (medium “MP5_24h”).
510 All parameters of SARTools that have default values were kept unchanged. Genes
511 with null read counts in the 37 samples were not taken into account for the analysis
512 with DESeq2. Genes corresponding to the second TIR (not used for the mapping) or
513 the rRNAs (RiboZero treatment) were excluded from further analysis.

514 **Supplementary Table 5** presents the normalized counts per gene in each growth
515 condition. To ensure that antisense-transcription did not affect the normalization of
516 sense-transcription data, the latter were normalized as follows: the percentage of
517 transcription in antisense orientation of each gene was determined with the raw data,
518 then the normalized antisense-transcription was evaluated by applying this

519 percentage to the normalized sense-transcript counts. To determine the category of
520 gene expression the normalized counts obtained by the SARTools DESeq2-based
521 pipeline were again normalized with respect to gene size (number of DESeq2
522 normalized reads/gene size x 1000 bp), and thereafter expressed in RPK (reads per
523 kb). This allowed us to compare directly the relative levels of gene expression of
524 individual genes. Data were analyzed with R software⁶⁷ and the Integrative Genomics
525 Viewer (IGV) tool was used to simultaneously visualize RNAseq data and genomic
526 annotations⁶⁸.

527 **Multidimensional analyses of the data**

528 We used the FactoShiny R package⁶⁹ to perform clustering, principal component
529 analyses and correspondence analyses. This overlay factor map presented in
530 **Extended Data Fig.2.C** results from the correspondence analysis performed using a
531 contingency table indicating the number of genes in each category of expression
532 level depending on the genome features of interest. Max', 'Mean' and 'Min' refer to
533 the maximal, mean and minimal expression levels in all studied condition, from the
534 lowest category ('0') to the highest ('4'). These categories were defined by
535 considering the distribution parameters of the normalized number of reads
536 (**Extended Fig.2.A.**). These gene expression categories as well as the number of
537 genes switched ON ('Switch', meaning that the expression level switches from
538 CAT_0 in at least one condition to CAT_3 or more in another condition), or
539 presenting a very low ('AS < 0.05') or high ('AS > 0.5') level of antisense index were
540 used to build the map, and projected on the plan. Here are some basic explanations
541 of how to interpret the results of correspondence analysis. On **Extended Data**
542 **Fig.2.C**, the squares correspond to the raw labels (genome features). The triangles

543 correspond to the column labels (e.g. category of level of mean, maximal, minimal
544 sense-transcription, antisense index < 0.05, antisense index > 0.5).

545 **Chromosome conformation capture (3C)**

546 Spores of *S. ambofaciens* ATCC 23877 ($\approx 4 \cdot 10^7$) were inoculated in 200 ml of
547 MP5 or YEME liquid medium. In these media, bacterial growth was monitored by
548 opacimetry. At the indicated time point, cells were harvested from 100 ml samples of
549 culture, adjusted to an OD_{600 nm} of about 0.15 (with the appropriate fresh medium)
550 and fixed by adding formaldehyde solution (3 % final concentration). Cells were then
551 incubated under moderate agitation for 30 min at RT and 30 min more at 4°C.
552 Glycine (250 mM final concentration) was added, and the bacteria incubated 30 min
553 at 4°C with moderate agitation. Cells were then harvested by centrifugation for 10
554 min at 4,000 g at 4°C. The cells were gently suspended in 50 ml of PBS 1X and then
555 again harvested by centrifugation for 10 min at 4,000 g at 4°C. This washing step
556 was repeated once before suspending the cells in 1 ml of PBS before final harvesting
557 by 10 min at 4,000 g at 4°C. The dry pellets were stored at -80°C until use.

558 Frozen pellets of exponentially grown cells were thawed, suspended in 600 µl Tris
559 10 mM EDTA 0.5 mM (TE) (pH 8) with 4 µl of lysozyme (35 U/µl; Tebu Bio) and
560 incubated at RT for 20 min. For the samples collected after 48 h, the pellets were
561 thawed, suspended in TE (pH 8) with 4 µl of lysozyme (35 U/µl; Tebu Bio) for 45 min
562 and then homogenized with a Bioruptor sonication device (3 cycles of 30 seconds,
563 with a pause of 30 second pause for each). Then, for both 24 h and 48 h samples,
564 SDS was added to the mix (final concentration 0.5%) of cells and incubated for 10
565 minutes at RT. 500 µl of lysed cells were transferred to a tube containing 4.5 ml of
566 digestion mix (1X NEB 3 buffer, 1% triton X-100) and 100 µl of the lysed cells were
567 transferred to a tube containing 0.9 ml of digestion mix. 800 units of Sall were added

568 to the 5 ml digestion mix. Both tubes were then incubated for 2 h at 37°C and 250
569 units of Sall were added to the 5 ml tube and further incubated, 2 h at 37°C. To stop
570 the digestion reaction, 4 ml of the digestion mix were immediately centrifuged for 20
571 min at 20,000 g, and pellets suspended in 4 ml of sterile water. The digested DNA (4
572 ml in total) was split into 4 aliquots and diluted in 8 ml ligation buffer (1X ligation
573 buffer NEB 3 (without ATP), 1 mM ATP, 0.1 mg/ml BSA, 125 Units of T4 DNA ligase
574 5 U/μl). Ligation was performed at 16°C for 4 h, followed by incubation overnight at
575 65°C with 100 μl of proteinase K (20 mg/ml) and 100μl EDTA 500 mM. DNA was
576 then precipitated with an equal volume of 3 M Na-Acetate (pH 5.2) and two volumes
577 of iso-propanol. After one hour at -80°C, DNA was pelleted and suspended in 500 μl
578 1X TE buffer. The remaining 1ml digestion mix with or without Sall were directly
579 incubated with 100 μl of proteinase K (20 mg/ml) overnight at 65°C. Finally, all the
580 tubes were transferred into 2 ml centrifuge tubes (8 tubes), extracted once with 400
581 μl phenol-chloroform pH 8.0, precipitated, washed with 1 ml cold ethanol 70% and
582 diluted in 30 μl 1X TE buffer in the presence of RNase A (1 μg/ml). Tubes containing
583 the ligated DNA (3C libraries) were pooled. The efficiency of the 3C preparation was
584 assayed by running aliquots of the 3C-librairies, the digested DNA or the non-
585 digested DNA on 1% agarose gel. Finally the 3C libraries were quantified on the gel
586 using QuantityOne software (BioRad).

587 **Processing of libraries for Illumina sequencing**

588 Approximately 5 μg of a 3C library was suspended in water (final volume 130 μL)
589 and sheared using a Covaris S220 instrument (Duty cycle 5, Intensity 5, cycles/burst
590 200, time 60 sec for 4 cycles). The DNA was purified using a Qiaquick® PCR
591 purification kit, DNA ends were prepared for adapter ligation following standard
592 protocols⁷⁰. Custom-made adapters²¹ were ligated overnight at 4°C. Ligase was

593 inactivated by incubating the tubes at 65°C for 20 min. To purify DNA fragments
594 ranging in size from 400 to 900 pb, a PippinPrep apparatus (SAGE Science) was
595 used. For each library, one PCR reaction of 12 cycles was performed using 2 to 3 µl
596 of 3C library, 0.2 µM Illumina primers PE1.0 and PE2.0 and 1 unit of Taq Phusion
597 [Finnzymes]. The PCR product was purified on Qiagen MinElute columns and dimers
598 of primers were removed from the 3C library by using AMPure XP beads following
599 the manufacturer's protocol (Beckman Coulter). Finally, libraries were subjected to 75
600 bp paired-end sequencing in an Illumina sequencer (NextSeq500).

601 **Processing of sequencing data**

602 PCR duplicates from each 3C library sequence dataset were discarded using the 6
603 Ns of custom-made adapters²¹. Reads were aligned independently using Bowtie 2 in
604 very sensitive mode⁷¹. Only reads with mapping quality > 30 were kept to establish
605 contact maps.

606 **Generation of contact maps**

607 Contact maps were built as described previously⁷². Briefly, each read was
608 assigned to a restriction fragment. Non-informative events such as self-circularized
609 restriction fragments, or uncut co-linear restriction fragments were discarded⁷³. The
610 chromosome of *S. ambofaciens* devoid of the TIRs was divided into 10 kbp bins and
611 the frequencies of contacts between genomic loci for each bin were assigned.
612 Contact frequencies were visualized as heatmaps. Raw contact maps were
613 normalized with the sequential component normalization procedure (SCN⁷³). To
614 facilitate visualization, contact matrices are visualized as log matrices. First we
615 applied to the SCN matrices the log₁₀ and then a Gaussian filter (H=1) to smooth the
616 image. The contact score for a pair of bins that due to mapping was identified as an

617 outlier (z-score > contact score) was replaced with the median value of the contact
618 matrix.

619 **Frontier Index determination**

620 To analyze the domain organization of *S. ambofaciens*, we used the frontier index
621 (FI) method that quantifies the involvement of each bin in the frontier of any domain,
622 i.e. at any scale – see technical details in³⁷. Briefly, this method consists, first, in
623 computing derivatives of the contact maps. To this end, we used normalized contact
624 maps where the average contact frequency, $P(s)$, was subtracted to reduce noise
625 coming from the natural overall decrease of contact frequencies as the genomic
626 distance (s) between loci increase. Here, we also added pseudocounts (equal to
627 $m/20$ where m is the maximal value of the contact frequencies) such that all values of
628 the contact maps are strictly positive and we considered the logarithm of the resulting
629 maps in order to mitigate the strongest variations close to the diagonal. We then
630 considered separately the derivatives of these maps along the vertical and horizontal
631 axes, whose large positive values (in the upper part of the maps) are respectively
632 associated with upstream and downstream frontiers of domains, respectively. After
633 setting negative values to zero (to reduce noise even more), for each bin the
634 corresponding upstream and downstream FIs were defined as the sum over all other
635 bins at a distance below 600 kb of the resulting signals. We then computed the
636 profile of upstream and downstream FIs. We considered peaks that are located
637 above the median of the peak values plus two times the standard deviation
638 (estimated by the 1.48 times the median absolute deviation to mitigate the impact of
639 outliers). These statistics were computed separately for the left terminal arm, the
640 central region and the right terminal arm as these compartments have distinct

641 statistical 3C-seq features. Altogether, this led us to a list of significant upstream
642 (orange) and downstream (green) peaks, respectively.

643 A boundary was then allocated to any bin (as well as to pairs of consecutive bins
644 to cope with uncertainty of peak positions) for which both the upstream and
645 downstream peaks were significant. To analyze the genetic environment associated
646 with a boundary, we considered the environment around the boundary identified by
647 the frontier index (± 2 bins). Note that due to intrinsically noisy nature of 3C-seq data,
648 we detected bins with only an upstream or a downstream significant peak
649 (**Supplementary Fig.3.D**). These bins with an ‘orphan’ peak were not associated
650 with a boundary, and hence, were removed from the analysis.

651 **The dispersion index**

652 The dispersion index of a signal reflects the range of variations relative to the
653 mean value of the signal. Here, we compute it for the frequency of contacts. More
654 precisely, it is defined as: $I(s) = var(s)/P(s)$, where $P(s)$ and $var(s)$ are, respectively,
655 the mean value and variance of the frequency of contacts between bins separated by
656 the genomic distance. These quantities are then computed separately in each
657 compartment of the genome. The larger the dispersion index is, the more variable are
658 the contact frequencies within a compartment.

659 **Statistical procedure**

660 Data were analyzed with R software⁶⁷. For the RNA-seq analyses, three independent
661 experiments (performed on different days) were carried out for each studied
662 condition, except for C3, C9 and C10 that were performed in duplicate and C5 in
663 quadruple. The statistical significance of RNA-seq analysis was assessed using the
664 SARTools DESeq2-based pipeline⁶⁶. For the 3C-seq analyses, two independent
665 experiments (performed on different days) were carried out for each condition. To

666 quantify gene regionalization, we defined the left and right terminal arms using as
667 limits the first (genome position: 1,469,670-1,474,823 bp) and last (genome position:
668 5,802,471-5,807,629 bp) rDNA operons, respectively. The statistical significance was
669 assessed by means of Fisher's exact test for count data, which is appropriated to the
670 analysis of contingency tables.

671 **Data availability**

672 The datasets generated during this study have been deposited in the NCBI Gene
673 Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) under the accession
674 number GSE162865.

675 **List of abbreviations**

676 3C: Chromosome Conformation Capture

677 “Actino. Sign.”: Actinobacterial Signature

678 BGC: Biosynthetic Gene Cluster

679 CDS: coding sequence

680 CID: Chromosome Interacting Domain

681 GI: Genomic island

682 IGV: Integrative Genomics Viewer

683 LHEG: long and highly expressed genes

684 NAP: Nucleoid Associated Protein

685 NAPSFs: Nucleoid Associated Proteins and structural factors

686 PCA: Principal Component Analysis

687 rDNA: ribosomal RNA encoding DNA

688 RPK: Read per kb

689 SARTools: Statistical Analysis of RNA-Seq data Tools

690 SMBGC: Specialized Metabolite Biosynthetic Gene Cluster

691 TIR: Terminal inverted repeat

692 VST: Variance Stabilizing Transformation

693

694 **Acknowledgments**

695 We acknowledge the High-throughput sequencing facility of I2BC for its sequencing
696 and bioinformatics expertise. We thank Barry Holland and Christophe Possoz for
697 careful reading of the manuscript and the members of F. B. and S. L. for fruitful
698 discussions and advice.

699

700

701 **Figure legend**

702 **Figure 1: The genetic compartmentalization of *S. ambofaciens* linear**
703 **chromosome**

704 The terminal inverted repeats (TIRs, 200 kb) and the origin of replication (*oriC*) are
705 indicated in black. The distribution of gene presenting features of interest are
706 represented in blue or red when enriched in the central region or within the extremities
707 (defined by the first and last rDNA operons), respectively (see **Supplementary Table**
708 **3** for a statistical analysis of the data). The dot and asterisk indicate the position of
709 pSAM2 and of a complete prophage, respectively. The position of the SMBGC
710 encoding the biosynthesis of all known antibacterial compounds of *S. ambofaciens*
711 ATCC 23877 is also indicated. The level of synteny along the chromosome of *S.*
712 *ambofaciens* ATCC 23877 is represented as a heat map of GOC (Gene Order
713 Conservation) scores using a sliding window (8 CDSs with 1 CDS steps). Each line
714 corresponds to the GOC profile of the reference against another species. The species
715 are organized from the phylogenetically closest to the furthest compared to *S.*
716 *ambofaciens*. The vertical black dotted line represents the location of the *dnaA* gene
717 delineating the two replichores. “NA” (in yellow) indicates the absence of orthologs in
718 the corresponding window. *S. ambofaciens* ATCC 23877 also harbors a circular
719 pSAM1 plasmid (≈ 89 kbp).

720 Abbreviations: ‘Actino. Sign. CDSs’ (coding sequences of the actinobacterial
721 signature); BGC (biosynthetic gene cluster); ‘C.’ (congocidine BGC); ‘Chr.’ (whole
722 chromosome); GIs (genes belonging to genomic islands); ‘K.’ (kinamycin BGC);
723 ‘Persist._{>0.95} CDSs’ (coding sequences of *S. ambofaciens* ATCC 23877 presenting a
724 gene persistence superior to 95 % in 124 other *Streptomyces* genomes); ‘Transl.
725 CDSs’ (genes encoding functions involved in translation process and/or RNA stability);

726 NAPSFs (nucleoid associated proteins and structural factors); SMBGCs (specialized
727 metabolite BGCs); 'Sp.' (spiramycin BGC); 'St.' (stambomycin BGC); 'rRNA'
728 (ribosomal RNA).

729

730 **Figure 2: Transcriptome dynamics depending on genome features and**
731 **metabolic differentiation**

732 **A. Transcriptome changes over growth.** Heatmaps of the normalized number of
733 reads per gene, ranked by their expression level at 24 h.

734 **B. Transcription along the chromosome over growth.** The normalized counts
735 (DESeq2 analysis) were mapped on *S. ambofaciens* chromosome. The two
736 clusters to which the different conditions belong were obtained by hierarchical
737 classification (**Extended Data Fig.1C**). Whereas the cluster 1 corresponds to
738 exponential phases (no ATB production), the cluster 2 gathers conditions
739 associated to the activation of SMBGC encoding at least two known antibacterial
740 activities (**Supplementary Fig.2**). 'C1' to 'C7' refer to the name of the studied
741 conditions: C1 to C5, cells grown in MP5 during 24 h, 30 h, 36 h, 48 h and 72 h,
742 respectively; C6 and C7: cells grown in YEME during 24h and 48h, respectively
743 (see **Table 1**, for details).

744 **C. Boxplot presenting the mean level of gene expression depending on**
745 **features of interest.** The distribution of the number of counts (normalized by
746 DESeq and on gene size) as the mean of all studied conditions is presented
747 depending on the genome features of interest (see the legend of **Figure 1** for
748 abbreviations). For clarity, outliers were excluded from this graphical
749 representation (but were taken into account for the numerical exploitation of the
750 data). The significant differences by Wilcoxon rank sum test with continuity

751 correction compared to the chromosome ('Chr.') are indicated by asterisks (**,
752 $p < 0.01$, ***, $p < 0.001$). Other abbreviation: 'Func. RNA (-rRNA)' (Functional
753 RNA excluding rRNA)

754

755 **Figure 3: Spatial organization of *S. ambofaciens* chromosome and**
756 **transcriptome in absence of metabolic differentiation**

757 3C-seq and RNA-seq were performed on *S. ambofaciens* grown in exponential phase
758 in YEME medium (24 h, C6 condition). The normalized contact map was obtained from
759 asynchronous populations. The x and y axes represent the genomic coordinates of the
760 reference genome. To simplify the analyses, TIRs were removed from the reference
761 genome. In **panel A**, the colour scale reflects the frequency of the contacts between
762 genome loci, from white (rare contacts) to dark purple (frequent contacts). In **panel B**,
763 the frontier index analysis is able to detect, for a given genome 10 kb bin, the change
764 in the contact bias with their neighbouring bins. Thus, a boundary is defined as any bin
765 in which there is a change in the right bias of contacts towards the left bias (± 2 bin,
766 green and orange peaks, respectively). Red and black circles indicate the position of
767 rDNAs and highly expressed genes (HEGs), respectively. In **panel C**, the normalized
768 counts (DESeq2 analysis) measured in cells grown in the same condition were mapped
769 on *S. ambofaciens* chromosome and binned in 10 kb. The **panel D** highlights genomic
770 and transcriptional features of interest. The right and left compartment were defined
771 owing the outer boundaries of the central compartment, which correspond to the first
772 and last rDNA operon position. The 'synteny-break' right arm corresponds to the
773 beginning of the spiramycin BGC. The replicate of this experiment is presented in the
774 **Supplementary Fig. 3.A.**

775

776 **Figure 4: Chromosome remodeling during metabolic differentiation**

777 **A. 3C-seq and RNA-seq performed after 24 h in MP5 growth medium (C1**
778 condition). Same legend than **Figure 3**. The replicate of this experiment is
779 presented in the **Supplementary Fig. 3.B**.

780 **B. Level of gene expression depending on gene location within boundaries**
781 **or domains**. The box-plot presents the sense-transcription (normalized number of
782 read per Kb) depending on gene location in exponential phase either in MP5 (C1)
783 or in YEME (C6), or after 48 h of growth in MP5 medium (C4). Boundaries in
784 common between MP5 24 h and YEME 24 h conditions ('Common C1 & C6') were
785 also analyzed separately. The significant differences by Wilcoxon rank sum test
786 with continuity correction (boundaries versus domains) are indicated by asterisks
787 (**, $p < 0.01$; ***, $p < 0.001$).

788 **C. Level of gene persistence depending on gene location within boundaries**
789 **or domains**. The box-plot presents the persistence index depending on gene
790 location in exponential phase (C1: MP5; C6: YEME) or after 48 h of growth in MP5
791 medium (C4). Boundaries in common between MP5 24 h and YEME 24 h
792 conditions ('Common C1 & C6') were also analyzed separately. The significant
793 differences by Wilcoxon rank sum test with continuity correction (boundaries versus
794 domains) are indicated by asterisks (***, $p < 0.001$).

795 **D. Dispersion index of the left, central and right compartments plotted as a**
796 **function of genomic distance**. The index of dispersion ($Var(s)/P(s)$) reflects the
797 range of variations relative to the mean value as a function of the genome distance
798 for the contact maps obtained for cells grown 24 h and 48 h in MP5 medium. Long-
799 range DNA contacts within the terminal compartments (> 100 kbp) are more
800 variable than within the central compartment.

801 **E. 3C-seq and RNA-seq performed after 48 h in MP5 growth medium (C4**
802 **condition).** Same legend than **Figure 3.** The arrow indicates the position of the
803 congocidine biosynthetic gene cluster (BGC). The replicate of this experiment is
804 presented in the **Supplementary Fig.3.C.**

805 **F. Schematic representation of genome dynamics during metabolic**
806 **differentiation.** The *Streptomyces* linear chromosome is represented as dark gray
807 (central compartment) and dark red sticks (left and right compartments). The origin
808 of replication is indicated as a light yellow circle (*oriC*). In exponential phase, the
809 chromosome of *Streptomyces* is organized into one transcriptionally active (light
810 green) central compartment and two rather silent compartments (pink grey). The
811 transcriptionally active compartment is enriched in core genes and is segmented in
812 multiple domains, flanked by long and highly expressed conserved genes. The low
813 frequency of contacts between the two arms is represented as an open
814 conformation of the chromosome. After 48 h of growth, the transcription of the
815 congocidine cluster is accompanied by a global remodeling of the chromosome.
816 The increase in the frequency of contacts between the two chromosomal regions
817 around the origin is represented as a closed conformation. Interestingly, a slight tilt
818 on the secondary diagonal towards the ends of the chromosome shows an
819 asymmetry between contacts within the terminal ends that was not schematized in
820 the figure. In addition, since TIRs cannot be distinguished at the sequence level,
821 they were neither considered in this analysis nor represented in this figure. However
822 they may be located close to each other⁷⁴, although this remains controversial⁷⁵.

823

824

825

826 **Extended Data Figure 1: The metabolic differentiation of *Streptomyces***

827 ***ambofaciens* analyzed at the transcriptional level**

828 **A. Growth curves of *Streptomyces ambofaciens* in MP5 and YEME liquid**

829 **media.** In these media, *S. ambofaciens* grows under a rather dispersed way,
830 so that growth can be monitored by optical density (OD_{600nm}). We although
831 refer to “pseudo-opacymetry” because of the presence of some clumps of
832 bacteria. Values are means ± standard errors (error bars) from 3 independent
833 experiments.

834 **B. Bioassays performed in conditions used in this study.** Bioassays

835 performed against *Micrococcus luteus* with the supernatants of *S.*
836 *ambofaciens* ATCC 23877 grown in MP5 or YEME liquid media and harvested
837 at different time points. Non-inoculated media were used as negative controls
838 (“T-“). ‘C1’ to ‘C7’ refer to the name of the studied conditions (Table 1). The
839 width of the plates is 12 cm. The presence of a halo indicates the presence of
840 one or more antibacterial activities. The expression of the SMBGCs encoding
841 known ATB is presented in **Supplementary Fig.2.**

842 **C. Multidimensional analyses of the studied conditions. Left panel - Axes 1**

843 **and 2 of the principal component analysis (PCA)** of the whole data set (21
844 samples), with percentages of variance associated with each axis. The
845 numbers indicate the condition (as indicated in Table 1) with a distinct letter (a,
846 b, c, d) for each replica. **Right panel- Ascending hierarchical classification**
847 (Euclidean distance, Ward criterion) and its inertia gain. The conditions form 2
848 clusters (1 and 2) and 5 subclusters which are boxed. The cluster 1
849 corresponds to the earliest growth time points (24 h, 30 h and 36 h) in which
850 no ATB production was detected by bioassay. The cluster 2 corresponds to

851 conditions associated to metabolic differentiation and the latest growth time
852 points. Its sub-clusters separate owing the growth time and medium: 48 h in
853 MP5 (subcluster 2.1), 72 h in MP5 (subcluster 2.2), 48 h in YEME (subcluster
854 2.3).

855 **D. MA-plot of the data for the comparison of transcriptomes over growth**
856 **and over media.** The MA-plot represents the log ratio of differential
857 expression as a function of the mean intensity for each pairwise comparison.
858 The differentially expressed features are highlighted in red, and the number of
859 up- and down-regulated genes indicated on each graph. Triangles correspond
860 to features having a too low/high $\log_2(\text{fold change})$ to be displayed on the plot.

861

862 **Extended Data Figure 2: The dynamic of *Streptomyces ambofaciens***
863 **transcriptome**

864 **A. Distribution of the normalized read counts per gene and definition of**
865 **gene expression level categories.** The boxplot represents the
866 distributions of the read counts per gene (normalized by DESeq and on
867 gene size, RPK) individually for each tested condition and as the mean
868 expression of each gene in all tested conditions. For clarity outliers were
869 excluded from this graphical representation (but were taken into account
870 for the numerical exploitation of the data). The genes were classified owing
871 their expression level by considering the distribution parameters as
872 indicated. These parameters were calculated both for each condition
873 individually, and for the mean expression in all tested condition.

874 **B. Pattern of gene expression depending on genome features of**
875 **interest.** The genes were classified owing their mean (upper panel) or their

876 minimal of maximal (lower panel) expression level in all studied conditions,
877 as described in panel A. The table presents the statistical analysis of the
878 data (Fisher's exact test for count data).

879 **C. Multidimensional analysis (overlay factor map) of genomic and**
880 **transcriptomic data.** The squares and triangles correspond to genomic and
881 transcriptomic features of interest, respectively (see abbreviations below).
882 The closer the squares or triangles are to the origin of the graph, the closer
883 they are likely to be to the mean data. Proximity between triangles or
884 between squares indicates similarity. A small angle (e.g. less than 30°)
885 connecting a triangle and a square to the origin indicate that they are
886 probably associated. If they are located at the opposite, they are probably
887 negatively associated. The proximity between the variables 'SMBGCs' and
888 'Switch' is due to the fact that a high proportion (19.3 %) of the SMBGC
889 genes switched on at the transcriptional level in at least one studied
890 condition (compared to 4.4 % at the level of the whole genome, odds ratios
891 5.2, p -value $< 2.2 \cdot 10^{-16}$; 7.6 % of genes within GIs, odds ratios 2.9, p -value
892 $2.2 \cdot 10^{-9}$, Fisher's exact test for count data).

893 Genomic feature abbreviations: 'Actino. Sign. CDSs' (coding sequences of
894 the actinobacterial signature); 'Chr.' (whole chromosome); 'Func. RNA (-
895 rRNA)' (Functional RNA excluded rRNA); GIs (genes belonging to genomic
896 islands); 'Persist._{0.95} CDSs' (coding sequences of *S. ambofaciens* ATCC
897 23877 presenting a gene persistence superior to 95 % in 124 other
898 *Streptomyces* genomes); 'Transl. CDSs' (genes encoding functions involved
899 in translation process and/or RNA stability); NAPSFs (nucleoid associated
900 proteins and structural factors); SMBGCs (specialized metabolite BGCs).

901 Expression feature abbreviations: ‘AS < 0.05’ and ‘AS > 0.5’ correspond to
902 very low (below 0.05) or very high (above 0.5) antisense-transcription
903 indices. ‘Max’, ‘Mean’ and ‘Min’ refer to the maximal, mean and minimal
904 expression levels in all studied condition, from the lowest category (‘0’) to
905 the highest (‘4’). ‘Switch’ correspond to the expression level switch from
906 CAT_0 in at least one condition to CAT_3 or more in another condition.

907 **D. Correlation between gene persistence and the transcription in sense**
908 **orientation.** The correlation between gene persistence and the mean
909 number of counts (normalized by DESeq2 and on gene size, RPK) in
910 sense- orientation for each gene (over all studied conditions) was analyzed
911 by a Spearman’s rank correlation test (not sensitive to the log2 conversion
912 of the data).

913 **E. Correlation between sense- and antisense-transcriptions.** The
914 correlation between the mean number of counts (normalized by DESeq2
915 and on gene size, RPK) in sense- and antisense- orientation for each gene
916 (over all studied conditions) was analyzed by a Spearman’s rank
917 correlation test (not sensitive to the log2 conversion of the data).

918 **F. Boxplot presenting the antisense index over growth in MP5 medium**
919 **depending on feature of interest.** The distribution of the antisense index
920 defined as the number of counts in antisense over total counts (in sense-
921 and antisense- orientation) is presented depending on the genome features
922 of interest (see panel C for abbreviations). For clarity outliers were excluded
923 from this graphical representation (but were taken into account for the
924 numerical exploitation of the data). The significant differences by Wilcoxon
925 rank sum test with continuity correction are indicated by asterisks (*, $p < 0.05$;

926 **, $p < 0.01$; ***, $p < 0.001$), in grey when comparing results after 24 h *versus*
927 48 h of growth for a given genome feature of interest, in violet and green
928 when comparing a genome feature of interest to the whole set of
929 chromosomal genes after 24 h *versus* 48 h of growth, respectively.

930

931 **Extended Data Figure 3: Data exploration linking the chromosome architecture,**
932 **transcription and gene persistence**

933 **A. Contact probability, $P(s)$, plotted as a function of genomic distance in**
934 **exponential phase in YEME medium (C6 condition).** The plot represent the
935 intra-chromosomal contact probability, $P(s)$, for pairs of loci separated by a
936 genomic distance (in base pairs) on the chromosome, showing that the
937 probability of contact extends up to approximately 1 Mb. The results of two
938 replicates (#1 and #2) are presented.

939 **B. Derivative of the contact matrix for exponentially growing WT cells.** The
940 map shows for each genome locus the tendency of contact frequencies to go
941 downwards or upwards. Boundaries are visualized as the black vertices of the
942 squares along the diagonal. Regions with no upstream/downstream bias show
943 no squares (*e.g.* terminal regions).

944 **C. Boxplot presenting the antisense index over growth depending of genes**
945 **depending on their architectural location.** The distribution of the antisense
946 index defined as the number of counts in antisense over total counts (in sense-
947 and antisense- orientation) is presented depending on the location of the genes
948 (within boundaries or domains). For clarity, outliers were excluded from this
949 graphical representation (but were taken into account for the numerical
950 exploitation of the data). The significant differences by Wilcoxon rank sum test

951 with continuity correction (boundaries *versus* domains) are indicated by
952 asterisks (***, $p < 0.001$).

953 **D. Contact probability, $P(s)$, of the left, central and right compartments**
954 **plotted as a function of genomic distance.** The plot represents the intra-
955 chromosomal contact probability, $P(s)$, for pairs of loci separated by a genomic
956 distance (in base pairs) on the three compartments.

957 **E. Dispersion index, $Var(s)/P(s)$, of the left, central and right compartments**
958 **plotted as a function of genomic distance in YEME after 24 h of growth (C6**
959 **condition).** The index of dispersion ($Var(s)/P(s)$) reflects the range of variations
960 relative to the mean value as a function of the genome distance for the contact
961 maps obtained for cells grown 24 h in YEME medium. Long-range DNA contacts
962 within the terminal compartments (> 100 kbp) are more variable than within the
963 central compartment.

964

965 **Supplementary Figure 1: Size distribution of *S. ambofaciens* ATCC 23877 GIs**

966 Using a minimal threshold of 15 CDS within the synteny break in at least one of the
967 compared genome, 50 GIs (composed of 2 to 158 CDSs) and 22 insertion points (≤ 1
968 CDS within the synteny break) were identified in *S. ambofaciens* ATCC 23877
969 genome.

970

971 **Supplementary Figure 2: Transcriptomes over growth of the four biosynthetic**
972 **gene clusters encoding all known antibacterial activities of *Streptomyces***
973 ***ambofaciens* ATCC 23877**

974 The genes expressed at the highest level is boxed in red for the congocidine,
975 spiramycin and kinamycin BGCs. They correspond to *cgc1* (SAM23877_RS39345,

976 encoding a transcriptional regulator), *srnB* (SAM23877_RS26680, encoding the
977 ABC-F type ribosomal protection protein SrmB) and *alpZ* (SAM23877_RS00890,
978 encoding a transcriptional regulator) genes, respectively. The gene encoding SrmS
979 (SAM23877_RS26595), the transcriptional regulator of the spiramycin BGC, is boxed
980 in black. The dots are joined by a line for clarity (but the lines do neither report a real
981 observation nor a model).

982

983 **Supplementary Figure 3: 3C-contact-maps for *S. ambofaciens* grown in the**
984 **studied conditions**

985 **A. Replicate of cells after 24 h in YEME growth medium, C6 condition**

986 **B. Replicate of after 24 h in MP5 growth medium, C1 condition**

987 **C. Replicate of after 48 h in MP5 growth medium, C4 condition.** The violet and
988 pink arrows indicate the location of boundaries formed at desferrioxamin and
989 congocidine BGCs, respectively.

990 **D. Results of the frontier index analysis including peaks that do not form**
991 **boundaries.** This panel presents the frontier index analysis of the 3C-maps
992 presented in **Fig.3** and **Fig.4**, showing single peaks that were not considered
993 as indicative of a boundary. At each locus, two indices are computed, reflecting
994 the intensity of the loss of contact frequencies when going downwards (green
995 peaks) or upwards (orange peaks), respectively, to the locus. A boundary is
996 defined as any bin in which there is a change in the right bias of contacts towards
997 the left bias (± 2 bin, green and orange peaks, respectively). The detection of
998 bins with only an upstream or a downstream significant peak reflects the
999 intrinsically noisy nature of 3C-seq data.

1000

1001 **Supplementary table 1: Collection of 125 *Streptomyces* genomes used in this**
1002 **study**

1003 **Name:** Species name

1004 **Genome Assembly ID:** Unique identifier used in the NCBI Genome Assembly
1005 database

1006 **Chromosome length (bp):** Chromosome length expressed in bp

1007 **Plasmid length (bp):** Cumulative length of plasmids [Number of plasmids] expressed
1008 in bp

1009 **CDSs nb:** Number of CDSs in chromosome [CDSs in plasmids]

1010 **TIR length (bp):** Size of one copy of a TIR expressed in bp.

1011

1012 **Supplementary table 2: List of the GIs identified in *S. ambofaciens* ATCC 23877**
1013 **genome**

1014

1015 **Supplementary table 3: Gene distribution in the central region *versus* the**
1016 **extremities (defined by the first and last rDNA operons) of *Streptomyces***
1017 ***ambofaciens* ATCC 23877 chromosome depending on features of interest**

1018

1019 **Supplementary table 4: List of the SMBGCs identified in *S. ambofaciens* ATCC**
1020 **23877 genome**

1021

1022

1023

1024

1025

1026 **Supplementary table 5: Overall results from the comparative genomics, RNA-**
1027 **seq and 3C-seq analyses**

1028 **ID:** Gene identifier (NCBI annotation).

1029 **SUPPORT:** genetic support, either chromosome (CHR) or pSAM1 plasmid.

1030 **TIR:** 'TIR' if the gene is located within a TIR, 'NA' otherwise.

1031 **START:** start position of the gene on the genetic support.

1032 **END:** end position of the gene on the genetic support.

1033 **STRAND:** orientation of the gene on the genetic support.

1034 **REPLICATION_ORIENTATION:** orientation of the gene regarding the origin of
1035 replication, either on the "leading" or the "lagging" strand.

1036 **GENE_SIZE:** size of the gene (in bp).

1037 **Protein_ID:** Protein identifier (NCBI annotation).

1038 **GO_ID:** Gene product identifier (Uniprot or RNA Central, suitable for GO analysis).

1039 **Gene_Name:** Gene name (from NCBI and manual annotation from SMBGCs
1040 encoding known antibiotics).

1041 **UNIPROT_Gene_Name:** Uniprot gene name, old locus tag.

1042 **Product (NCBI annotation):** Product annotation from NCBI.

1043 **Product (Uniprot annotation):** Product annotation from Uniprot.

1044 **ACTINO_SIGN:** 'ACTINO_SIGN' if the gene encodes a protein of the actinobacterial
1045 signature, 'NA' otherwise.

1046 **CORE:** 'CORE' if the gene belongs to the core CDSs identified by comparing 125
1047 *Streptomyces* genomes (see Methods section for details, as well as Supplementary
1048 Table 1).

1049 **PERSIST95:** 'PERSIST95' if the gene persistence is superior to 95 %,
1050 'NOT_PERSIST95' otherwise.

1051 **GI:** annotation as to whether or not the gene belongs to a GI (see details about GI
1052 names on Supplementary Table 2).

1053 **Strain-specific_CDSs:** 'UNIQ' if the gene is only present in *S. ambofaciens*
1054 ATCC23877 (in a panel of 125 *Streptomyces* genomes described in Supplementary
1055 Table 1), 'NOT_UNIQ' otherwise.

1056 **Func_RNA:** annotation as to whether or not the gene encodes a functional RNA
1057 (tRNA, tmRNA, SRP RNA, RNase P RNA) excluding rRNA.

1058 **Transl._CDS:** 'translation' if the CDS encodes a function related to translation and/or
1059 RNA stability, 'NA' otherwise

1060 **NAPSF:** 'NAPSF' if the gene encodes a nucleoid-associated protein or a
1061 chromosome structural protein and/or a protein experimentally found associated to
1062 *Streptomyces* chromatin (see Methods for details), 'NA' otherwise.

1063 **SMBGC:** annotation as to whether or not the gene belongs to a SMBGC (see details
1064 about SMBGC names on Supplementary Table 4).

1065 **pSAM2:** 'pSAM2' if the gene belongs to pSAM2 integrative plasmid, 'NA' otherwise.

1066 **PROPHAGE:** 'PROPHAGE' if the gene belongs to the prophage integrated in *S.*
1067 *ambofaciens* ATCC23877 genome, 'NA' otherwise

1068 **PERSIST_INDEX:** Persistence index (N_{orth}/N , where N_{orth} is the number of genomes
1069 carrying a given orthologue and N the number of genomes searched, 124 in our
1070 case). This persistence index was calculated only for CDSs (functional RNAs were
1071 not considered).

1072 **'SENSE' columns:** Number of normalized counts (DESeq2 normalization and
1073 normalized on gene size) per kb for each gene in the sense orientation, as the mean
1074 of the number of counts in all tested conditions ('SENSE_Mean') or of the replicates
1075 in each condition considered individually ('SENSE_C1' to 'SENSE_C7')

1076 **'ANTISENSE INDEX' columns:** Number of reads in the antisense orientation over
1077 the total (sense plus antisense) number of reads for each gene in all tested conditions
1078 ('ANTISENSE_INDEX_Mean') or in each condition considered individually
1079 ('ANTISENSE_INDEX_C1' to 'ANTISENSE_INDEX_C7')

1080 **'DIFF' columns:** annotation as to whether or not the gene is statistically differentially
1081 expressed compared to C1 condition, with an adjusted p value ≤ 0.05 ('DIFF0.05'),
1082 0.01 ('DIFF0.01') or 0.001 ('DIFF0.001').

1083 **'CAT' columns:** Gene expression level category in all tested conditions ('CAT
1084 _Mean'), in each condition considered individually ('CAT_C1' to 'CAT_C7'), maximal
1085 category in all tested conditions ('CAT_MAX') and minimal category in all tested
1086 conditions ('CAT_MIN').

1087 **SWITCH:** annotation as to whether or not the gene expression level switches from
1088 CAT_0 to a maximal CAT_3 ('SWITCH0->3') or CAT_4 ('SWITCH0->4') category,
1089 when considering all studied conditions.

1090 **'BORDER' columns:** annotation as to whether or not the gene belongs to a boundary
1091 observed on the map contact obtained in C1 ('C1_BORDER'), C4 ('C4_BORDER')
1092 and/or C6 condition ('C4_BORDER'). The boundaries are numbered according to their
1093 order in the genome. 'rRNA' indicates that the boundary also contains an rDNA operon.
1094

1095 References

- 1096 1 Berdy, J. Thoughts and facts about antibiotics: Where we are now and where we are
1097 heading. *J Antibiot (Tokyo)* **65**, 441, doi:10.1038/ja.2012.54 (2012).
- 1098 2 Yague, P., Lopez-Garcia, M. T., Rioseras, B., Sanchez, J. & Manteca, A. New insights on the
1099 development of *Streptomyces* and their relationships with secondary metabolite
1100 production. *Curr Trends Microbiol* **8**, 65-73 (2012).
- 1101 3 Hodgson, D. A. Primary metabolism and its control in streptomycetes: a most unusual
1102 group of bacteria. *Adv Microb Physiol* **42**, 47-238, doi:10.1016/s0065-2911(00)42003-5
1103 (2000).
- 1104 4 Barka, E. A. *et al.* Taxonomy, Physiology, and Natural Products of Actinobacteria.
1105 *Microbiol Mol Biol Rev* **80**, 1-43, doi:10.1128/MMBR.00019-15 (2016).
- 1106 5 Liu, G., Chater, K. F., Chandra, G., Niu, G. & Tan, H. Molecular regulation of antibiotic
1107 biosynthesis in streptomycetes. *Microbiol Mol Biol Rev* **77**, 112-143,
1108 doi:10.1128/MMBR.00054-12 (2013).
- 1109 6 Kong, D., Wang, X., Nie, J. & Niu, G. Regulation of Antibiotic Production by Signaling
1110 Molecules in *Streptomyces*. *Front Microbiol* **10**, 2927, doi:10.3389/fmicb.2019.02927
1111 (2019).
- 1112 7 McLean, T. C., Wilkinson, B., Hutchings, M. I. & Devine, R. Dissolution of the Disparate: Co-
1113 ordinate Regulation in Antibiotic Biosynthesis. *Antibiotics (Basel)* **8**,
1114 doi:10.3390/antibiotics8020083 (2019).
- 1115 8 Redenbach, M. *et al.* A set of ordered cosmids and a detailed genetic and physical map for
1116 the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. *Mol Microbiol* **21**, 77-96 (1996).
- 1117 9 Omura, S. *et al.* Genome sequence of an industrial microorganism *Streptomyces*
1118 *avermitilis*: deducing the ability of producing secondary metabolites. *Proc Natl Acad Sci U*
1119 *S A* **98**, 12215-12220, doi:10.1073/pnas.211433198 (2001).
- 1120 10 Karoonuthaisiri, N., Weaver, D., Huang, J., Cohen, S. N. & Kao, C. M. Regional organization
1121 of gene expression in *Streptomyces coelicolor*. *Gene* **353**, 53-66,
1122 doi:10.1016/j.gene.2005.03.042 (2005).
- 1123 11 Ikeda, H. *et al.* Complete genome sequence and comparative analysis of the industrial
1124 microorganism *Streptomyces avermitilis*. *Nat Biotechnol* **21**, 526-531,
1125 doi:10.1038/nbt820 (2003).
- 1126 12 Choulet, F. *et al.* Evolution of the terminal regions of the *Streptomyces* linear
1127 chromosome. *Mol Biol Evol* **23**, 2361-2369, doi:10.1093/molbev/msl108 (2006).
- 1128 13 Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete *Streptomyces*
1129 *coelicolor* A3(2). *Nature* **417**, 141-147, doi:10.1038/417141a (2002).
- 1130 14 Aigle, B. *et al.* Genome mining of *Streptomyces ambofaciens*. *J Ind Microbiol Biotechnol*
1131 **41**, 251-263, doi:10.1007/s10295-013-1379-y (2014).
- 1132 15 Fischer, G., Wenner, T., Decaris, B. & Leblond, P. Chromosomal arm replacement
1133 generates a high level of intraspecific polymorphism in the terminal inverted repeats of
1134 the linear chromosomal DNA of *Streptomyces ambofaciens*. *Proc Natl Acad Sci U S A* **95**,
1135 14296-14301 (1998).
- 1136 16 Hoff, G., Bertrand, C., Piotrowski, E., Thibessard, A. & Leblond, P. Genome plasticity is
1137 governed by double strand break DNA repair in *Streptomyces*. *Sci Rep* **8**, 5272,
1138 doi:10.1038/s41598-018-23622-w (2018).
- 1139 17 Tidjani, A. R. *et al.* Massive Gene Flux Drives Genome Diversity between Sympatric
1140 *Streptomyces* Conspecifics. *MBio* **10**, doi:10.1128/mBio.01533-19 (2019).
- 1141 18 Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation.
1142 *Science* **295**, 1306-1311, doi:10.1126/science.1067799 (2002).
- 1143 19 Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals
1144 folding principles of the human genome. *Science* **326**, 289-293,
1145 doi:10.1126/science.1181369 (2009).

- 1146 20 Le, T. B., Imakaev, M. V., Mirny, L. A. & Laub, M. T. High-resolution mapping of the spatial
1147 organization of a bacterial chromosome. *Science* **342**, 731-734,
1148 doi:10.1126/science.1242059 (2013).
- 1149 21 Marbouty, M. *et al.* Condensin- and Replication-Mediated Bacterial Chromosome Folding
1150 and Origin Condensation Revealed by Hi-C and Super-resolution Imaging. *Mol Cell* **59**,
1151 588-602, doi:10.1016/j.molcel.2015.07.020 (2015).
- 1152 22 Lioy, V. S. *et al.* Multiscale Structuring of the E. coli Chromosome by Nucleoid-Associated
1153 and Condensin Proteins. *Cell* **172**, 771-783 e718, doi:10.1016/j.cell.2017.12.027 (2018).
- 1154 23 Le, T. B. & Laub, M. T. Transcription rate and transcript length drive formation of
1155 chromosomal interaction domain boundaries. *EMBO J* **35**, 1582-1595,
1156 doi:10.15252/embj.201593561 (2016).
- 1157 24 Wang, X. *et al.* Condensin promotes the juxtaposition of DNA flanking its loading site in
1158 *Bacillus subtilis*. *Genes Dev* **29**, 1661-1675, doi:10.1101/gad.265876.115 (2015).
- 1159 25 Duigou, S. & Boccard, F. Long range chromosome organization in *Escherichia coli*: The
1160 position of the replication origin defines the non-structured regions and the Right and
1161 Left macrodomains. *PLoS Genet* **13**, e1006758, doi:10.1371/journal.pgen.1006758
1162 (2017).
- 1163 26 Val, M. E. *et al.* A checkpoint control orchestrates the replication of the two chromosomes
1164 of *Vibrio cholerae*. *Sci Adv* **2**, e1501914, doi:10.1126/sciadv.1501914 (2016).
- 1165 27 Bohm, K. *et al.* Chromosome organization by a conserved condensin-ParB system in the
1166 actinobacterium *Corynebacterium glutamicum*. *Nat Commun* **11**, 1485,
1167 doi:10.1038/s41467-020-15238-4 (2020).
- 1168 28 Fang, G., Rocha, E. & Danchin, A. How essential are nonessential genes? *Mol Biol Evol* **22**,
1169 2147-2156, doi:10.1093/molbev/msi211 (2005).
- 1170 29 Acevedo-Rocha, C. G., Fang, G., Schmidt, M., Ussery, D. W. & Danchin, A. From essential to
1171 persistent genes: a functional approach to constructing synthetic life. *Trends Genet* **29**,
1172 273-279, doi:10.1016/j.tig.2012.11.001 (2013).
- 1173 30 Koehorst, J. J. *et al.* Comparison of 432 *Pseudomonas* strains through integration of
1174 genomic, functional, metabolic and expression data. *Sci Rep* **6**, 38699,
1175 doi:10.1038/srep38699 (2016).
- 1176 31 Lorenzi, J. N., Lespinet, O., Leblond, P. & Thibessard, A. Subtelomeres are fast-evolving
1177 regions of the *Streptomyces* linear chromosome. (Submitted).
- 1178 32 Chandra, G. & Chater, K. F. Developmental biology of *Streptomyces* from the perspective
1179 of 100 actinobacterial genome sequences. *FEMS Microbiol Rev* **38**, 345-379,
1180 doi:10.1111/1574-6976.12047 (2014).
- 1181 33 Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining
1182 pipeline. *Nucleic Acids Res* **47**, W81-W87, doi:10.1093/nar/gkz310 (2019).
- 1183 34 Pernodet, J. L., Alegre, M. T., Blondelet-Rouault, M. H. & Guerineau, M. Resistance to
1184 spiramycin in *Streptomyces ambofaciens*, the producer organism, involves at least two
1185 different mechanisms. *J Gen Microbiol* **139**, 1003-1011, doi:10.1099/00221287-139-5-
1186 1003 (1993).
- 1187 35 Juguet, M. *et al.* An iterative nonribosomal peptide synthetase assembles the pyrrole-
1188 amide antibiotic congocidine in *Streptomyces ambofaciens*. *Chem Biol* **16**, 421-431,
1189 doi:10.1016/j.chembiol.2009.03.010 (2009).
- 1190 36 Jeong, Y. *et al.* The dynamic transcriptional and translational landscape of the model
1191 antibiotic producer *Streptomyces coelicolor* A3(2). *Nat Commun* **7**, 11605,
1192 doi:10.1038/ncomms11605 (2016).
- 1193 37 Varoquaux, N., Lioy, V. S., Boccard, F. & Junier, I. Computational tools for the multiscale
1194 analysis of Hi-C data in bacterial chromosomes. *arXiv: Genomics* (2020).
- 1195 38 Nicolas, P. *et al.* Condition-dependent transcriptome reveals high-level regulatory
1196 architecture in *Bacillus subtilis*. *Science* **335**, 1103-1106, doi:10.1126/science.1206848
1197 (2012).
- 1198 39 Huang, J., Lih, C. J., Pan, K. H. & Cohen, S. N. Global analysis of growth phase responsive
1199 gene expression and regulation of antibiotic biosynthetic pathways in *Streptomyces*

- 1200 coelicolor using DNA microarrays. *Genes Dev* **15**, 3183-3192, doi:10.1101/gad.943401
1201 (2001).
- 1202 40 Gehrke, E. J. *et al.* Silencing cryptic specialized metabolism in *Streptomyces* by the
1203 nucleoid-associated protein Lsr2. *Elife* **8**, doi:10.7554/eLife.47691 (2019).
- 1204 41 Dekker, J. & Heard, E. Structural and functional diversity of Topologically Associating
1205 Domains. *FEBS Lett* **589**, 2877-2884, doi:10.1016/j.febslet.2015.08.044 (2015).
- 1206 42 Fang, G., Rocha, E. P. & Danchin, A. Persistence drives gene clustering in bacterial
1207 genomes. *BMC Genomics* **9**, 4, doi:10.1186/1471-2164-9-4 (2008).
- 1208 43 Vargas-Blanco, D. A. & Shell, S. S. Regulation of mRNA Stability During Bacterial Stress
1209 Responses. *Front Microbiol* **11**, 2111, doi:10.3389/fmicb.2020.02111 (2020).
- 1210 44 Naumova, N. *et al.* Organization of the mitotic chromosome. *Science* **342**, 948-953,
1211 doi:10.1126/science.1236083 (2013).
- 1212 45 Gruber, S. & Errington, J. Recruitment of condensin to replication origin regions by
1213 ParB/SpoOJ promotes chromosome segregation in *B. subtilis*. *Cell* **137**, 685-696,
1214 doi:10.1016/j.cell.2009.02.035 (2009).
- 1215 46 Sullivan, N. L., Marquis, K. A. & Rudner, D. Z. Recruitment of SMC by ParB-parS organizes
1216 the origin region and promotes efficient chromosome segregation. *Cell* **137**, 697-707,
1217 doi:10.1016/j.cell.2009.04.044 (2009).
- 1218 47 Wang, X., Brandao, H. B., Le, T. B., Laub, M. T. & Rudner, D. Z. *Bacillus subtilis* SMC
1219 complexes juxtapose chromosome arms as they travel from origin to terminus. *Science*
1220 **355**, 524-527, doi:10.1126/science.aai8982 (2017).
- 1221 48 Banigan, E. J. & Mirny, L. A. Loop extrusion: theory meets single-molecule experiments.
1222 *Curr Opin Cell Biol* **64**, 124-138, doi:10.1016/j.ccb.2020.04.011 (2020).
- 1223 49 Liroy, V. S., Junier, I., Lagage, V., Vallet, I. & Bocard, F. Distinct Activities of Bacterial
1224 Condensins for Chromosome Management in *Pseudomonas aeruginosa*. *Cell Rep* **33**,
1225 108344, doi:10.1016/j.celrep.2020.108344 (2020).
- 1226 50 Tran, N. T., Laub, M. T. & Le, T. B. K. SMC Progressively Aligns Chromosomal Arms in
1227 *Caulobacter crescentus* but Is Antagonized by Convergent Transcription. *Cell Rep* **20**,
1228 2057-2071, doi:10.1016/j.celrep.2017.08.026 (2017).
- 1229 51 Marko, J. F., De Los Rios, P., Barducci, A. & Gruber, S. DNA-segment-capture model for
1230 loop extrusion by structural maintenance of chromosome (SMC) protein complexes.
1231 *Nucleic Acids Res* **47**, 6956-6972, doi:10.1093/nar/gkz497 (2019).
- 1232 52 Ganji, M. *et al.* Real-time imaging of DNA loop extrusion by condensin. *Science* **360**, 102-
1233 105, doi:10.1126/science.aar7831 (2018).
- 1234 53 Kois-Ostrowska, A. *et al.* Unique Function of the Bacterial Chromosome Segregation
1235 Machinery in Apically Growing *Streptomyces* - Targeting the Chromosome to New
1236 Hyphal Tubes and its Anchorage at the Tips. *PLoS Genet* **12**, e1006488,
1237 doi:10.1371/journal.pgen.1006488 (2016).
- 1238 54 Kim, H. J., Calcutt, M. J., Schmidt, F. J. & Chater, K. F. Partitioning of the linear chromosome
1239 during sporulation of *Streptomyces coelicolor* A3(2) involves an oriC-linked parAB locus.
1240 *J Bacteriol* **182**, 1313-1320, doi:10.1128/jb.182.5.1313-1320.2000 (2000).
- 1241 55 Donczew, M. *et al.* ParA and ParB coordinate chromosome segregation with cell
1242 elongation and division during *Streptomyces* sporulation. *Open Biol* **6**, 150263,
1243 doi:10.1098/rsob.150263 (2016).
- 1244 56 Kois, A., Swiatek, M., Jakimowicz, D. & Zakrzewska-Czerwinska, J. SMC protein-dependent
1245 chromosome condensation during aerial hyphal development in *Streptomyces*. *J*
1246 *Bacteriol* **191**, 310-319, doi:10.1128/JB.00513-08 (2009).
- 1247 57 Dedrick, R. M., Wildschutte, H. & McCormick, J. R. Genetic interactions of *smc*, *ftsK*, and
1248 *parB* genes in *Streptomyces coelicolor* and their developmental genome segregation
1249 phenotypes. *J Bacteriol* **191**, 320-332, doi:10.1128/JB.00858-08 (2009).
- 1250 58 Brandao, H. B. *et al.* RNA polymerases as moving barriers to condensin loop extrusion.
1251 *Proc Natl Acad Sci U S A* **116**, 20489-20499, doi:10.1073/pnas.1907009116 (2019).
- 1252 59 Kieser T., B. M. J., Buttner M. J., Chater K. F., Hopwood D. A. *Practical Streptomyces*
1253 *genetics*. John Innes Foundation, Norwich, United Kingdom edn, (2000).

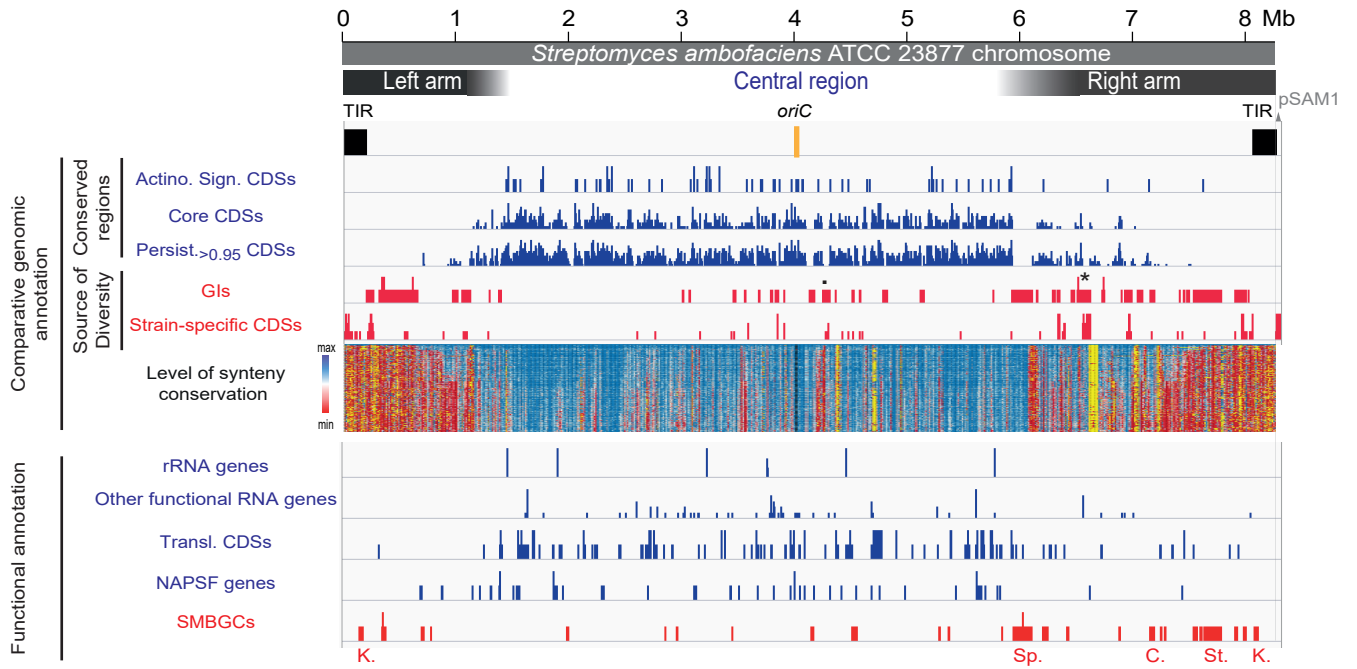
- 1254 60 Haas, D. *et al.* Synteruptor, a new comparative genomics-based approach for identifying
1255 non-classical specialized metabolite gene clusters. (Submitted).
- 1256 61 R. Dame, C. D. *Bacterial chromatin*. (Springer, 2010).
- 1257 62 Bradshaw, E., Saalbach, G. & McArthur, M. Proteomic survey of the *Streptomyces*
1258 *coelicolor* nucleoid. *J Proteomics* **83**, 37-46, doi:10.1016/j.jprot.2013.02.033 (2013).
- 1259 63 Fang, G., Bhardwaj, N., Robilotto, R. & Gerstein, M. B. Getting started in gene orthology
1260 and functional analysis. *PLoS Comput Biol* **6**, e1000703,
1261 doi:10.1371/journal.pcbi.1000703 (2010).
- 1262 64 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21,
1263 doi:10.1093/bioinformatics/bts635 (2013).
- 1264 65 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for
1265 assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930,
1266 doi:10.1093/bioinformatics/btt656 (2014).
- 1267 66 Varet, H., Brillet-Gueguen, L., Coppee, J. Y. & Dillies, M. A. SARTools: A DESeq2- and
1268 EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. *PLoS*
1269 *One* **11**, e0157022, doi:10.1371/journal.pone.0157022 (2016).
- 1270 67 RDC, T. R. *A language and environment for statistical computing* (2008).
- 1271 68 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26,
1272 doi:10.1038/nbt.1754 (2011).
- 1273 69 Pauline Vaissie, A. M., Francois Husson. Factoshiny: Perform Factorial Analysis from
1274 'FactoMineR' with a Shiny Application. (2020).
- 1275 70 Cournac, A., Marbouty, M., Mozziconacci, J. & Koszul, R. Generation and Analysis of
1276 Chromosomal Contact Maps of Yeast Species. *Methods Mol Biol* **1361**, 227-245,
1277 doi:10.1007/978-1-4939-3079-1_13 (2016).
- 1278 71 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**,
1279 357-359, doi:10.1038/nmeth.1923 (2012).
- 1280 72 Lioy, V. S. & Boccard, F. Conformational Studies of Bacterial Chromosomes by High-
1281 Throughput Sequencing Methods. *Methods Enzymol* **612**, 25-45,
1282 doi:10.1016/bs.mie.2018.07.007 (2018).
- 1283 73 Cournac, A., Marie-Nelly, H., Marbouty, M., Koszul, R. & Mozziconacci, J. Normalization of
1284 a chromosomal contact map. *BMC Genomics* **13**, 436, doi:10.1186/1471-2164-13-436
1285 (2012).
- 1286 74 Yang, M. C. & Losick, R. Cytological evidence for association of the ends of the linear
1287 chromosome in *Streptomyces coelicolor*. *J Bacteriol* **183**, 5180-5186,
1288 doi:10.1128/jb.183.17.5180-5186.2001 (2001).
- 1289 75 Ventura, M. *et al.* Genomics of Actinobacteria: tracing the evolutionary history of an
1290 ancient phylum. *Microbiol Mol Biol Rev* **71**, 495-548, doi:10.1128/MMBR.00005-07
1291 (2007).

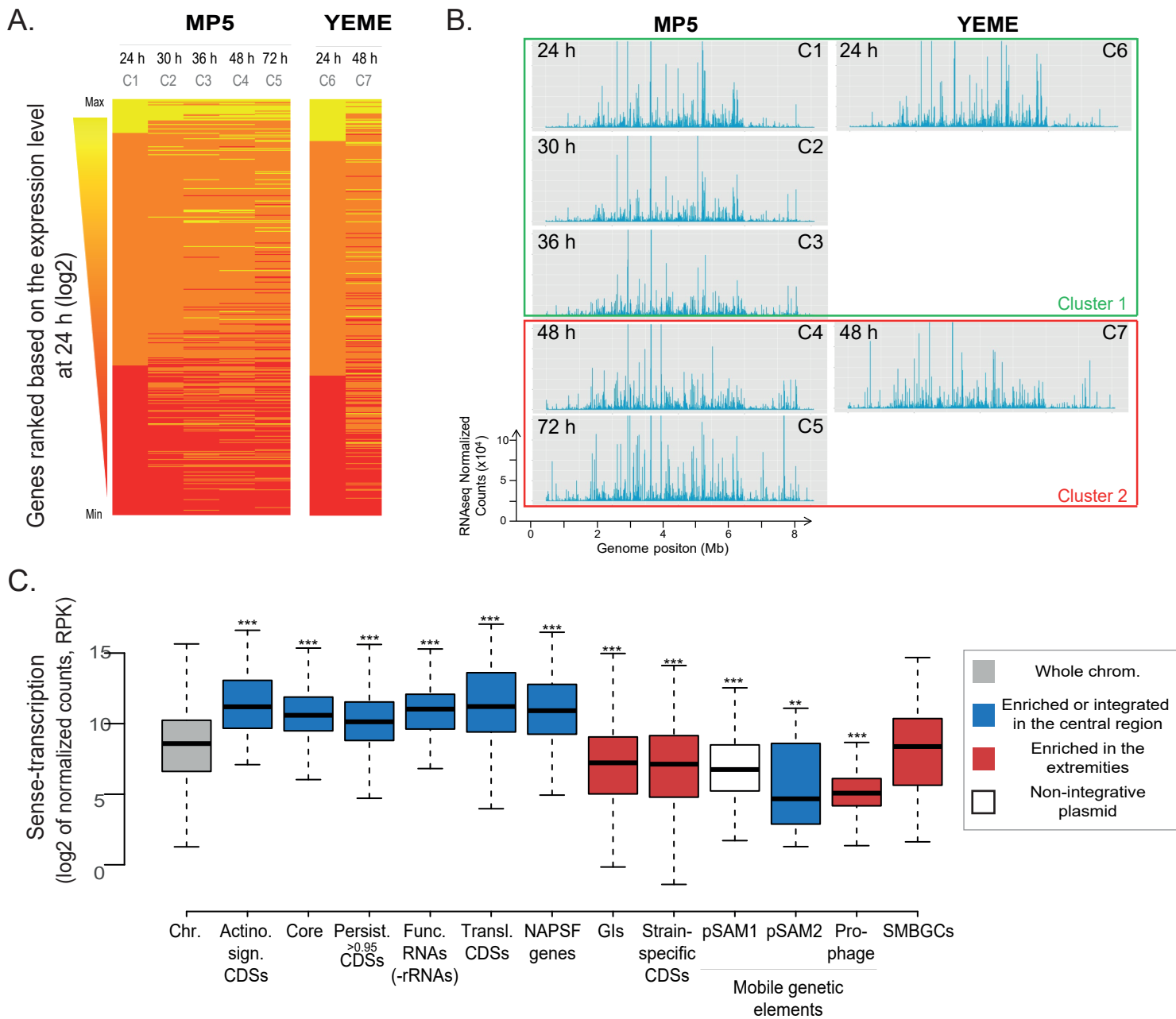
1292

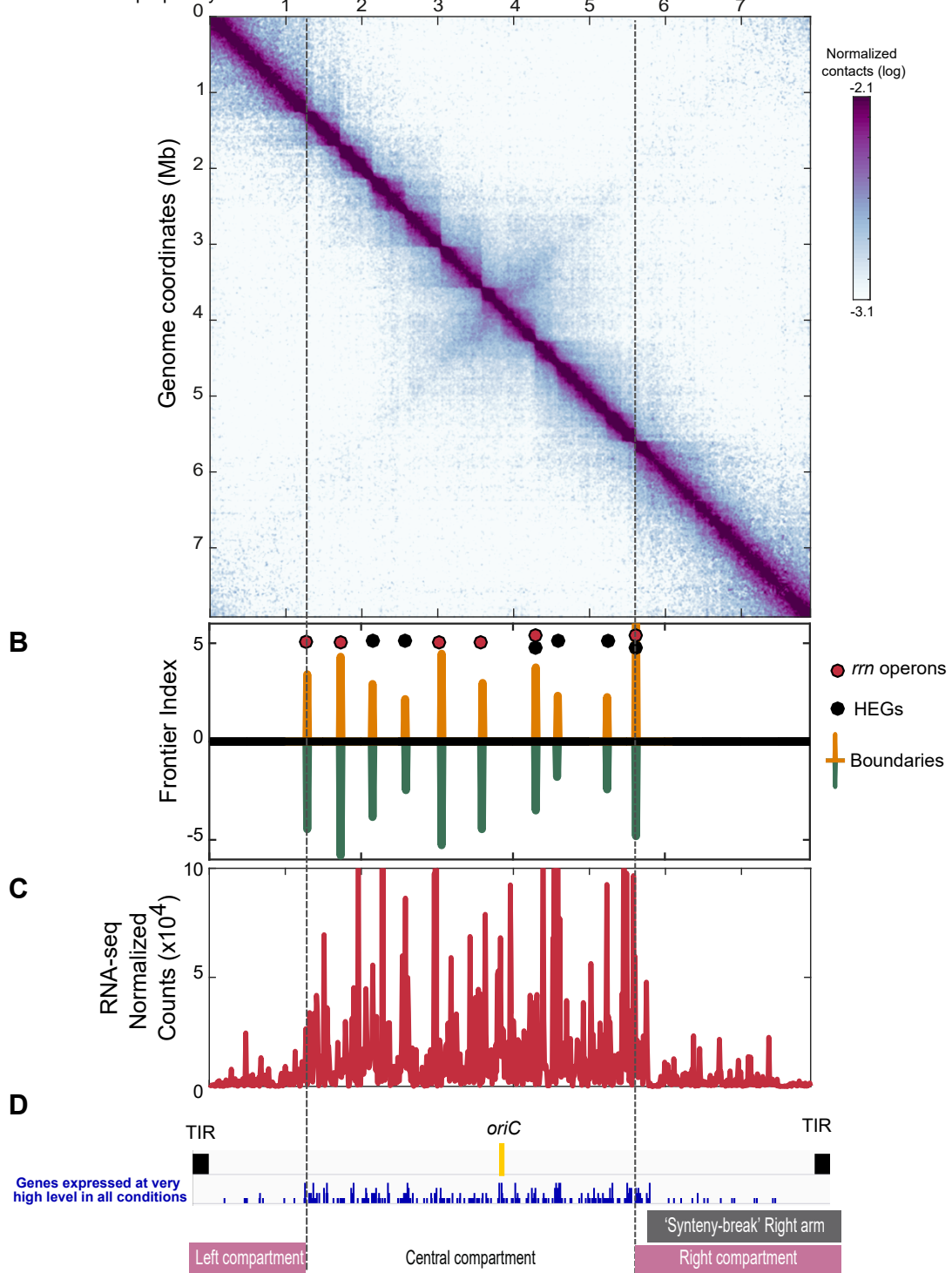
Table 1: Growth conditions used to performed -omics analyses

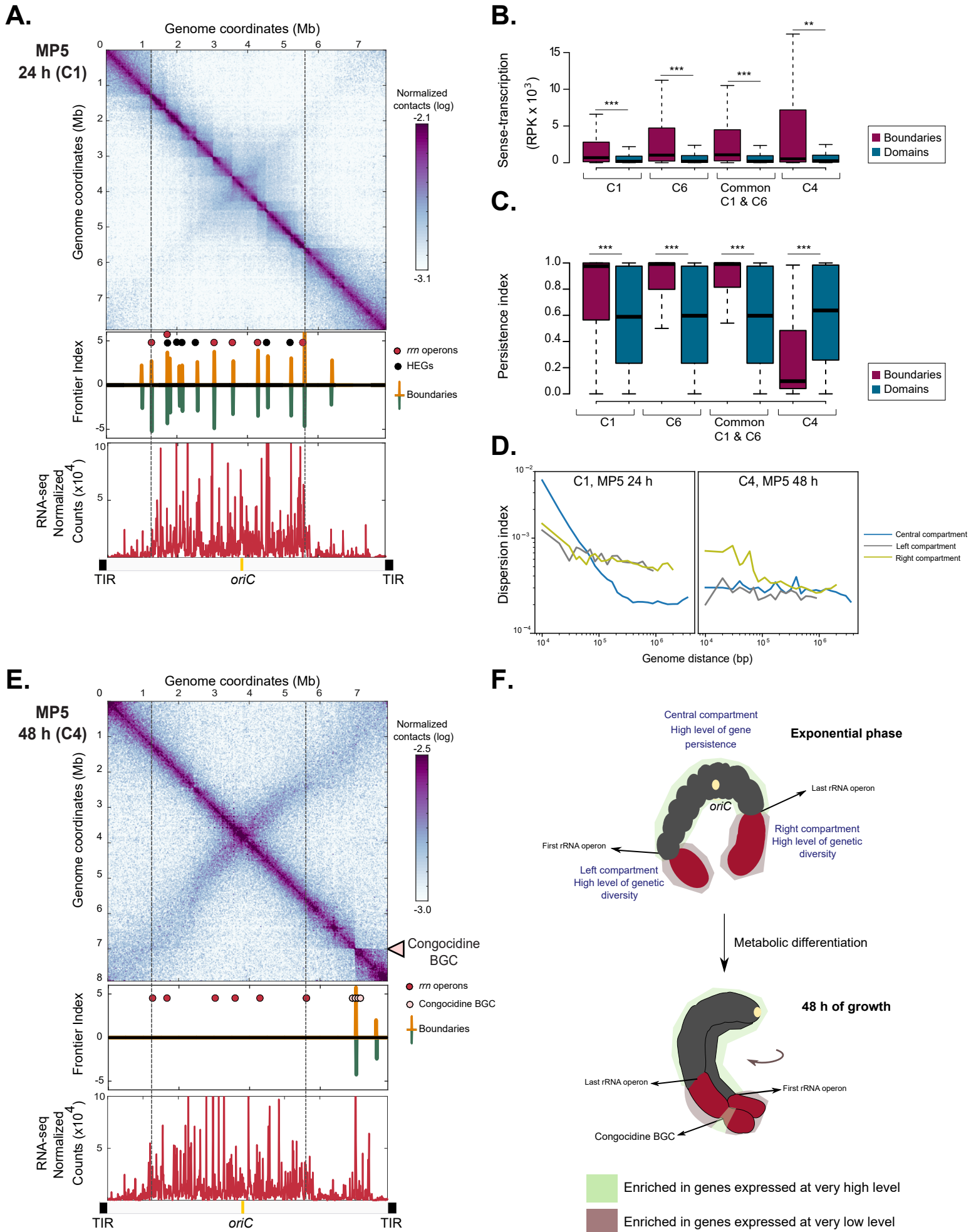
Conditions	Media	Brief description	Time (h)	Antibacterial activity*
C1	MP5	Liquid medium based on yeast extract containing glycerol (3.6 %), NaNO ₃ (1 g/l), MOPS - optimized for congocidine and spiramycin production	24	-
C2			30	-
C3			36	-
C4			48	++
C5			72	+++
C6	YEME	Liquid medium based on yeast extract containing saccharose (10.3 %), glucose (1 %), and malt extract	24	-
C7			48	+

*Size of the inhibition halo: “-“ (no detectable inhibition), “+” (< 20 mm), “++” ([20; 30] mm), “+++” (> 30 mm) – See **Extended Data Fig.1B**.

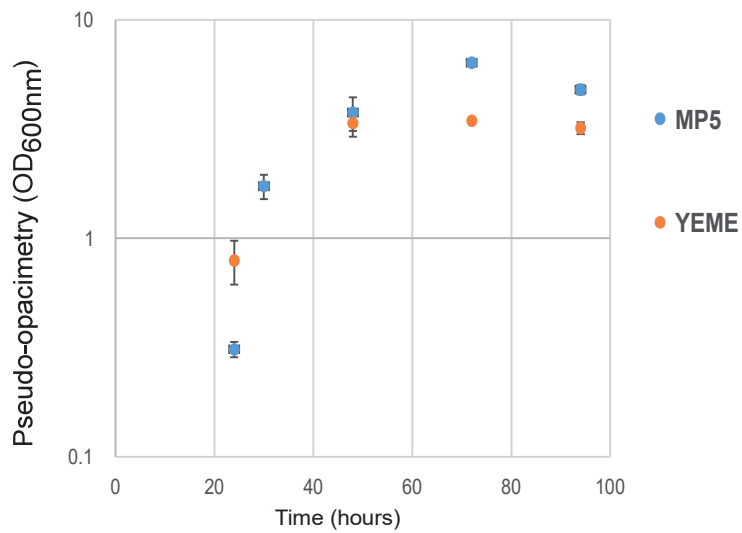




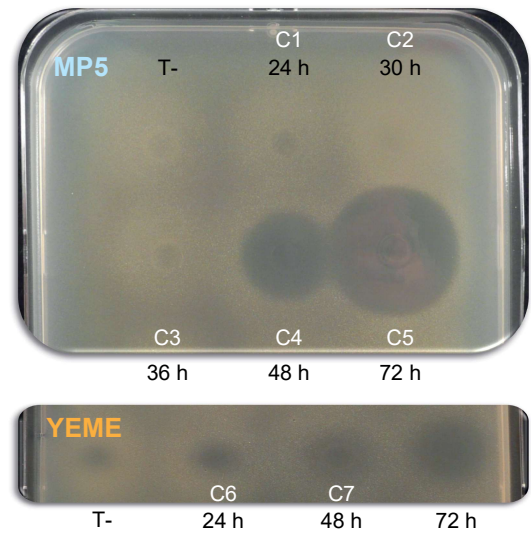




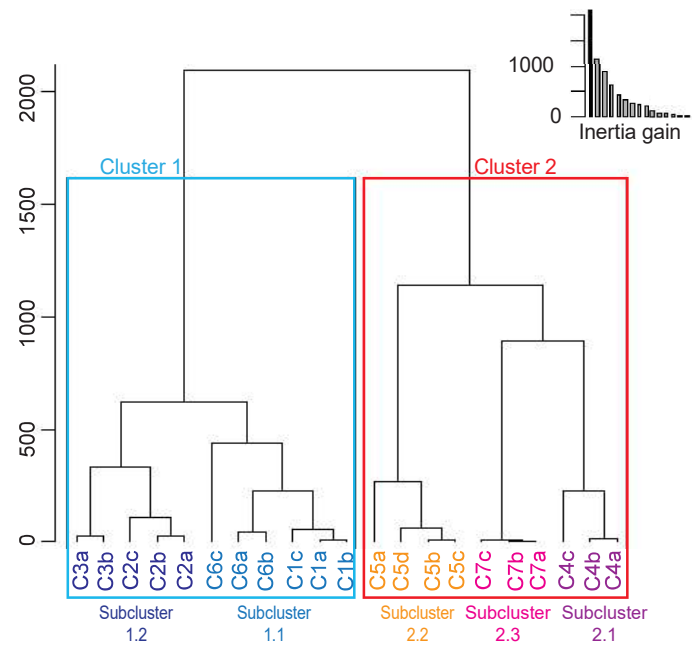
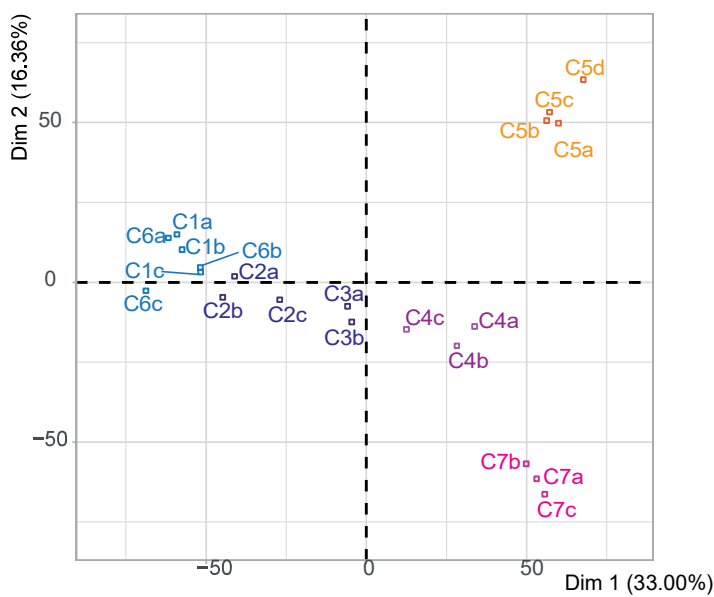
A.



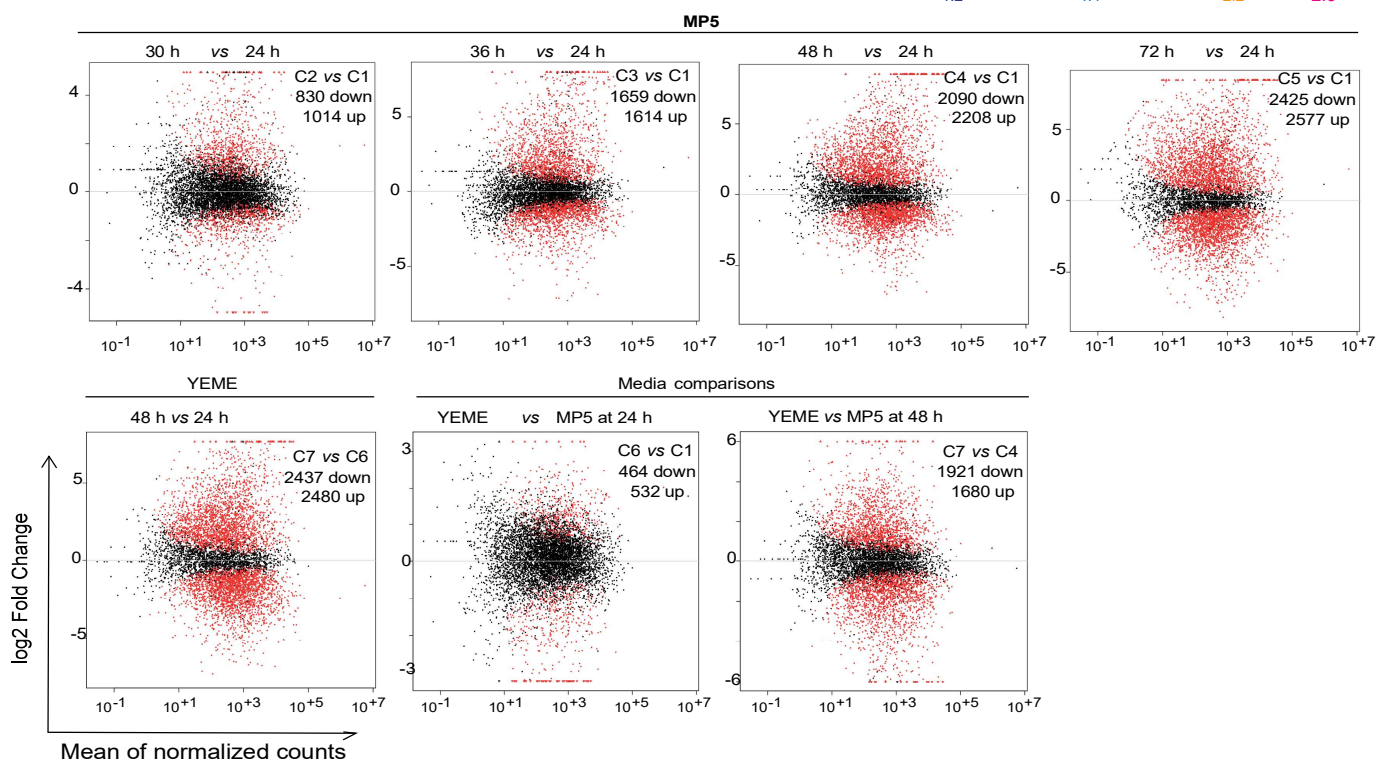
B.



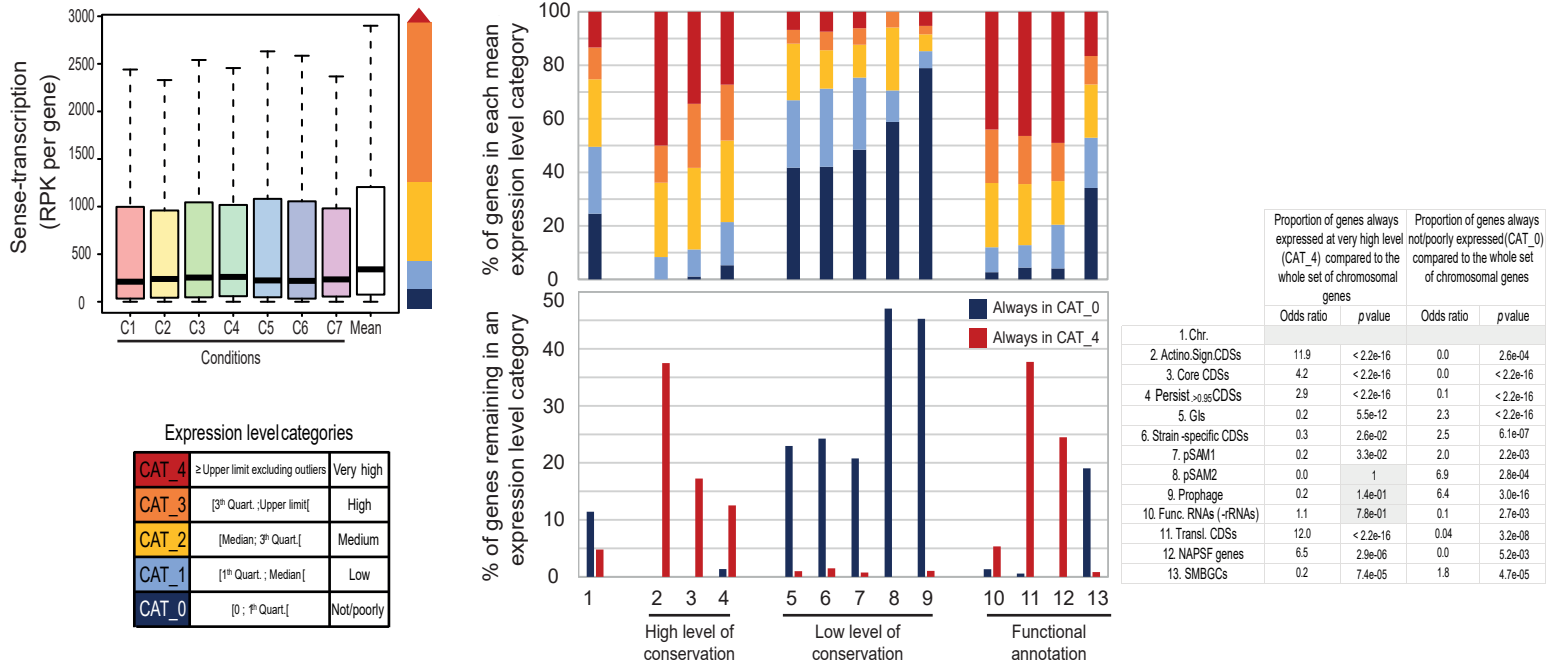
C.



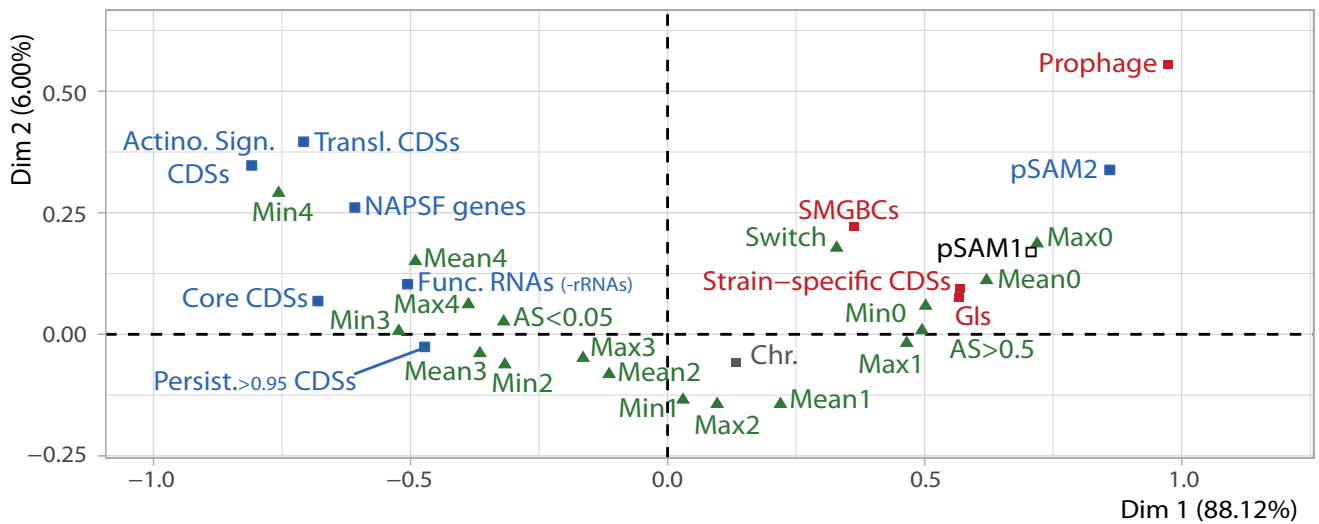
D.



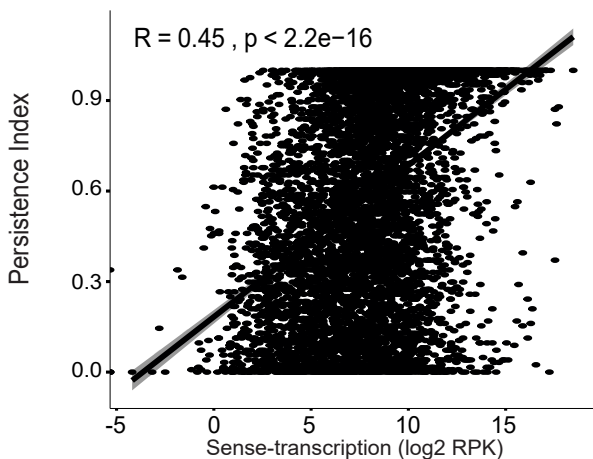
A.



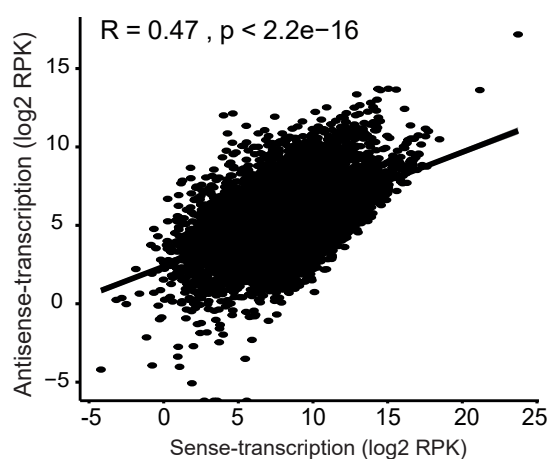
C.



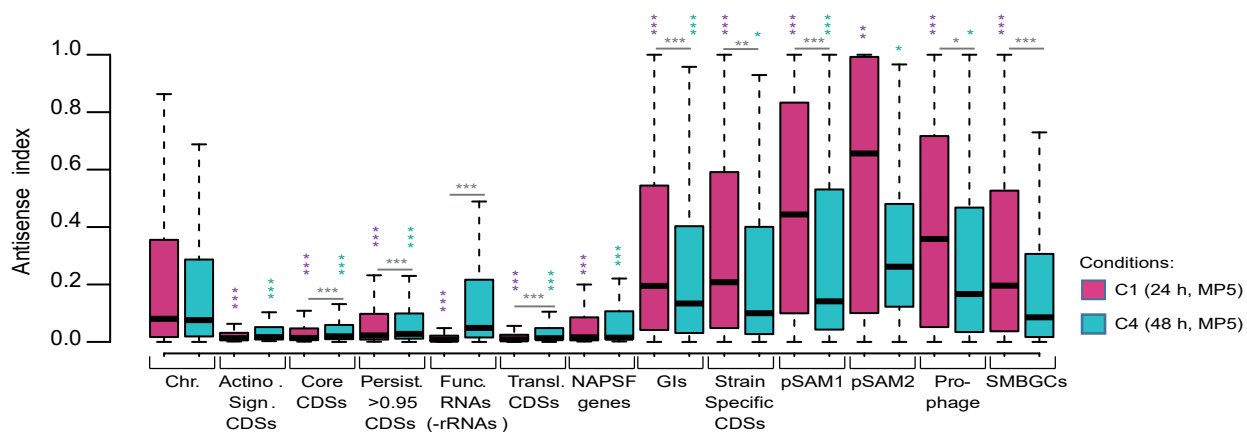
D.



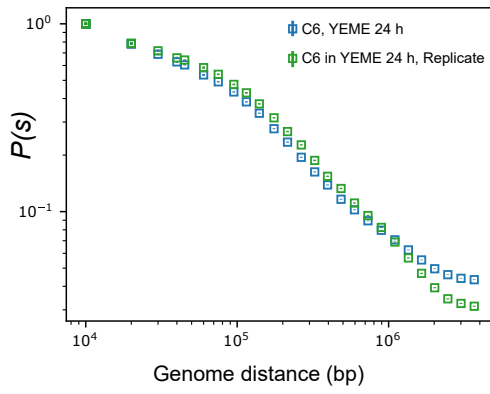
E.



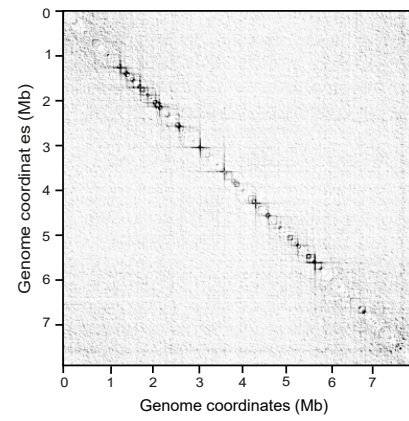
F.



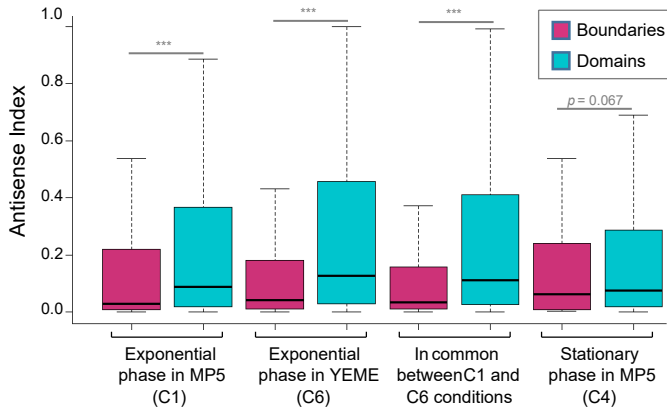
A.



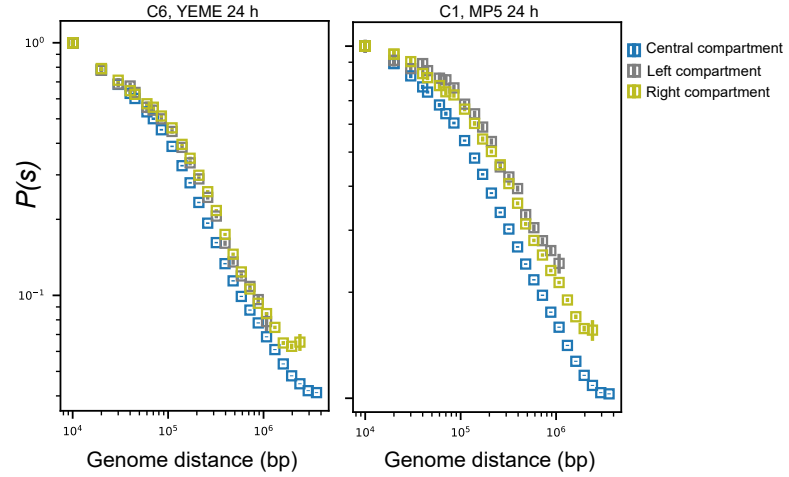
B.



C.



D.



E.

