



**HAL**  
open science

# Gene level resolution of bacteria and archaea genome folding

Charlotte Cockram, Agnès Thierry, Roxane Lestini, Romain Koszul

► **To cite this version:**

Charlotte Cockram, Agnès Thierry, Roxane Lestini, Romain Koszul. Gene level resolution of bacteria and archaea genome folding. 2021. hal-03106900

**HAL Id: hal-03106900**

**<https://hal.science/hal-03106900>**

Preprint submitted on 12 Jan 2021

**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.

## Gene-level resolution of bacterial and archaeal genomes folding

Charlotte Cockram<sup>1</sup>, Agnès Thierry<sup>1</sup>, Roxane Lestini<sup>2</sup>, Romain Koszul<sup>1,\*</sup>

Affiliations:

1 Institut Pasteur, Unité Régulation Spatiale des Génomes, CNRS, UMR 3525, F-75015 Paris, France

2 Laboratoire d'Optique et Biosciences, École Polytechnique, CNRS UMR7645 – INSERM U1182, IP Paris, 91128 Palaiseau Cedex, France

\* All correspondence should be addressed to [romain.koszul@pasteur.fr](mailto:romain.koszul@pasteur.fr)

## 1 **Summary:**

2 During the past decade, chromosome conformation capture (3C/Hi-C)-based methods have  
3 been used to probe the average 3D structure and organization of prokaryotic genomes,  
4 revealing fundamental aspects of chromosome dynamics. However, the current protocols are  
5 relatively expensive and inefficient, limiting the resolution of bacterial contact maps. Here we  
6 present a simple, cost-effective Hi-C approach that allows the exploration of bacterial and  
7 archaeal chromosome folding at the gene or operon level. We first apply it to the well-studied  
8 *E. coli* and *V. cholera* bacterial species, generating sub-kilobase resolution contact maps, and  
9 unveiling previously undetected gene-level chromosomal structures. We then apply it to the  
10 halophilic euryarchaeal species, *H. volcanii*, generating highly reproducible Hi-C matrices at  
11 a resolution of up to 1kb. The increase in resolution compared to available archaeal Hi-C maps  
12 facilitated the identification of a novel chromosome conformation that is distinct to that seen  
13 in crenarchaea, and more similar to bacterial chromosome conformations. We then explore  
14 the processes responsible for this conformation, identifying a novel role for the *H. volcanii*  
15 SMC protein, and propose how these results may be reflective of the Euryarchaea phylum as  
16 a whole.

17

## 18 **Introduction**

19 Due to their small size and dynamic architecture, the three-dimensional (3D) organization of  
20 microbial genomes have historically proved challenging to investigate. Archaeal genomes are  
21 similar to bacteria in terms of size and overall organization, with both possessing circular  
22 genomes that are unconfined by a membrane or nucleus<sup>1,2</sup>. Yet in terms of DNA replication,

23 genome segregation, and cell division, these two kingdoms are quite different. Bacterial  
24 genomes typically consist of a single chromosome with a unique origin of replication (*oriC*),  
25 although a few species also contain large secondary replicons comparable in size to the main  
26 chromosome<sup>3</sup>, but relying on plasmid-based mechanisms for their replication<sup>4</sup>. In contrast,  
27 archaea chromosomes commonly have multiple origins of replication and utilize the  
28 Orc1/Cdc6 initiator proteins, which are homologous to eukaryotic systems<sup>5</sup>. They also often  
29 have large secondary replicons, but unlike bacteria, these mini-chromosomes rely on the same  
30 replication machinery as the main chromosome<sup>6</sup>.

31         Recently, the application of chromosome conformation capture (3C/Hi-C) approaches  
32 has provided the opportunity to further highlight the myriad of different processes responsible  
33 for chromosome organization within these two domains. Notably, all bacteria chromosomes  
34 display self-interacting chromosomal interaction domains (CIDs), ranging in size from 30 -  
35 420kb, and whose precise nature and function remain elusive<sup>7-10</sup>. In addition, molecular  
36 complexes called condensin, and belonging to the structural maintenance of chromosomes  
37 (SMC) proteins family, bridge the chromosome arms of *Bacillus subtilis* and many other  
38 species<sup>9-12</sup>. Yet, despite the presence of the SMC homolog MukBEF, this phenomenon is  
39 absent in *E. coli*<sup>8</sup>. More recently, the genomes of the hyperthermophile crenarchaeal species,  
40 *S. acidocaldarius*, and *S. islandicus*, were investigated by Hi-C, resulting in the first archaeal  
41 contact maps<sup>13</sup>. These maps show that the *Sulfolobus* species display a chromosomal  
42 conformation distinct to that of bacteria, with the presence of compartments that somehow  
43 correlate with their gene expression activity. These crenarchaea also appear to have evolved  
44 a condensin-independent mechanism of chromosome organization in which a distant SMC  
45 homolog protein, called coalescin (ClsN), maintains this bi-compartmentalization. Whether  
46 these results are common to all archaea or if, as in bacteria, different species display  
47 significantly different genome organization remains unknown.

48           Taken together, the variability already observed across different prokaryotic species  
49 demonstrate the need for more in-depth investigations of chromosome architecture in different  
50 clades of bacteria and archaea. But this is not necessarily easy, as many bacterial and archaeal  
51 species exist in extreme environments which are difficult to manipulate in a laboratory setting.  
52 In this respect, Hi-C has already proved a readily applicable tool to gain a first glance at the  
53 genome folding of species that are less amenable than the standard laboratory models.  
54 However, despite key improvements to eukaryotic Hi-C protocols leading to the development  
55 of commercial kits, the prokaryotic assay remains highly inefficient and costly. Early Hi-C  
56 experiments performed in the bacteria *Caulobacter crescentus* and *B. subtilis* used restriction  
57 enzymes that recognize six base-pair sequences to digest the genome into fragments of several  
58 kilobases in size<sup>7,9</sup> and although these contact maps were informative, their resolution was  
59 limited to ~10 kb (**Supplementary Table 1**). Our team concomitantly reached a slightly  
60 better resolution by using four base-pair cutting enzymes which digest the genomes into  
61 smaller fragments but at a much higher cost (**Supplementary Table 1**)<sup>8,10</sup>. Despite these  
62 efforts, the highest resolution of published prokaryotic contact maps remains limited to ~4 kb  
63 for bacteria and 30 kb for archaea, well above the average size of a gene or operon.

64           In the present work, we bypass the current limitations in bacterial and archaeal Hi-C  
65 assays, by identifying several crucial steps and forging an optimized Hi-C protocol that results  
66 in significantly improved contact maps at a fraction (1/5th) of the sequencing cost. Not only  
67 is this approach more cost and time-effective compared to existing protocols, but most  
68 importantly the resulting contact matrices display up to a 30-fold improvement in resolution.  
69 We validate this protocol on two model bacteria (*E. coli* and *V. cholera*), utilizing the 500 bp  
70 resolution to reveal features previously unseen in published contact maps, and then apply it  
71 to the Euryarchaea *Haloferax volcanii*. This halophilic archaeon, first described in 1975  
72 following isolation from the bottom sediment of the Dead Sea<sup>14</sup>, is one of the most widely

73 studied archaeal models, and as a polyploid is representative of the Euryarchaea phylum as a  
74 whole<sup>15</sup>. We show that *H. volcanii* has a novel chromosome conformation, distinct to that of  
75 *S. acidocaldarius* and *S. islandicus*, and more similar to bacterial chromosome conformations.  
76 We then explore the processes responsible for this conformation, identifying a novel role for  
77 the *H. volcanii* SMC protein, and then propose how these results may be reflective of the  
78 Euryarchaeal phylum as a whole.

## 79 **Results**

### 80 **Improvements of microbial Hi-C contact maps**

81 We pursued the development of a Hi-C protocol more suited to bacteria but that could  
82 potentially be applied to archaea. We then benchmarked this protocol with published 3C  
83 contact maps of *E. coli* and *V. cholera*<sup>8,16</sup> to determine the extent to which it facilitated higher-  
84 resolution studies of prokaryotic chromosomes. The level of improvement was measured in  
85 two ways; firstly, by comparing the number of reads in the final contact map with the total  
86 number of reads sequenced (**Supplementary Table 1**), and secondly by determining the  
87 maximum resolution (defined here as bin size) of the final contact map. We tested and  
88 combined two different four-base cutting restriction enzymes with biotinylation to generate  
89 Hi-C contact maps of *E. coli* (**Figure 1a, b; Supplementary Figure 1 and Methods**).  
90 Compared to previously published 3C data, the use of 4-base cutting enzymes in a Hi-C  
91 experiment did improve the number of informative reads, with HpaII and MluCI both  
92 performing well (**Supplementary Table 1**). The HpaII digestion gave a better signal-to-noise  
93 ratio for *E. coli* (**Supplementary Figure 1**) as well as the highest increase in informative  
94 reads in the contact map (**Supplementary Table 1**) and therefore was retained to perform all  
95 subsequent experiments. Although it is worth noting that if a genome is particularly AT-rich,  
96 or if a specific AT-rich region is under investigation, we would recommend using MluCI. We

97 also added modifications to the blunt-end ligation step (**Figure 1a and Methods**) to obtain a  
98 ~3-fold increase in informative reads compared to that previously reported<sup>8</sup>, allowing the  
99 generation of higher resolution matrices with fewer empty bins (**Supplementary Figure 1**).  
100 Finally, we modified the library preparation to include a streptavidin-mediated pull-down of  
101 biotinylated interactions immediately after sonication and size-selection of the library (**Figure**  
102 **1a**). The preparation of the sequencing library was then done directly on the streptavidin beads  
103 carrying biotinylated chimeric DNA fragments, facilitating a further 3.5-fold enrichment of  
104 informative reads<sup>17,18</sup> and a ~10-fold increase overall (**Supplementary Figure 1**).

105 To evaluate the quality of the data generated through this approach, the resulting  
106 bacterial contact maps (**Figure 1b and c**) were compared to published 3C/Hi-C datasets from  
107 our lab and others, processed using the same computational pipeline (**Supplementary Table**  
108 **1**). Firstly, we compared the total number of reads sequenced with the number of reads  
109 retained in the final contact maps after alignment and filtering<sup>19</sup>. We found that, starting with  
110 a tenth of material, the protocol retained between five and twenty-fold more reads compared  
111 to former experiments using the same strains and restriction enzymes (**Supplementary Table**  
112 **1**)<sup>8,10,16</sup>. This increase in informative reads means that for less than a fifth of the sequencing  
113 cost, one can obtain Hi-C matrices as good as those published, making prokaryotic Hi-C  
114 experiments much more cost-effective. More importantly, this increase in informative reads  
115 means that contact maps can now easily be more covered and “filled-in” with informative  
116 contacts, with an improved resolution and bins down to 500 bp resulting in up to a 10-fold  
117 increase in resolution compared to published bacterial 3C contact maps (**Figure 1d-g,**  
118 **Supplementary Figure 1 and Supplementary Table 1**)<sup>8,16</sup>.

119 Although all bacteria showed the same overall global chromosome organization as  
120 previously reported, our high-resolution Hi-C matrices were much crisper and carried more  
121 contrasted features<sup>8,16</sup>. In *E. coli* (binned at 5 kb and 500 bp), we observed that the thickness

122 of the main-diagonal was much more heterogeneous compared to previous contact maps  
123 **(Figure 1b)**<sup>8</sup>. This improved visualization was particularly beneficial in highlighting local  
124 changes in chromosome organization. For example, we observed that the 800 kb region  
125 surrounding *dif* and constituting the Ter domain<sup>20</sup> exhibited enrichment in short to mid-range  
126 contacts compared with the rest of the chromosome **(Figure 1b, c, and d)**. The 500 bp  
127 resolution contact maps also allowed us to further disclose new sub-structures in the Ter MD,  
128 such as the splitting of a domain exactly at the *dif* site.

129         The most exciting feature of a sub-kilobase resolution was the ability to study bacterial  
130 chromosomes at the gene level. The 500 bp binned *E. coli* Hi-C maps facilitated the  
131 identification of new chromosomal borders the *fliF-R* and *flgB-J* flagella operons, **(Figure 1f**  
132 **and g)**. These operons are 7-10 kb in size, and appear to have similar structures, which are  
133 characterized as dense signals along the diagonal, capable of insulating the neighboring  
134 regions of the chromosome. Furthermore, we also observed signals corresponding to other  
135 highly expressed genes and smaller operons in the surrounding genome, indicating that this  
136 method is capable of gene-level studies of bacterial chromosomes. This increased resolution  
137 was also beneficial in studying *V. cholera*, facilitating an increased definition of structures  
138 throughout the genome, such as the superintegron on chromosome 2 **(Figure 1h)**. This  
139 identification of these bacterial structures is only possible because, with the new Hi-C method,  
140 these regions had fewer empty bins and a higher signal to noise ratio, making the visualization  
141 of these structures much easier. Finally, to determine if our new method introduced biases  
142 regarding the type of reads that were enriched, we compared the intra vs. inter-chromosomal  
143 contact ratio of the two chromosomes of *V. cholera*. We found that both inter- and intra-  
144 chromosomal contacts showed approx. a 10-fold increase compared to previously published  
145 data **(Figure 1i)**, conserving the ratio of the two and indicating that the protocol does not bias  
146 the type of contacts enriched. The increase in inter-chromosomal contacts, as seen for *V.*

147 *cholera* (**Figure 1j**), will be particularly important for studying prokaryotes with multiple  
148 chromosomes or extra-chromosomal elements.

### 149 **Features of chromosome organization in *H. volcanii***

150 The optimized Hi-C protocol was next applied to the archaeal species *H. volcanii*  
151 growing in exponential phase. Despite the polyploid nature of these cells, highly reproducible  
152 contact matrices with a resolution of up to 1 kb were generated (**Figure 2a and**  
153 **Supplementary Figure 2**). The genome is composed of a 2.85Mb main chromosome, three  
154 smaller chromosomes ((pHV4 (636 kb), pHV3 (438 kb), pHV1 (85.1kb)), and the pHV2  
155 plasmid (6.4kb)<sup>21</sup>. The main chromosome has three origins of replication (*oriC1–3*) and the  
156 smaller chromosomes are distinct from the plasmid because they each carry a unique *oriC*<sup>1</sup>,  
157 whereas the plasmid has none. As expected, the contact maps of each chromosome displayed  
158 a single, strong diagonal signal resulting from the enrichment of contacts between neighboring  
159 loci. In agreement with previous reports<sup>22</sup>, our H26 strain no longer contained the pHV2  
160 plasmid<sup>23</sup> and we observed that the pHV4 chromosome had integrated into the main  
161 chromosome (coordinates: 249,185bp - 884,970 bp). The ~30-fold increase in resolution,  
162 compared to previously published archaeal Hi-C maps<sup>13</sup>, facilitated the visualization of much  
163 more detailed structures (**Figure 2a and Supplementary Figure 2**), notably:

#### 164 *Self-interacting domains*

165 The Hi-C matrix displays self-interacting regions, that appear as squares along the  
166 main diagonal, reminiscent of CIDs in bacteria<sup>7–10,13</sup>. A directional index analysis performed  
167 at 100 kb and 300 kb identified ~10 larger domains ranging from ~70 kb - 570 kb in size  
168 (**Figure 2b**) and ~23 domains ranging from ~25 kb to 200 kb in size (**Figure 2c**), scales that  
169 very similar to those observed previously in bacteria<sup>7–10</sup>. To gain insight into the mechanisms  
170 responsible for this organization, we compared the Hi-C maps and DI analysis with the  
171 transcription profile of the cells and found a good correlation between the CID boundaries

172 (100 kb scale) and gene expression (**Figure 2c and d**). Correlations between domain  
173 boundaries and transcription level have previously been shown in both prokaryotes and  
174 eukaryotes and are in good agreement with the idea that gene expression creates plectonemic-  
175 free regions that insulate adjacent regions of the chromosome<sup>24-26</sup>

#### 176 *Plaid-like contact pattern*

177 The *H. volcanii* contact matrix displays a striking plaid pattern involving very short  
178 (~2-15 kb) DNA segments throughout the genome. Similar patterns, but involving much  
179 larger regions of DNA, have been found in both crenarchaea and some eukaryotes<sup>13,27</sup>. This  
180 bipartite organization is associated with the compartmentalization of chromatin into active  
181 and inactive compartments. To determine if the same is true for *H. volcanii*, Pearson  
182 correlation matrices of Hi-C maps were computed and used to perform principal component  
183 analysis (PCA; **Figure 2e and Supplementary Figure 3**). Although the effect was not as  
184 sharp as previously reported, a separation of the *H. volcanii* genome into compartment-like  
185 structures emerged from the approach. Upon comparing the analysis to gene expression levels  
186 characterized by RNA-seq (**Figure 2d**) we observed that the majority of the main  
187 chromosome appeared to belong to the transcriptionally active compartment, whereas the  
188 small chromosomes pHV1, pHV3, and the integrated pHV4 seemed to mostly belong to a less  
189 transcriptionally active compartment (**Figure 2e**). The exception was the region surrounding  
190 the origin of replication of the integrated pHV4 (*ori-pHV4*) chromosome, which, like the rest  
191 of the origins of replication on the main chromosome (*oriCI-3*), appeared to fall into the A  
192 compartment. To test whether replication initiation could directly mediate compartment  
193 organization, we performed Hi-C on a H26 strain lacking all origins of replication on the main  
194 chromosome (*OriCI-3* and *ori-pHV4*)<sup>22</sup>. This strain has previously been shown to grow  
195 significantly faster than wild-type cells because cells are capable of initiating replication at  
196 sites dispersed throughout the genome<sup>22</sup>. When we compared the  $\Delta ori$  mutant to WT cells,

197 we found that not only were the compartment-like structures unchanged (**Supplementary**  
198 **Figure 4**) but that there was very little overall difference in chromosome structuring.  
199 (**Supplementary Figure 4**). We, therefore, concluded that replication is not implicated in the  
200 chromosome structuring of *H. volcanii*.

201 The *Haloferax* genome has an average GC content of ~65%, but there is extensive  
202 variation between coding and non-coding regions (**Figure 2f**). In addition, 102 AT-rich  
203 regions corresponding to insertion sequence (IS) elements are scattered throughout the  
204 genome<sup>21</sup>. A comparison between the GC content and the Hi-C contacts (**Figure 2a**) showed  
205 a strong correlation between the plaid patterns, that could be attributed to compartment-like  
206 structures, and the AT-rich regions of the genome which correspond to lowly transcribed  
207 regions (**Figure 2d**). This correlation was even more apparent for the pHV4 chromosome and  
208 the surrounding main chromosome (**Figure 2g**). Indeed, the pHV4 chromosome, which  
209 recently inserted into the main chromosome, is much more AT-rich than the surrounding  
210 sequence and contains nearly half of the 102 IS elements found in the genome. In fact, the  
211 genomic rearrangement that led to the integration of the pHV4 chromosome occurred between  
212 two identical ISH18 sequences<sup>22</sup> located in either chromosome. When we looked at the pHV1  
213 and pHV3 chromosomes, we found a similar correlation. pHV3 is more GC-rich than the  
214 other chromosomes, has fewer IS elements, and consequently has much fewer instances of  
215 the plaid pattern within its ~438 kb (**Figure 2h**). The pHV1 chromosome, on the other hand,  
216 is AT-rich, poorly transcribed, and contains 16 IS elements over 85 kb, resulting in a dense  
217 plaid pattern and many empty bins within the Hi-C matrix (**Figure 2i**). Since all Hi-C  
218 experiments were performed with HpaII (C<sup>^</sup>CGG), a possibility is that these patterns  
219 associate with AT-rich sequences simply because they are devoid of HpaII sites and are  
220 therefore under-represented in the final Hi-C contact map. This hypothesis is not supported  
221 by the fact that the 636 kb pHV4 chromosome contains over 4,900 HpaII sites (out of the

222 31,557 in the 4Mb genome) and that these sites occur at a similar density to the surrounding  
223 chromosome. Furthermore, the same plaid pattern persisted in the Hi-C matrix when Hi-C  
224 was done using a combination of HpaII and MluCI (^AATT). Taken together these  
225 observations suggest that the pattern likely results from the nature of the DNA in these regions  
226 and not the experimental system.

#### 227 *Discrete DNA loops scattered throughout the chromosomes*

228 Finally, the increased resolution afforded by our improved Hi-C technique facilitated  
229 the visualization and characterization of finer structures such as DNA loops and chromosomal  
230 borders (**Figure 2j**). To detect and quantify these structures we utilized Chromosight, a  
231 custom-made, computer-vision based program recently developed to call *de novo* DNA motifs  
232 in Hi-C maps, and efficient on compact genomes<sup>28</sup>. Chromosight detected an average of 26  
233 borders in Hi-C contact maps of WT *H. volcanii* cells across three biological replicates  
234 (**Figure 2k**), in good agreement with the 23 CIDs detected by DI analysis (**Figure 2c**).  
235 Furthermore, and more intriguingly, Chromosight detected an average of 64 loop-like  
236 structures in the *H. volcanii* genome (**Figure 2k**), most of which were detected on the main  
237 chromosome.

#### 238 **Transcription drives chromosome structuring in *H. volcanii***

239 Since CID boundaries correlated well with gene expression (**Figure 2a, c, and d**) we sought  
240 to determine the influence of transcription on the different levels of chromosome structuring.  
241 First, we compared the chromosome conformation of exponentially growing cells to cells in  
242 stationary phase. RNA-seq analysis showed a significant reduction in transcription in  
243 stationary phase, while Hi-C displayed a concomitant reduction in short- to mid-range  
244 contacts throughout the genome (**Figure 3a-c**). Also, a loss of chromosomal loops was  
245 observed throughout the main chromosome and the smaller chromosomes (**Figure 3d**)  
246 suggesting that these structures form as a consequence of gene expression, presumably

247 through the loading of a protein involved in the maintenance of the contacts between pairs of  
248 loci. To test for this hypothesis, we inhibited transcription in exponentially growing cells by  
249 treating them with actinomycin D (**Figure 3e and f**). In agreement with stationary phase  
250 cultures, we observed a decrease in short-mid range contacts, chromosomal borders, and a  
251 disappearance of DNA loops throughout the genome compared to the untreated (DMSO)  
252 control (**Figure 3g-i**). We applied Chromosight to the different Hi-C matrices to quantify the  
253 loss of these structures, showing that stationary phase and actinomycin-D treated cells  
254 displayed a ~90% decrease in the number of chromosomal loops and a significant reduction  
255 in loop score (strength) for the remaining 10% <sup>28</sup>(**Figure 3j**). Chromosomal borders, on the  
256 other hand, were less affected with a 32% and 65% reduction in the number of borders for  
257 actinomycin D-treated and stationary phase cells respectively. No decrease in border score  
258 was seen for those that remained (**Supplementary Figure 5**). Interestingly, despite the  
259 reduction in the number of chromosomal borders and loops in stationary phase cultures and  
260 actinomycin D-treated cells (**Supplementary Figure 5**), there were very little differences in  
261 the plaid patterning and assignment of A and B compartments across the genome. This is  
262 contrary to that seen in crenarchaea, where the large plaid patterns were reduced in stationary  
263 phase cultures and eliminated in cells treated with actinomycin D<sup>13</sup>. This further strengthens  
264 the conclusion that the plaid-like patterns seen in *H. volcanii* Hi-C maps are different from  
265 those seen in eukaryotes and crenarchaea and don't correspond to a partition of the genome  
266 into active and inactive compartments. Instead, they are most likely due to the AT-rich nature  
267 and increased density of IS-elements in these regions of the genome.

## 268 **Role of SMC in chromosome organization**

269 DNA loops have been identified in all eukaryotic genomes investigated using Hi-C so far, as  
270 well as in bacteria that possess a canonical member of the SMC family, such as SMC-ScpAB  
271 in *B. subtilis*<sup>10</sup>. Recently, SMC proteins have also been identified in archaea, with condensin

272 shown to be highly conserved throughout archaeal species, the notable exception being  
273 Crenarchaea, which instead encodes a distant homolog coalescin (ClSN; **Figure 4a**)<sup>13</sup>. *H.*  
274 *volcanii* contains two SMC proteins; SMC (HVO\_0689) and RAD50 (HVO\_0854), as well  
275 as the SMC-like Sph4 (HVO\_B0173) protein. The SMC protein is highly conserved across  
276 the Halobacteria (>65% identity) and other Euryarchaea, whilst also possessing some (25-  
277 30%) sequence identity with members of the Asgard and TACK superphyla (**Figure 4a**). We  
278 hypothesized that the link between transcription and DNA-loops in *H. volcanii* may be  
279 mediated by the SMC protein and so generated a strain in which the *smc* gene (HVO\_0689)  
280 had been deleted. Interestingly, and contrary to many species, the *H. volcanii smc* mutant was  
281 viable and grew similar to WT cells (**Figure 4b**). It also exhibited a normal cellular  
282 morphology and DNA content by microscopy (**Figure 4c**). Hi-C experiments performed on  
283 the *smc* mutant revealed a complex pattern of changes across the genome (**Figure 4d**). First,  
284 there was an overall decrease in short-range contacts in  $\Delta smc$  cells compared to wild-type.  
285 We also observed a reduction in both the number and strength of some loops across the  
286 genome, although the effect was less than that seen in actinomycin D-treated or stationary  
287 phase cultures. However, the most striking observation was the complete loss of several  
288 topological borders throughout the main chromosome (**Figure 4e**). Despite close inspection  
289 of the sequences and gene annotations present at these borders (**Supplementary Figure 6**),  
290 no pattern as to why these specific sites were affected in the  $\Delta smc$  mutant could be identified.  
291 To determine whether this pattern was conserved across the SMC family members present in  
292 *Haloferax*, we performed Hi-C on the published  $\Delta rad50 mre11$  mutant (**Figure 4f**)<sup>29</sup>. Like  
293 the SMC protein, Rad50 is conserved across most Euryarchaea but is less-well conserved in  
294 the Asgard and TACK groups. We chose to focus on the *rad50 mre11* double mutant because,  
295 in *H. volcanii*, both genes are expressed from the same operon, and the Mre11 Rad50 proteins  
296 have previously been shown to work together as a complex that binds DNA to prevent the

297 repair of double-strand breaks by homologous recombination<sup>29</sup>. The  $\Delta rad50 mre11$  mutant  
298 grows similar to WT and  $\Delta smc$  cells (**Figure 4b and c**) and exhibits the same loss of short-  
299 range contacts as the  $\Delta smc$  mutant (**Figure 4g**). However, we did not observe the same loss  
300 of borders in the  $\Delta rad50 mre11$  mutant, indicating that this phenotype was specific to the  
301  $\Delta smc$  cells. When we applied Chromosight to these Hi-C matrices (**Figure 4h**), these results  
302 were confirmed, the *smc* mutant showed a 50% reduction in both borders and DNA loops,  
303 whereas the *rad50 mre11* mutant was very similar to WT cells (**Figure 4i and j**). Taken  
304 together these results highlight a new role for SMC, along with transcription, in shaping  
305 Euryarchaeal genomes.

306

## 307 **Discussion**

308 The simple, cost-effective Hi-C approach described above allows the exploration of bacterial  
309 and archaeal chromosome conformations at the gene or operon level. Applied to the well-  
310 studied *E. coli* and *V. cholera* bacterial species, it resulted in sub-kilobase resolution contact  
311 maps unveiling previously undetected gene-level structures<sup>8,16</sup>. Applied to the less-studied  
312 halophilic euryarchaeal species, *H. volcanii*, it generated reproducible Hi-C matrices at a  
313 resolution of up to 1kb. This 30-fold increase in resolution, compared to available archaeal  
314 Hi-C maps of hyperthermophile species<sup>13</sup>, facilitated the identification of new chromosomal  
315 architectures in the *Haloferax* species.

316

### 317 ***The Haloferax genome is not organized into active and inactive compartments***

318 The *Haloferax* genome is organized into CIDs, similar in size to those previously seen in  
319 bacteria. In addition, the contact map of *Haloferax* displays a plaid pattern reminiscent of that  
320 seen in the Hi-C contact maps of *Sulfolobus* and eukaryotes<sup>13</sup>. In *Sulfolobus*, this patterning  
321 was shown to correspond to the transcription-dependent organization of the genome into two,

322 active and inactive, spatial compartments. However, this type of compartmentalization seems  
323 absent in *H. volcanii*, as the plaid patterns correspond to AT-rich regions of the chromosomes  
324 and likely result from the IS elements contained within these regions. A possible explanation  
325 to these structures may come from *Halobacterium salinarum*, a halophilic archaeon isolated  
326 from salted fish. This archaeon is similar to *H. volcanii* because it too has a GC-rich genome  
327 that is interspersed with AT-rich islands containing IS elements<sup>30</sup>. The AT-rich regions of  
328 *Hbt. salinarum* are hypermutable due to random transposition and recombination events  
329 between the IS elements. Although this hypermutable phenotype is yet to be demonstrated for  
330 *H. volcanii*<sup>31</sup>, it is quite conceivable that similar mechanisms exist. Many of the IS elements  
331 between the two species contain similar inverted repeats<sup>32</sup> and we know that the pHV4  
332 chromosome integrated into the main chromosome via a recombination event between two  
333 ISH18 elements<sup>22</sup>. These results suggest that the segregation of the hyperthermophile genome  
334 into different chromatin domains could be a trait specific to this lineage, linked to the extreme  
335 temperature in its environment or the fact that, unlike many archaeal species including *H.*  
336 *volcanii*, the crenarchaea do not contain histones<sup>15</sup>. This highlights that, as in bacteria,  
337 different lineages present different types of genome arrangements. The availability of an  
338 efficient protocol will allow the rapid sampling of multiple species to determine the specificity  
339 of these different organizations in this kingdom.

340

### 341 ***Transcription and SMC organize the Haloferax genome into loops and domains***

342 Discrete DNA loops structures were identified throughout the genome, whose strength and  
343 visibility strongly correlate with the transactional status of the cells. We hypothesized that the  
344 link between transcription and DNA-loops in *H. volcanii* may be mediated by an SMC  
345 protein. In most species, disruption of the SMC proteins results in lethality or a severely  
346 impaired growth phenotype. A *Haloferax* strain carrying a deletion of *smc* displays no obvious

347 defect, a phenotype previously described in several bacteria species. Indeed, in *M.*  
348 *tuberculosis*, *M. smegmatis*, *C. glutamicum*, and *D. radiodurans smc* mutations have also been  
349 shown to have very little or no effect on cell growth or the frequency of anucleate cells<sup>11,33,34</sup>.  
350 This would indicate that there is a degree of redundancy in the proteins that ensure the correct  
351 organization and compaction of the nucleoid, for instance, through distant SMC homologs  
352 such as the recently discovered condensin MksB<sup>11,35</sup>. *D. radiodurans* is of particular interest  
353 as, like *H.volcanii*, it is an extremophile containing a multipartite genome present in several  
354 copies<sup>36</sup>. *D. radiodurans smc* mutants are similar to WT cells in every way, except that they  
355 are hypersensitive to DNA gyrase inhibitors, suggesting SMC is required to maintain the  
356 supercoiling equilibrium in these cells<sup>34</sup>. Furthermore, in WT cells it was shown that SMC is  
357 located at several foci on the nucleoid periphery and that these foci are unevenly distributed  
358 throughout the cell, signifying that SMC may bind to a few distinct DNA regions throughout  
359 the *D. radiodurans* genome<sup>34</sup>. Our data in *H. volcanii* agree with these observations, and the  
360 chromosomal borders that are most severely disrupted in *smc* mutant cells may correspond to  
361 specific SMC binding sites in the *Haloferax* genome. This would also explain why we don't  
362 see the same phenotype in *rad50 mre11* mutant cells, and it would be interesting to see if the  
363 SMC-like protein, Sph4 is also implicated at these sites. The peculiarities of these sites remain  
364 an open question, as no conserved sequences could be found at the boundaries. A possibility  
365 could be that these sites are implicated in chromosome segregation or cohesion between  
366 homologous chromosomes, but this is difficult to test by Hi-C as the technique is not able to  
367 discern between homologous chromosomes. The SMC protein is well conserved throughout  
368 the Euryarchaeal phylum, which consists mainly of polyploids. Here too, further studies will  
369 determine whether these structures are conserved in these species.

370

371 This study pinpoints that understanding the chromosomal architecture of *H. volcanii* and  
372 archaea, in general, is still in its infancy. The stark contrast between this archaeon and the  
373 *Sulfolobus* species may be pondered once high-resolution maps of the latter will be made  
374 available, as some of the *Haloferax* features such as DNA loops could exist in *Sulfolobus* as  
375 well. However, the extreme and diverse environment of archaea also forecasts the  
376 identification of even more discrepancies between the different lineages. Since Hi-C can be  
377 applied to cells directly crosslinked in their natural environment, there should be no  
378 limitation to lab-based model organisms, and it would be of interest to generate high-  
379 resolution contact maps of species spanning the archaeal tree to reach a first broad overview  
380 of their differences and similarities. The availability of high-resolution metagenomic Hi-C  
381 assays will further allow the sampling of genome folding of hundreds of species from the  
382 wild, including archaea. This technique will further boost the exploration of the evolutionary  
383 role of genome folding in the microbial world.

384

385

## 386 **Methods:**

### 387 **Strains, media and growth conditions**

388 Detailed information about all strains, plasmids, and primers used in this study can be found  
389 in Supplementary Table 2. *Escherichia coli* MG1655 K12<sup>37</sup> and *Vibrio cholera* N16961<sup>38</sup>  
390 were both grown at 37°C in 1 x Minimal Media A supplemented with 0.2 % casamino acids  
391 and 0.5 % glucose. The archaea, *Haloferax volcanii* H26 strain ( $\Delta pyrE$ ) is referred to as “wild-  
392 type” during this study. *H. volcanii* strains were grown using enriched Hv-YPC media at  
393 45°C, as described previously<sup>39</sup>. To block transcription, cells were exposed for 30 min to  
394 Actinomycin D (Sigma Aldrich, 5  $\mu$ g/ml final concentration, prepared in 100% DMSO). A  
395 negative control was prepared in parallel by adding the same volume of 100% DMSO (Sigma  
396 Aldrich, 0.5% final concentration). *Escherichia coli* strains XL1-Blue MRF' and GM121  
397 were used for cloning. GM121 was also used to prepare unmethylated plasmid DNA for  
398 efficient transformation of *H. volcanii*.

### 399 **Construction of the *H. volcanii* $\Delta smc$ strain**

400 To generate the  $\Delta smc$  (HVO\_0689) construct, the upstream region (US) and downstream  
401 region (DS) of *smc* were cloned into pTA131<sup>39</sup> to generate the pRL93 plasmid used to  
402 construct the deletion strain by a gene knockout system<sup>40</sup>. The US region was generated by  
403 PCR on H26 genomic DNA using primers RL292 and RL296. The DS region was generated  
404 by PCR on H26 genomic DNA using primers RL294 and RL295. Each PCR product  
405 contained 30 bp homology with adjacent fragments for SLIC cloning<sup>41</sup>: (i) with the pTA131  
406 linearized fragments after NotI and EcoRI double-digestion (contained in RL292) in the US  
407 fragment, (ii) with the US fragment (contained in RL294) and with the pTA131 linearized  
408 fragments after NotI and EcoRI double-digestion (contained in RL295) in the DS fragment.  
409 Following the SLIC method, PCR fragments and linearized plasmid were digested using T4

410 DNA polymerase for 45 minutes at 22°C to generate 3'-single-stranded extremities, then all  
411 amplification products were mixed in a 1:1:1 molar ratio and incubated for 30 minutes at 37°C  
412 before transformation into *E. coli* XL1-blue. Transformants were selected on LB plates  
413 containing 100 µg/ml ampicillin, 0.5 µM IPTG and 80 µg/ml X-Gal. The presence of the  
414 correct insert in the plasmid, as determined by white colonies, was tested by colony PCR  
415 using primers pBSF2 and pBSR3. The sequence of one selected plasmid, dubbed pRL93, was  
416 further confirmed by Sanger sequencing. Then pRL93 was used to transform H26 strain using  
417 the pop-in/pop-out method as described previously<sup>40</sup>. Pop-out colonies were plated on Hv-Ca  
418 plates containing 5-FOA and thymidine. The absence of the *smc* was tested by colony lift on  
419 100 colonies from the pop-out plates using a probe targeting the *smc* gene. The DIG-labeled  
420 probe was generated by PCR on H26 genomic DNA using primers RL305 and RL310 and the  
421 PCR DIG (digoxigenin) labeling Mix (Roche). Probe hybridization was detected using the  
422 DIG Luminescent Detection Kit (Roche) and a ChemiDoc MP (BioRad). 10 colonies out of  
423 100 were deleted for the *smc* gene. One  $\Delta smc$  construct was conserved for further studies and  
424 dubbed HvRL138.

## 425 **Optimized microbial Hi-C protocol**

### 426 **Cell fixation**

427 All Hi-C experiments were performed with  $\sim 1 \times 10^8$  cells growing in the exponential growth  
428 phase ( $OD_{600nm} \sim 0.2$ ). Protein-DNA interactions were chemically crosslinked with fresh  
429 formaldehyde (Sigma Aldrich; 3% final concentration) for 30 min at room temperature with  
430 gentle agitation. Crosslinking was quenched by the addition of glycine (Sigma Aldrich; 0.5  
431 M final concentration) for 20 min at room temperature. For *H. volcanii*, which is very sensitive  
432 to salt concentrations, 2.5M glycine was prepared in 18 % (w/v) saltwater solution ( 2.5 M  
433 NaCl, 90 mM MgCl<sub>2</sub>.6H<sub>2</sub>O, 90 mM MgSO<sub>4</sub>.7H<sub>2</sub>O, 60 mM KCl, 10 mM Tris-HCl pH 7.5) to  
434 avoid cell lysis. Cells were then collected by centrifugation (4000 x g, 10 min, room

435 temperature) resuspended in 50 ml of 1 x PBS (18% (w/v) saltwater for *H. volcanii*) and  
436 centrifuged again. The pellet was resuspended in 1 ml 1 x PBS and transferred to a 1.5 ml  
437 Eppendorf tube before a final centrifugation step (4000 x g, 5 min, room temperature), the  
438 supernatant was then removed and the pellet stored at -80°C.

### 439 **Cell lysis and DNA digestion**

440 Cell pellets were removed from the -80°C freezer and allowed to completely thaw on ice. The  
441 pellet was then resuspended in 1 ml of 1x TE + cOmplete protease inhibitor cocktail (EDTA-  
442 free, Sigma Aldrich) and transferred to a 2 ml Precellys tube containing 0.5 mm glass beads  
443 (VK05, Ozyme). Cells were then subject to mechanical disruption using the Precellys  
444 Evolution tissue homogenizer (V7500: 5 x 30s, 20s pause). Please note, that for Precellys  
445 machines without the Cryolys cooling attachment (Ozyme), tubes need to be removed every  
446 3 cycles and placed on ice for 5 min to stop sample overheating and subsequent degradation.  
447 The lysate (~1ml in volume) was then carefully transferred to a 5 ml Eppendorf tube, avoiding  
448 the transfer of any glass beads. Proteins that were not crosslinked to DNA during  
449 formaldehyde fixation were then degraded by the addition of SDS (Thermo Fisher, 0.5% final  
450 concentration) for 10 min at room temperature. DNA was then prepared for digestion by the  
451 addition of 3 ml dH<sub>2</sub>O, 500 µl 10X Digestion buffer (200 mM Tris-HCl pH 7.5, 100 mM  
452 MgCl<sub>2</sub>, 10 mM DTT, 1 mg/ml BSA) and 500 µl 10 % Triton-X-100 (Thermo-Fisher). After  
453 thoroughly mixing the reaction, 400 µl was removed and transferred to a 1.5 ml Eppendorf  
454 tube as a non-digested (ND) control. The restriction enzyme, HpaII (New England Biolabs,  
455 1000 U) was then added to the remaining sample and the tube incubated with gentle agitation  
456 for 3h at 37°C.

### 457 **Choice of restriction enzymes for Hi-C**

458 Previously, it has been shown that the frequency of restriction sites is correlated with the GC-  
459 content of a genome, and thus certain restriction enzymes can introduce significant bias into  
460 3C/Hi-C contact maps. For this reason, eukaryotic Hi-C protocols typically use DpnII  
461 ( $\text{^GATC}$ ) to offer more uniform coverage, but this is not possible for bacterial species, such  
462 as *E. coli* and *V. cholera*, whose genomes are subject to Dam methylation. We have previously  
463 utilized both HpaII ( $\text{GGC^C}$ ) and MluCI ( $\text{A^ATT}$ ) for bacterial 3C and metagenomic  
464 experiments<sup>12,42</sup> and both these enzymes have between  $\sim 7,700 - 31,000$  sites in all three of  
465 organisms in this study, generating an average fragment length of 188 bp (**Supplementary**  
466 **Table 1**). We decided to test these two enzymes in combination with biotinylation to see if  
467 we could improve our Hi-C resolution, we also combined both HpaII and MluCI in the same  
468 experiment, which would, in theory, reduce the average fragment size to  $\sim 82$  bp, further  
469 increasing the resolution of the Hi-C data and covering both AT and GC-rich regions of the  
470 genome in the same experiment. The biotinylation step and subsequent streptavidin-  
471 enrichment was done using Biotin-14-dCTP and Biotin-14-dATP for HpaII and MluCI,  
472 respectively. For the experiments performed using both enzymes, we used the same quantity  
473 of enzyme as the single-enzyme experiments and combined the two biotinylated nucleotides  
474 in equal amounts for the enrichment step. Compared to previously published data, the  
475 combination of the four base cutters with biotinylation did improve the number of informative  
476 reads for each experiment (**Supplementary Figure S1**), with HpaII and MluCI performing  
477 well. As expected, MluCI failed to digest the GC-rich *H. volcanii* genome and in all  
478 organisms, the combination of the two enzymes didn't add to the resolution, if anything it  
479 made the contact maps noisier than using either enzyme alone (**Supplementary Figure S1**).

#### 480 **Biotinylation of DNA ends**

481 Following digestion, 400 $\mu$ l of the sample was removed and transferred to a 1.5ml Eppendorf  
482 as the digested (D) control. The remaining lysate was centrifuged (16,000 x g, 20 min, room

483 temperature) to pellet the insoluble fraction containing the protein-DNA complexes of  
484 interest. After removing the supernatant, the pellet was resuspended in 400  $\mu$ l dH<sub>2</sub>O. To this  
485 the following was added; 50  $\mu$ l 10x Ligation Buffer (500 mM Tris-HCl pH 7.5, 100 mM  
486 MgCl<sub>2</sub>, 100 mM DTT), 4.5  $\mu$ l 10 mM dAGTTP, 37.5  $\mu$ l Biotin-14-dCTP (Thermo Fisher),  
487 50 Units of DNA Polymerase I - Large Klenow Fragment (New England Biolabs). After  
488 briefly mixing, the reaction was incubated with agitation for 45 min at 37°C.

#### 489 **Blunt-end ligation**

490 For further gains, we took inspiration from published human protocols and reduced the  
491 volume of the blunt-end ligation step by 20-fold to mimic the conditions of 'in situ Hi-C'  
492 protocols<sup>17,43</sup>. Our rationale was that in a smaller volume there would be less spurious events  
493 than in dilute conditions and, as a result, less noise in the resulting Hi-C matrices. Following  
494 further trial-and-error optimization, we performed the ligation at room temperature and  
495 reduced its duration (**Figure 1a and Supplementary Figure 1**). Once the fill-in reaction was  
496 complete, the ligation was set up by adding the following; 120 $\mu$ l 10x Ligation Buffer, 12  $\mu$ l  
497 10 mg/ml BSA, 12  $\mu$ l 100mM ATP, 540  $\mu$ l dH<sub>2</sub>O, 480 U T4 DNA ligase (Thermo Fisher).  
498 The reaction was mixed gently and then incubated with gentle agitation for 3h at room  
499 temperature.

#### 500 **Reverse crosslinking and DNA purification**

501 Following ligation, proteins were denatured by the addition of 20  $\mu$ l 500 mM EDTA, 20  $\mu$ l  
502 10 % SDS, and 100  $\mu$ l 20 mg/ml proteinase K (EuroBio). The non-digested (ND) and digested  
503 (D) controls were also treated using 20  $\mu$ l 500 mM EDTA, 20  $\mu$ l 10 % SDS and 10  $\mu$ l 20mg/ml  
504 proteinase K (EuroBio). All samples were incubated at 65°C overnight to reverse  
505 formaldehyde-mediated protein-DNA crosslinks.

506 The following day, DNA was purified by the addition of an equal volume of  
507 Phenol:Chloroform: Isoamyl alcohol (25:24:1, Amresco), the mixture was vortexed for 30s  
508 and then centrifuged (12,000 x g, 5 min, room temperature). Following centrifugation, the  
509 upper aqueous phase was carefully removed and transferred to a new Eppendorf tube. DNA  
510 was precipitated by the addition of 2.5 x volume of ice-cold 100 % EtOH and 1/10 volume of  
511 3 M NaOAc (pH 5.0). Samples were then incubated at -80°C for 30 min. Precipitated DNA  
512 was then pelleted by centrifugation (12,000 x g, 20 min, 4°C), the supernatant removed and  
513 the pellets washed in 500 µl 70% EtOH. After a second centrifugation (12,000 x g, 5 min,  
514 4°C), the EtOH was removed and the pellets dried on a 37°C heat block for 5-10 min. Once  
515 the remaining EtOH had evaporated, pellets were resuspended in 140 µl 1X TE buffer +  
516 1mg/ml RNase (Euromedex) and incubated for 30 min at 37°C with agitation. When the DNA  
517 had been completely resuspended, 10 µl of HiC libraries and the ND and D controls were  
518 checked on a 1 % agarose gel (Supplementary Figure 2e). This step is included to ensure the  
519 DNA isn't degraded, was successfully digested by the restriction enzyme, and subsequently  
520 ligated after biotin integration. Once the quality of the HiC library has been confirmed, the  
521 ND and D controls can be discarded.

## 522 **Preparation of sequencing libraries**

### 523 **DNA sonication and size-selection**

524 For efficient sonication, a maximum of 5 µg of DNA should be used, if the HiC library  
525 exceeds this, then an aliquot should be taken and the remaining DNA stored at -20°C, as a  
526 backup. Once the correct quantity of DNA was obtained, the sample volume was adjusted to  
527 130 µl using 1x TE buffer. HiC samples were transferred to a sonication tube (microTUBE  
528 AFA Fiber Pre-Slit Snap-Cap, Covaris). The DNA was then sheared using the Covaris S220  
529 Focused Ultrasonicator to yield an average fragment size of 300 bp (140 W peak incidence  
530 power, 10 % duty factor, 200 cycles per burst, 7°C). Following sonication, the DNA was

531 transferred to a 1.5 ml Eppendorf tube and an equal volume of AmPure XP beads (Beckman)  
532 added. The sample was mixed 10x by gentle pipetting and then incubated for 5 min at room  
533 temperature to allow the DNA fragments to bind to the magnetic beads. The tube was then  
534 transferred to a magnetic rack for 1 min or until the beads separated to the wall of the tube.  
535 The supernatant was then carefully removed and the beads washed 2x using freshly-prepared  
536 70 % EtOH. The pellet was then allowed to air-dry for 30s, to remove residual ethanol, before  
537 being resuspended in 320  $\mu$ l of elution buffer (10 mM Tris-HCl, pH 8,5). Size-selection was  
538 checked by running 18  $\mu$ l of the DNA on a 1% agarose gel (Supplementary Figure 2e) and  
539 DNA quantity was checked by QuBit analysis, the remaining 300  $\mu$ l was taken for sequencing  
540 library preparation.

#### 541 **Biotin pull-down**

542 The biotin-streptavidin pull-down maximizes the number of “true” HiC contacts in an  
543 experiment by selecting for chimeric DNA molecules that have been digested, biotin-filled  
544 and blunt-end ligated. To do this, the HiC library was bound to Streptavidin C1 Dynabeads  
545 (Thermo Fisher). First, the beads were mixed well by vortexing and then 30  $\mu$ l of beads were  
546 aliquoted to a 1.5 ml Eppendorf. The beads were placed on a magnet and left for 1 min to  
547 clear the supernatant, the supernatant was then removed and 500  $\mu$ l of 1x Tween Wash Buffer  
548 (TWB: 5 mM Tris-HCl pH7.5, 0.5 mM EDTA, 1M NaCl, 0.05% Tween 20) was added to  
549 rinse the beads. The solution was then separated on a magnet and the supernatant discarded,  
550 the beads were then re-suspended in 300  $\mu$ l of 2x Binding Buffer (BB: 10 mM Tris-HCl  
551 pH7.5, 1 mM EDTA, 2 M NaCl), 300  $\mu$ l of HiC sample was added and the mixture incubated  
552 on a tube rotator for 15 min at room temperature. The tube was then placed back on the magnet  
553 for 1 min to separate the biotinylated DNA bound to the streptavidin beads from the  
554 supernatant containing unbiotinylated DNA. The beads were then washed twice in 500  $\mu$ l  
555 TWB in a thermomixer (2 min, 55°C, 1,000 rpm). Finally, the beads were resuspended in 100

556  $\mu$ l of 1x T4 ligase buffer (New England Biolabs), and the mixture transferred to a new 1.5 ml  
557 Eppendorf.

### 558 **End-repair**

559 The end-repair mix was prepared as follows: 85  $\mu$ l 1x T4 ligase buffer, 5  $\mu$ l 10mM dNTPs,  
560 50 U T4 PNK, 12 U T4 DNA polymerase, 5 U DNA Polymerase I, Large Klenow Fragment.  
561 After the HiC bank was placed on the magnet and the supernatant removed, the beads were  
562 resuspended in the end repair mix and the reaction incubated for 30 min at room temperature.

### 563 **A-tailing**

564 Following end-repair, the beads were washed twice in 500  $\mu$ l TWB and once in 100  $\mu$ l 1x  
565 NEB2 buffer using a thermomixer (2 min, 55°C, 1000 rpm). The beads containing the end-  
566 repaired DNA were then resuspended in 100  $\mu$ l of A-tailing Mix (90  $\mu$ l 1xNEB 2 buffer, 0.5  
567 mM dATP, 25 U Klenow Fragment (3'-5'exo-)) and incubated for 30 min at 37°C.

568

### 569 **Adapter ligation**

570 After A-tailing, the beads were washed twice in 500  $\mu$ l TWB and once in 50  $\mu$ l 1x Quick  
571 Ligase Buffer (New England Biolabs) using a thermomixer (2 min, 55°C, 1000 rpm). The  
572 beads were then resuspended in the Ligation Mix (50  $\mu$ l 1x Quick Ligase Buffer, 2  $\mu$ l Quick  
573 DNA Ligase) and 2  $\mu$ l of sequencing adapter was added (NEXTflex Illumina Barcodes, Bioo  
574 Scientific). The reaction was mixed well and then incubated at room temperature for 10 min,  
575 the beads were washed twice in 500  $\mu$ l TWB and once in 100  $\mu$ l of 10 mM Tris-HCl pH 8  
576 using a thermomixer (2 min, 55°C, 1000 rpm). Finally, the beads were resuspended in 50  $\mu$ l  
577 of 10 mM Tris-HCl pH 8 and transferred to a new tube.

### 578 **Library amplification by PCR**

579 The PCR reaction was set up as follows: 40  $\mu$ l of Phusion 2x High Fidelity Master Mix (New  
580 England Biolabs), 5  $\mu$ l of 2  $\mu$ M Primer Mix (NEXTflex, Bioo Scientific), 32  $\mu$ l dH<sub>2</sub>O and 3  
581  $\mu$ l Streptavidin-beads containing the HiC library. The library was amplified for 12 cycles  
582 following the manufacturers' instructions. After PCR, the reaction was placed on a magnet  
583 for 1 min to separate the streptavidin beads from the PCR reaction containing the amplified  
584 library. The supernatant was transferred to a new tube and the beads discarded. The PCR  
585 reaction was then purified by adding an equal volume of AmPure XP beads, the reaction was  
586 then mixed well and incubated at room temperature for 5 min. Following two washes with 80  
587 % EtOH, the beads were mixed with 50  $\mu$ l of 10 mM Tris-HCl pH 8 and incubated for 5 min  
588 at room temperature. Finally, the reaction was placed on the magnet and the DNA-containing  
589 supernatant transferred to a new Eppendorf tube. The DNA was then checked on a 1% agarose  
590 gel to determine the size of the final library and to check for the presence of primer dimers.  
591 Once this was confirmed, the Hi-C library was ready to prepare for paired-end sequencing  
592 using the 75 cycle High-Output Kit v2.5. Sequencing was performed on a Nextseq500  
593 according to Illumina's instructions.

#### 594 **Processing of Hi-C sequencing data**

595 All data were processed as previously described<sup>8,19</sup>. So that our datasets could be compared  
596 to those previously published, the raw reads were downloaded from the SRA database  
597 (<https://www.ncbi.nlm.nih.gov/sra>) and re-processed using our pipeline (Supplementary  
598 Table 1). Briefly, reads were aligned independently by Bowtie2 using a very-sensitive and  
599 iterative approach. Each read was then assigned to a restriction fragment, with uninformative  
600 events, such as self-circularized and uncut fragments discarded by filtering. The filtered  
601 fragments were then binned into 500 bp, 1 kb or 5 kb segments, and the corresponding contact  
602 maps generated and normalized using the sequential component normalization (SCN)  
603 procedure<sup>19</sup>.

#### 604 **Calculating generation time of wild-type and mutant strains**

605 Overnight cultures were diluted to  $OD_{600nm} \sim 0.05$  and incubated at 45°C. Every 2h, aliquots  
606 were collected and serial dilutions made in 18% saltwater (18% SW) containing 2.5M NaCl,  
607 90 mM  $MgCl_2 \cdot 6H_2O$ , 90 mM  $MgSO_4 \cdot 7H_2O$ , 60 mM KCl, 10 mM Tris HCl pH 7.5. 20  $\mu$ l cell  
608 aliquots were then spotted on Hv-YPC plates and individual colonies were counted after 4  
609 days of incubation at 45 °C. The curve CFU/ml = f(time) was drawn. One early time-point  
610 and one late time point were then selected to calculate generation time (G) using the formula:  
611  $G = [(t_{Late} - t_{Early})] / [(\text{Log}_{10}(N_{Late}) - \text{Log}_{10}(N_{Early}) / \text{Log}_{10}(2))]$ , t designating time and N the  
612 number of colonies per ml. Generation time measurements were performed at three  
613 independent times for each strain, to provide associated standard errors.

#### 614 **Wide-field microscopy and DNA staining of *H. volcanii* cells**

615 DNA was labeled using the fluorescent Hoechst 33342 dye (excitation 361nm / emission 486  
616 nm). Exponentially growing cells were centrifuged and resuspended in an equal volume of  
617 18% SW containing 5  $\mu$ g *per* ml of Hoechst 33342. Cells were incubated for 10 min in the  
618 dark and then mounted onto glass slides covered with a thin layer of 1% agarose (suspended  
619 in 18% SW solution). DIC and fluorescent images were obtained at room temperature using  
620 a ZEISS Axio Observer microscope equipped with a 40x, 1.4 NA oil immersion objective.  
621 365 nm excitation at a maximum available intensity (2  $W \cdot cm^{-2}$ ) and filter set 49 (EX G 365,  
622 BS FT 395, EM BP 445/50) were used for fluorescence imaging of Hoechst signal. A Z-stack  
623 (30 slices) centered on DIC focus was performed and data was collected sequentially for DIC  
624 and Hoechst signals at each plane. The optical slice with the maximum number of cells at the  
625 focus was chosen for further study.

#### 626 **RNA-seq**

627 Total RNA was extracted from *H. volcanii* using the Nucleospin RNA Extraction Kit  
628 (Macherey-Nagel) according to manufacturers' instructions. After treatment with Turbo  
629 DNase (Thermo Fisher), the RNA was purified by phenol extraction (pH 4.5, Amresco) and  
630 ethanol precipitation. The RNA was then resuspended in DEPC-treated water. Ribosomal  
631 RNA depletion, cDNA library preparation, and illumina PE150 sequencing was performed  
632 by Novogene. Three biological replicates were generated for each condition and on average  
633 ~7million reads were generated per sample. Clean reads were aligned to the reference genome  
634 with HISAT2 using default parameters<sup>44</sup>. Reads were then processed using bamCoverage  
635 (implemented in deepTools<sup>45</sup>) to calculate Reads per Kilobase region per Million mapped  
636 reads (RPKM) for genomic bins of a fixed size. Bin size was determined by the Hi-C map  
637 that the RNA-seq data was being compared to (typically 1 kb or 5 kb bins).

### 638 **Directional index analysis**

639 The directional index (DI) is a statistical parameter that quantifies the degree of upstream and  
640 downstream contact biases for a genomic region<sup>46</sup>. It is based on a t-test between vectors to  
641 the left and right of each bin, up to a certain scale. We utilized a workflow that has been  
642 described previously<sup>8</sup> to perform a DI analysis for domains consistent with the scale of  
643 bacterial CIDs (100 kb) and larger structures such as macrodomains (300 kb).

### 644 **Identification of compartments**

645 The compartment index was calculated as described previously<sup>47</sup>. For each Hi-C matrix, a  
646 distance-normalized matrix was generated. These matrices were then converted into Pearson  
647 correlation matrices and from this, the eigenvalues and first eigenvector were calculated. The  
648 eigenvector allows the A and B compartments to be attributed. Positive and negative  
649 eigenvalues were arbitrarily chosen to correspond to the A and B compartments respectively.

### 650 **Detection of chromosomal loops and borders using Chromosight**

651 Chromosight is an algorithm based on computer vision approaches that are capable of  
652 automatically detecting common patterns seen in Hi-C contact maps<sup>28</sup>. These patterns are then  
653 associated with a score (Pearson coefficient) which allows for the interpretation of results and  
654 the comparison of different mutants/treatments. We applied Chromosight to each of our *H.*  
655 *volcanii* matrices (5 kb bin) and searched for loops (Pearson: 0.38), borders (Pearson: 0.2),  
656 and hairpins (default parameters). We detected both loops and borders in our Hi-C matrices  
657 but failed to detect any hairpin structures. We then used Chromosight to compare loop and  
658 border scores between wild-type and mutant/treated cells to determine the effect on these finer  
659 chromosome structures.

#### 660 **Phylogenetic tree of SMCs, Rad50, and ClsN in different archaeal lineages**

661 The tree shown in Figure 4 has been adapted from two previous studies<sup>13,48</sup> with the addition  
662 of details about the conservation of Rad50 in the different archaeal lineages. To identify  
663 probable homologs of Rad50, we used the sequence of human Rad50 as a query (accession  
664 number: AAB07119.1) to perform a BLAST search against available archaeal genomes<sup>49</sup>. We  
665 then classified the hits into one of four groups: i) present, ii) absent, iii) distant homologs  
666 present in all species, iv) distant homologs present in some species. From this, we conclude  
667 that, like canonical SMCs, Rad50 is present in many euryarchaeal species and although Rad50  
668 has been observed in the Asgard and Tack groups, it seems to be less conserved for the species  
669 within these groups.

#### 670 **Acknowledgments**

671 We thank Olivier Espéli, Frédéric Bocard, and Vicky Lioy for careful reading of the  
672 manuscript. We also thank the members of the RSG lab for insightful discussions about  
673 experiments, especially Cyril Matthey-Doret and Axel Cournac for their guidance regarding  
674 computational analysis as well as for providing direct access to the Chromosight program.

675 This research was supported by funding to R.K from the European Research Council under  
676 the Horizon 2020 Program (ERC grant agreement 260822).

677 **Author Contributions**

678 C.C, R.L, and R.K conceived of the study and designed the experiments. C.C. and A.T,  
679 adapted the Hi-C protocol. C.C, A.T, and R.L performed experiments. C.C did the analysis.  
680 C.C and R.K wrote the manuscript.

681

682 **Declaration of interest**

683 The authors declare no competing interests.

684

685 **Data Availability**

686 Sample description and raw sequences are accessible on SRA database through the following  
687 accession number: PRJNA587586.

## References:

1. Ausiannikava, D. *et al.* Evolution of Genome Architecture in Archaea: Spontaneous Generation of a New Chromosome in *Haloferax volcanii*. *Mol. Biol. Evol.* **35**, 1855–1868 (2018).
2. Koonin, E. V. & Wolf, Y. I. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* **36**, 6688–6719 (2008).
3. Harrison, P. W., Lower, R. P. J., Kim, N. K. D. & Young, J. P. W. Introducing the bacterial ‘chromid’: not a chromosome, not a plasmid. *Trends Microbiol.* **18**, 141–148 (2010).
4. Egan, E. S., Fogel, M. A. & Waldor, M. K. Divided genomes: negotiating the cell cycle in prokaryotes with multiple chromosomes. *Mol. Microbiol.* **56**, 1129–1138 (2005).
5. Makarova, K. S. & Koonin, E. V. Archaeology of eukaryotic DNA replication. *Cold Spring Harb. Perspect. Biol.* **5**, a012963 (2013).
6. Ng, W. V. *et al.* Snapshot of a large dynamic replicon in a halophilic archaeon: megaplasmid or minichromosome? *Genome Res.* **8**, 1131–1141 (1998).
7. Le, T. B. K., Imakaev, M. V., Mirny, L. A. & Laub, M. T. High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome. *Science* **342**, 731–734 (2013).
8. Lioy, V. S. *et al.* Multiscale Structuring of the *E. coli* Chromosome by Nucleoid-Associated and Condensin Proteins. *Cell* **172**, 771–783.e18 (2018).
9. Wang, X. *et al.* Condensin promotes the juxtaposition of DNA flanking its loading site in *Bacillus subtilis*. *Genes Dev.* **29**, 1661–1675 (2015).
10. Marbouty, M. *et al.* Condensin- and Replication-Mediated Bacterial Chromosome Folding and Origin Condensation Revealed by Hi-C and Super-resolution Imaging. *Mol. Cell* **59**, 588–602 (2015).
11. Böhm, K. *et al.* Chromosome organization by a conserved condensin-ParB system in the actinobacterium *Corynebacterium glutamicum*. *Nat. Commun.* **11**, 1–17 (2020).

12. Marbouty, M., Baudry, L., Cournac, A. & Koszul, R. Scaffolding bacterial genomes and probing host-virus interactions in gut microbiome by proximity ligation (chromosome capture) assay. *Sci. Adv.* **3**, e1602105 (2017).
13. Takemata, N., Samson, R. Y. & Bell, S. D. Physical and Functional Compartmentalization of Archaeal Chromosomes. *Cell* **179**, 165-179.e18 (2019).
14. Mullakhanbhai, M. F. & Larsen, H. *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Arch. Microbiol.* **104**, 207–214 (1975).
15. Barillà, D. Driving Apart and Segregating Genomes in Archaea. *Trends Microbiol.* **24**, 957–967 (2016).
16. Val, M.-E. *et al.* A checkpoint control orchestrates the replication of the two chromosomes of *Vibrio cholerae*. *Sci. Adv.* **2**, e1501914 (2016).
17. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).
18. Orlando, G., Kinnersley, B. & Houlston, R. S. Capture Hi-C Library Generation and Analysis to Detect Chromatin Interactions. *Curr. Protoc. Hum. Genet.* **98**, e63 (2018).
19. Cournac, A., Marie-Nelly, H., Marbouty, M., Koszul, R. & Mozziconacci, J. Normalization of a chromosomal contact map. *BMC Genomics* **13**, 436 (2012).
20. Boccard, F., Esnault, E. & Valens, M. Spatial arrangement and macrodomain organization of bacterial chromosomes. *Mol. Microbiol.* **57**, 9–16 (2005).
21. Hartman, A. L. *et al.* The Complete Genome Sequence of *Haloferax volcanii* DS2, a Model Archaeon. *PLOS ONE* **5**, e9605 (2010).
22. Hawkins, M., Malla, S., Blythe, M. J., Nieduszynski, C. A. & Allers, T. Accelerated growth in the absence of DNA replication origins. *Nature* **503**, 544–547 (2013).
23. Wendoloski, D., Ferrer, C. & Dyall-Smith, M. L. A new simvastatin (mevinolin)-resistance marker from *Haloarcula hispanica* and a new *Haloferax volcanii* strain cured of plasmid pHV2The GenBank accession number for the sequence reported in this paper is AF123438. *Microbiology*, vol. 147 959–964 (2001).

24. Le, T. B. & Laub, M. T. Transcription rate and transcript length drive formation of chromosomal interaction domain boundaries. *EMBO J.* e201593561 (2016) doi:10.15252/embj.201593561.
25. Sexton, T. *et al.* Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell* **148**, 458–472 (2012).
26. Hsieh, T.-H. S. *et al.* Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **162**, 108–119 (2015).
27. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).
28. Chromosight: A computer vision program for pattern detection in chromosome contact maps | bioRxiv. <https://www.biorxiv.org/content/10.1101/2020.03.08.981910v2.abstract>.
29. Delmas, S., Shunburne, L., Ngo, H.-P. & Allers, T. Mre11-Rad50 Promotes Rapid Repair of DNA Damage in the Polyploid Archaeon *Haloferax volcanii* by Restraining Homologous Recombination. *PLOS Genet.* **5**, e1000552 (2009).
30. Hackett, N. R., Bobovnikova, Y. & Heyrovská, N. Conservation of chromosomal arrangement among three strains of the genetically unstable archaeon *Halobacterium salinarum*. *J. Bacteriol.* **176**, 7711–7718 (1994).
31. López-García, P., Abad, J. P., Smith, C. & Amils, R. Genomic organization of the halophilic archaeon *Haloferax mediterranei*: physical map of the chromosome. *Nucleic Acids Res.* **20**, 2459–2464 (1992).
32. Hofman, J. D., Schalkwyk, L. C. & Doolittle, W. F. ISH51: a large, degenerate family of insertion sequence-like elements in the genome of the archaeobacterium, *Halobacterium volcanii*. *Nucleic Acids Res.* **14**, 6983–7000 (1986).
33. Güthlein, C., Wanner, R. M., Sander, P., Böttger, E. C. & Springer, B. A Mycobacterial *smc* Null Mutant Is Proficient in DNA Repair and Long-Term Survival. *J. Bacteriol.* **190**, 452–456 (2008).

34. Bouthier de la Tour, C. *et al.* The *Deinococcus radiodurans* SMC protein is dispensable for cell viability yet plays a role in DNA folding. *Extrem. Life Extreme Cond.* **13**, 827–837 (2009).
35. Pradhan, S. *et al.* MksB, an alternate condensin from *Mycobacterium smegmatis* is involved in DNA binding and condensation. *Biochimie* **171–172**, 136–146 (2020).
36. Cox, M. M. & Battista, J. R. *Deinococcus radiodurans* — the consummate survivor. *Nat. Rev. Microbiol.* **3**, 882–892 (2005).
37. Blattner, F. R. *et al.* The Complete Genome Sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1462 (1997).
38. Heidelberg, J. F. *et al.* DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**, 477–483 (2000).
39. Allers, T., Ngo, H.-P., Mevarech, M. & Lloyd, R. G. Development of Additional Selectable Markers for the Halophilic Archaeon *Haloferax volcanii* Based on the *leuB* and *trpA* Genes. *Appl. Environ. Microbiol.* **70**, 943–953 (2004).
40. Bitan-Banin, G., Ortenberg, R. & Mevarech, M. Development of a Gene Knockout System for the Halophilic Archaeon *Haloferax volcanii* by Use of the *pyrE* Gene. *J. Bacteriol.* **185**, 772–778 (2003).
41. Li, M. Z. & Elledge, S. J. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* **4**, 251–256 (2007).
42. Marbouty, M. *et al.* Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. *eLife* <https://elifesciences.org/articles/03318> (2014) doi:10.7554/eLife.03318.
43. Gavrilov, A. A. *et al.* Disclosure of a structural milieu for the proximity ligation reveals the elusive nature of an active chromatin hub. *Nucleic Acids Res.* **41**, 3563–3575 (2013).
44. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

45. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
46. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
47. Moreau, P. *et al.* Tridimensional infiltration of DNA viruses into the host genome shows preferential contact with active chromatin. *Nat. Commun.* **9**, 1–14 (2018).
48. Spang, A., Caceres, E. F. & Ettema, T. J. G. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* **357**, eaaf3883 (2017).
49. Hopfner, K.-P. *et al.* Mre11 and Rad50 from *Pyrococcus furiosus*: Cloning and Biochemical Characterization Reveal an Evolutionarily Conserved Multiprotein Machine. *J. Bacteriol.* **182**, 6036–6041 (2000).
50. Allers, T., Barak, S., Liddell, S., Wardell, K. & Mevarech, M. Improved Strains and Plasmid Vectors for Conditional Overexpression of His-Tagged Proteins in *Haloferax volcanii*. *Appl. Environ. Microbiol.* **76**, 1759–1769 (2010).

## Figure Legends

### Figure 1: Improved bacterial contact maps using an optimized Hi-C assay

(a) Overview of the Hi-C protocol, highlighting the steps that have been optimized. Normalized Hi-C contact maps of asynchronously growing populations of WT bacterial cells. The X and Y axes represent the coordinates of the chromosome and the colorscale reflects the frequency of contacts between two regions of the genome (arbitrary units), from white (rare contacts) to dark red (frequent contacts). Features of interest are indicated along the top axis.

(b) Hi-C contact matrix of WT *E. coli* binned at 5 kb. (c) Hi-C contact matrix of WT *V. cholera* binned at 5 kb. (d and e) Side-by-side comparison of the 800 kb region surrounding the terminus of replication *dif*, from both the new and originally published contact maps (Lioy *et al.*, Cell, 2018). Matrices are binned at either 5 kb or 0.5 kb. (f) Side-by-side comparison of a 50 kb region surrounding the *fliF-R* operon from both the new and originally published contact maps (Lioy *et al.* 2018), with 0.5 kb binning. (g) Side-by-side comparison of a 50 kb region surrounding the *flgB-J* operon from both the new and originally published contact maps (Lioy *et al.* 2018), with 0.5 kb binning. (h) Side-by-side comparison of 250 kb region surrounding the superintegron located on chromosome 2 of *V. cholera* from both the new and originally published contact maps (Val *et al.*, Science Advances, 2016). (i) Percentage of 3D intra- and 3D inter-chromosomal events obtained from *V. cholera* obtained through the new and published approaches. (j) Side-by-side comparison of inter-chromosomal contact maps (bin: 5 kb) centered on *dif* (top panels) or *ori* (bottom panels) of chromosomes 1 and 2 obtained from data obtained using the new and published approaches (Val *et al.*, Science Advances, 2016). The crtS site is indicated.

### Figure 2: Chromosomal organization of *H. volcanii*

Normalized Hi-C contact maps of asynchronously growing populations of WT *H. volcanii* cells. The X and Y axes represent the coordinates of the chromosome and the colorscale reflects the frequency of contacts between two regions of the genome (arbitrary units), from

white (rare contacts) to dark red (frequent contacts). Features of interest are indicated along the top axis. (a) Hi-C contact matrix of WT *H. volcanii* binned at 5 kb. (b) Directionality index analysis (300 kb scale), downstream (yellow), and upstream (grey) biases are indicated. (c) Directionality index analysis similar to that shown in (b) but at the scale of 100 kb. (d) RNA-seq depicting the transcription levels across the genome. (e) Directionality index analysis similar to that shown in (b) but at the scale of 100 kb. (f) Principle component analysis of the matrices in (b and c) based on the Pearson correlation matrices. The first principle component is shown as a compartment index. (g) GC content (50 kb sliding window) of the genome. (h-i) Magnifications of the 85 kb pHV1 chromosome, 437.9 kb pHV3 chromosome, and a 1 Mb region (coordinates: 65 kb – 1,065 kb) surrounding the pHV4 chromosome which has integrated into the main chromosome. For each magnification, the location of insertion (IS) elements, GC percentage (1 kb binning), and RNA-seq are shown. (j) Magnification of 1 Mb region of the main chromosome (coordinates 800kb – 1800 kb). The position of chromosomal loops and borders that have been detected with Chromosight are indicated in black and blue respectively. (k) Bar plot showing the average number of borders and loops detected in the entire genome of an exponentially growing culture of *H. volcanii* (detected using Chromosight, n=3). Associated pileup plots of windows centered on the detected loops in each condition are shown in the panels above.

### **Figure 3: Effect of transcription on the chromosome organization of *H. volcanii***

Normalized Hi-C contact maps of asynchronously growing populations of WT and mutant *H. volcanii* cells (5 kb bin). The X and Y axes represent the coordinates of the chromosome and the colorscale reflects the frequency of contacts between two regions of the genome (arbitrary units), from white (rare contacts) to dark red (frequent contacts). Features of interest are indicated along the top axis. (a and b) Hi-C contact matrix of WT *H. volcanii* growing in either exponential or stationary phase. For both conditions, the compartment index and RNA-seq are shown. (c) Differential contact map corresponding to the log<sub>2</sub> ratio of Hi-C interactions

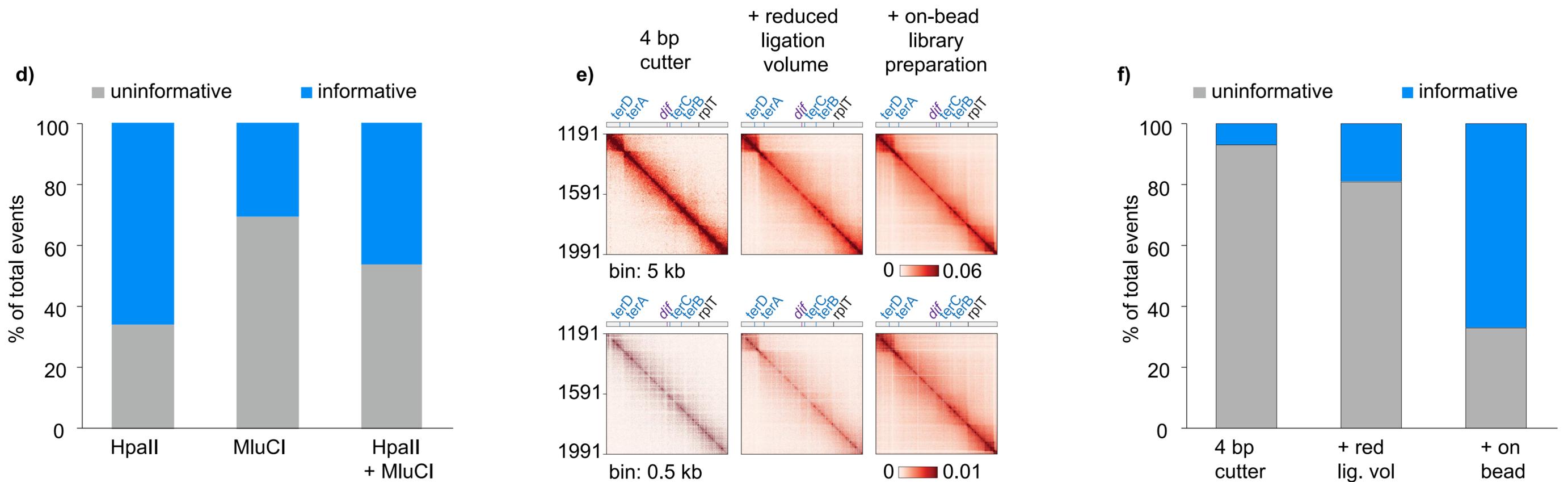
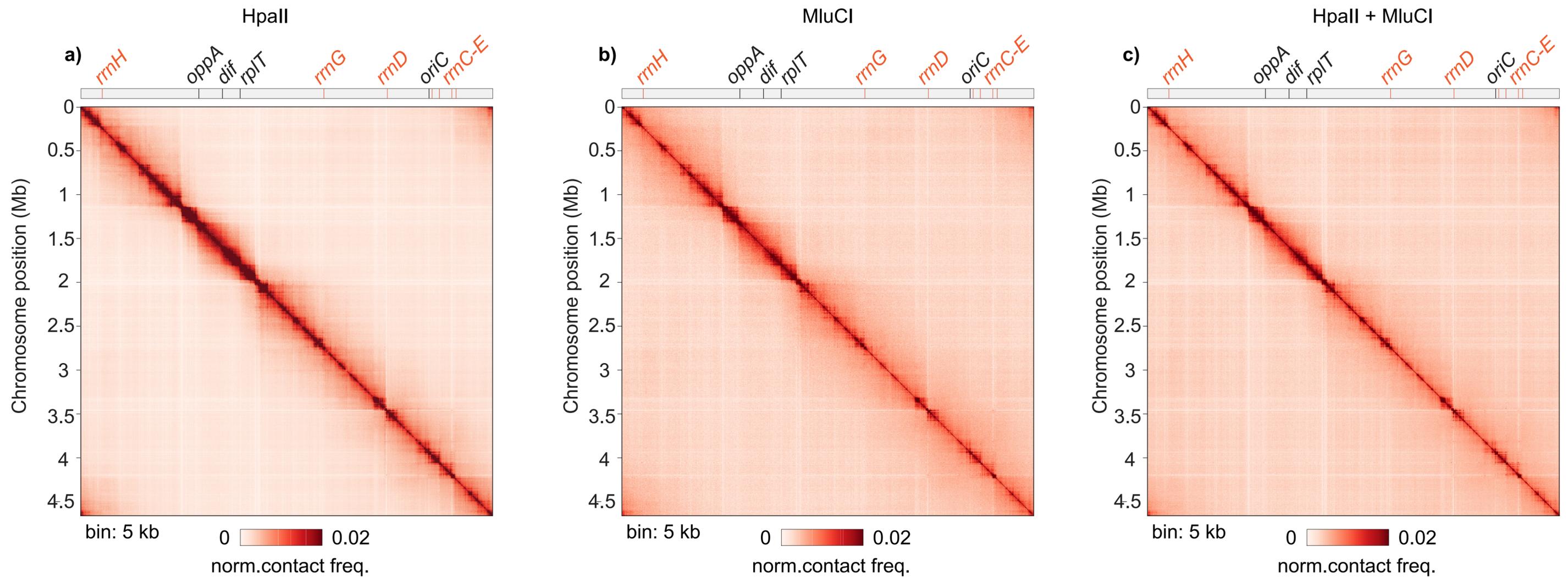
between exponential and stationary phase cultures. Blue to red colorscale reflects the enrichment in contacts in one population with respect to the other, white indicates no difference. (d) Magnifications of the 437.9 kb pHV3 chromosome, and a 1 Mb region of the main chromosome (coordinates: 800 kb – 1,800 kb) for exponential and stationary phase cultures. (e and f) Hi-C contact matrix of WT *H. volcanii* growing in the presence of DMSO (control) or actinomycin D. The compartment index is also shown for both conditions. (g) Differential contact map corresponding to the log<sub>2</sub> ratio of Hi-C interactions between actinomycin D treated cells and the DMSO control. (h and i) Magnifications of the 437.9 kb pHV3 chromosome, and a 1 Mb region of the main chromosome (coordinates: 800 kb – 1,800 kb) for actinomycin D treated cells and the DMSO control. (j) Comparison of loop score distributions obtained using Chromosight for the Hi-C matrices shown in panels (a), (b), (e), and (f). The number of loops detected and the mean of the Pearson coefficient for each condition is shown above (n=2).

**Figure 4: Role of SMCs in *H. volcanii* chromosome organization.**

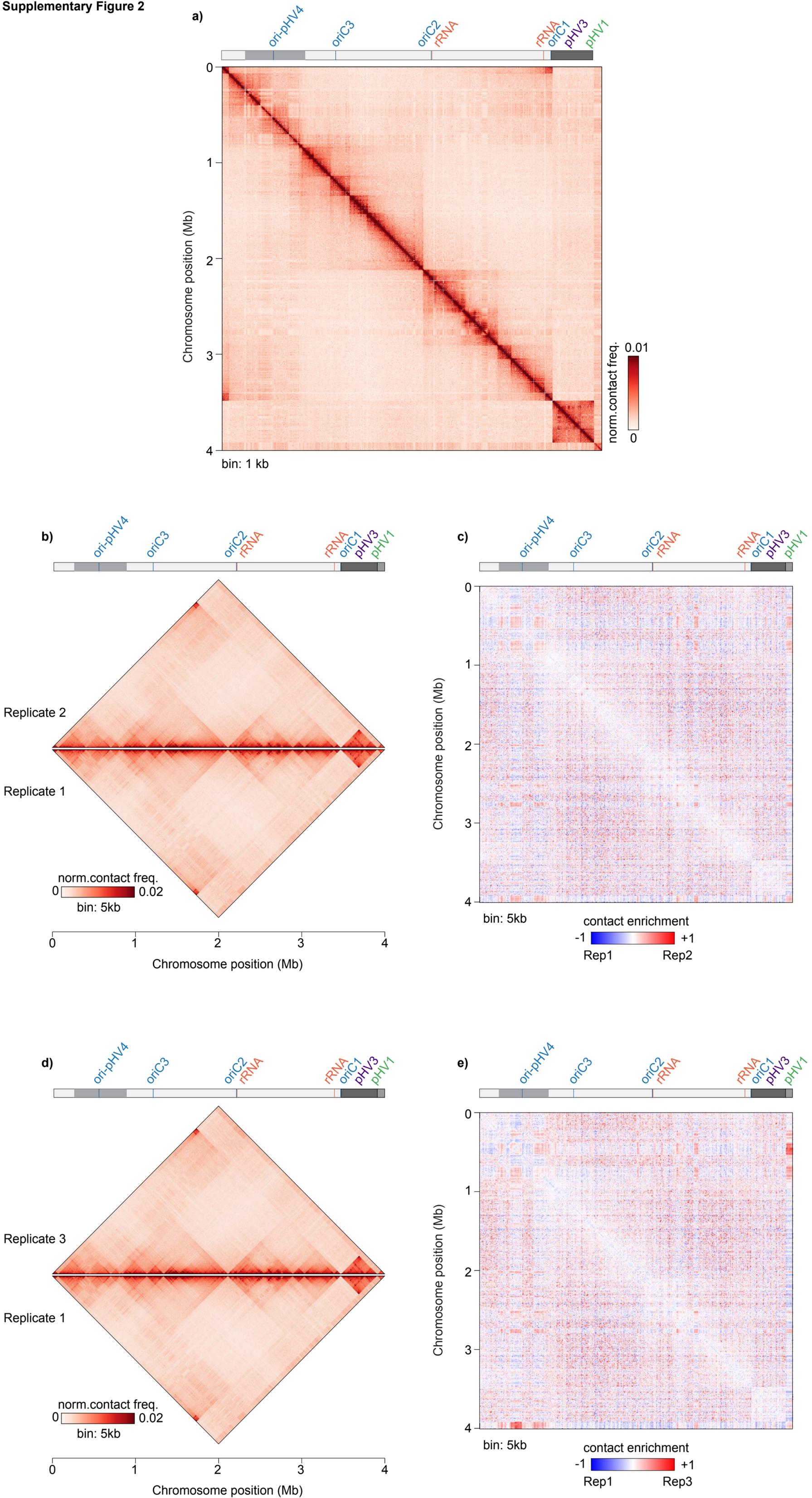
(a) Phylogenetic tree showing the conservation of canonical SMCs, Rad50, and the recently identified Coalescin (ClSN) proteins between different archaeal lineages. Adapted from <sup>13,48</sup> and explained in the Materials and Methods. (b) Generation times of wild-type (WT),  $\Delta smc$ , and  $\Delta mre11 rad50$ . Error bars represent the standard deviation of the mean where  $n > 3$ . p values are based on t-test comparisons between each mutant and the WT strain. (c) Images of DIC, Hoechst signal and the merge image (DIC signal in grey and the DNA signal in cyan) of WT,  $\Delta smc$ , and  $\Delta mre11 rad50$  cells. Scale bar = 5  $\mu$ m. (d) Normalized Hi-C contact maps of asynchronously growing populations of WT and  $\Delta smc$  mutant cells (5 kb bin). The X and Y axes represent the coordinates of the chromosome and the colorscale reflects the frequency of contacts between two regions of the genome (arbitrary units), from white (rare contacts) to dark red (frequent contacts). Features of interest are indicated along the top axis. (e) Differential contact map corresponding to the log<sub>2</sub> ratio of Hi-C interactions between WT and

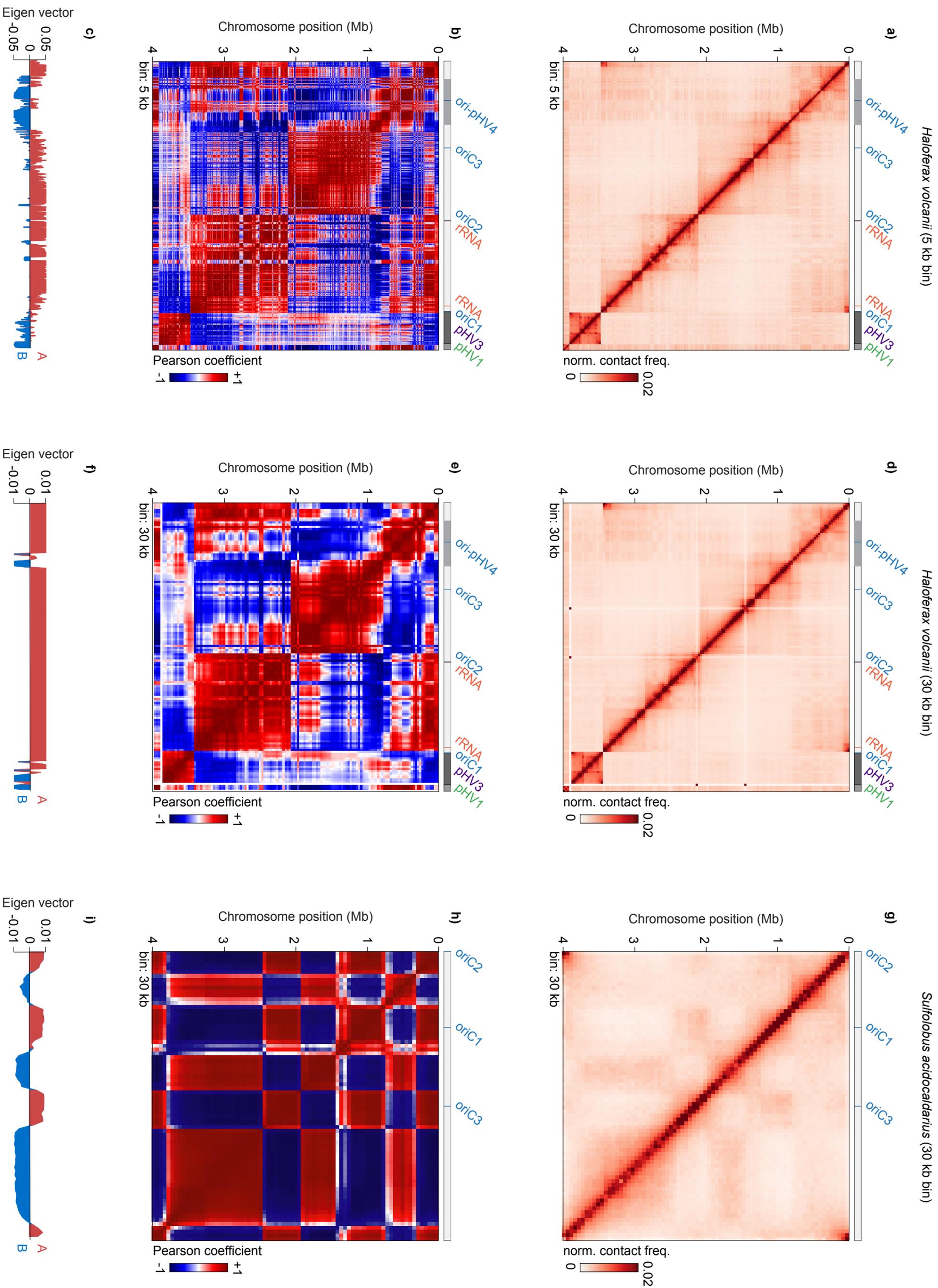
$\Delta smc$  mutant cells. Blue to red colorscale reflects the enrichment in contacts in one population with respect to the other, white indicates no difference. (f) Normalized Hi-C contact maps of asynchronously growing populations of WT and  $\Delta mre11 rad50$  mutant cells (5 kb bin). (g) Differential contact map corresponding to the log<sub>2</sub> ratio of Hi-C interactions between WT and  $\Delta mre11 rad50$  mutant cells. (h) Magnification of 1 Mb region of the main chromosome (coordinates 1620 kb – 2620 kb) for WT,  $\Delta smc$ , and  $\Delta mre11 rad50$  mutant cells. The position of chromosomal loops and borders that have been detected with Chromosight<sup>28</sup> are indicated in black and blue respectively. (i) Comparison of loop score distributions obtained using Chromosight for the Hi-C matrices shown in panels (d) and (f). The number of loops detected and the mean of the Pearson coefficient for each condition is shown above (n=2).

Supplementary Figure 1

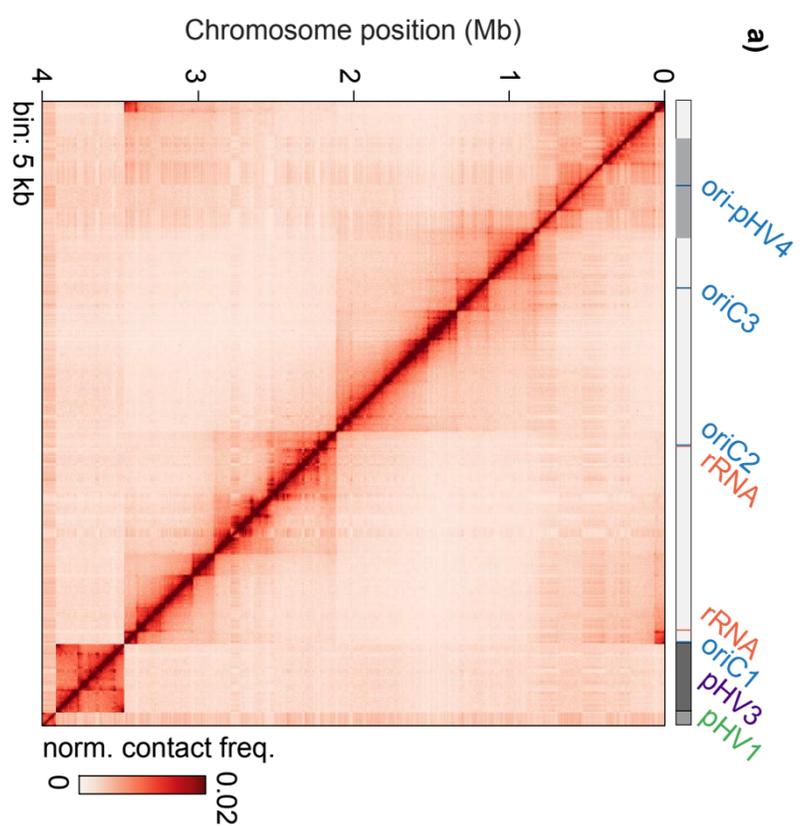


Supplementary Figure 2

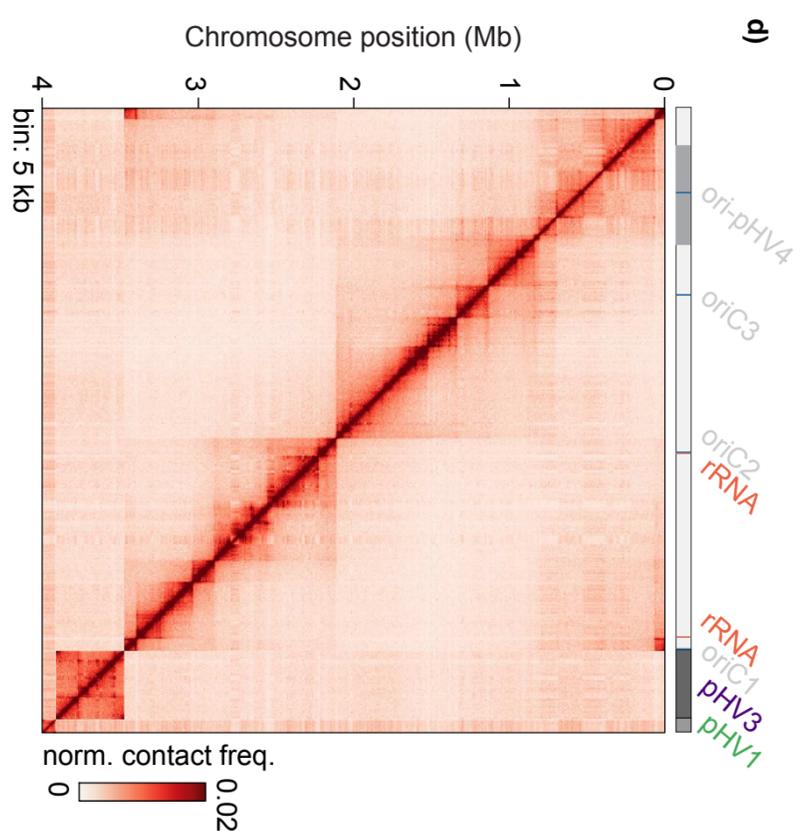




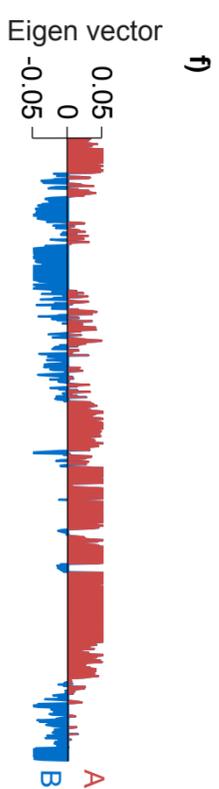
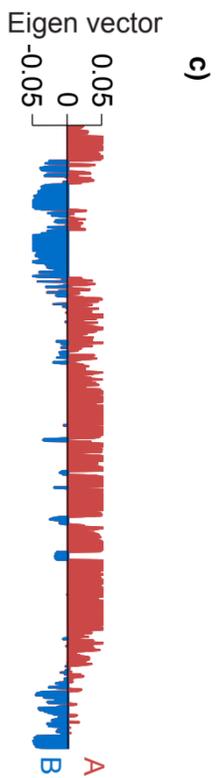
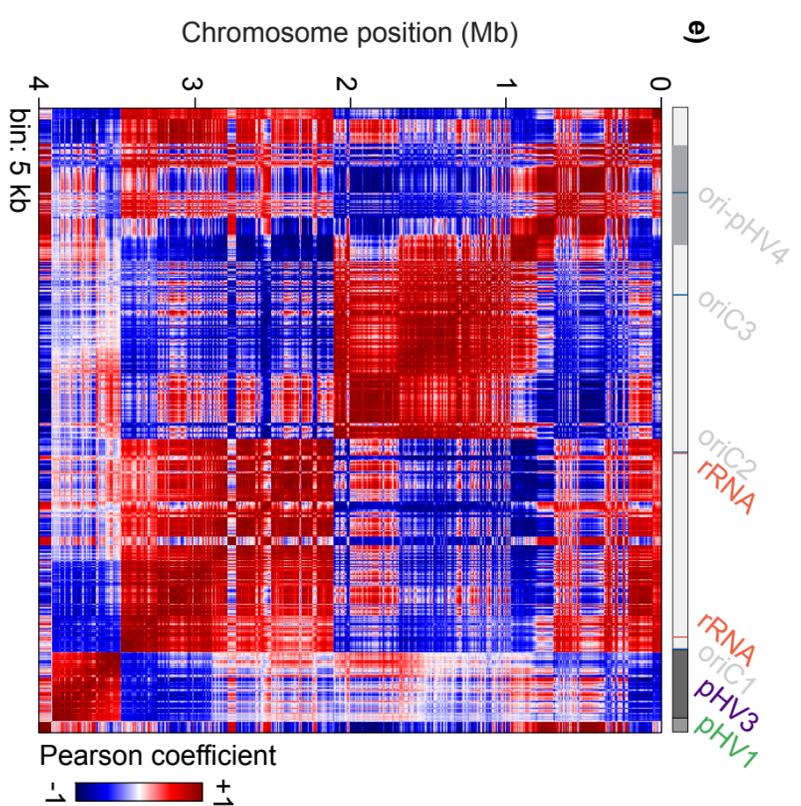
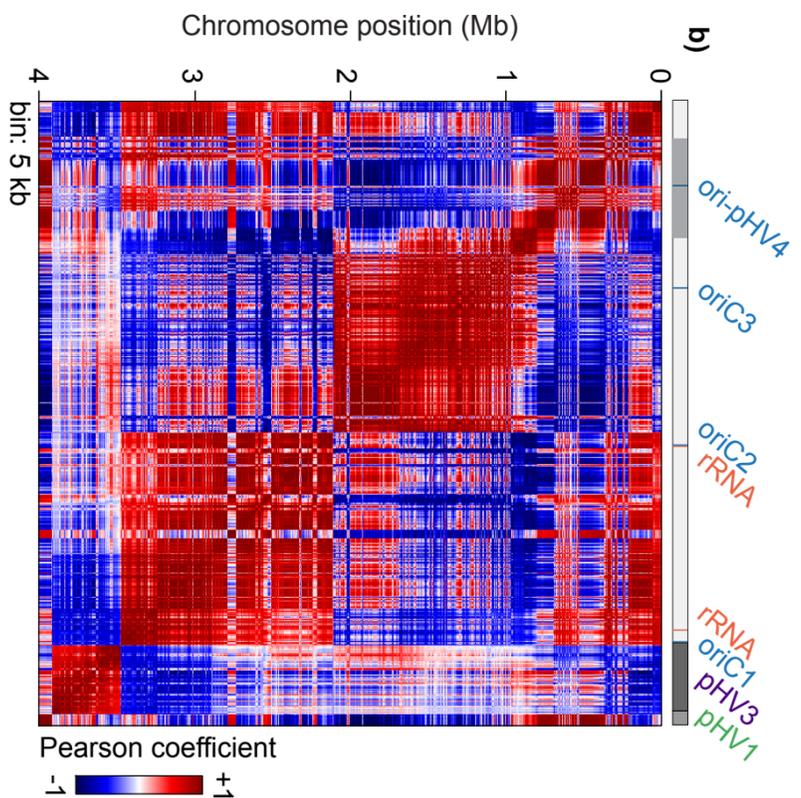
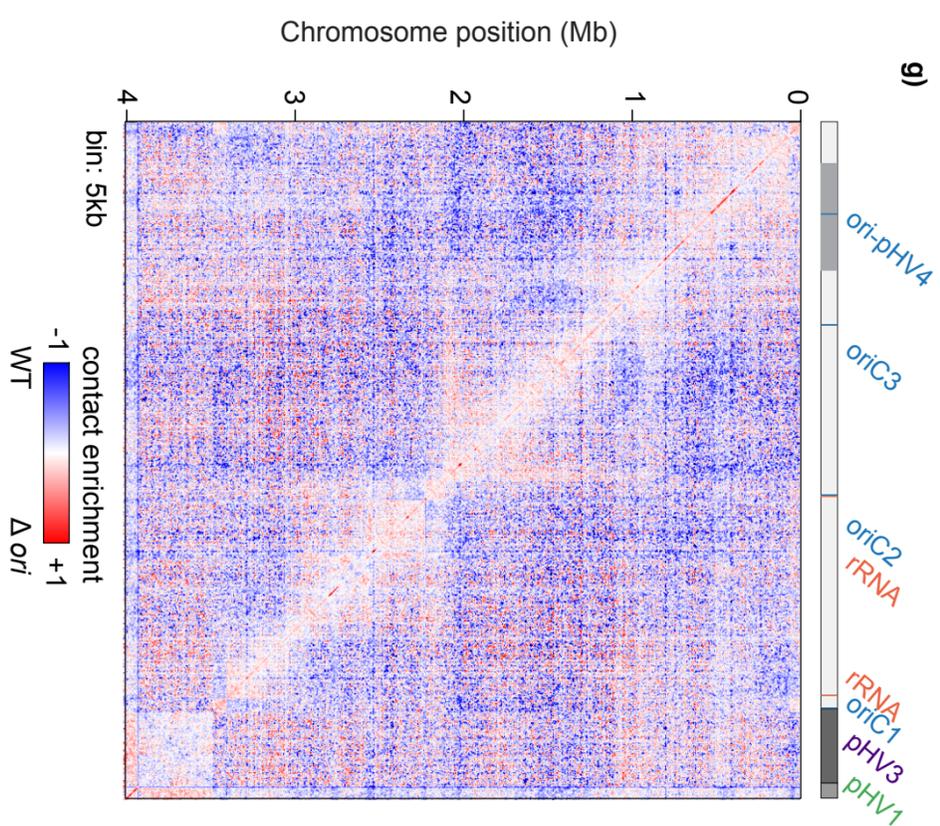
WT



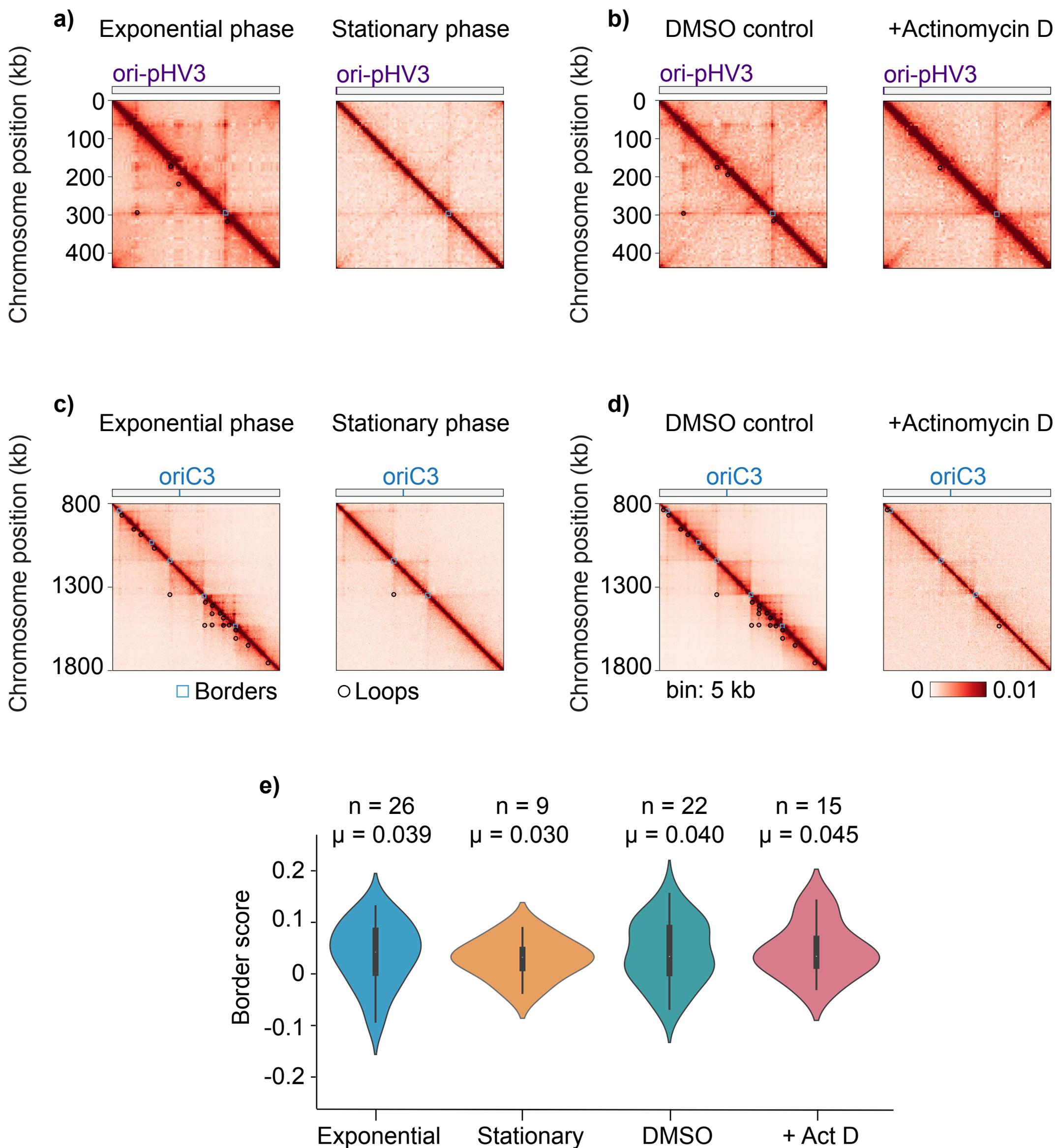
$\Delta ori$



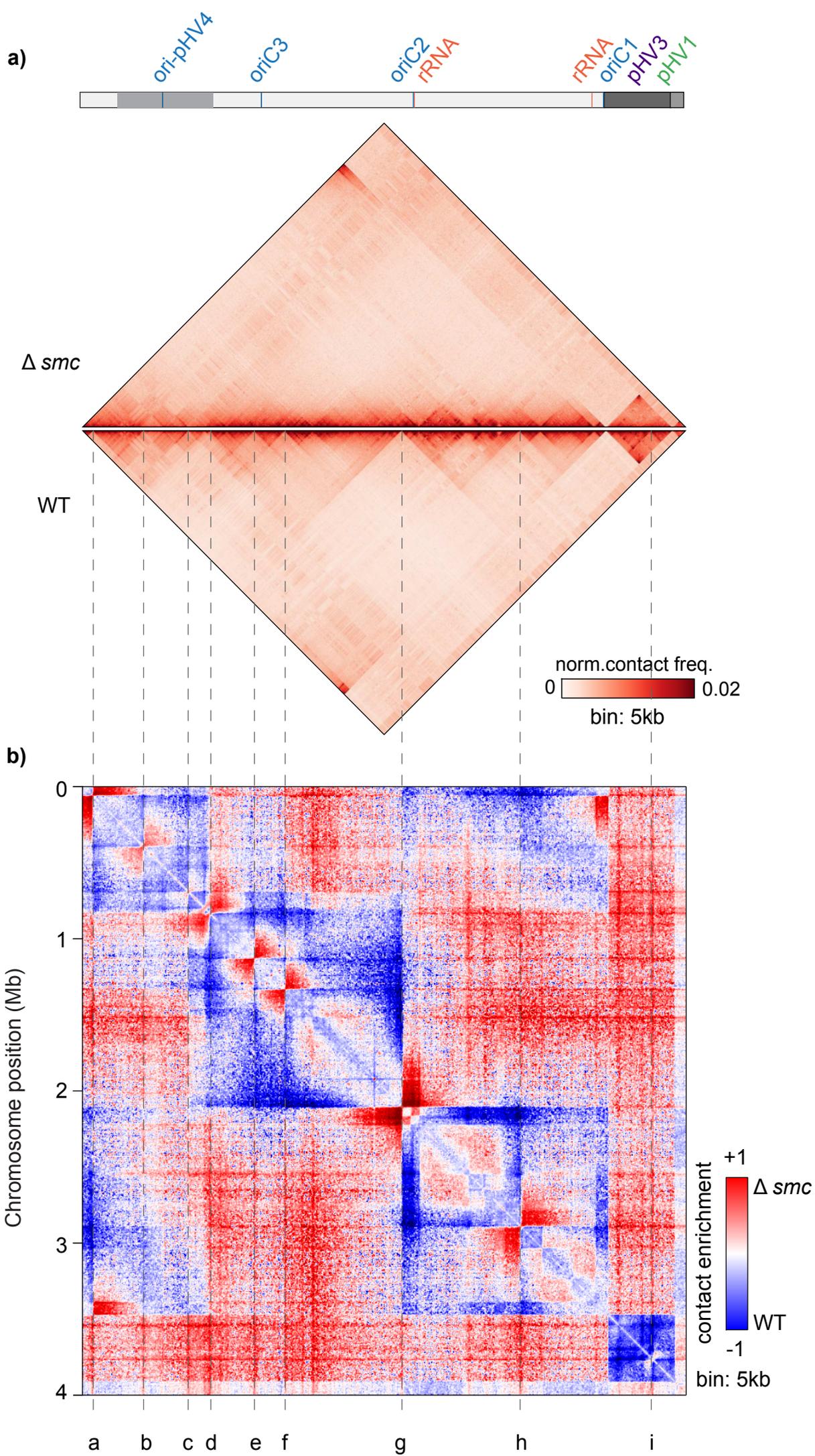
$\Delta ori$  / WT



# Supplementary Figure 5



Supplementary Figure 6



c)

Border	Chromosome	Approx position (bp)	Closest gene	Proposed function
a	Main	71,500	HVO_0069	arylsulphatase
b	Main - pHV4	405,500	HVO_A0132	ISH16 transposase
c	Main - pHV4	701,000	HVO_A0463	rnh RNaseH
d	Main - pHV4	846,000	HVO_A0311	HalC8 putative halocin C8
e	Main	1,141,000	HVO_0568	Transcription regulator
f	Main	1,345,500	HVO_0791	ogt DNA methyltransferase
g	Main	2,120,000	HVO_1615	RNase H-like domain-containing protein
h	Main	2,904,500	HVO_2401	Glycine cleavage system P-protein
i	pHV3	295,500	HVO_B0248	SDR oxidoreductase

### **Supplementary Figure 1: Comparison of different 4bp cutting restriction enzymes on Hi-C contact maps**

Normalized Hi-C matrices of asynchronously growing populations of WT *E. coli*. The X and Y axes represent the coordinates of the chromosome and the colorscale reflects the frequency of contacts between two regions of the genome (arbitrary units), from white (rare contacts) to dark red (frequent contacts). Features of interest are indicated along the top axis. (a-c) Whole-genome plot of *E. coli* Hi-C experiments using either HpaII, MluCI, or a combination of the two restriction enzymes (5 kb bin). (d) Percentage of uninformative (grey), and 3D intra-chromosomal (blue) events obtained using each of the enzymes described in panels a-c. (e) Side-by-side comparison of the 800 kb region surrounding the terminus of replication *dif*, demonstrating the contribution of each optimization step (shown in Figure 1a) to the resulting Hi-C contact maps and comparing them to previously published results (Lioy *et al.*, Cell, 2018). Matrices are binned at either 5 kb or 0.5 kb. (f) Bar chart showing the combined contribution of each improvement (shown in e) to the percentage of informative reads obtained from a given Hi-C experiment in *E. coli*.

### **Supplementary Figure 2: Reproducibility of *H. volcanii* Hi-C contact maps**

(a) Normalized Hi-C contact map of asynchronously growing populations of WT *H. volcanii* cells (1 kb bin). The X and Y axes represent the coordinates of the chromosome and the colorscale reflects the frequency of contacts between two regions of the genome (arbitrary units), from white (rare contacts) to dark red (frequent contacts). Features of interest are indicated along the top axis. (b) Normalized Hi-C contact maps of asynchronously growing populations *H. volcanii* wild-type replicates 1 and 2 (5 kb bin). (c) Differential contact map corresponding to the log<sub>2</sub> ratio of Hi-C interactions between replicate 1 and replicate 2 WT cultures. Blue to red colorscale reflects the enrichment in contacts in one population with respect to the other, white indicates no difference. (d) Normalized Hi-C contact maps of asynchronously growing populations *H. volcanii* wild-type replicates 1 and 3 (5 kb bin). (e)

Differential contact map corresponding to the log<sub>2</sub> ratio of Hi-C interactions between replicate 1 and replicate 3 WT cultures.

**Supplementary Figure 3: Comparison of compartment-like structures in *H. volcanii* and *S. acidocaldarius***

(a) Normalized Hi-C contact map of an asynchronously growing population of WT *H. volcanii* cells. The X and Y axes represent the coordinates of the chromosome and the colorscale reflects the frequency of contacts between two regions of the genome (arbitrary units), from white (rare contacts) to dark red (frequent contacts). Features of interest are indicated along the top axis (5 kb bin). (b) The matrix in (a) converted into a Pearson correlation matrix. (c) Principal component analysis was performed on the matrix in (b) and the first principle component is shown as a compartment index. (d-f) Hi-C contact matrix, Pearson correlation matrix, and principal component analysis using the same data as (a-c) but re-binned (30 kb bin). (g-i) Hi-C contact matrix, Pearson correlation matrix, and principal component analysis of *S. acidocaldarius*<sup>13</sup> using the same workflow as used for (a-f).

**Supplementary Figure 4: The genome organization of *H. volcanii* is unaffected by replication inactivation**

(a) Normalized Hi-C contact map of an asynchronously growing population of WT *H. volcanii* cells. The X and Y axes represent the coordinates of the chromosome and the colorscale reflects the frequency of contacts between two regions of the genome (arbitrary units), from white (rare contacts) to dark red (frequent contacts). Features of interest are indicated along the top axis (5 kb bin). (b) The matrix in (a) converted into a Pearson correlation matrix. (c) Principal component analysis was performed on the matrix in (b) and the first principle component is shown as a compartment index. (d-f) Hi-C contact matrix, Pearson correlation matrix, and principal component analysis of a  $\Delta ori$  mutant (*oriCl-3* and *ori-pHV4* have been deleted). (g) Differential contact map corresponding to the log<sub>2</sub> ratio of Hi-C interactions between WT and  $\Delta ori$  mutant cells. Blue to red colorscale reflects the enrichment in contacts in one population with respect to the other, white indicates no difference.

### **Supplementary Figure 5: Measuring the effect of transcription on the *H. volcanii* chromosomal borders**

Normalized Hi-C contact maps of asynchronously growing populations of WT and mutant *H. volcanii* cells (5 kb bin). The X and Y axes represent the coordinates of the chromosome and the colorscale reflects the frequency of contacts between two regions of the genome (arbitrary units), from white (rare contacts) to dark red (frequent contacts). Features of interest are indicated along the top axis. (a) Magnification of the 437.9 kb pHV3 chromosome from exponentially growing and stationary phase cultures. The position of chromosomal loops and borders that have been detected with Chromosight<sup>28</sup> are indicated in black and blue respectively. (b) Magnification of the 437.9 kb pHV3 chromosome from cells grown in the presence/absence of actinomycin D. (c) Magnification of a 1 Mb region of the main chromosome (coordinates: 800 kb – 1,800 kb) from exponentially growing and stationary phase cultures. (d) Magnification of a 1 Mb region of the main chromosome (coordinates: 800 kb – 1,800 kb) from cells grown in the presence/absence of actinomycin D. (e) Comparison of border score distributions obtained using Chromosight for the Hi-C matrices shown in panels (d) and (f). The number of borders detected and the mean of the Pearson coefficient for each condition is shown above (n=2).

### **Supplementary Figure 6: Mapping chromosomal borders in $\Delta smc$ mutant cells**

(a) Normalized Hi-C contact maps of asynchronously growing populations of WT and  $\Delta smc$  mutant cells (5 kb bin). The X and Y axes represent the coordinates of the chromosome and the colorscale reflects the frequency of contacts between two regions of the genome (arbitrary units), from white (rare contacts) to dark red (frequent contacts). Features of interest are indicated along the top axis. (b) Differential contact map corresponding to the log<sub>2</sub> ratio of Hi-C interactions between WT and  $\Delta smc$  mutant cells. Blue to red colorscale reflects the enrichment in contacts in one population with respect to the other, white indicates no difference. Chromosomal borders that have been lost in  $\Delta smc$  mutant are highlighted by dashed lines. (c) Table showing the location and approximate genomic coordinates of the borders lost in  $\Delta smc$  mutant cells. The closest gene to each border and its proposed function is also noted.

Supplementary Table 1: Statistics of Hi-C experiments performed in this study compared to published data

Organism	ID	Total no. of cells in experiment	Enzyme	Total no. of fragments	Median fragment size (bp)	Max. resolution of contact map	+/- biotin	Total no. reads	Percentage of events			Percentage of reads in final matrix/total no. reads	Reference
									uncut	3D intra	3D inter		
<i>E. coli</i>	CC328	1 x 10 <sup>8</sup>	HpaII	24312	122	0.5 kb	+	136318986	35	65	0	44%	This study
	CC334						+	65791890	33	67	0	42%	This study
	CC335		MluCI	19660	138		+	47649120	70	30	0	26%	This study
	CC336		HpaII + MluCI	43971	72		+	53225732	49	51	0	41%	This study
	SRR6354565	1-2 x 10 <sup>9</sup>	HpaII	24312	122	5 kb	-	88615212	93	7	0	6%	<sup>8</sup>
<i>V. cholera</i>	CC340	1 x 10 <sup>8</sup>	HpaII	7713	318	1 kb	+	229263798	73	22	5	20%	This study
	SRR3180994	1 x 10 <sup>9</sup>				5 kb	-	170426044	97	3	0	1%	<sup>16</sup>
<i>C. crescentus</i>	SRR824843	2.5 x 10 <sup>8</sup>	BglII	701	3711	10 kb	+	34822386	30	70	0	37%	<sup>7</sup>
	SRR824846		NcoI	2026	1299		+	47673122	32	68	0	48%	<sup>7</sup>
<i>H. volcanii</i>	CC363	1 x 10 <sup>8</sup>	HpaII	31557	72	1 kb	+	296443022	46	43	11	26%	This study
	CC364						+	54669004	78	18	4	9%	This study
	CC375						+	140191394	58.22	35.78	6	20%	This study
	CC378						+	259909492	65.4	29.46	5.15	19%	This study
	CC397						+	137148304	64.6	29.08	6.32	21%	This study
	CC398						+	131549348	51.59	39.83	8.58	31%	This study
	CC399						+	132842414	68.03	25.74	6.23	17%	This study
	CC400						+	157435870	73.35	22.69	3.96	14%	This study
	CC401						+	130977656	76.93	19.77	3.31	11%	This study
	CC402						+	135893168	60.43	32.57	7	25%	This study
	CC403						+	92187088	85.46	11.3	3.06	8%	This study
	CC404						+	110172672	61.58	30.96	7.46	21%	This study
	<i>S. acidocaldarius</i>						SRR8699877	4 x 10 <sup>8</sup>	HindIII	808	1838	30 kb	+
SRR8699887		1179	1477	+	139639604	23	77			0	42%		<sup>13</sup>

**Supplementary Table 2: List of strains, plasmids, and primers used in this study**

**Strains**

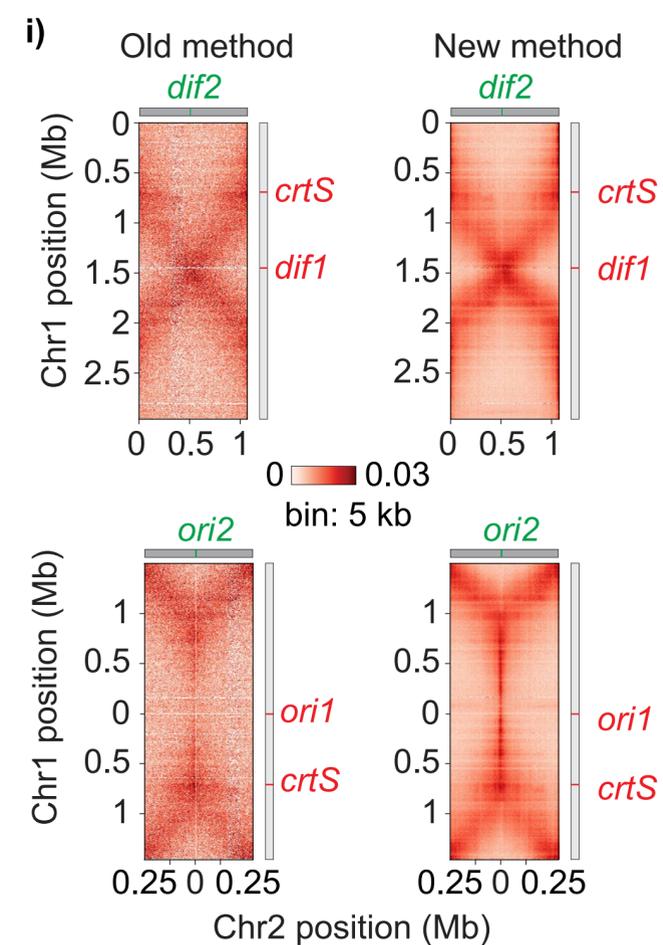
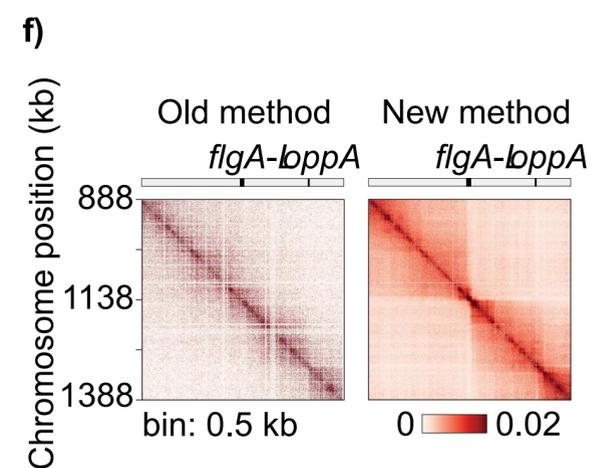
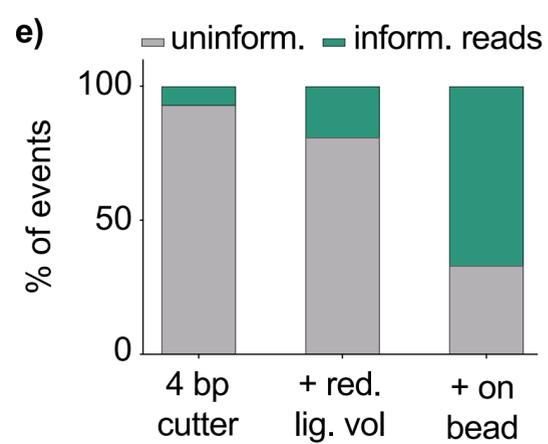
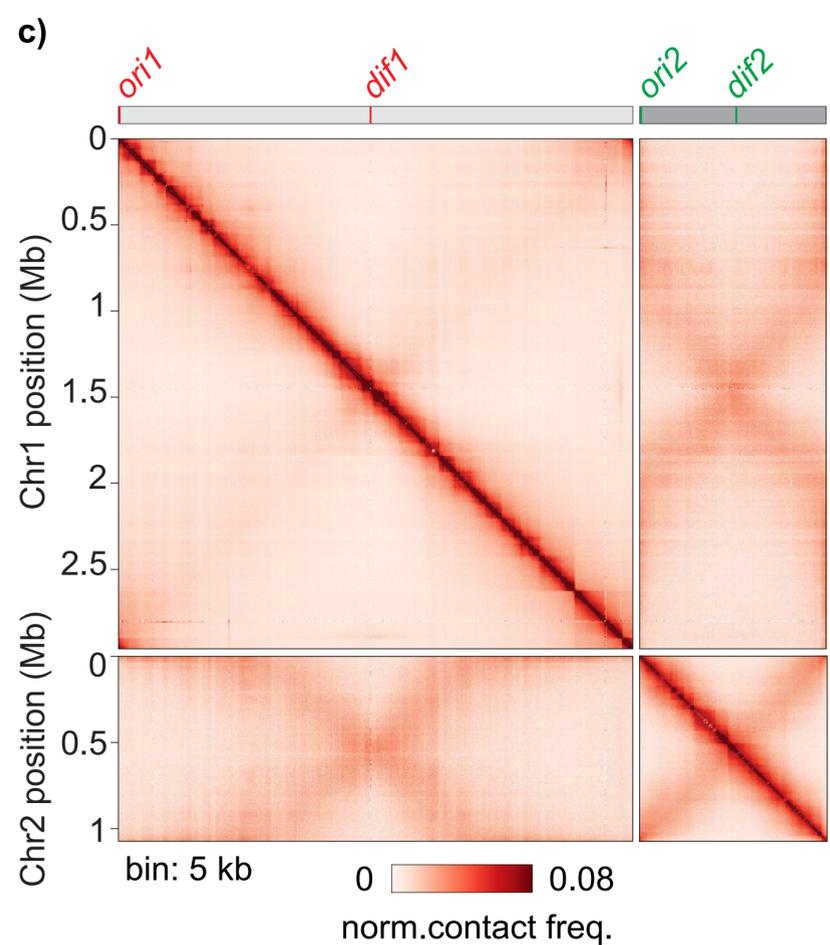
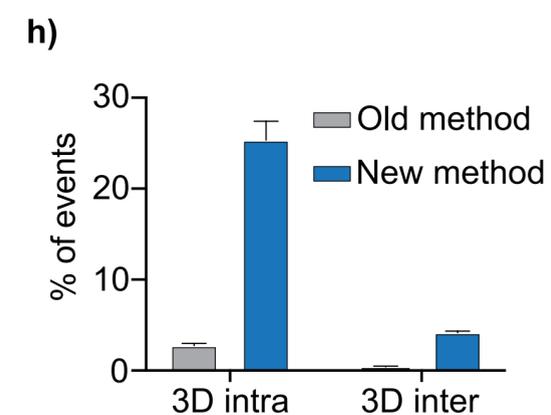
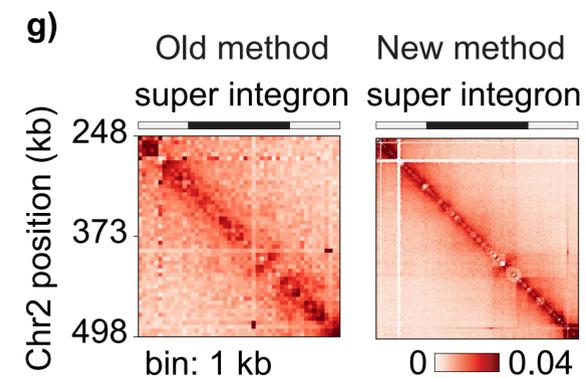
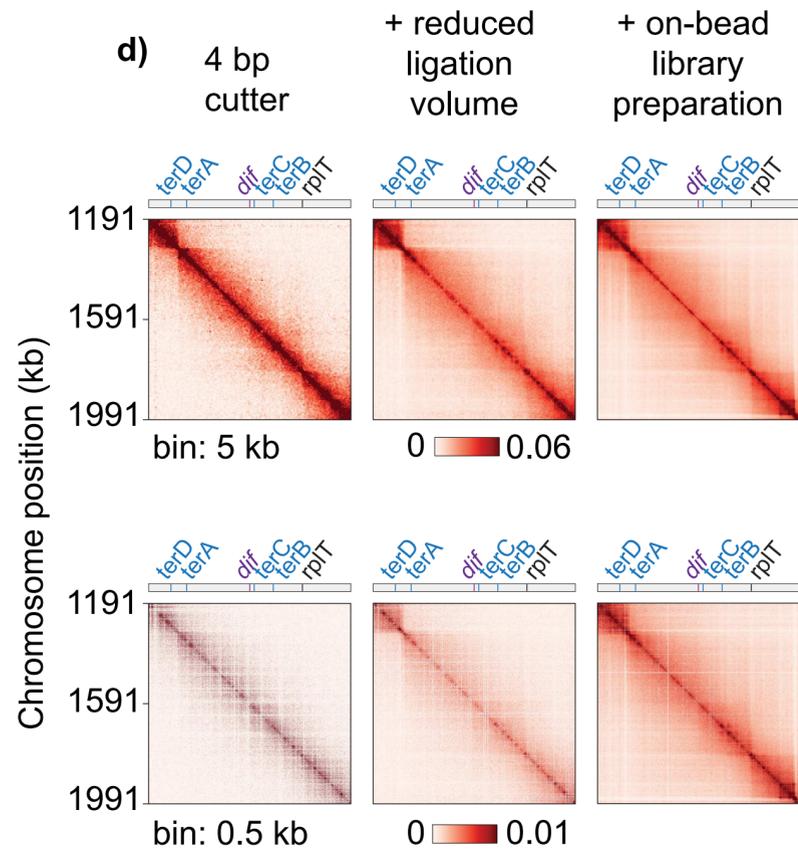
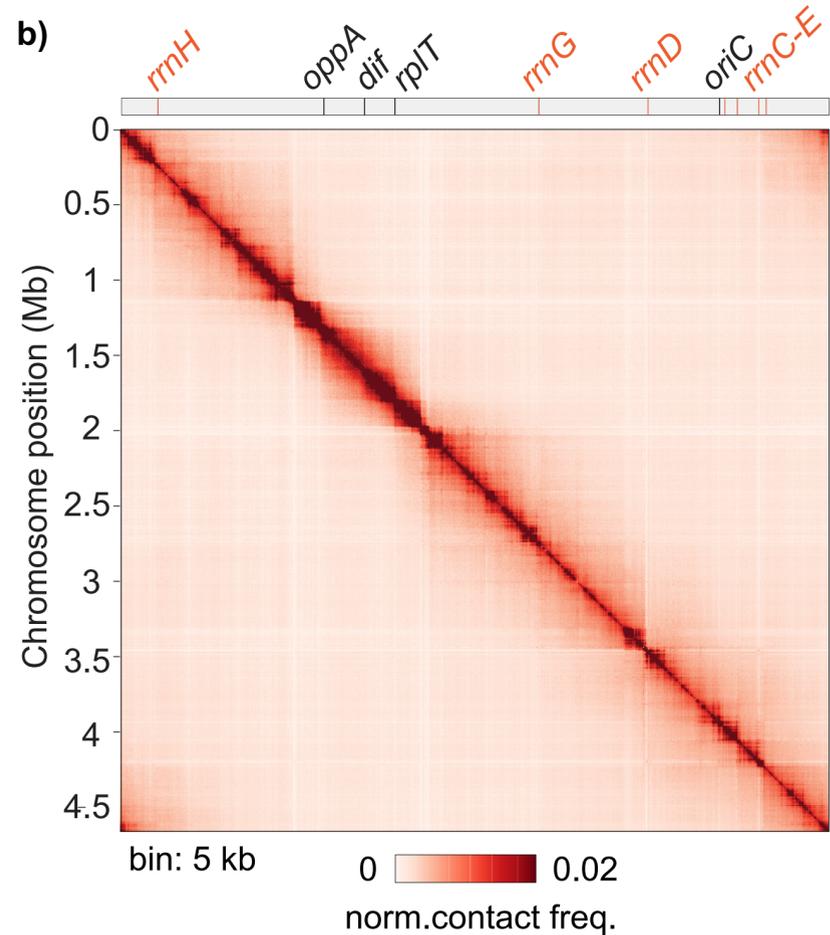
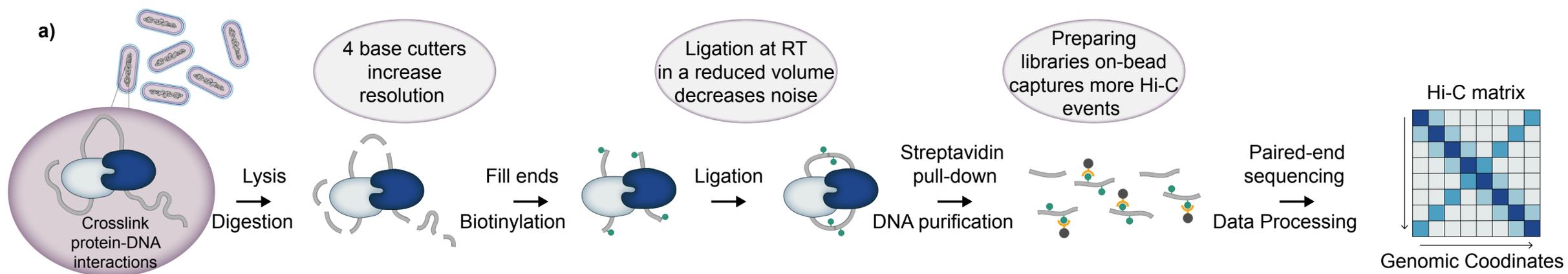
<b>Name</b>	<b>Genotype</b>	<b>Source</b>
<i>E. coli</i> K12 MG1655	<i>F-lambda- ilvG- rfb-50 rph-1</i>	37
<i>V. cholera</i> N16961	Vibrio cholera O1 El Tor N16961	38
<i>H. volcanii</i> (wild-type)	<i>H. volcanii</i> H26 $\Delta$ <i>pyrE2</i>	39
<i>H. volcanii</i> $\Delta$ <i>mre11 rad50</i> (H161)	H26 $\Delta$ <i>pyrE2</i> $\Delta$ <i>mre11</i> $\Delta$ <i>rad50</i>	29
<i>H. volcanii</i> $\Delta$ <i>ori</i> (H1546)	H26 $\Delta$ <i>pyrE2</i> $\Delta$ <i>trpA</i> $\Delta$ <i>oriC1</i> $\Delta$ <i>oriC2</i> $\Delta$ <i>oriC3</i> $\Delta$ <i>ori-pHV4::trpA</i>	22
H26 $\Delta$ <i>smc</i> (HvRL138)	H26 $\Delta$ <i>pyrE2</i> $\Delta$ <i>smc</i>	This study
XL1-Blue MRF'	$\Delta$ <i>mcrA183</i> $\Delta$ <i>mcrCB-hsdSMR-mrr173</i> <i>endA1 supE44 thi-1 recA1 gyrA96 relA1 lac</i> [ <i>F'</i> <i>proAB lacIqZ</i> $\Delta$ <i>M15 Tn10</i> ]	Agilent technologies
GM121	<i>F-dam</i> 3 <i>dcm</i> 6 <i>ara</i> 14 <i>fhuA31 galK2 galT22 hdsR3 lacY1 leu-6 thi-1 thr-1 tsx-78</i>	50

**Plasmids**

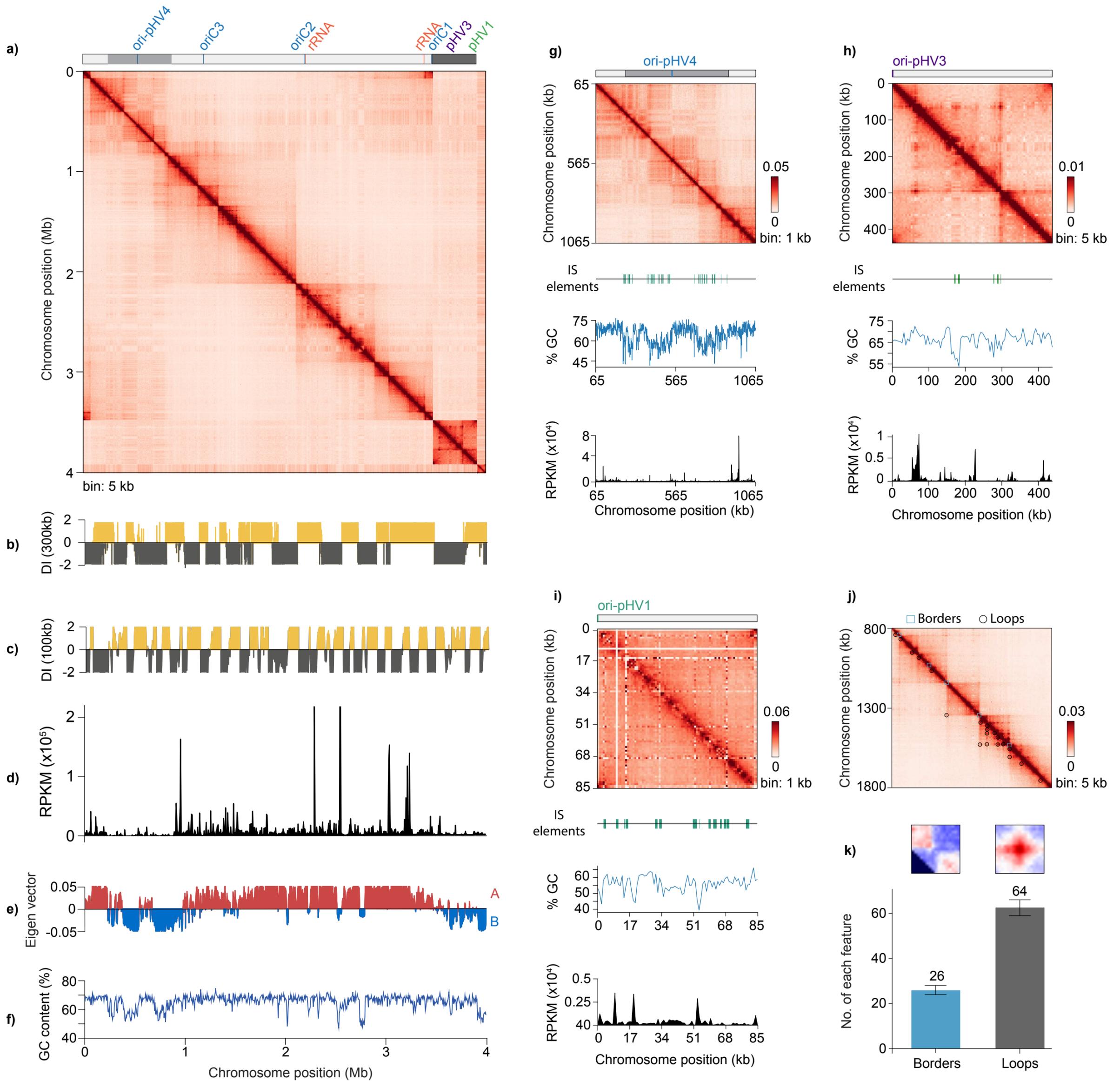
<b>Name</b>	<b>Genotype</b>	<b>Source</b>
pTA131	pBluescript II containing <i>pyrE2</i> under ferredoxin promoter	39
pRL93	PTA131 $\Delta$ <i>smc</i>	This study

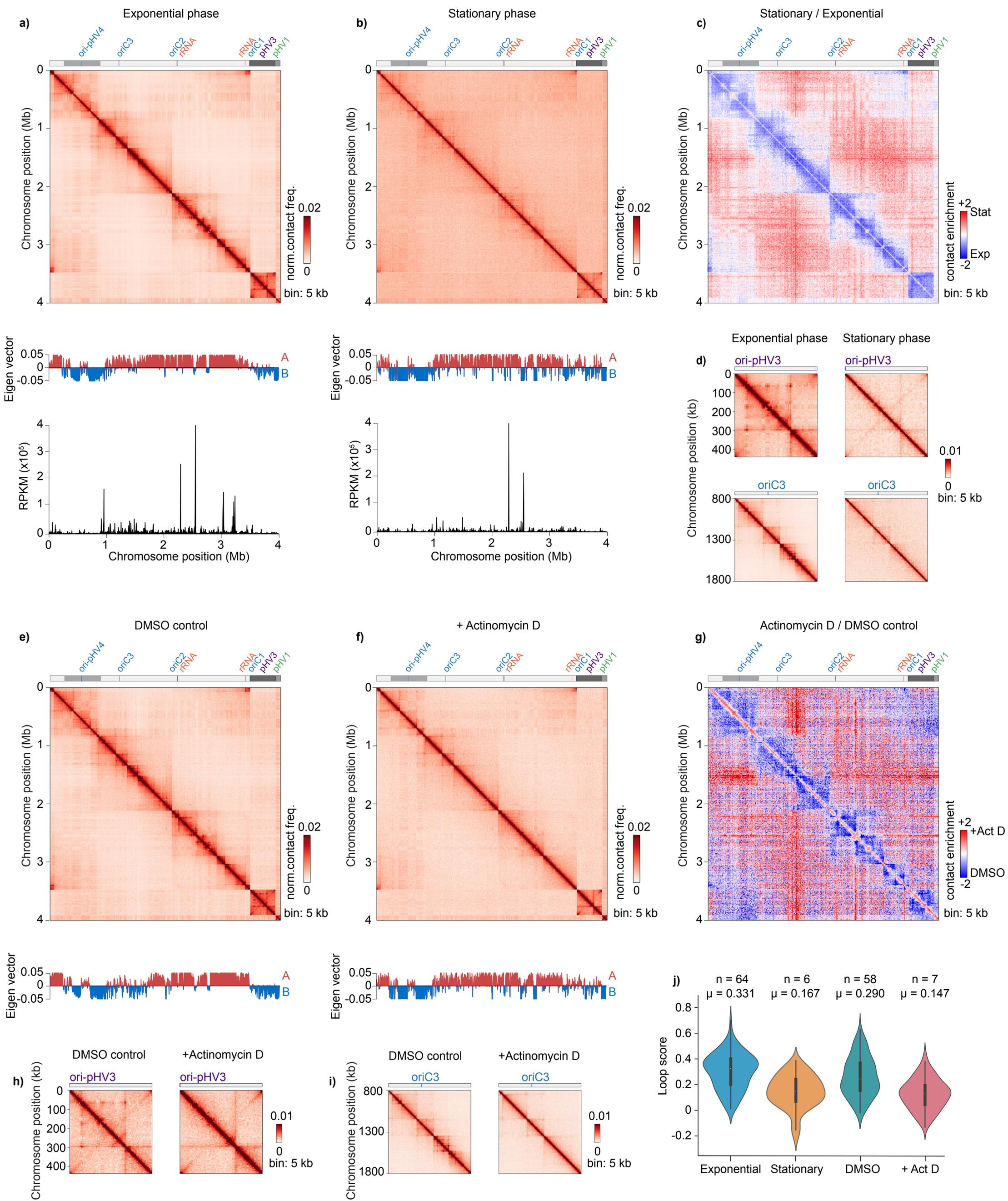
**Primers**

<b>Name</b>	<b>Sequence (5' - 3')</b>	<b>Source</b>
RL292	AGGTTCGACGGTATCGATAAGCTTGATATCGGACGCACATTTATGTCTCACG	This study
RL294	TGCGGGCGTCGCCCGCGTGAACACCAACCCATGACTGACGACATCCCCGAAC	This study
RL295	AAAAGCTGGAGCTCCACCGCGGTGGCGGCCTCCAGCGCGTTCGATTGGGTTCG	This study
RL296	GGGTTGGTGTTCACGCGGGCG	This study
pBSF2	TTAAGTTGGGTAACGCCAGGG	This study
pBSF3	ACCCAGGCTTTACACTTTATGC	This study
RL305	GCCGTCACGACCATCCTCAAC	This study
RL310	GTTGTCGGGCTCGCGGGCAG	This study

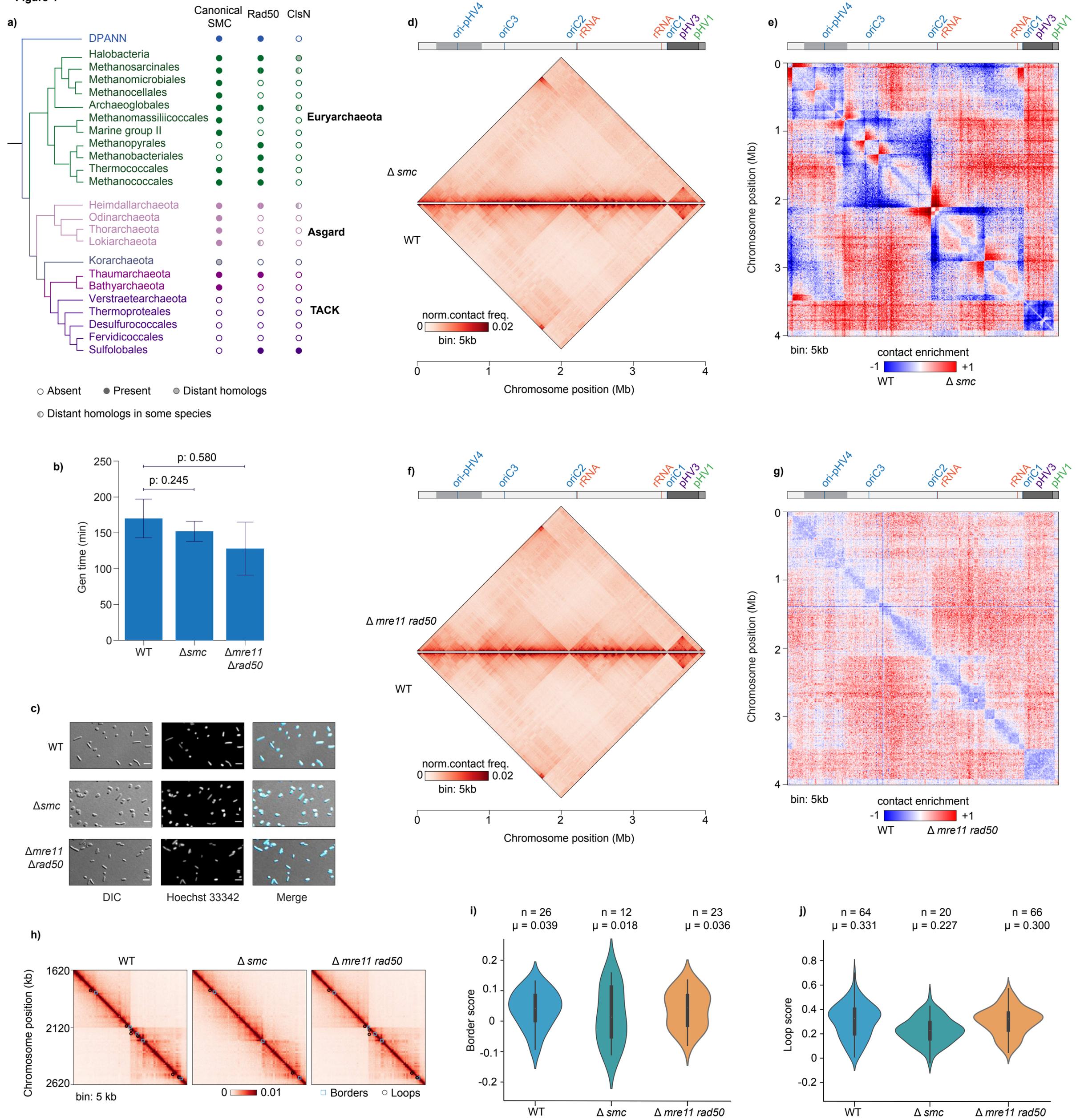
**Figure 1**

**Figure 2**

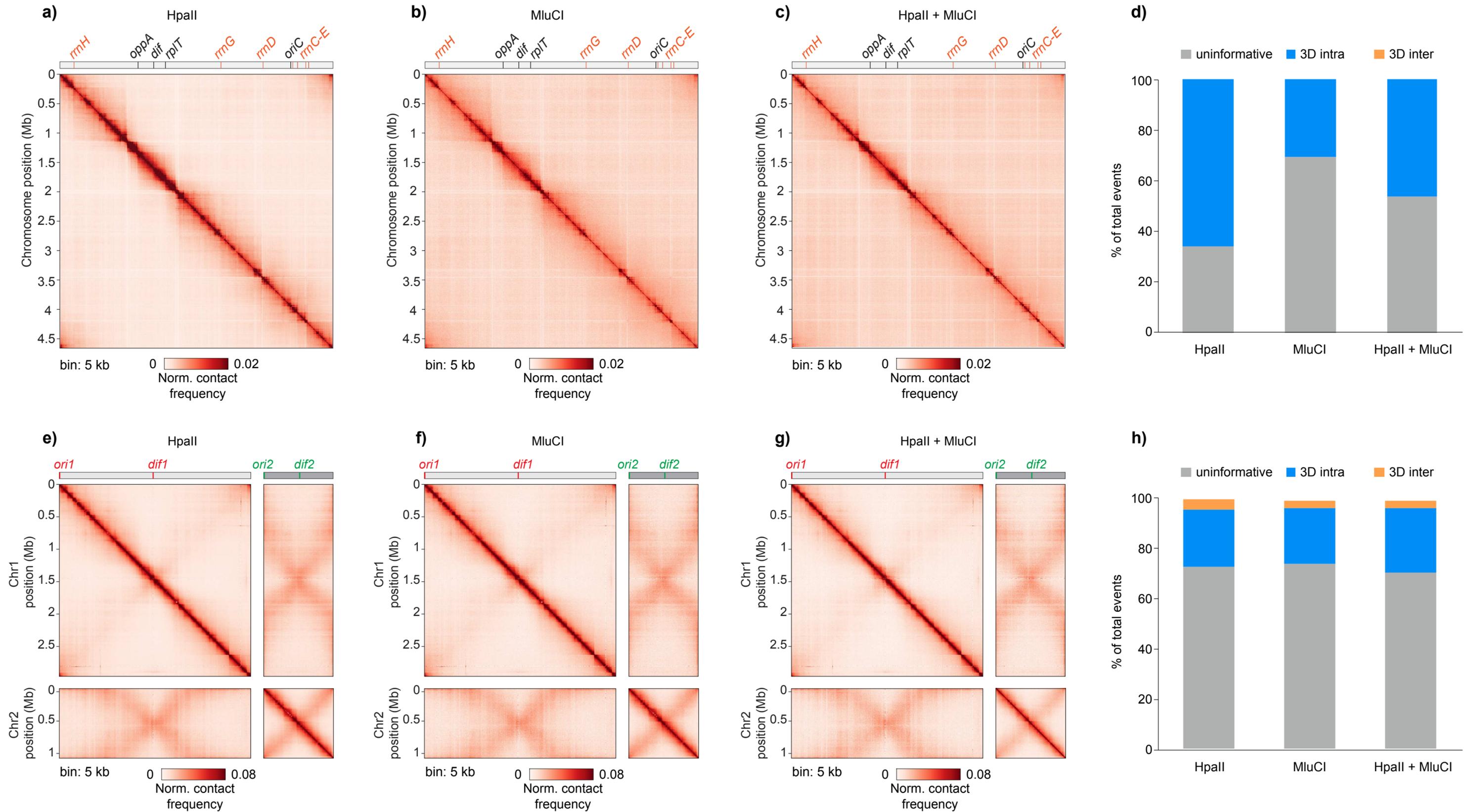


**Figure 3**

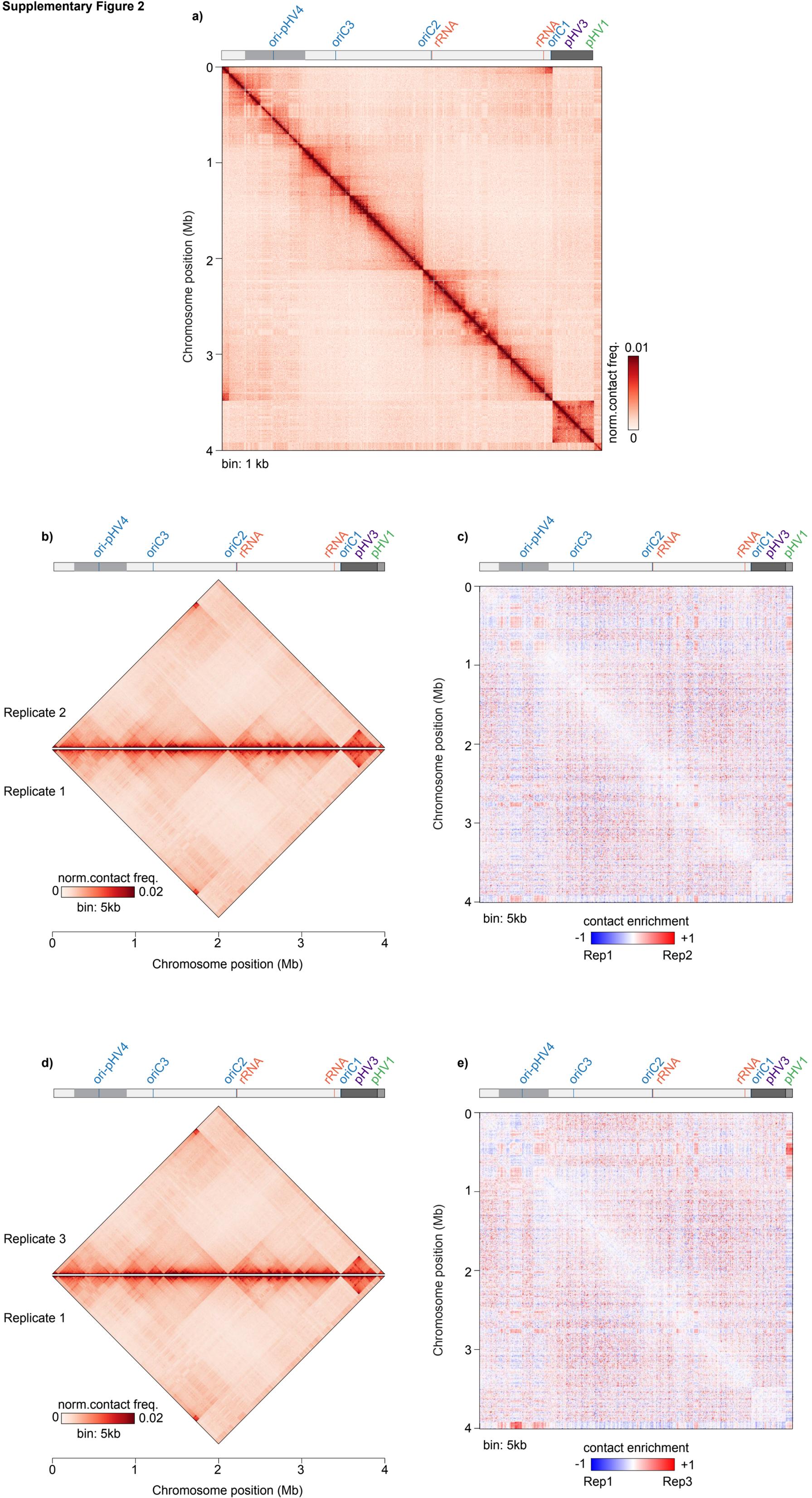
**Figure 4**



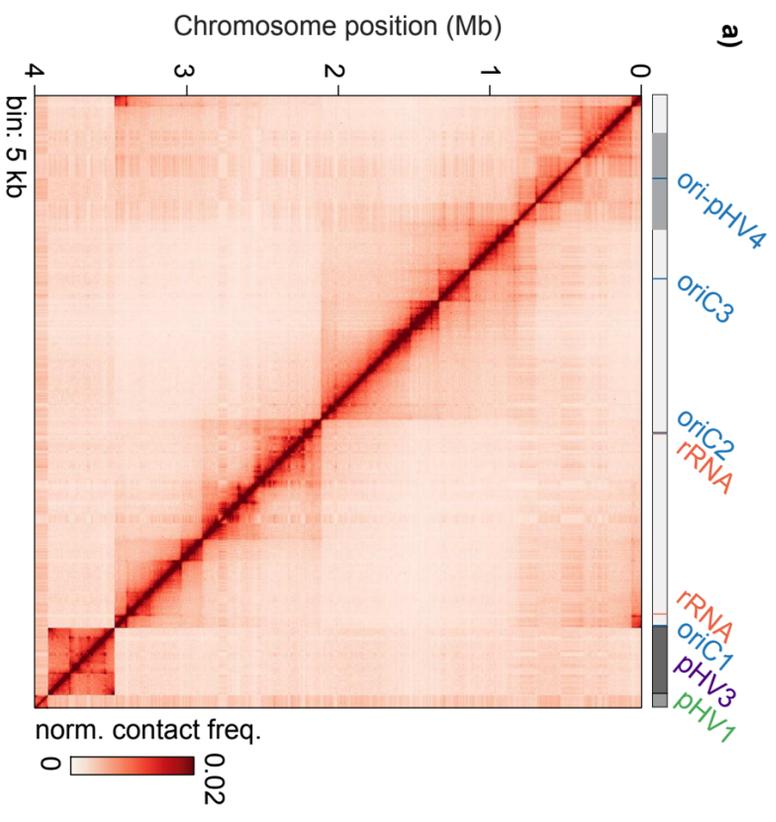
Supplementary Figure 1



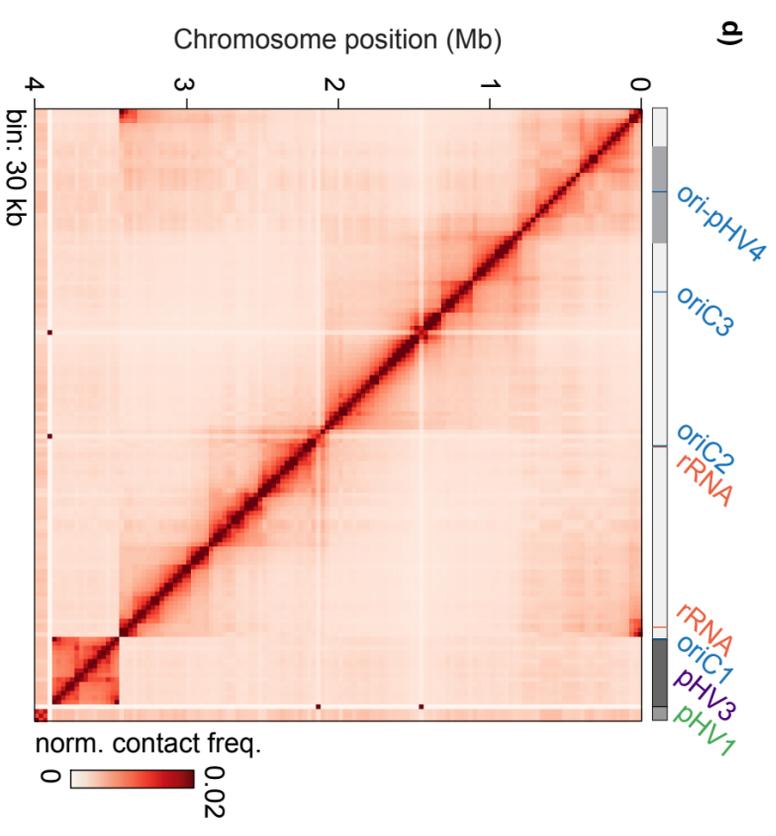
Supplementary Figure 2



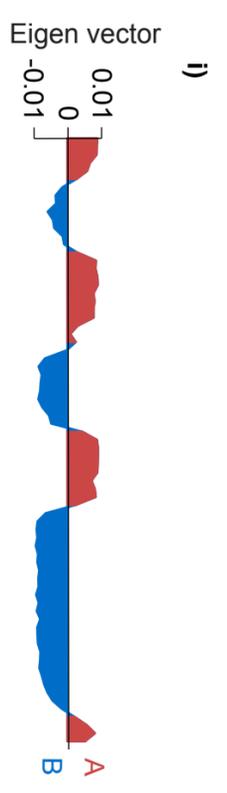
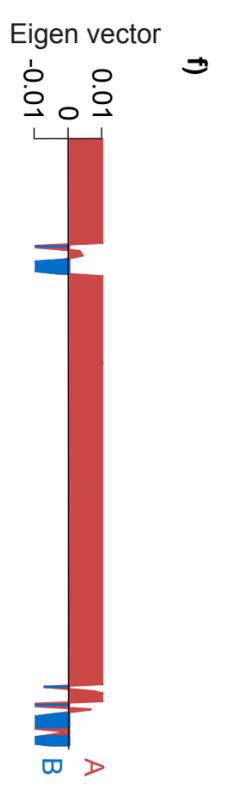
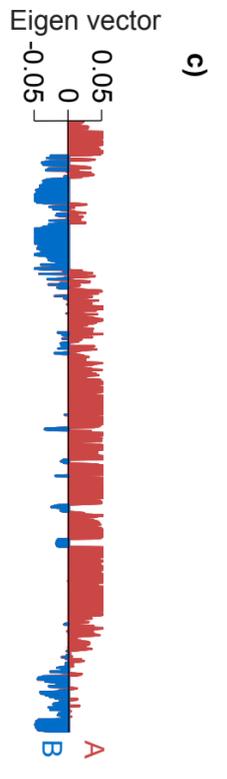
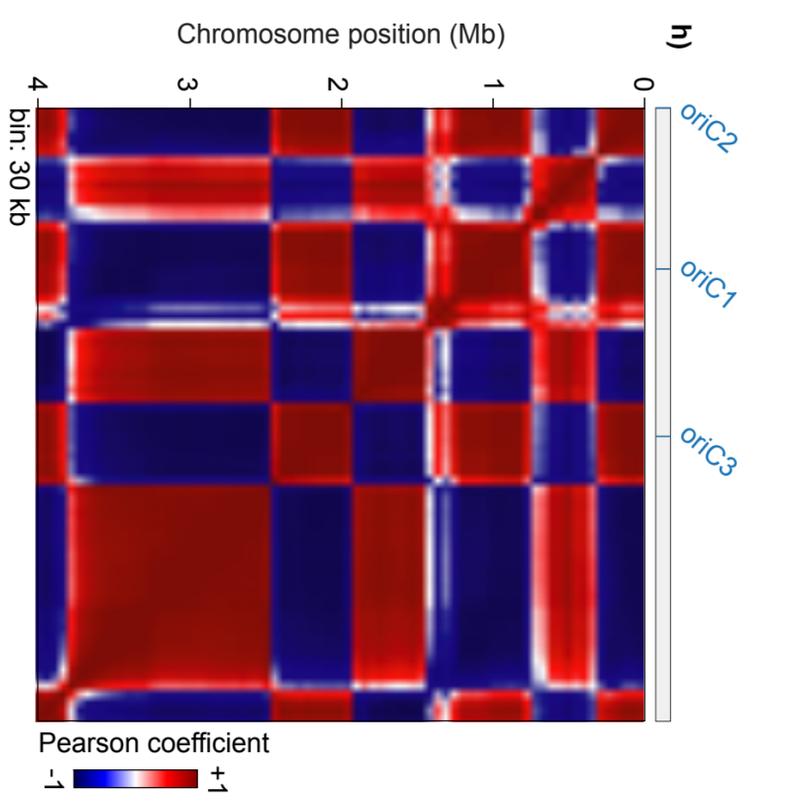
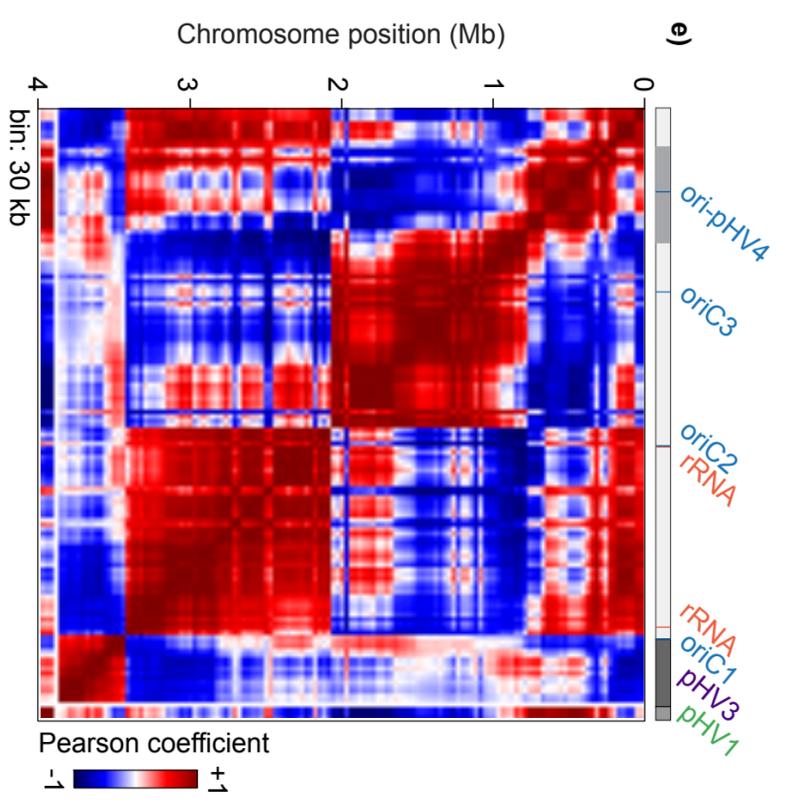
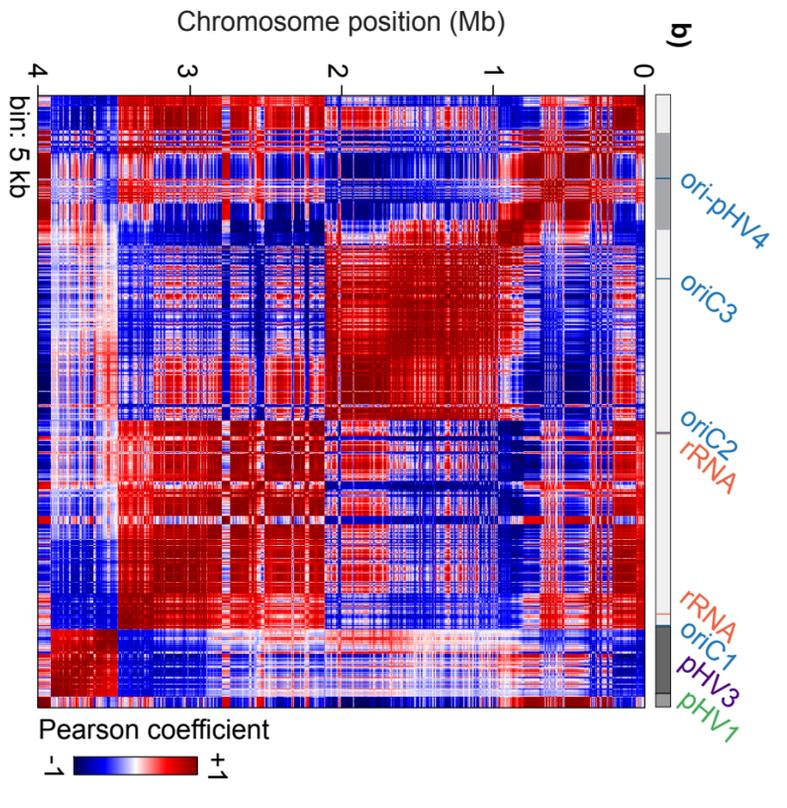
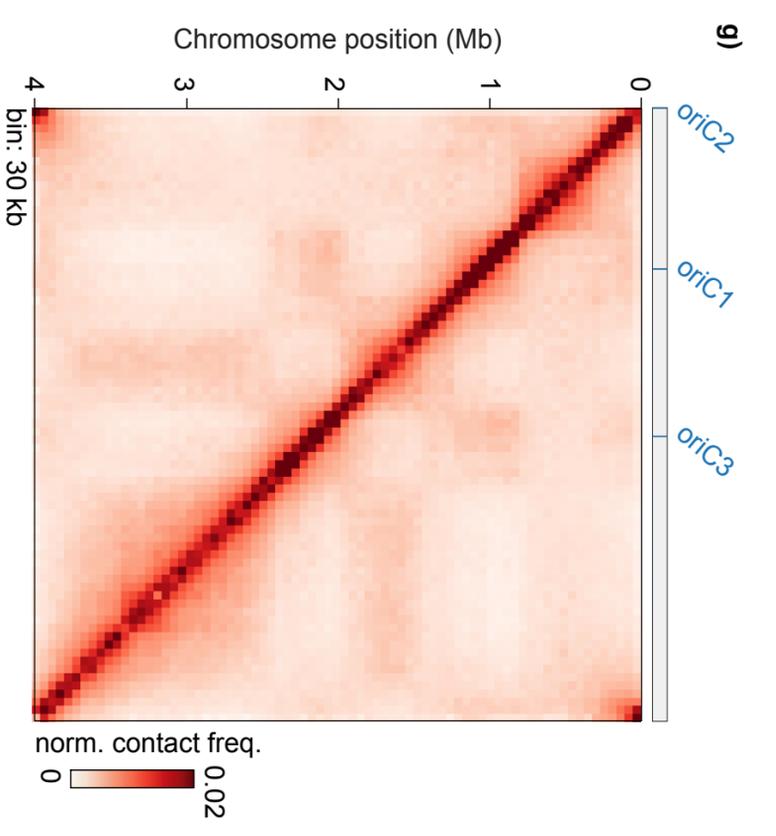
*Haloflex volcanii* (5 kb bin)



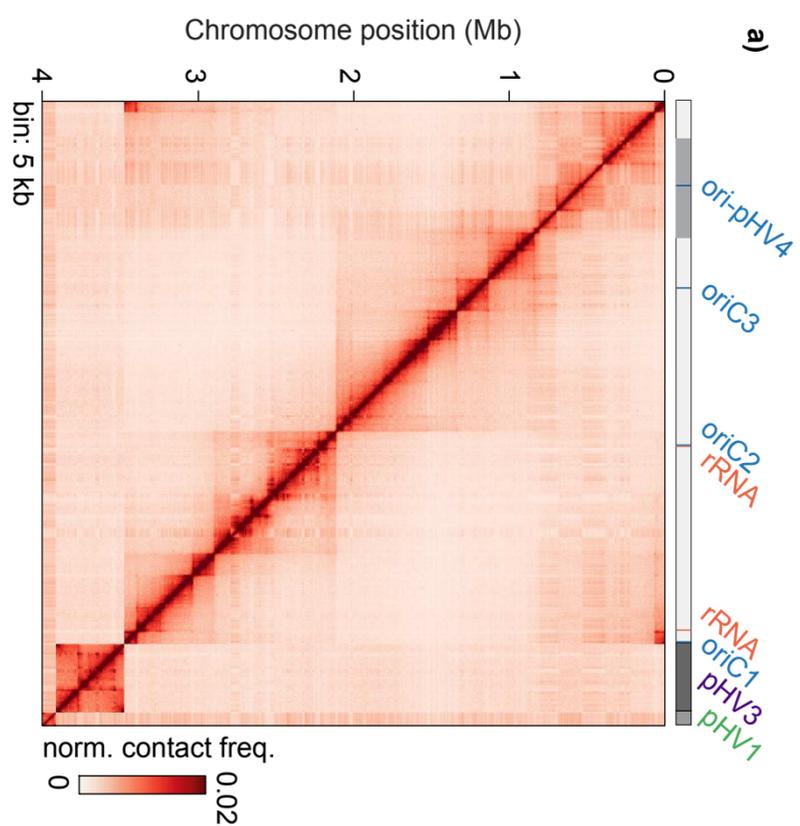
*Haloflex volcanii* (30 kb bin)



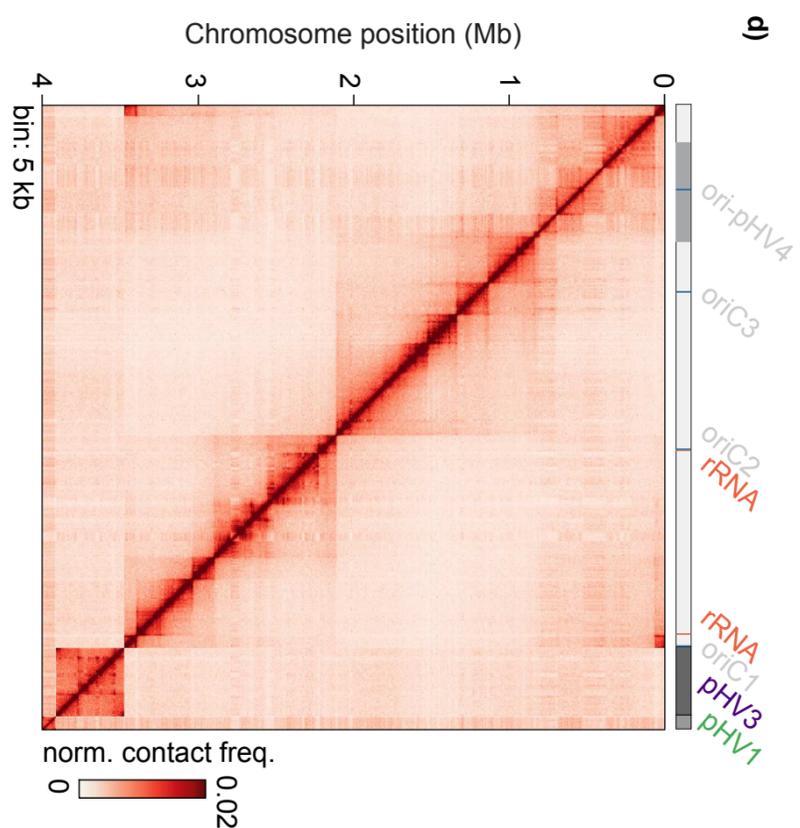
*Sulfolobus acidocaldarius* (30 kb bin)



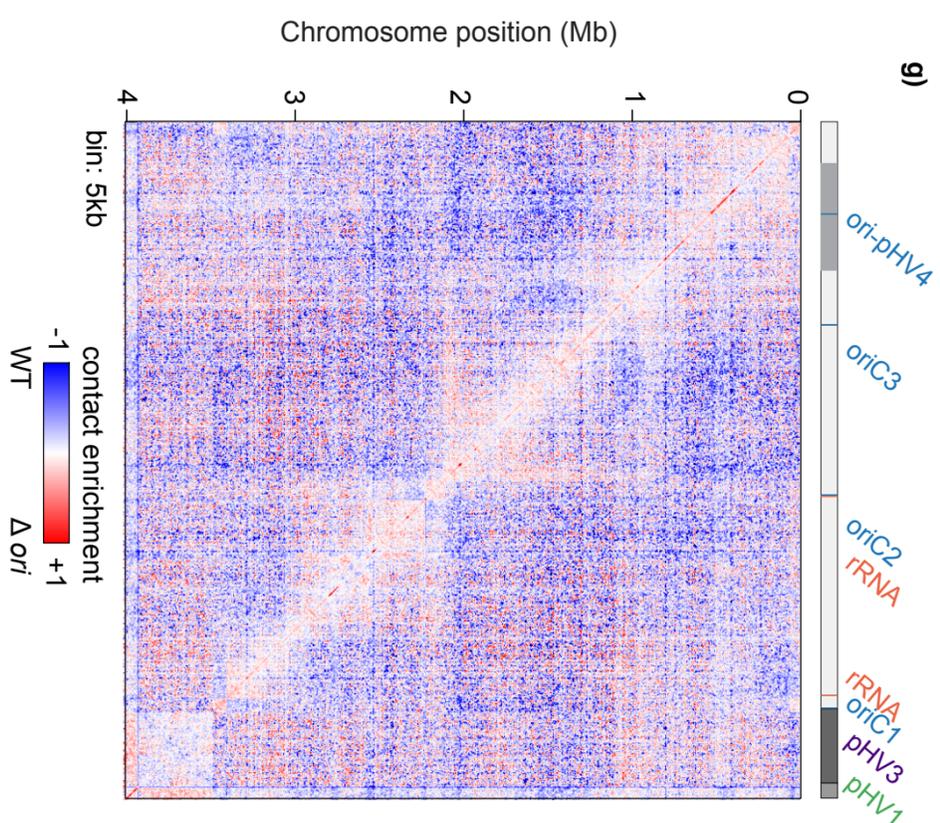
WT



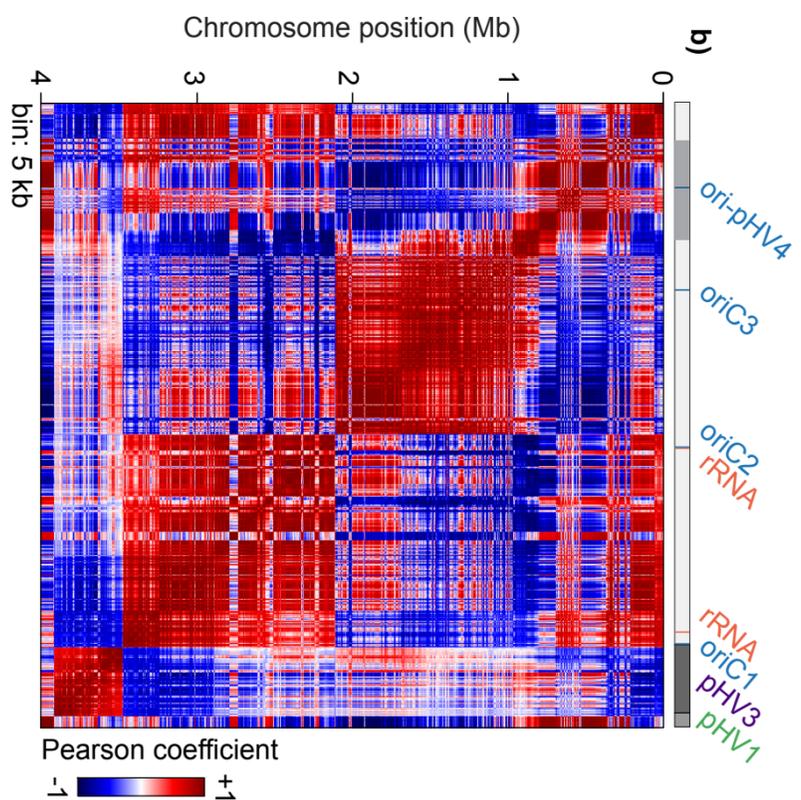
$\Delta ori$



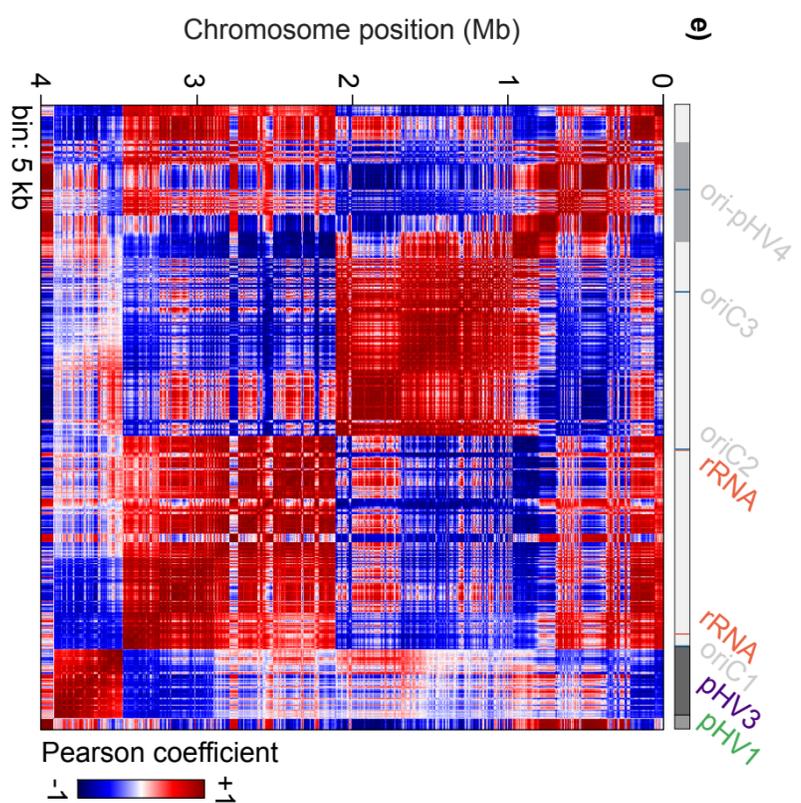
$\Delta ori$  / WT



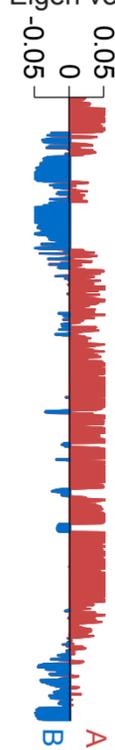
b)



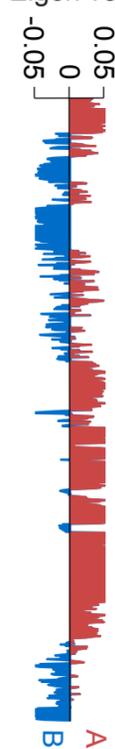
e)



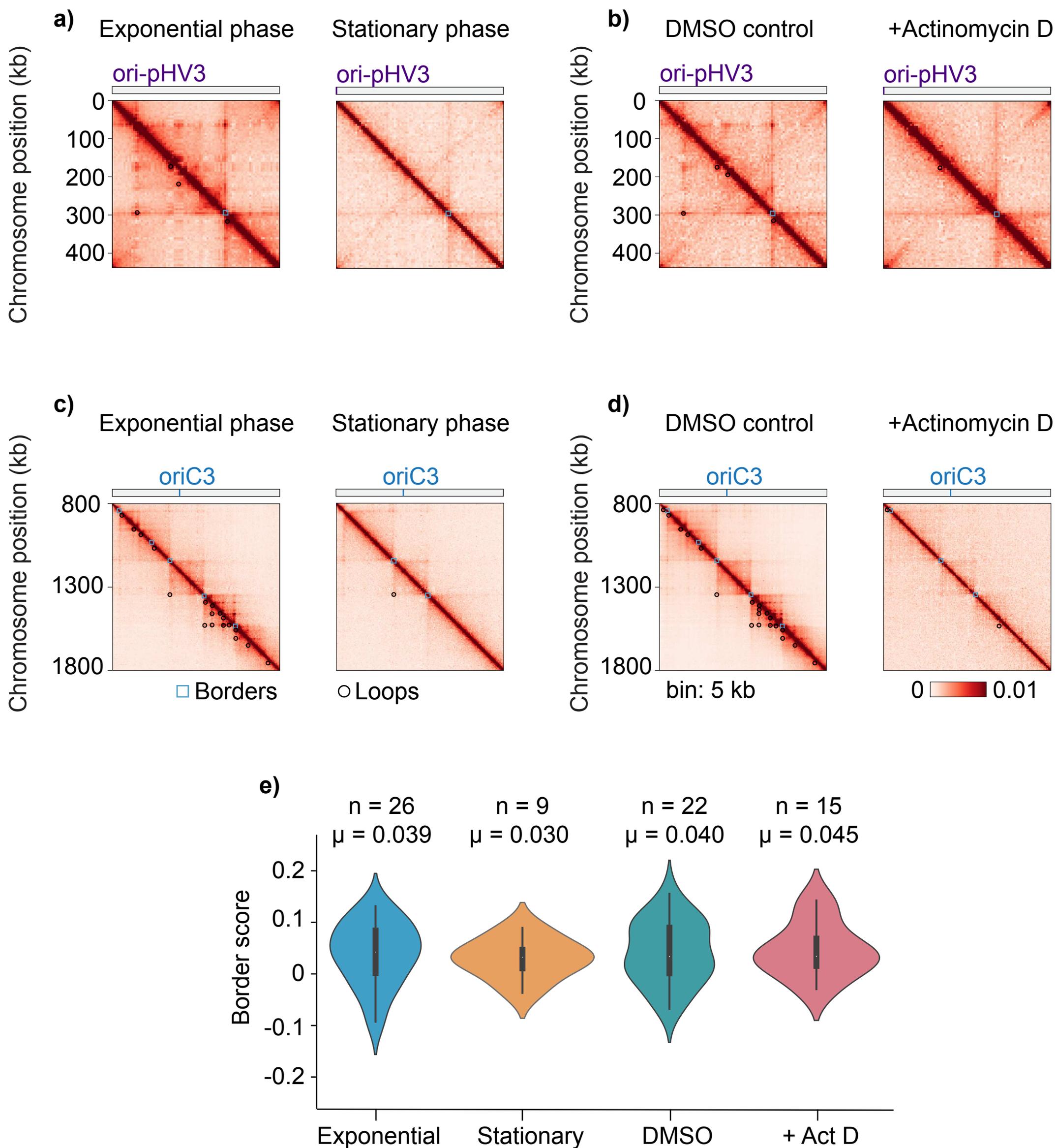
Eigen vector



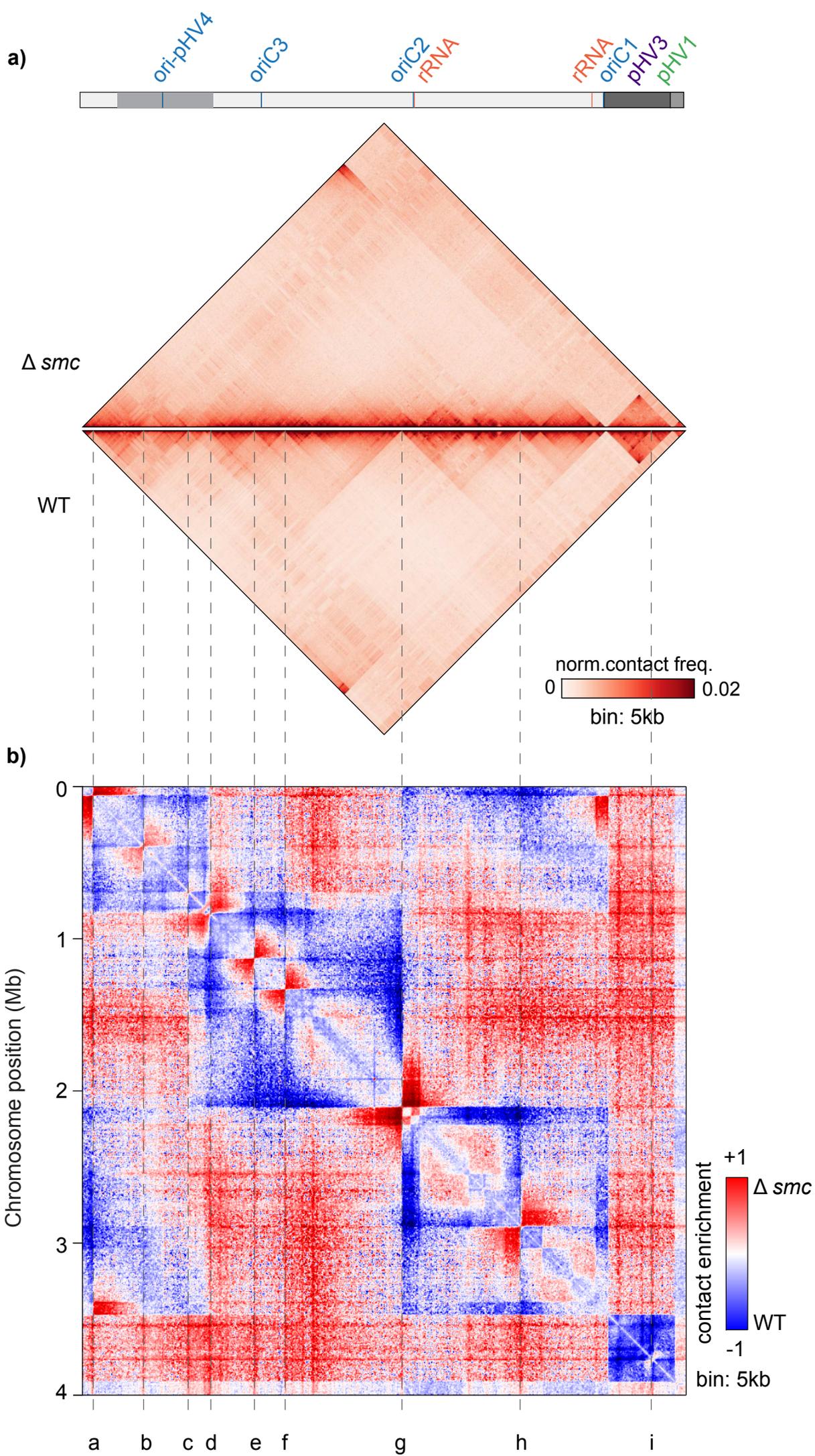
Eigen vector



# Supplementary Figure 5



Supplementary Figure 6



c)

Border	Chromosome	Approx position (bp)	Closest gene	Proposed function
a	Main	71,500	HVO_0069	arylsulphatase
b	Main - pHV4	405,500	HVO_A0132	ISH16 transposase
c	Main - pHV4	701,000	HVO_A0463	rnh RNaseH
d	Main - pHV4	846,000	HVO_A0311	HalC8 putative halocin C8
e	Main	1,141,000	HVO_0568	Transcription regulator
f	Main	1,345,500	HVO_0791	ogt DNA methyltransferase
g	Main	2,120,000	HVO_1615	RNase H-like domain-containing protein
h	Main	2,904,500	HVO_2401	Glycine cleavage system P-protein
i	pHV3	295,500	HVO_B0248	SDR oxidoreductase