Gene level resolution of bacteria and archaea genome folding

Charlotte Cockram, Agnès Thierry, Roxane Lestini, Romain Koszul

Gene-level resolution of bacterial and archaeal genomes folding

Charlotte Cockram[1], Agnès Thierry[1], Roxane Lestini[2], Romain Koszul[1,*]

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

Summary:

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

Introduction

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

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

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

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

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

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

Results

Improvements of microbial Hi-C contact maps

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

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

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

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

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

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

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

Features of chromosome organization in *H. volcanii*

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

Self-interacting domains

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

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

Plaid-like contact pattern

The *H. volcanii* contact matrix displays a striking plaid pattern involving very short (~2-15 kb) DNA segments throughout the genome. Similar patterns, but involving much larger regions of DNA, have been found in both crenarchaea and some eukaryotes [13,27]. This bipartite organization is associated with the compartmentalization of chromatin into active and inactive compartments. To determine if the same is true for *H. volcanii*, Pearson correlation matrices of Hi-C maps were computed and used to perform principal component analysis (PCA; **Figure 2e and Supplementary Figure 3)**. Although the effect was not as sharp as previously reported, a separation of the *H. volcanii* genome into compartment-like structures emerged from the approach. Upon comparing the analysis to gene expression levels characterized by RNA-seq **(Figure 2d)** we observed that the majority of the main chromosome appeared to belong to the transcriptionally active compartment, whereas the small chromosomes pHV1, pHV3, and the integrated pHV4 seemed to mostly belong to a less transcriptionally active compartment **(Figure 2e)**. The exception was the region surrounding the origin of replication of the integrated pHV4 (*ori-pHV4*) chromosome, which, like the rest of the origins of replication on the main chromosome (*oriC1-3*), appeared to fall into the A compartment. To–test whether replication initiation could directly mediate compartment organization, we performed Hi-C on a H26 strain lacking all origins of replication on the main chromosome (*OriC1-3* and *ori-pHV4)*[22]. This strain has previously been shown to grow significantly faster than wild-type cells because cells are capable of initiating replication at sites dispersed throughout the genome[22]. When we compared the Δ*ori* mutant to WT cells,

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

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

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

Discrete DNA loops scattered throughout the chromosomes

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

Transcription drives chromosome structuring in *H. volcanii*

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

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

Role of SMC in chromosome organization

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

shown to be highly conserved throughout archaeal species, the notable exception being Crenarchaea, which instead encodes a distant homolog coalescin (ClsN; **Figure 4a)**[13]. *H. volcanii* contains two SMC proteins; SMC (HVO_0689) and RAD50 (HVO_0854), as well as the SMC-like Sph4 (HVO_B0173) protein. The SMC protein is highly conserved across the Halobacteria (>65% identity) and other Euryarchaea, whilst also possessing some (25-30%) sequence identity with members of the Asgard and TACK superphyla **(Figure 4a)**. We hypothesized that the link between transcription and DNA-loops in *H. volcanii* may be mediated by the SMC protein and so generated a strain in which the *smc* gene (HVO_0689) had been deleted. Interestingly, and contrary to many species, the *H. volcanii smc* mutant was viable and grew similar to WT cells **(Figure 4b)**. It also exhibited a normal cellular morphology and DNA content by microscopy **(Figure 4c)**. Hi-C experiments performed on the *smc* mutant revealed a complex pattern of changes across the genome **(Figure 4d)**. First, there was an overall decrease in short-range contacts in Δ*smc* cells compared to wild-type. We also observed a reduction in both the number and strength of some loops across the genome, although the effect was less than that seen in actinomycin D-treated or stationary phase cultures. However, the most striking observation was the complete loss of several topological borders throughout the main chromosome (**Figure 4e**). Despite close inspection of the sequences and gene annotations present at these borders **(Supplementary Figure 6)**, no pattern as to why these specific sites were affected in the Δ*smc* mutant could be identified. To determine whether this pattern was conserved across the SMC family members present in *Haloferax*, we performed Hi-C on the published Δ*rad50 mre11* mutant (**Figure 4f)**[29]. Like the SMC protein, Rad50 is conserved across most Euryarchaea but is less-well conserved in the Asgard and TACK groups. We chose to focus on the *rad50 mre11* double mutant because, in *H. volcanii*, both genes are expressed from the same operon, and the Mre11 Rad50 proteins have previously been shown to work together as a complex that binds DNA to prevent the

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

Discussion

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

The Haloferax genome is not organized into active and inactive compartments

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

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

Transcription and SMC organize the Haloferax genome into loops and domains

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

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

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

Methods:

Strains, media and growth conditions

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

Construction of the H. *volcanii* Δsmc strain

To generate the Δ*smc* (HVO_0689) construct, the upstream region (US) and downstream region (DS) of *smc* were cloned into pTA131[39] to generate the pRL93 plasmid used to construct the deletion strain by a gene knockout system[40]. The US region was generated by PCR on H26 genomic DNA using primers RL292 and RL296. The DS region was generated by PCR on H26 genomic DNA using primers RL294 and RL295. Each PCR product contained 30 bp homology with adjacent fragments for SLIC cloning[41]: (i) with the pTA131 linearized fragments after NotI and EcoRI double-digestion (contained in RL292) in the US fragment, (ii) with the US fragment (contained in RL294) and with the pTA131 linearized fragments after NotI and EcoRI double-digestion (contained in RL295) in the DS fragment. Following the SLIC method, PCR fragments and linearized plasmid were digested using T4

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

Optimized microbial Hi-C protocol

Cell fixation

All Hi-C experiments were performed with ~1 x 10^8 cells growing in the exponential growth phase (OD_{600nm} ~0.2). Protein-DNA interactions were chemically crosslinked with fresh formaldehyde (Sigma Aldrich; 3% final concentration) for 30 min at room temperature with gentle agitation. Crosslinking was quenched by the addition of glycine (Sigma Aldrich; 0.5 M final concentration) for 20 min at room temperature. For *H.volcanii,* which is very sensitive to salt concentrations, 2.5M glycine was prepared in 18 % (w/v) saltwater solution (2.5 M NaCl, 90 mM $MgCl_2.6H_2O$, 90 mM $MgSO_4.7H_2O$, 60 mM KCl, 10 mM Tris–HCl pH 7.5) to avoid cell lysis. Cells were then collected by centrifugation (4000 x g, 10 min, room

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

Cell lysis and DNA digestion

Cell pellets were removed from the -80°C freezer and allowed to completely thaw on ice. The pellet was then resuspended in 1 ml of 1x TE + cOmplete protease inhibitor cocktail (EDTA-free, Sigma Aldrich) and transferred to a 2 ml Precellys tube containing 0.5 mm glass beads (VK05, Ozyme). Cells were then subject to mechanical disruption using the Precellys Evolution tissue homogenizer (V7500: 5 x 30s, 20s pause). Please note, that for Precellys machines without the Cryolys cooling attachment (Ozyme), tubes need to be removed every 3 cycles and placed on ice for 5 min to stop sample overheating and subsequent degradation. The lysate (~1ml in volume) was then carefully transferred to a 5 ml Eppendorf tube, avoiding the transfer of any glass beads. Proteins that were not crosslinked to DNA during formaldehyde fixation were then degraded by the addition of SDS (Thermo Fisher, 0.5% final concentration) for 10 min at room temperature. DNA was then prepared for digestion by the addition of 3 ml dH2O, 500 µl 10X Digestion buffer (200 mM Tris-HCl pH 7.5, 100 mM MgCl2, 10 mM DTT, 1 mg/ml BSA) and 500 µl 10 % Triton-X-100 (Thermo-Fisher). After thoroughly mixing the reaction, 400 µl was removed and transferred to a 1.5 ml Eppendorf tube as a non-digested (ND) control. The restriction enzyme, HpaII (New England Biolabs, 1000 U) was then added to the remaining sample and the tube incubated with gentle agitation for 3h at 37°C.

Choice of restriction enzymes for Hi-C

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

Biotinylation of DNA ends

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

temperature) to pellet the insoluble fraction containing the protein-DNA complexes of interest. After removing the supernatant, the pellet was resuspended in 400 µl dH2O. To this the following was added; 50 µl 10x Ligation Buffer (500 mM Tris-HCl pH 7.5, 100 mM MgCl2, 100 mM DTT), 4.5 µl 10 mM dAGTTp, 37.5 µl Biotin-14-dCTP (Thermo Fisher), 50 Units of DNA Polymerase I - Large Klenow Fragment (New England Biolabs). After briefly mixing, the reaction was incubated with agitation for 45 min at 37°C.

Blunt-end ligation

For further gains, we took inspiration from published human protocols and reduced the volume of the blunt-end ligation step by 20-fold to mimic the conditions of 'in situ Hi-C' protocols[17,43]. Our rationale was that in a smaller volume there would be less spurious events than in dilute conditions and, as a result, less noise in the resulting Hi-C matrices. Following further trial-and-error optimization, we performed the ligation at room temperature and reduced its duration (**Figure 1a and Supplementary Figure 1**). Once the fill-in reaction was complete, the ligation was set up by adding the following; 120µl 10x Ligation Buffer, 12 µl 10 mg/ml BSA, 12 µl 100mM ATP, 540 µl dH2O, 480 U T4 DNA ligase (Thermo Fisher). The reaction was mixed gently and then incubated with gentle agitation for 3h at room temperature.

Reverse crosslinking and DNA purification

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

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

Preparation of sequencing libraries

DNA sonication and size-selection

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

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

Biotin pull-down

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

µl of 1x T4 ligase buffer (New England Biolabs), and the mixture transferred to a new 1.5 ml Eppendorf.

End-repair

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

A-tailing

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

Adapter ligation

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

Library amplification by PCR

The PCR reaction was set up as follows: 40 µl of Phusion 2x High Fidelity Master Mix (New England Biolabs), 5 µl of 2 µM Primer Mix (NEXTflex, Bioo Scientific), 32 µl dH2O and 3 µl Streptavidin-beads containing the HiC library. The library was amplified for 12 cycles following the manufacturers' instructions. After PCR, the reaction was placed on a magnet for 1 min to separate the streptavidin beads from the PCR reaction containing the amplified library. The supernatant was transferred to a new tube and the beads discarded. The PCR reaction was then purified by adding an equal volume of AmPure XP beads, the reaction was then mixed well and incubated at room temperature for 5 min. Following two washes with 80 % EtOH, the beads were mixed with 50 µl of 10 mM Tris-HCl pH 8 and incubated for 5 min at room temperature. Finally, the reaction was placed on the magnet and the DNA-containing supernatant transferred to a new Eppendorf tube. The DNA was then checked on a 1% agarose gel to determine the size of the final library and to check for the presence of primer dimers. Once this was confirmed, the Hi-C library was ready to prepare for paired-end sequencing using the 75 cycle High-Output Kit v2.5. Sequencing was performed on a Nextseq500 according to Illumina's instructions.

Processing of Hi-C sequencing data

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

Calculating generation time of wild-type and mutant strains

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

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

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

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[44]. Reads were then processed using bamCoverage
635 (implemented in deepTools2[45]) 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[46]. 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[8] 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[47]. 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**

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

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

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

Acknowledgments

We thank Olivier Espéli, Frédéric Boccard, and Vicky Lioy for careful reading of the manuscript. We also thank the members of the RSG lab for insightful discussions about experiments, especially Cyril Matthey-Doret and Axel Cournac for their guidance regarding 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. & Heyrovska, N. Conservation of chromosomal arrangement among three strains of the genetically unstable archaeon Halobacterium salinarium. *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 archaehacterium, 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. (d) Directionality index analysis similar to that shown in (b) but at the scale of 100 kb. (e) 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. (f) GC content (50 kb sliding window) of the genome. (g-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 log2 ratio of Hi-C interactions

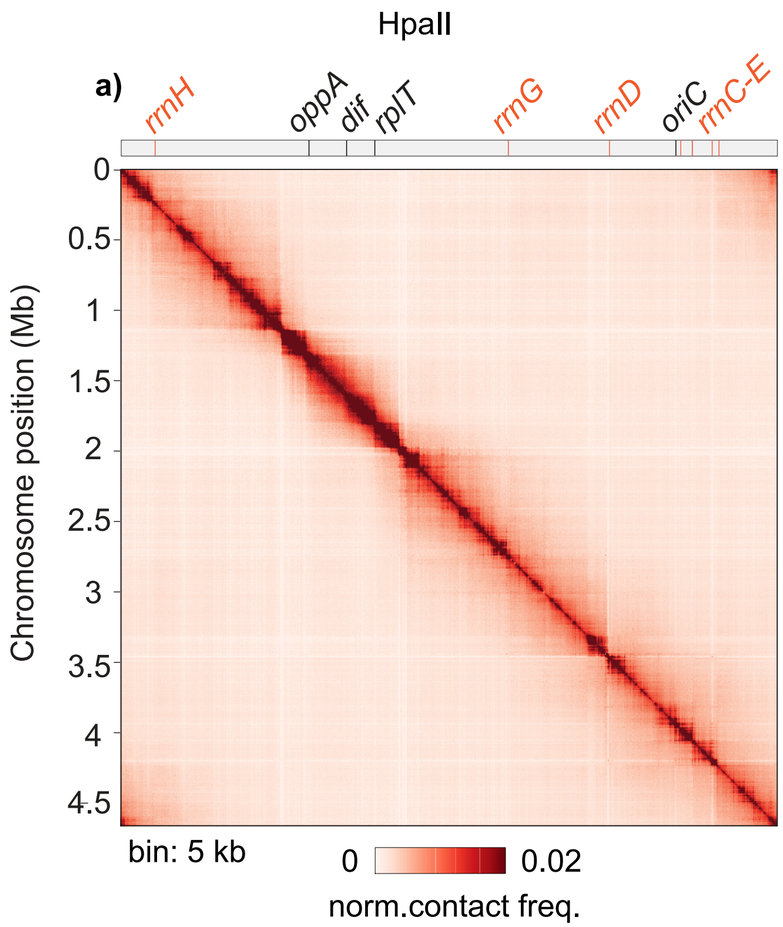
between exponential and stationary phase cultures. Blue to red colorscale reflects the enrichment in contacts in one population with respect to the other, white indicates no difference. (d) Magnifications of the 437.9 kb pHV3 chromosome, and a 1 Mb region of the main chromosome (coordinates: 800 kb – 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 log2 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).

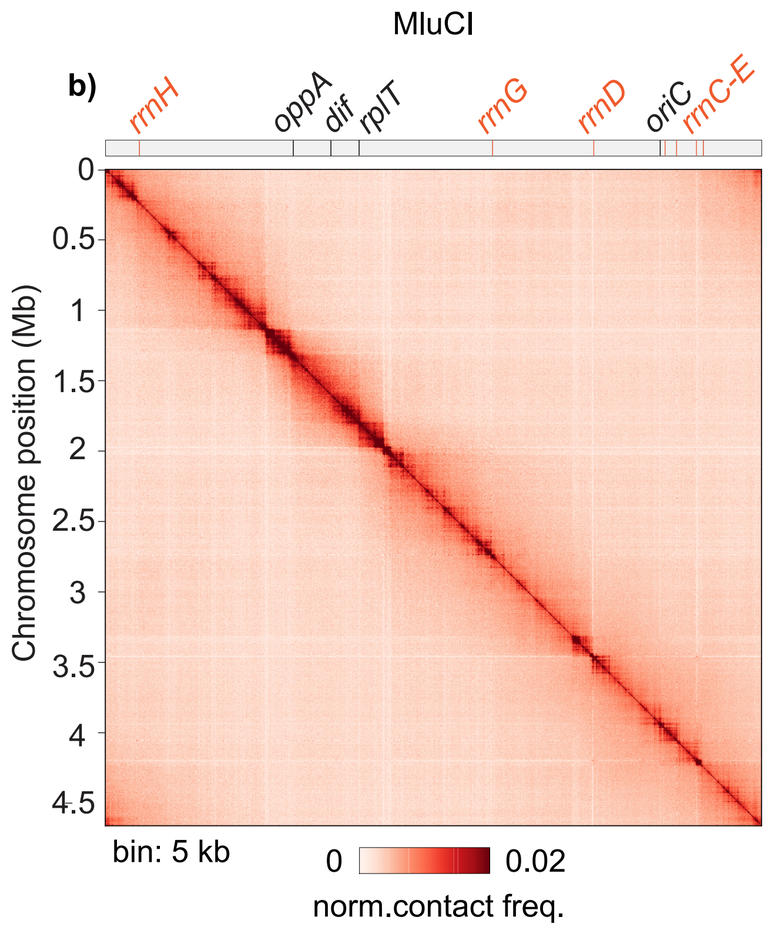
Figure 4: Role of SMCs in *H. volcanii* chromosome organization.

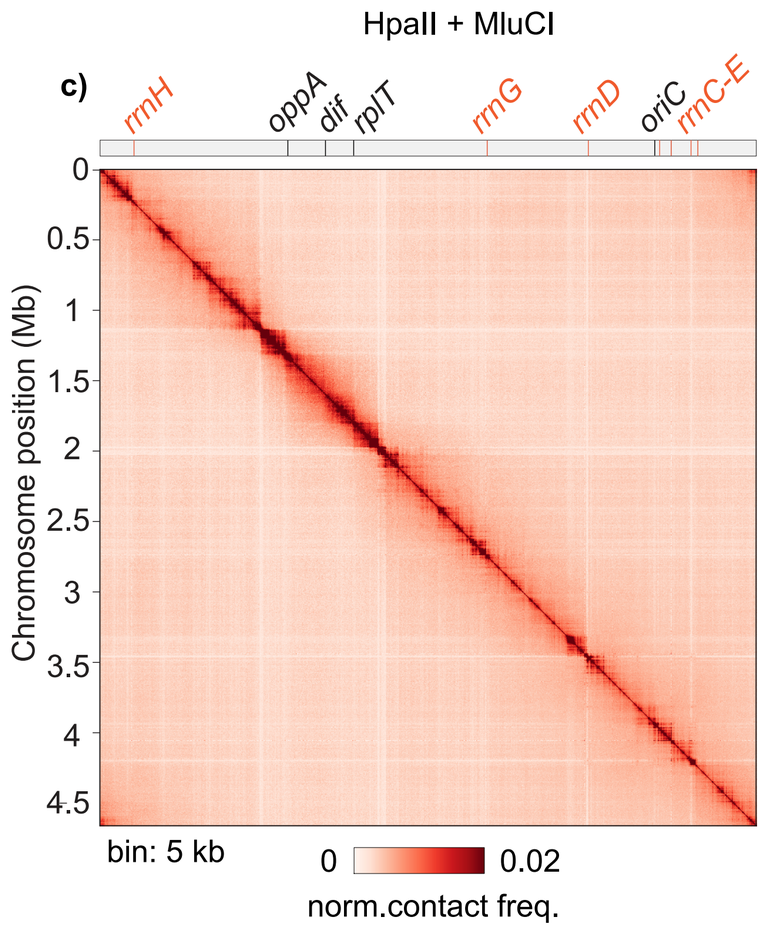
(a) Phylogenetic tree showing the conservation of canonical SMCs, Rad50, and the recently identified Coalescin (ClsN) proteins between different archaeal lineages. Adapted from [13,48] and explained in the Materials and Methods. (b) Generation times of wild-type (WT), Δ*smc,* and Δ*mre11 rad50*. Error 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, Δ*smc,* and Δ*mre11 rad50* cells. Scale bar = 5 µm. (d) Normalized Hi-C contact maps of asynchronously growing populations of WT and Δ*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 log2 ratio of Hi-C interactions between WT and

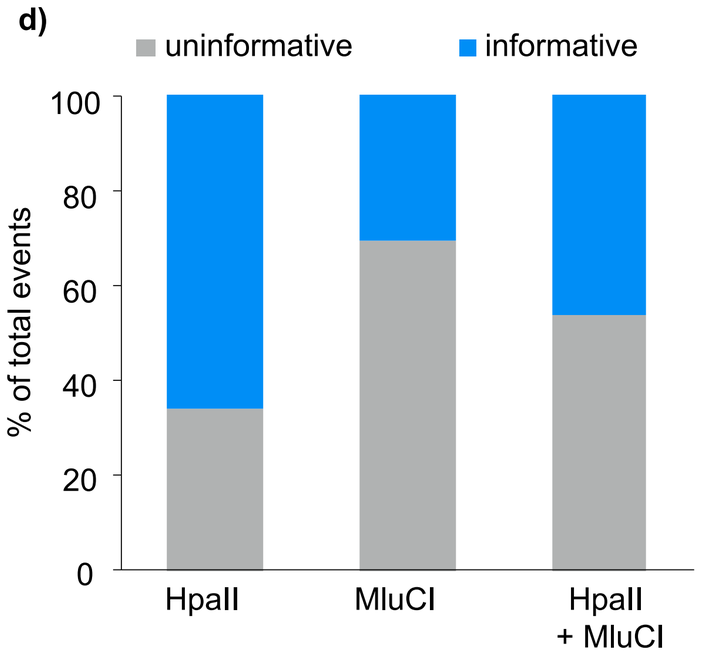
Δ*smc* mutant cells. Blue to red colorscale reflects the enrichment in contacts in one population with respect to the other, white indicates no difference. (f) Normalized Hi-C contact maps of asynchronously growing populations of WT and Δ*mre11 rad50* mutant cells (5 kb bin). (g) Differential contact map corresponding to the log2 ratio of Hi-C interactions between WT and Δ*mre11 rad50* mutant cells. (h) Magnification of 1 Mb region of the main chromosome (coordinates 1620 kb – 2620 kb) for WT, Δ*smc,* and Δ*mre11 rad50* mutant cells. The position of chromosomal loops and borders that have been detected with Chromosight[28] 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).

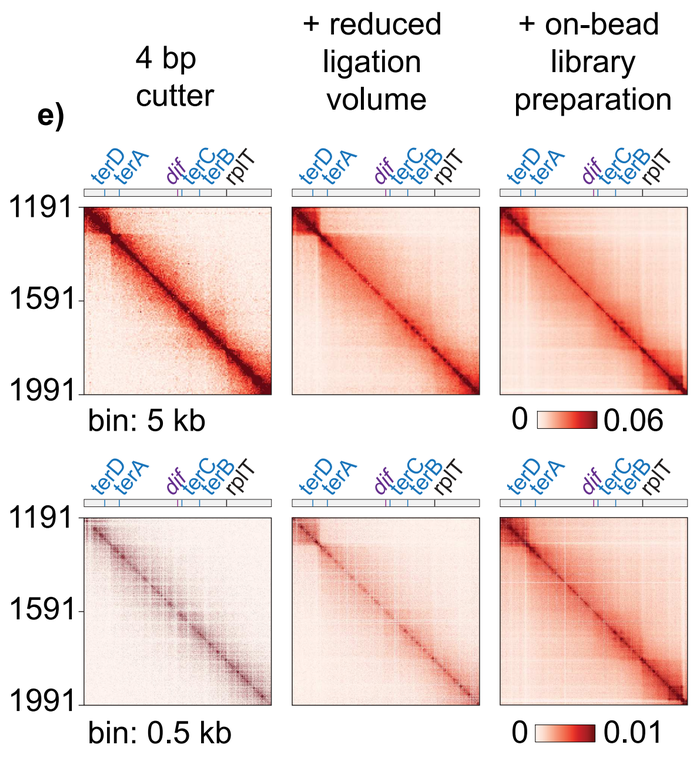
Supplementary Figure 1

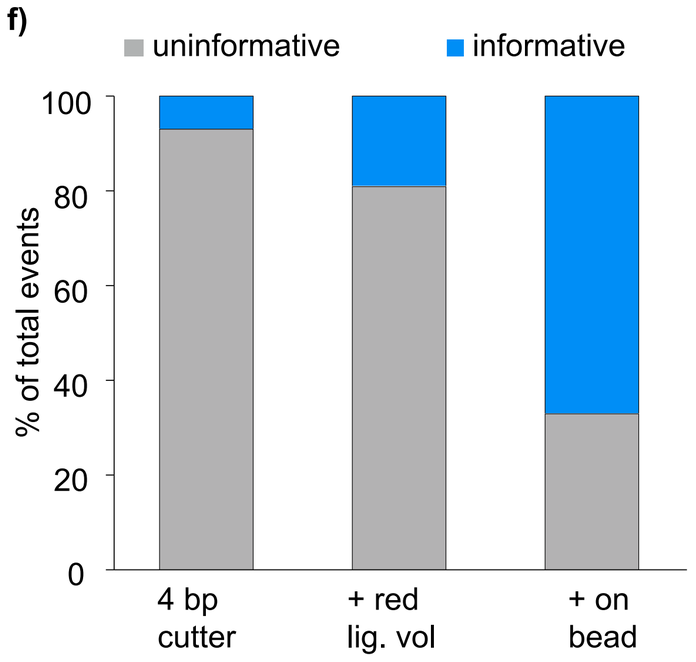




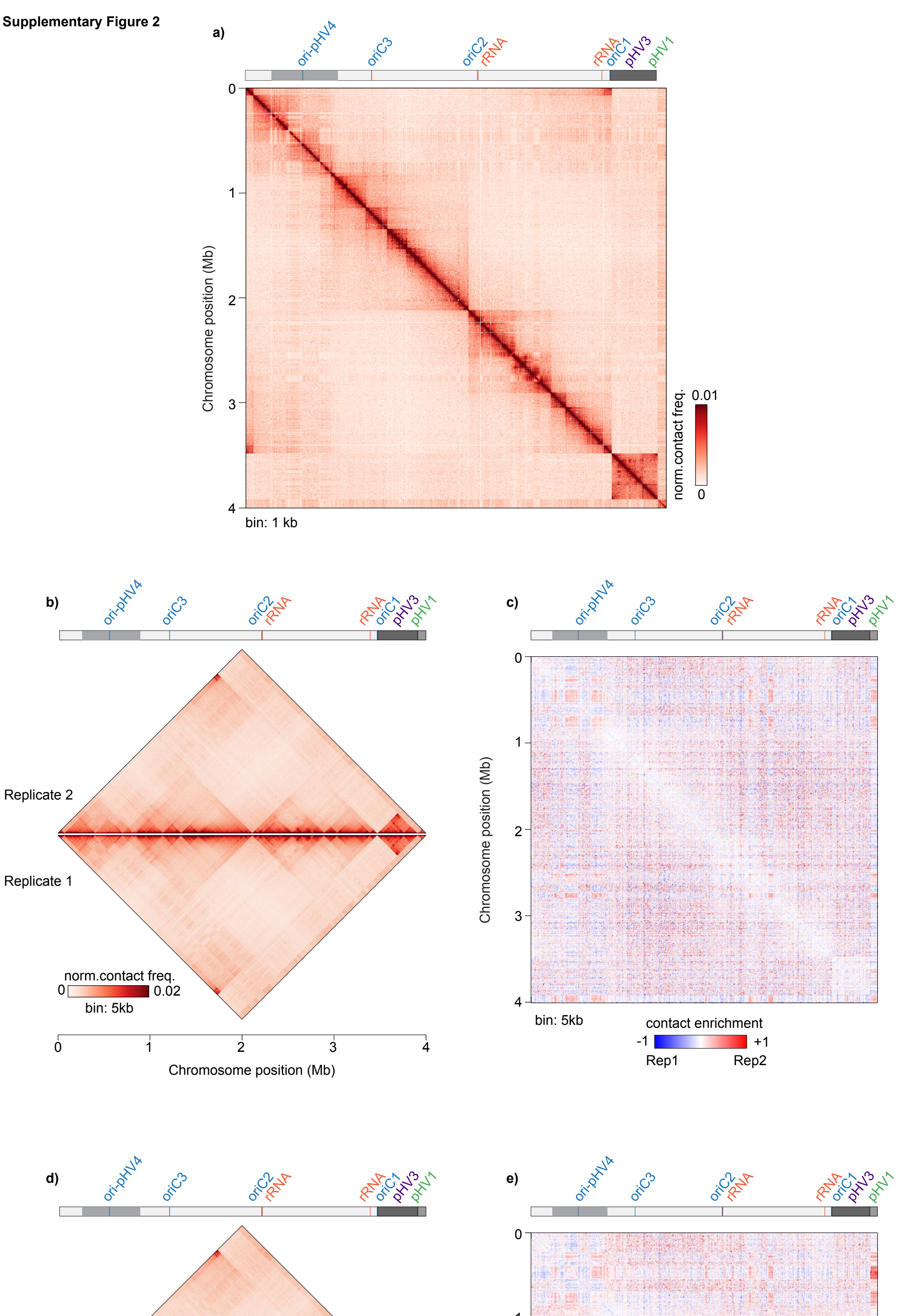




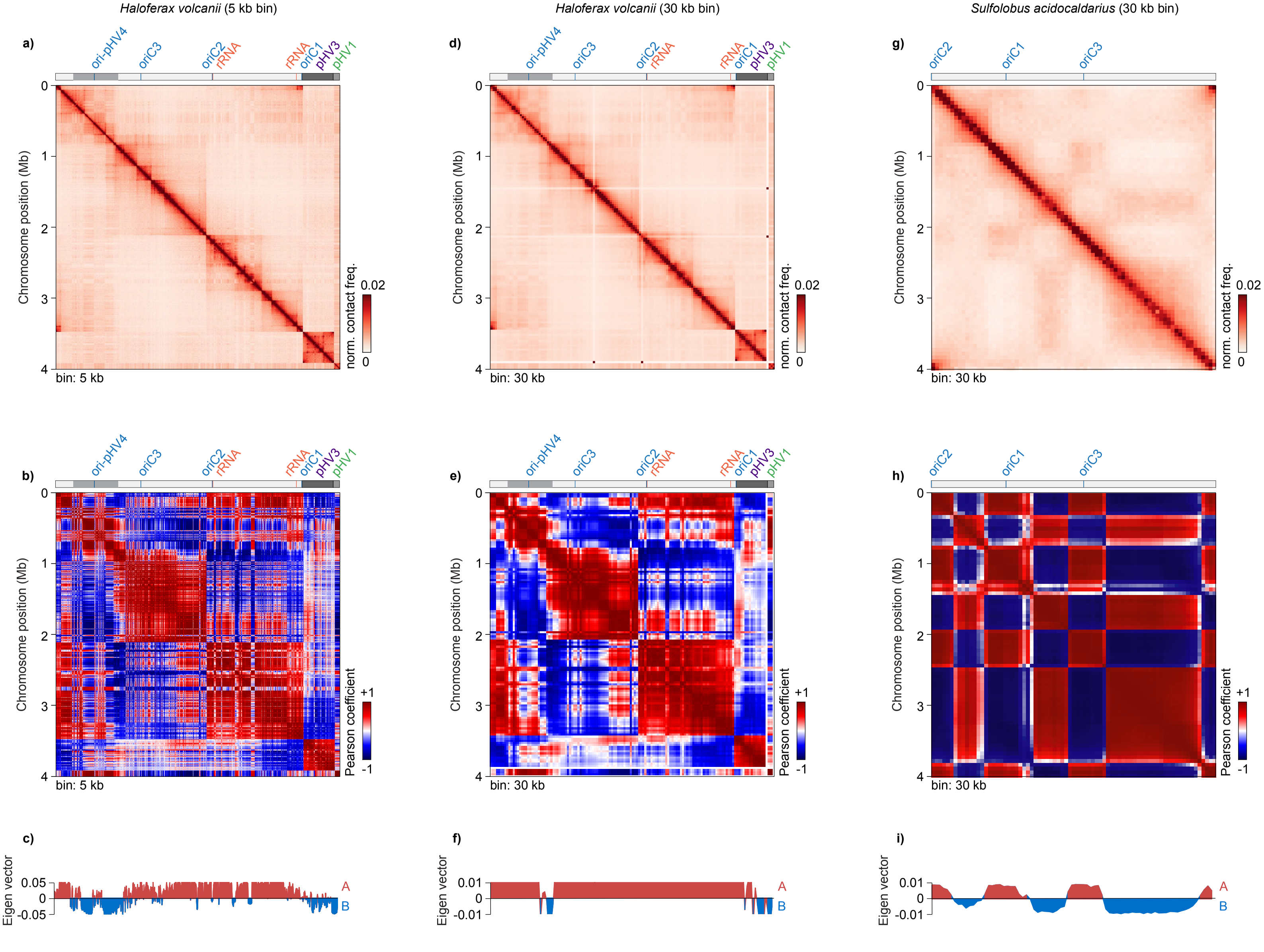




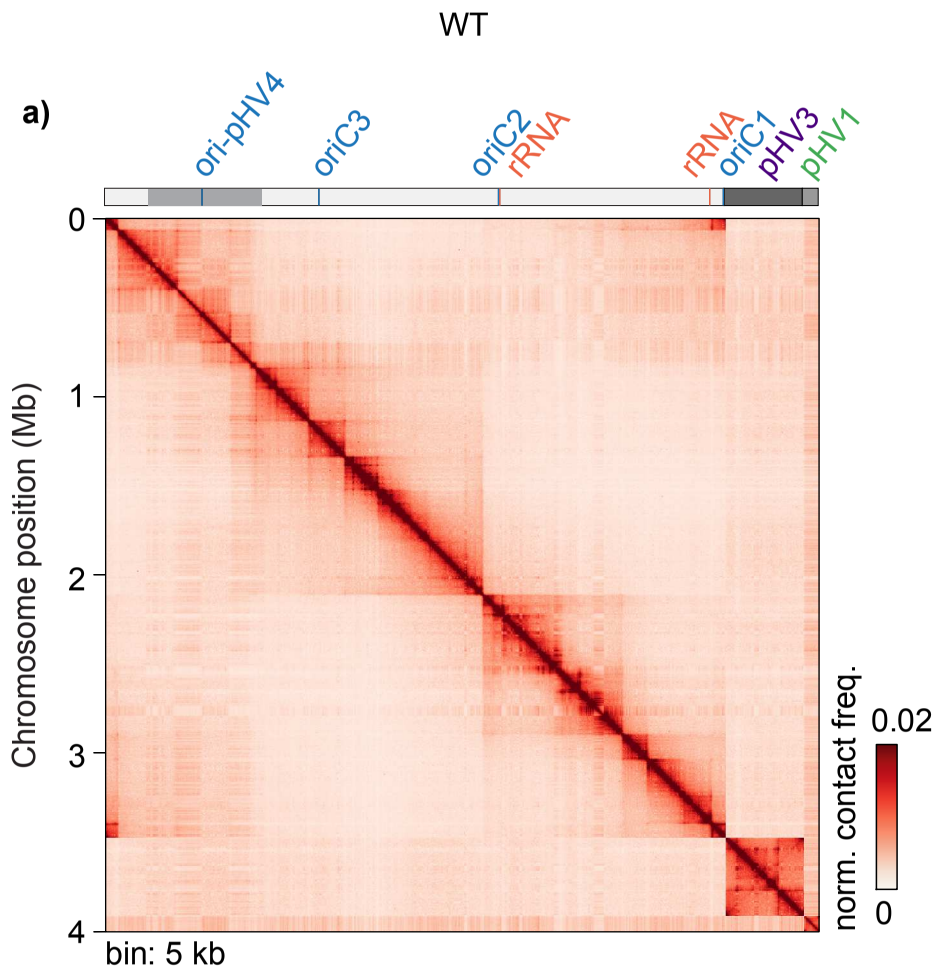
Supplementary Figure 2

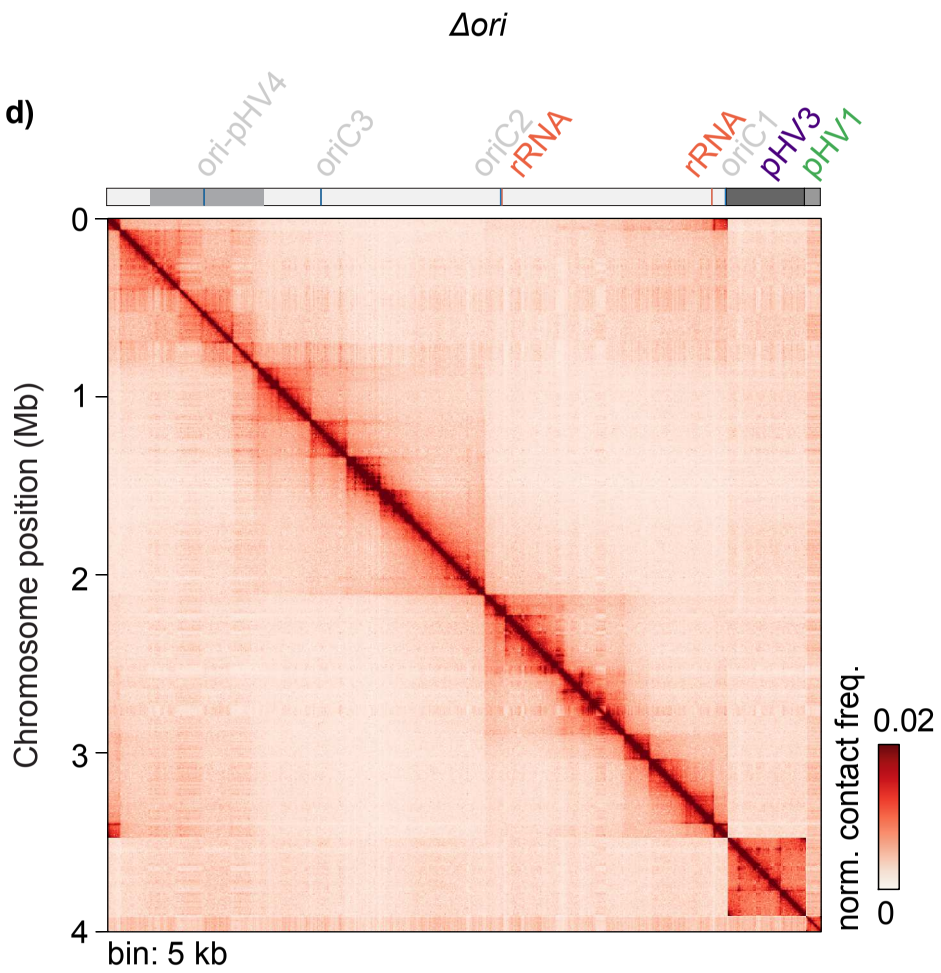


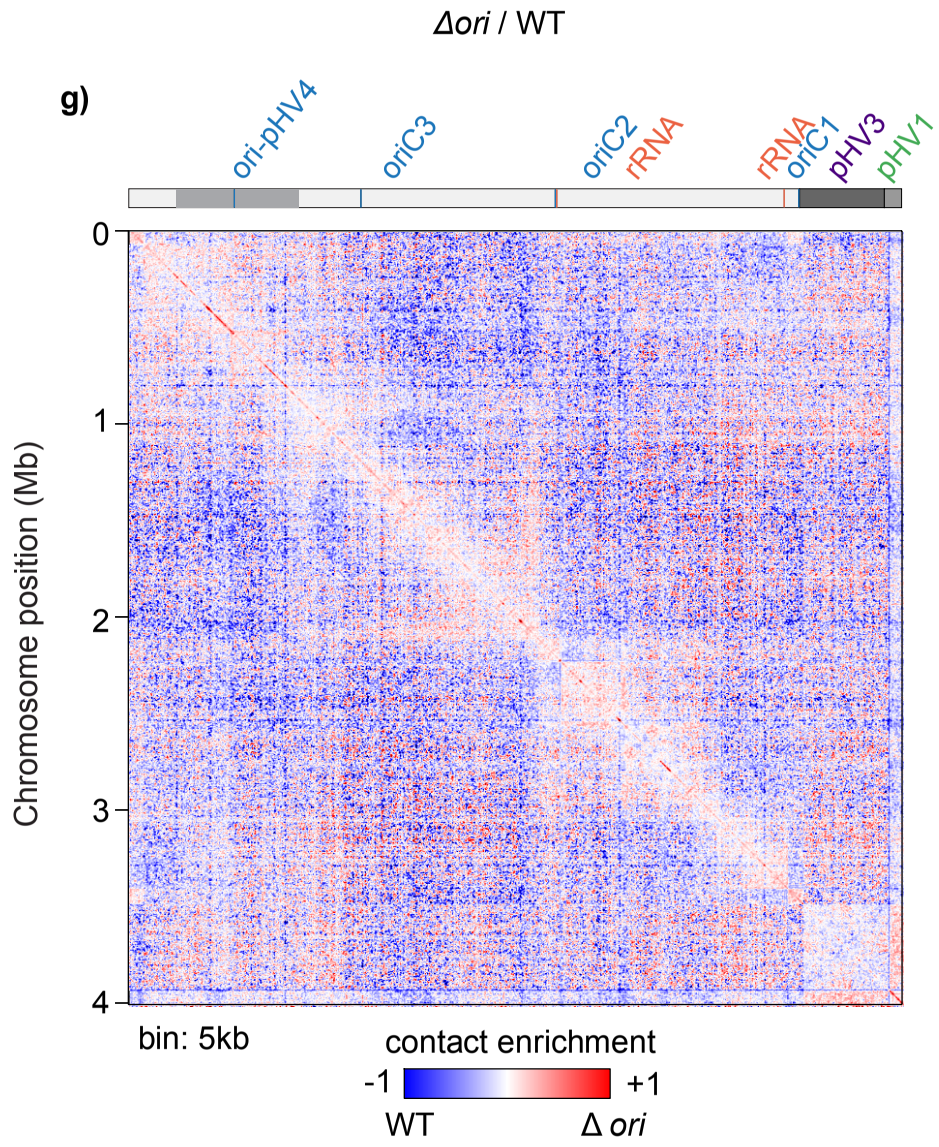
Supplementary Figure 3

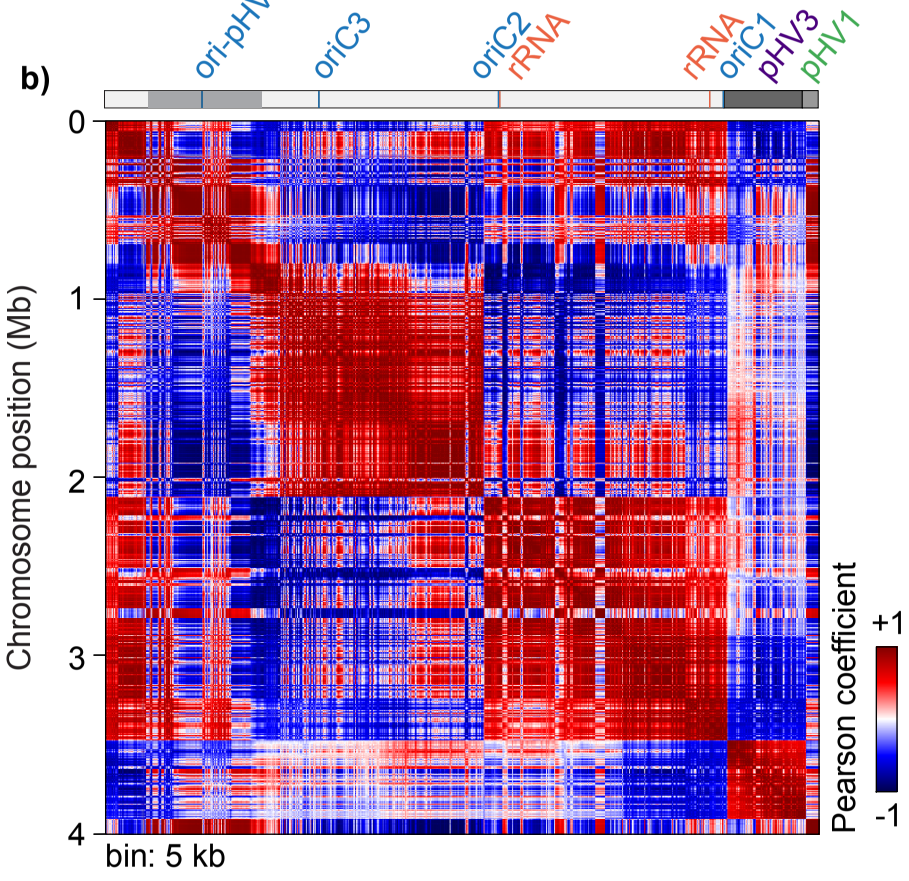


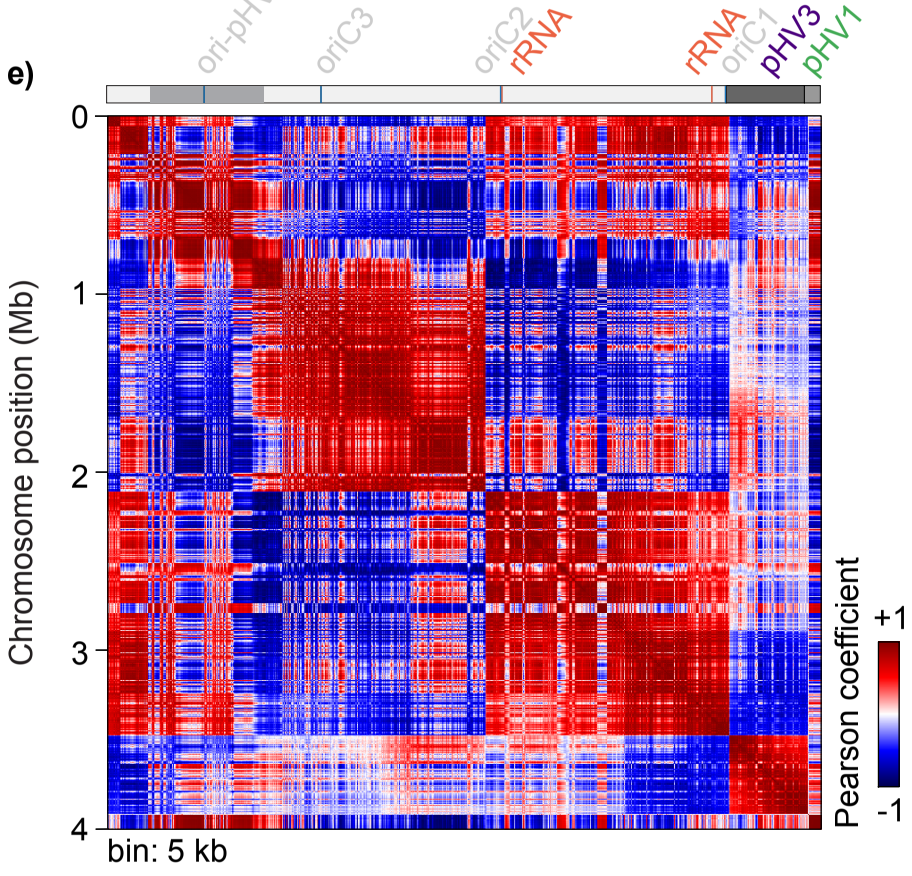
Supplementary Figure 4

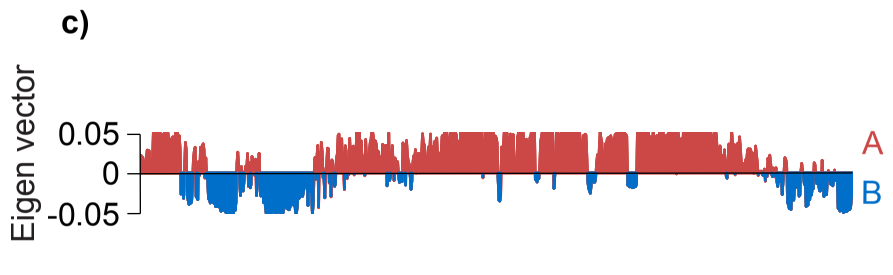


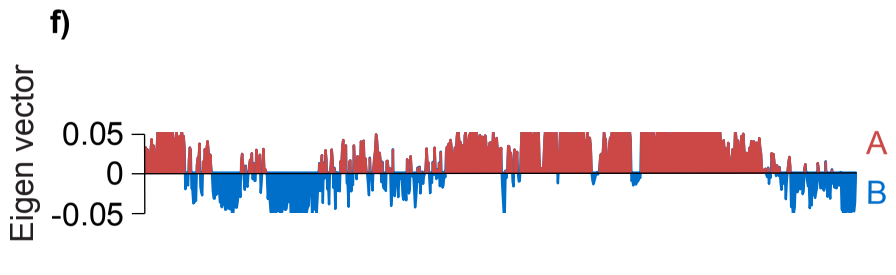




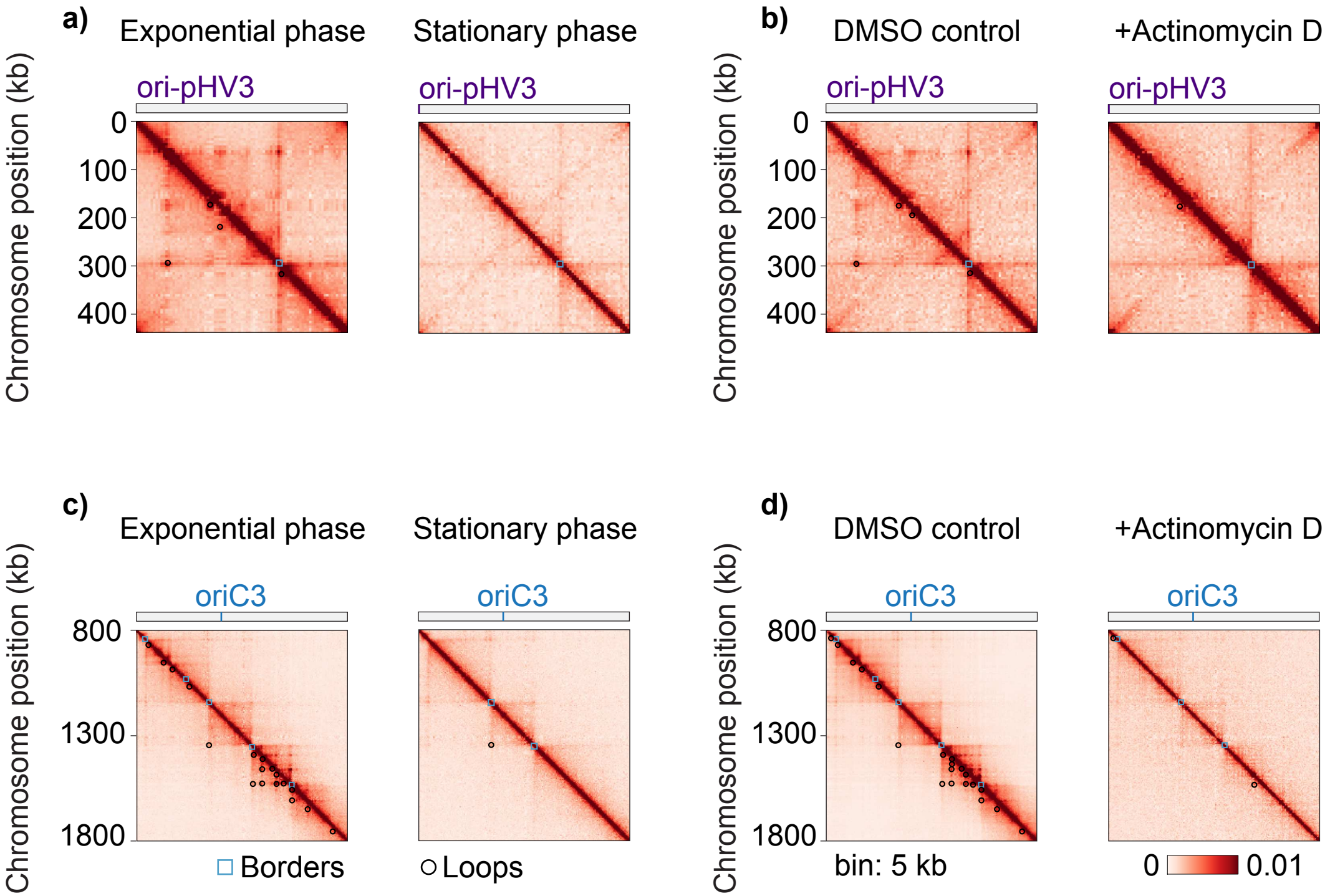


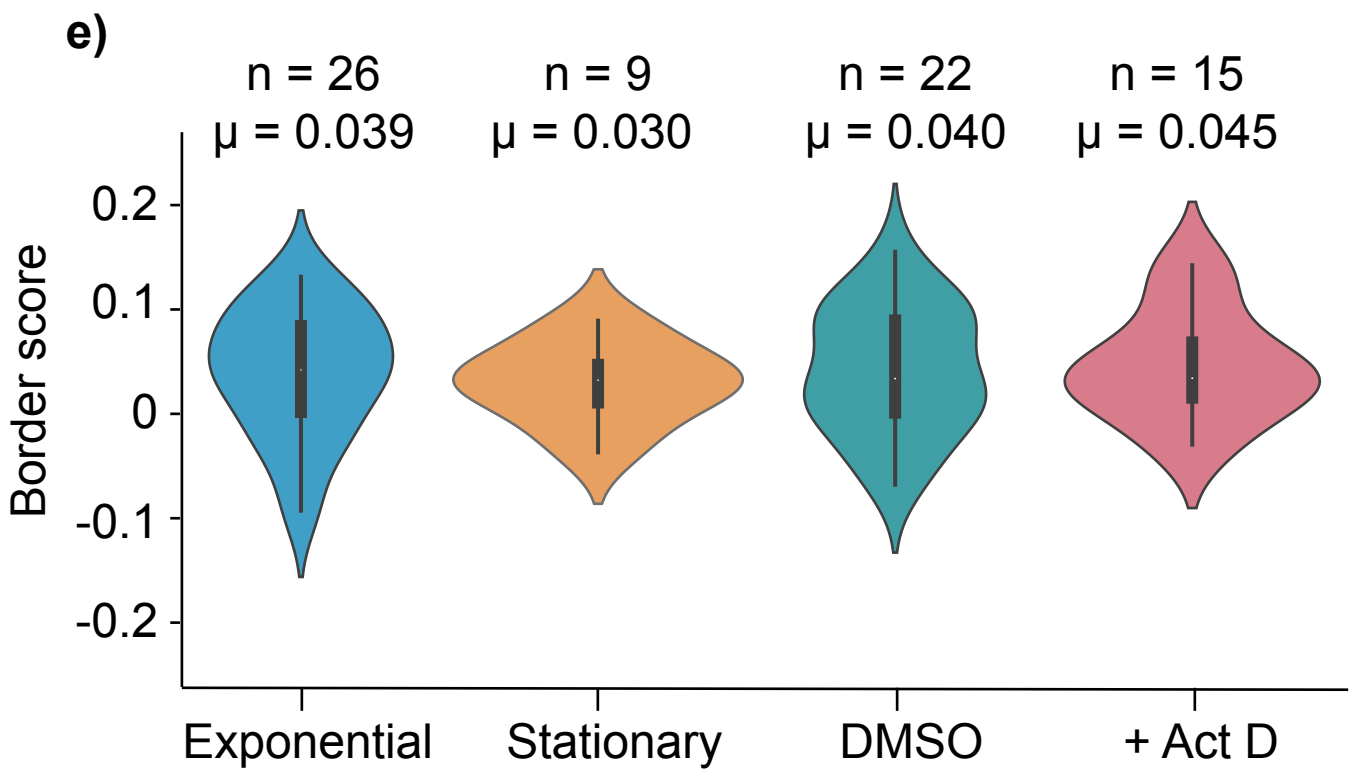




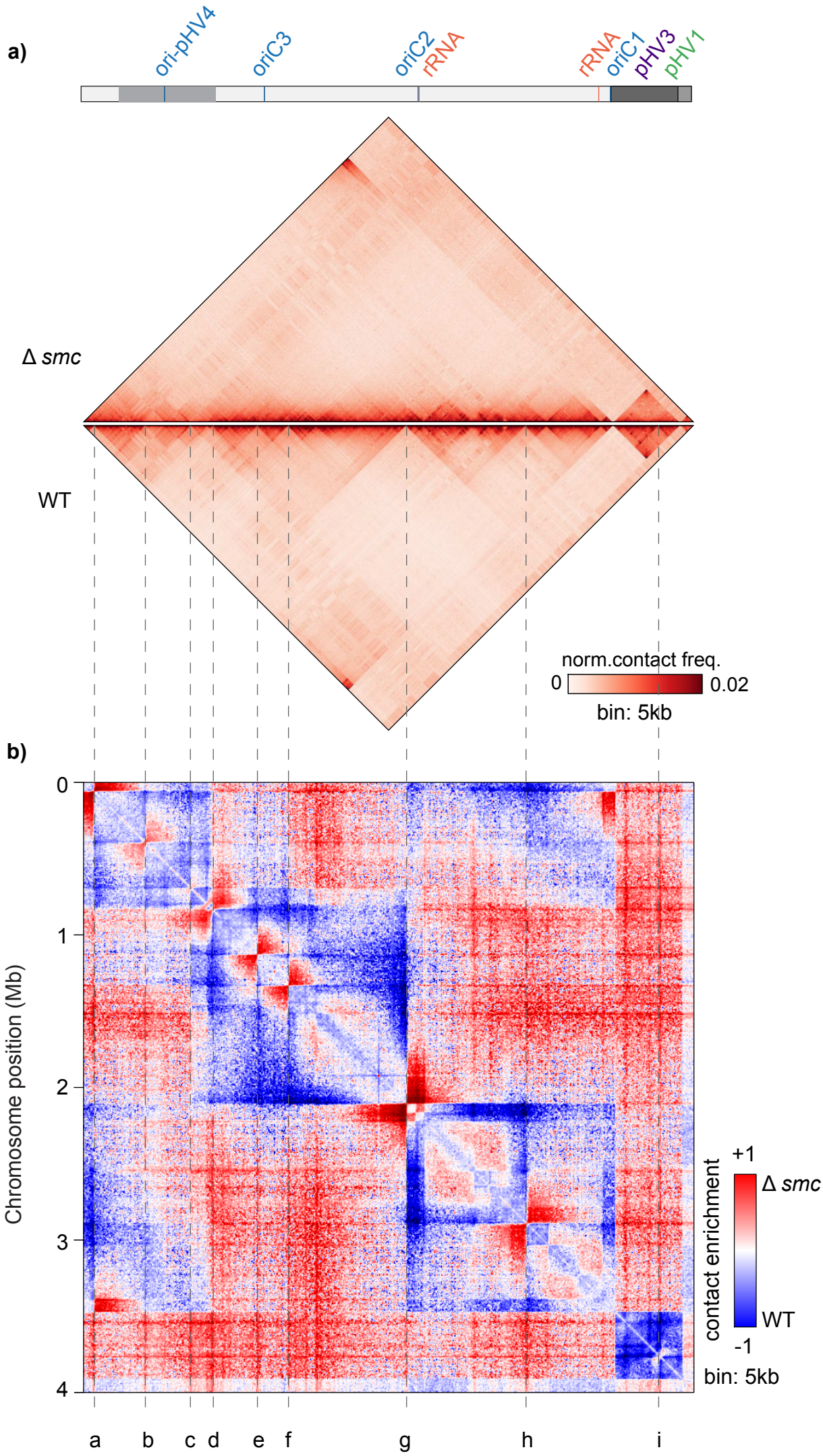


Supplementary Figure 5





Supplementary Figure 6



c)

Border	Chromsome	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 log2 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 log2 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. acidcaldarius*

(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*[13] 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 Δ*ori* mutant (*oriC1-3* and *ori-pHV4* have been deleted). (g) Differential contact map corresponding to the log2 ratio of Hi-C interactions between WT and Δ *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[28] 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 Δ*smc* mutant cells

(a) Normalized Hi-C contact maps of asynchronously growing populations of WT and Δ*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 log2 ratio of Hi-C interactions between WT and Δ*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 Δ*smc* mutant are highlighted by dashed lines. (c) Table showing the location and approximate genomic coordinates of the borders lost in Δ*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		
E. coli	CC328	1×10^8	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	$1\text{-}2 \times 10^9$	HpaII	24312	122	5 kb	-	88615212	93	7	0	6%	[8]
V. cholera	CC340	1×10^8	HpaII	7713	318	1 kb	+	229263798	73	22	5	20%	This study
	SRR3180994	1×10^9				5 kb	-	170426044	97	3	0	1%	[16]
C. crescentus	SRR824843	2.5×10^8	BglII	701	3711	10 kb	+	34822386	30	70	0	37%	[7]
	SRR824846		NcoI	2026	1299		+	47673122	32	68	0	48%	[7]
H. volcanii	CC363	1×10^8	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
S. acidocaldarius	SRR8699877	4×10^8	HindIII	808	1838	30 kb	+	117648700	47	53	0	10%	[13]
S. islandicus	SRR8699887			1179	1477		+	139639604	23	77	0	42%	[13]

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

Strains

Name	Genotype	Source
E.coli K12 MG1655	*F- lambda- ilvG- rfb-50 rph-1*	[37]
V. cholera N16961	Vibrio cholera O1 El Tor N16961	[38]
H.volcanii (wild-type)	H. volcanii H26 *ΔpyrE2*	[39]
H. volcanii Δ*mre11 rad50* (H161)	H26 *ΔpyrE2 Δmre11 Δrad50*	[29]
H. volcanii Δ*ori* (H1546)	H26 *ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC2 ΔoriC3 Δori-pHV4::trpA*	[22]
H26 Δ*smc* (HvRL138)	H26 *ΔpyrE2 Δsmc*	This study
XL1-Blue MRF'	*ΔmcrA183 ΔmcrCB-hsdSMR-mrr173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac [F' proAB lacIqZΔM15 Tn10*	Agilent technologies
GM121	F^{-}*dam*$^{-}$*3 dcm*$^{-}$*6 ara*$^{-}$*14 fhuA31 galK2 galT22 hdsR3 lacY1 leu-6 thi-1 thr-1 tsx-78*	[50]

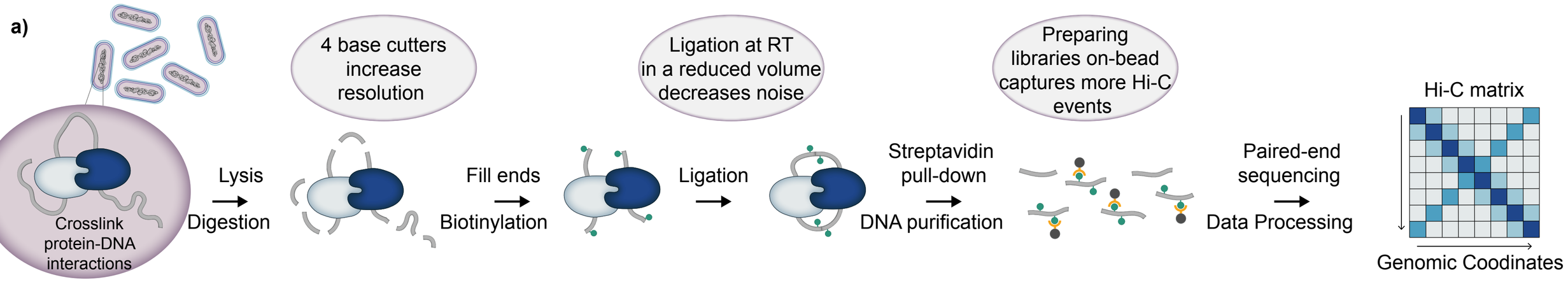
Plasmids

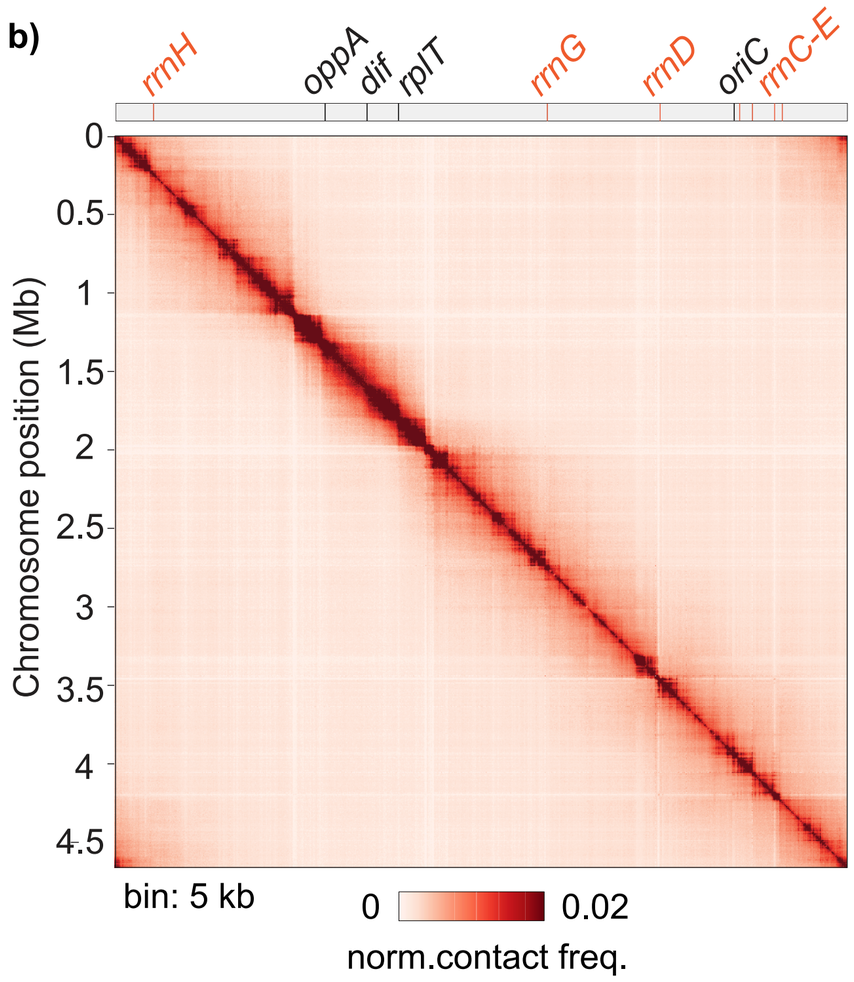
Name	Genotype	Source
pTA131	pBluescript II containing *pyrE2* under ferredoxin promoter	[39]
pRL93	PTA131 Δ*smc*	This study

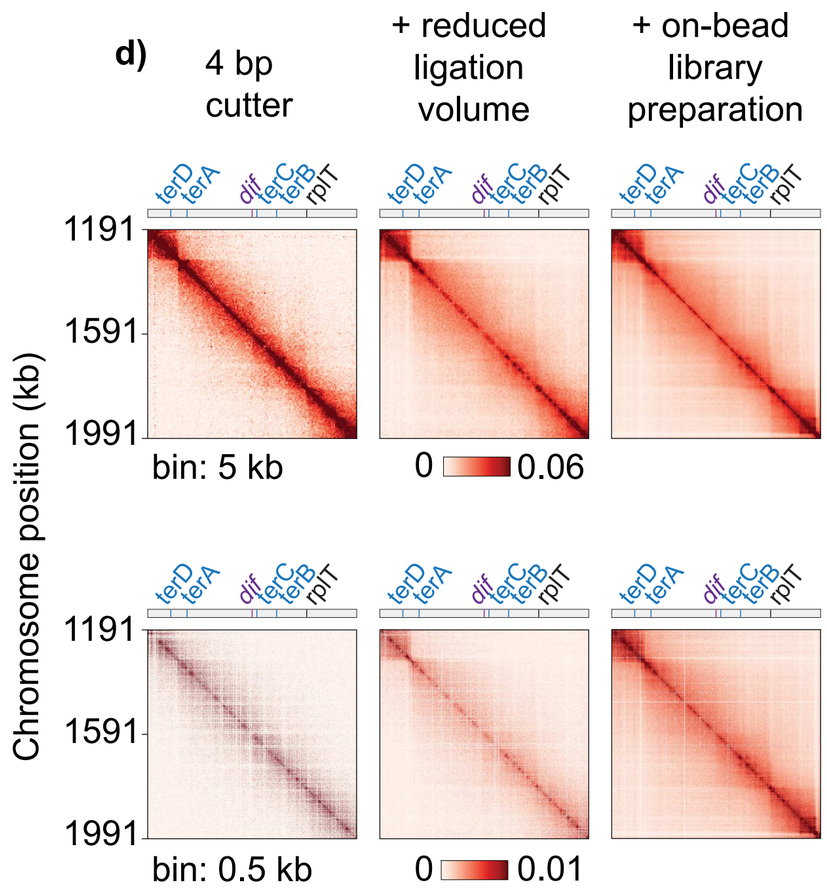
Primers

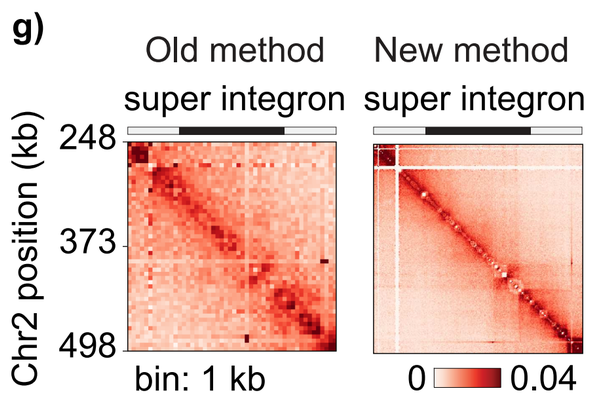
Name	Sequence (5' - 3')	Source
RL292	AGGTCGACGGTATCGATAAGCTTGATATCGGACGCACATTTATGTCTCACG	This study
RL294	TGCGGGCGTCGCCCGCGTGAACACCAACCCATGACTGACGACATCCCCGAAC	This study
RL295	AAAAGCTGGAGCTCCACCGCGGTGGCGGCCTCCAGCGCGTCGATTGGGTCG	This study
RL296	GGGTTGGTGTTCACGCGGGCG	This study
pBSF2	TTAAGTTGGGTAACGCCAGGG	This study
pBSF3	ACCCCAGGCTTTACACTTTATGC	This study
RL305	GCCGTCACGACCATCCTCAAC	This study
RL310	GTTGTCGGGCTCGCGCGGCAG	This study

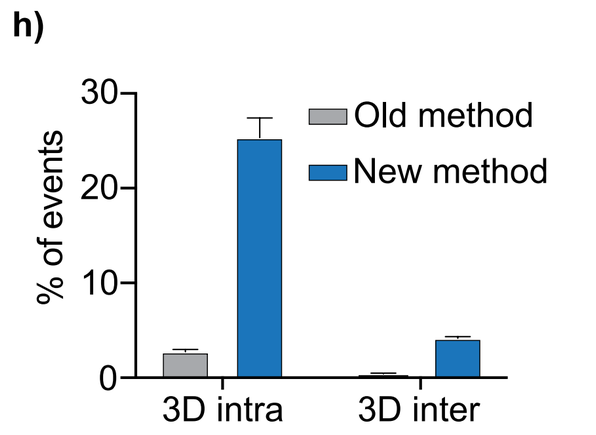
Figure 1

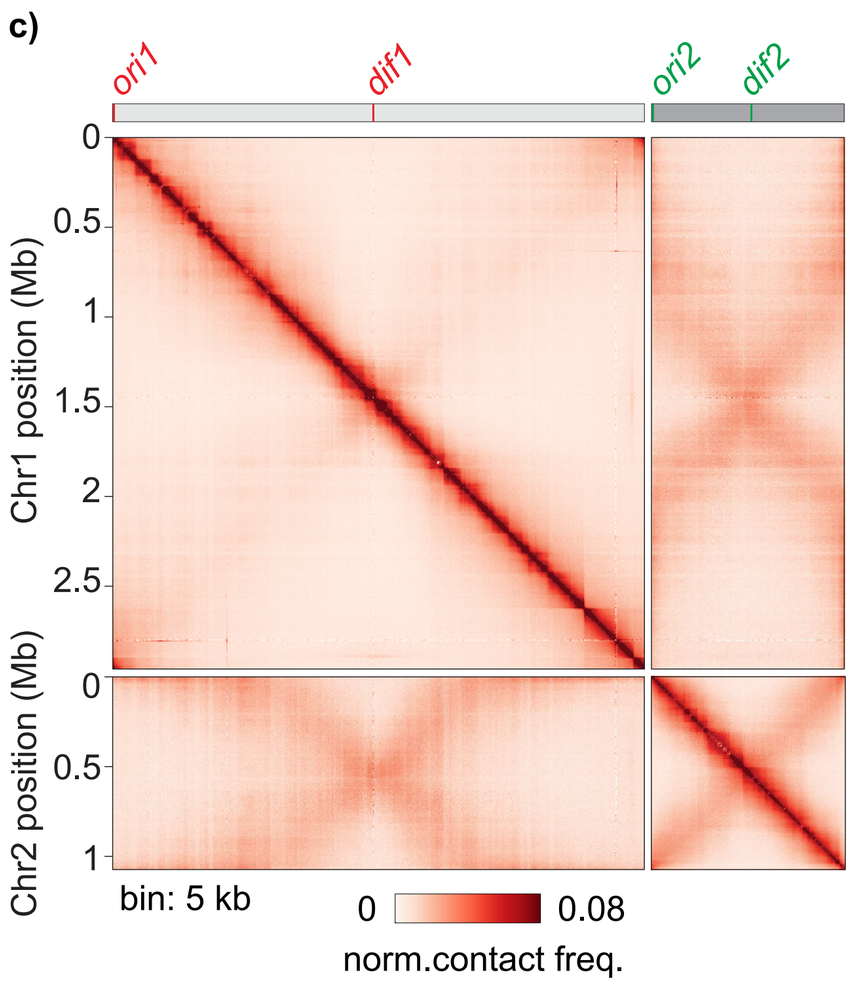

a)
Crosslink protein-DNA interactions
Lysis
Digestion
4 base cutters increase resolution
Fill ends
Biotinylation
Ligation at RT in a reduced volume decreases noise
Ligation
Streptavidin pull-down
DNA purification
Preparing libraries on-bead captures more Hi-C events
Paired-end sequencing
Data Processing
Hi-C matrix
Genomic Coodinates


b)
rrnH
oppA
dif
rplT
rrnG
rrnD
oriC
rrnC-E
Chromosome position (Mb)
bin: 5 kb
0
0.02
norm.contact freq.

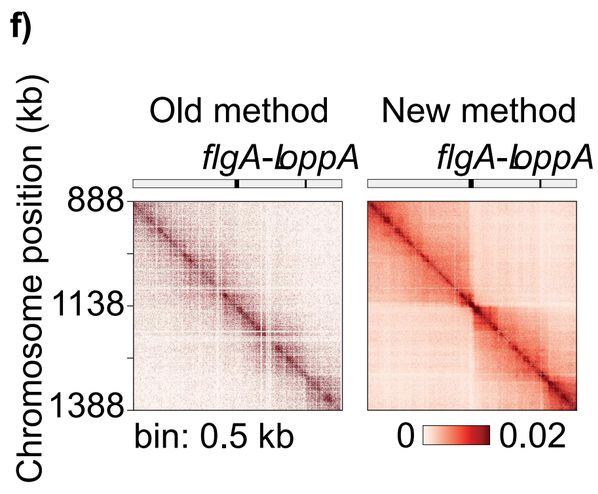

d)
4 bp cutter
+ reduced ligation volume
+ on-bead library preparation
terD
terA
dif
terC
terB
rplT
Chromosome position (kb)
1191
1591
1991
bin: 5 kb
0
0.06
bin: 0.5 kb
0.01


g)
Old method super integron
New method super integron
Chr2 position (kb)
248
373
498
bin: 1 kb
0
0.04


h)
% of events
30
20
10
0
Old method
New method
3D intra
3D inter


c)
ori1
dif1
ori2
dif2
Chr1 position (Mb)
Chr2 position (Mb)
bin: 5 kb
0
0.08
norm.contact freq.


e)
uninform.
inform. reads
% of events
100
50
0
4 bp cutter
+ red. lig. vol
+ on bead


f)
Old method
New method
flgA-loppA
Chromosome position (kb)
888
1138
1388
bin: 0.5 kb
0
0.02

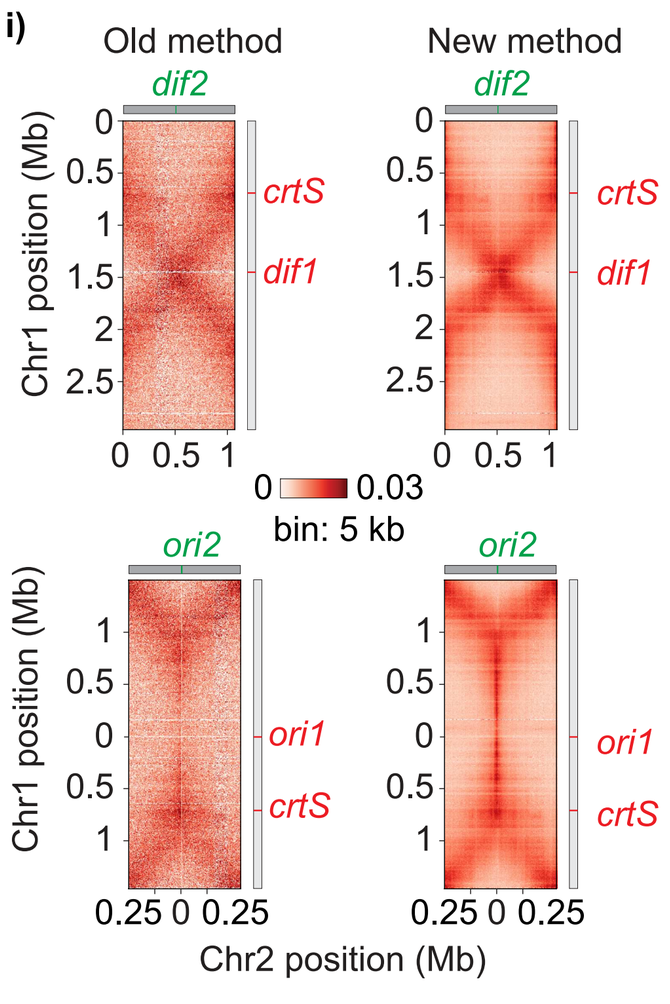

i)
Old method
New method
dif2
crtS
dif1
Chr1 position (Mb)
0
0.03
bin: 5 kb
ori2
ori1
crtS
Chr2 position (Mb)

Figure 2

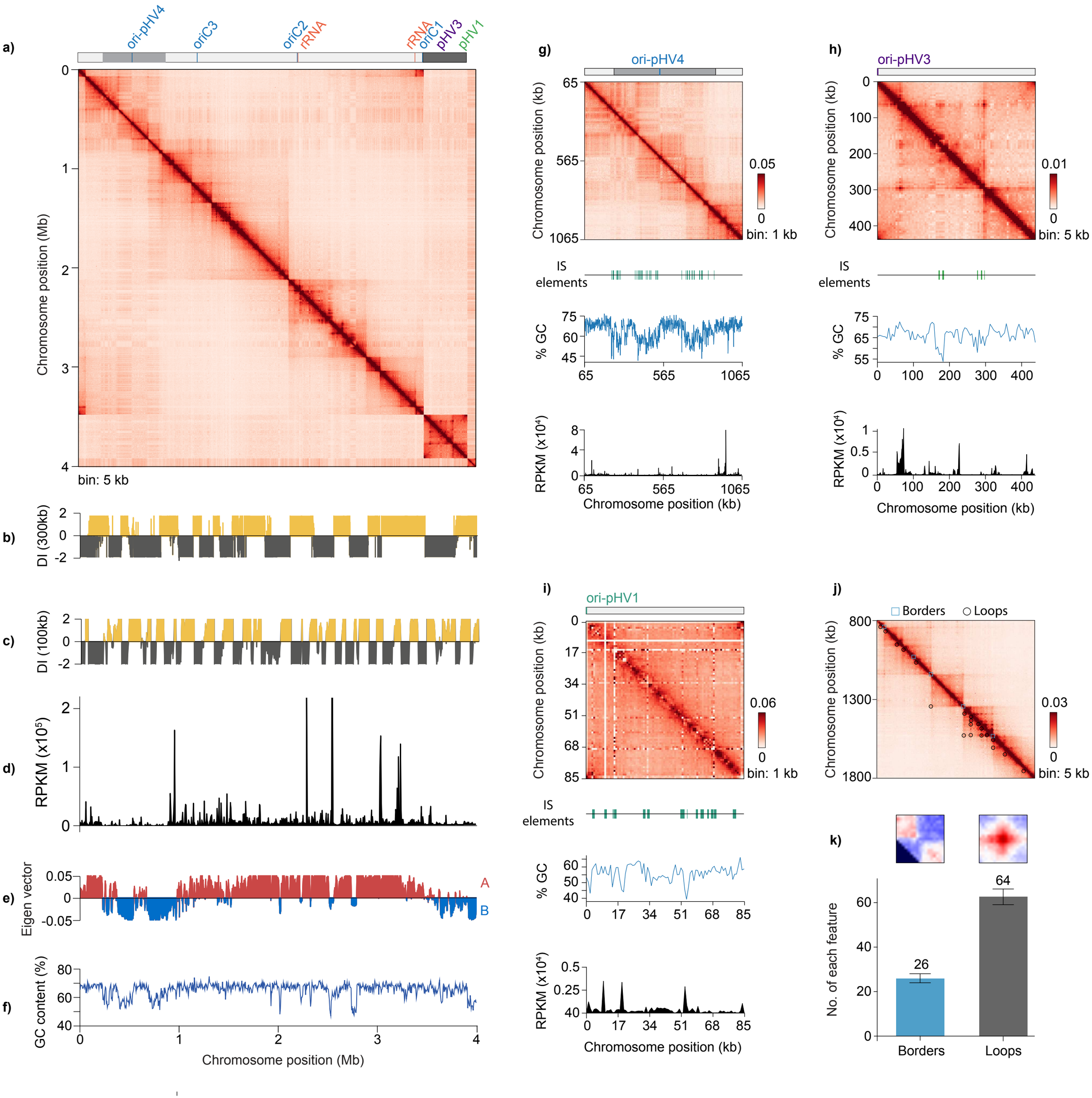
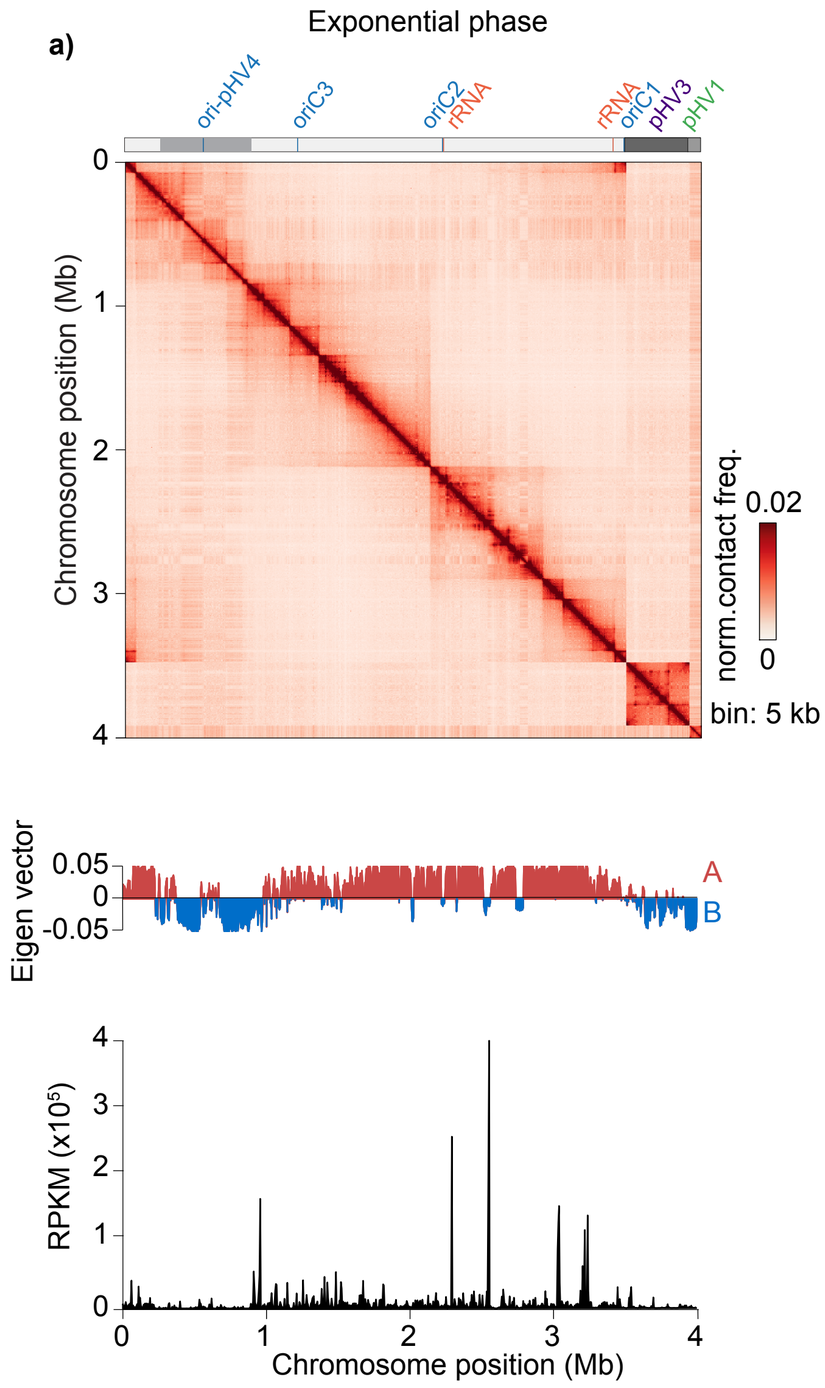
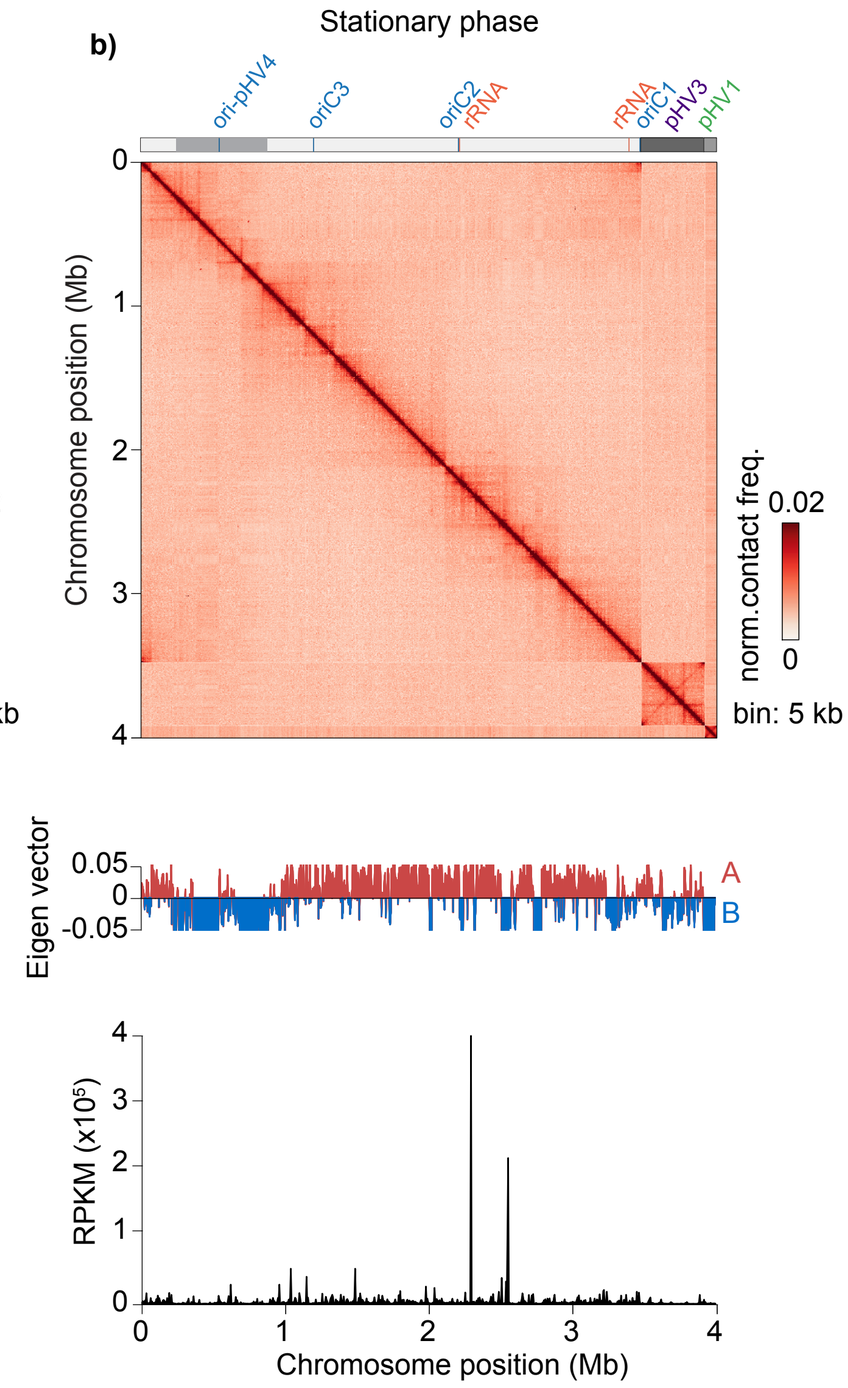
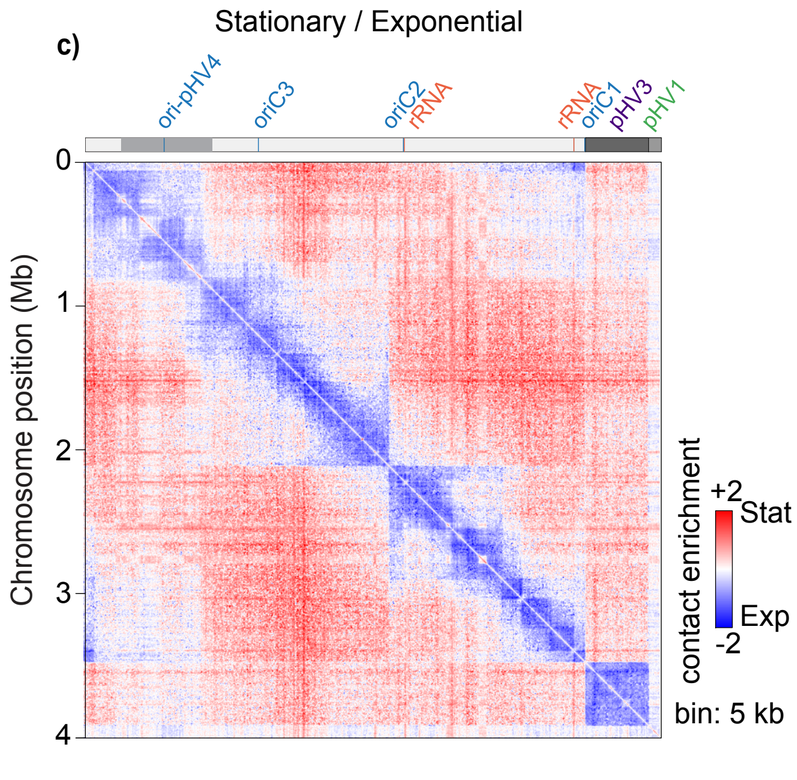
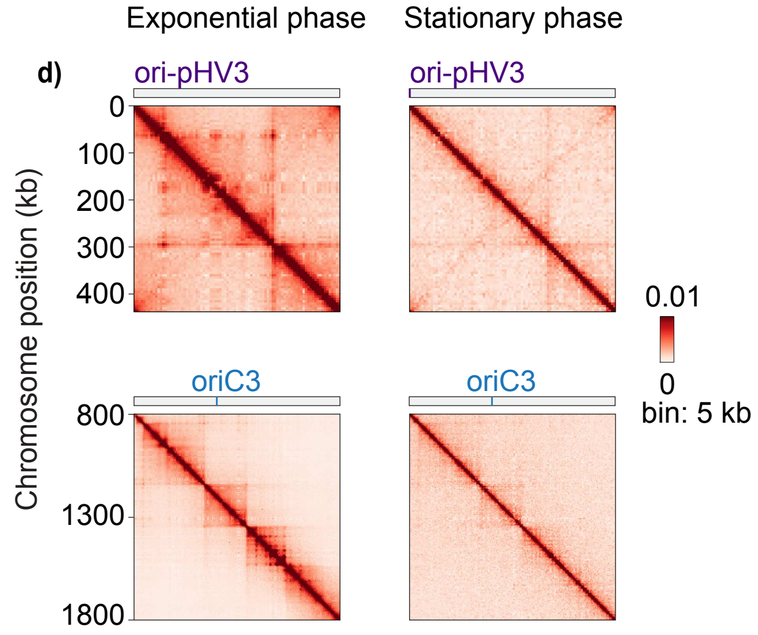

a)
ori-pHV4
oriC3
oriC2
rRNA
rRNA
oriC1
pHV3
pHV1
Chromosome position (Mb)
bin: 5 kb
b)
DI (300kb)
c)
DI (100kb)
d)
RPKM (x10⁵)
e)
Eigen vector
A
B
f)
GC content (%)
Chromosome position (Mb)
g)
ori-pHV4
Chromosome position (kb)
0.05
bin: 1 kb
IS elements
% GC
RPKM (x10⁴)
Chromosome position (kb)
h)
ori-pHV3
Chromosome position (kb)
0.01
bin: 5 kb
IS elements
% GC
RPKM (x10⁴)
Chromosome position (kb)
i)
ori-pHV1
Chromosome position (kb)
0.06
bin: 1 kb
IS elements
% GC
RPKM (x10⁴)
Chromosome position (kb)
j)
Borders
Loops
Chromosome position (kb)
0.03
bin: 5 kb
k)
No. of each feature
26
64
Borders
Loops

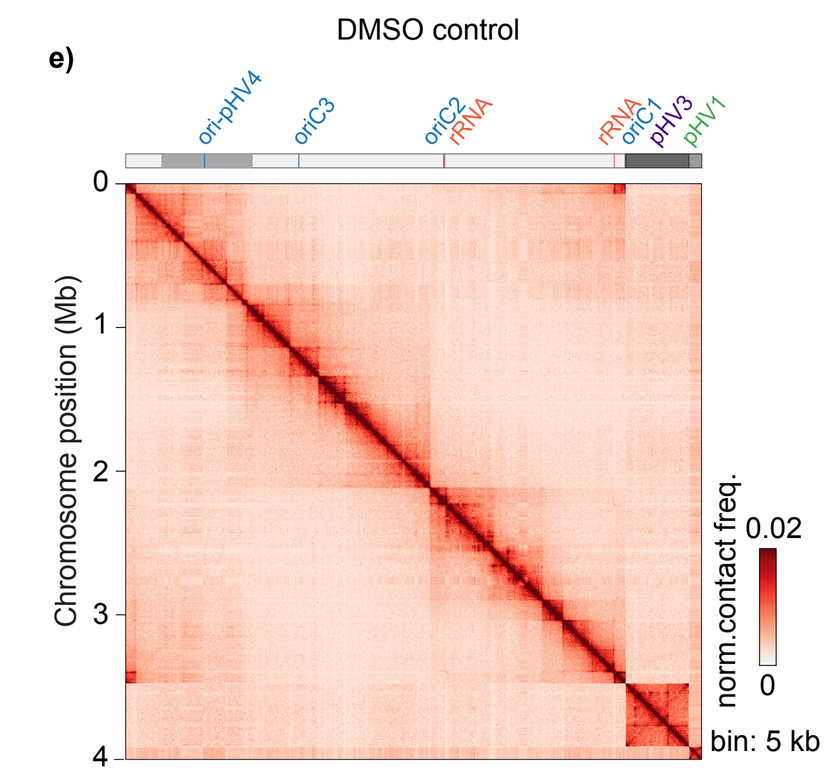
Figure 3

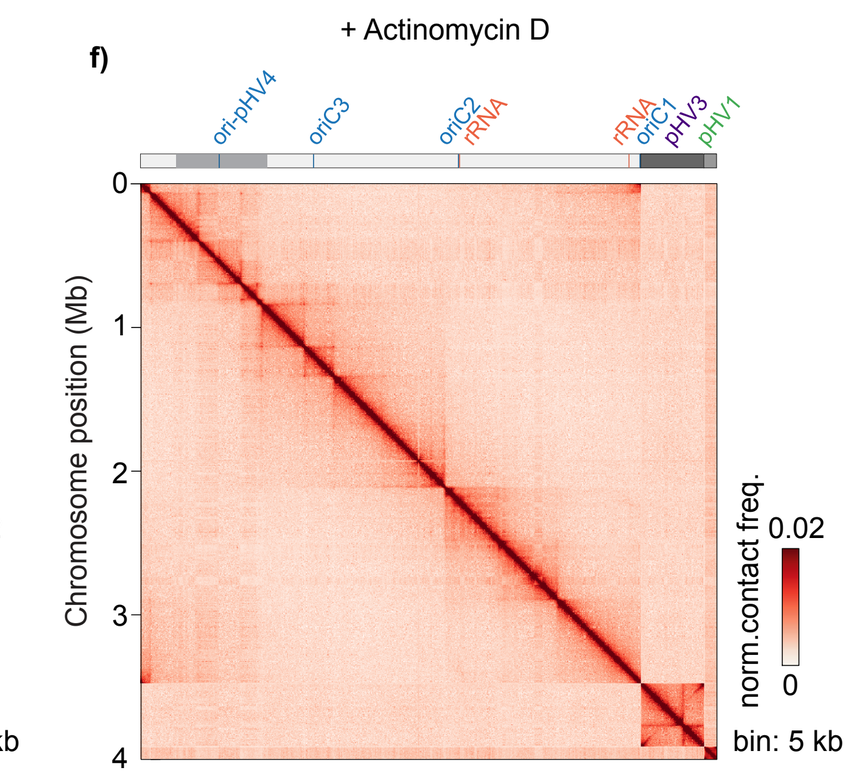

a)
Exponential phase
ori-pHV4
oriC3
oriC2
rRNA
rRNA
oriC1
pHV3
pHV1
Chromosome position (Mb)
norm.contact freq.
0.02
0
bin: 5 kb
Eigen vector
A
B
RPKM (x10^5)
Chromosome position (Mb)

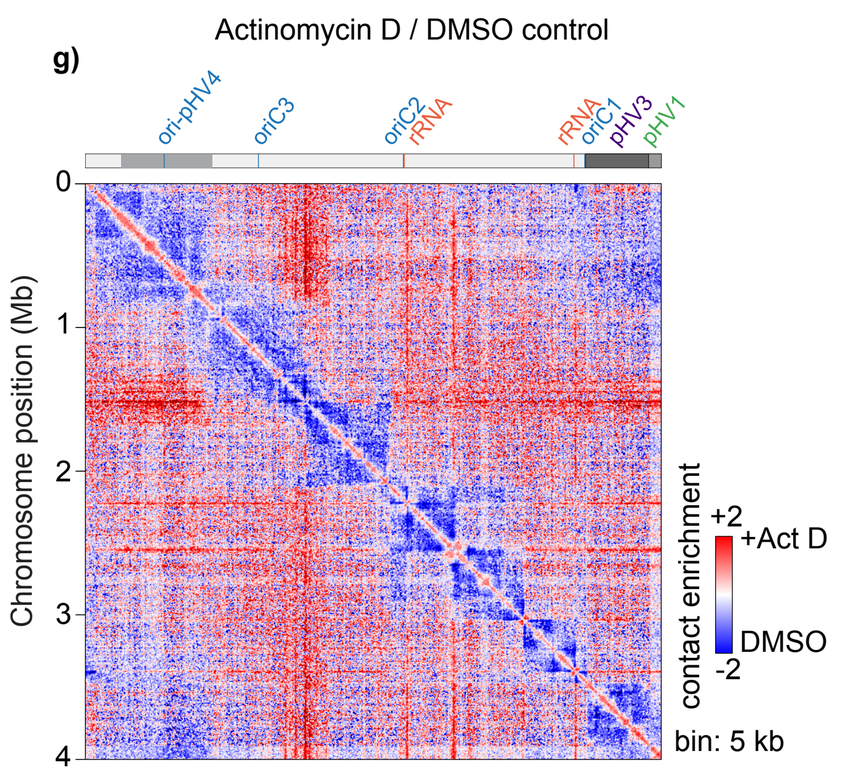

b)
Stationary phase
ori-pHV4
oriC3
oriC2
rRNA
rRNA
oriC1
pHV3
pHV1
Chromosome position (Mb)
norm.contact freq.
0.02
0
bin: 5 kb
Eigen vector
A
B
RPKM (x10^5)
Chromosome position (Mb)

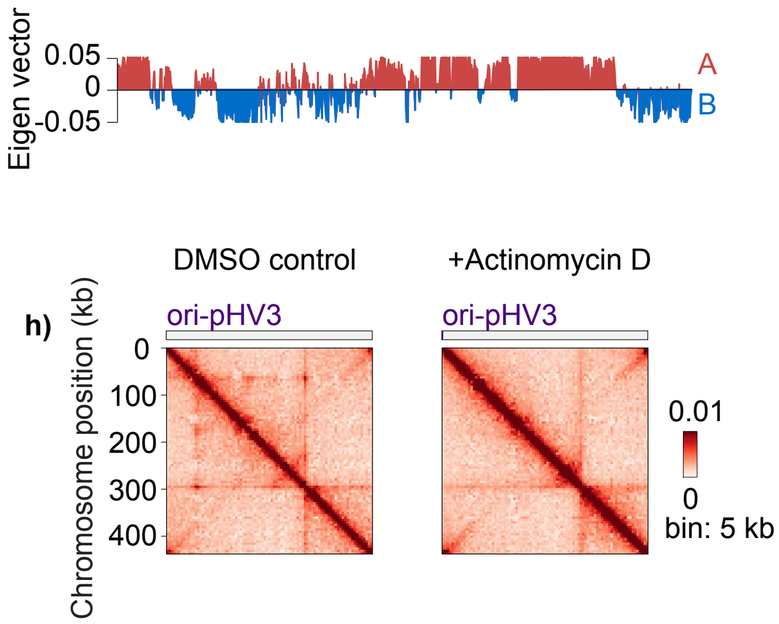

c)
Stationary / Exponential
ori-pHV4
oriC3
oriC2
rRNA
rRNA
oriC1
pHV3
pHV1
Chromosome position (Mb)
contact enrichment
+2
Stat
Exp
-2
bin: 5 kb

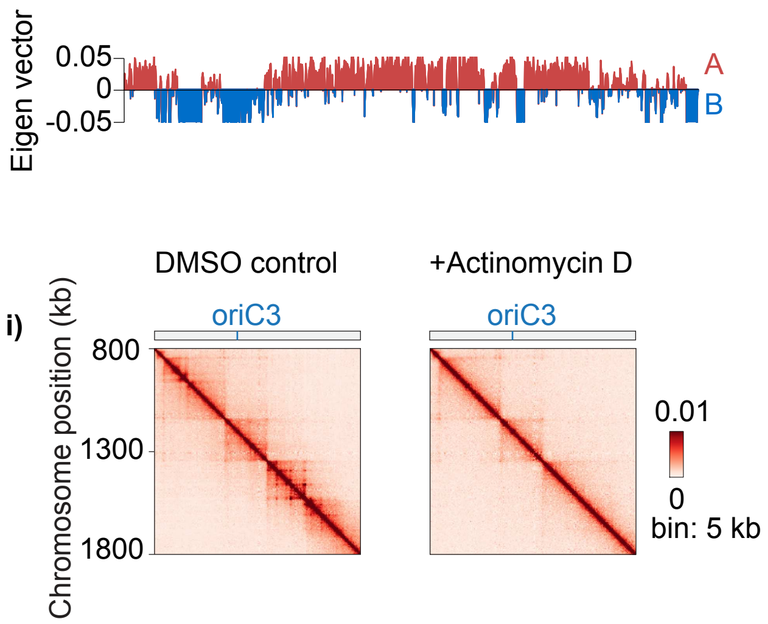

Exponential phase
Stationary phase
d)
ori-pHV3
ori-pHV3
oriC3
oriC3
Chromosome position (kb)
0.01
0
bin: 5 kb


e)
DMSO control
ori-pHV4
oriC3
oriC2
rRNA
rRNA
oriC1
pHV3
pHV1
Chromosome position (Mb)
norm.contact freq.
0.02
0
bin: 5 kb
Eigen vector
A
B


f)
+ Actinomycin D
ori-pHV4
oriC3
oriC2
rRNA
rRNA
oriC1
pHV3
pHV1
Chromosome position (Mb)
norm.contact freq.
0.02
0
bin: 5 kb
Eigen vector
A
B


g)
Actinomycin D / DMSO control
ori-pHV4
oriC3
oriC2
rRNA
rRNA
oriC1
pHV3
pHV1
Chromosome position (Mb)
contact enrichment
+2
+Act D
DMSO
-2
bin: 5 kb


h)
DMSO control
+Actinomycin D
ori-pHV3
ori-pHV3
Chromosome position (kb)
0.01
0
bin: 5 kb


i)
DMSO control
+Actinomycin D
oriC3
oriC3
Chromosome position (kb)
0.01
0
bin: 5 kb

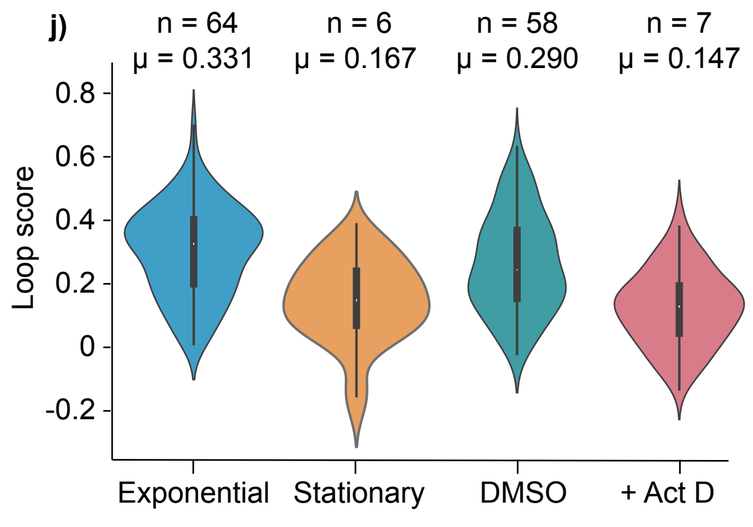
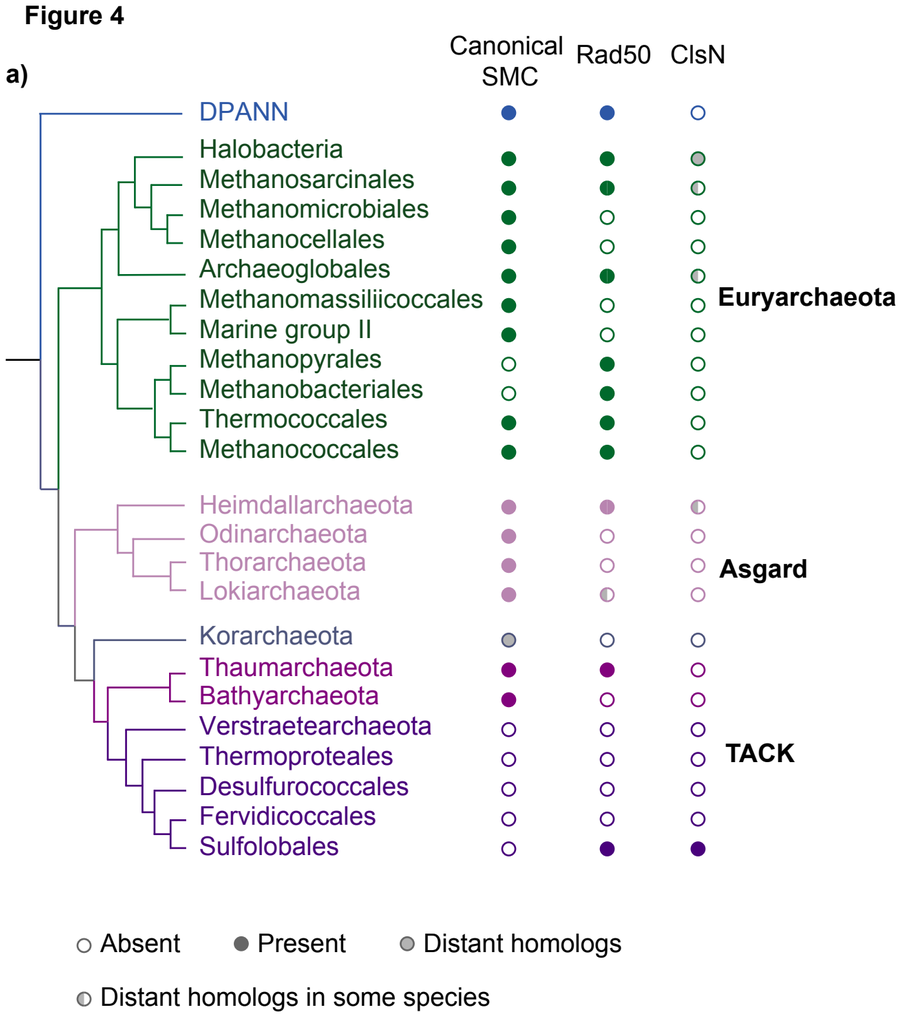
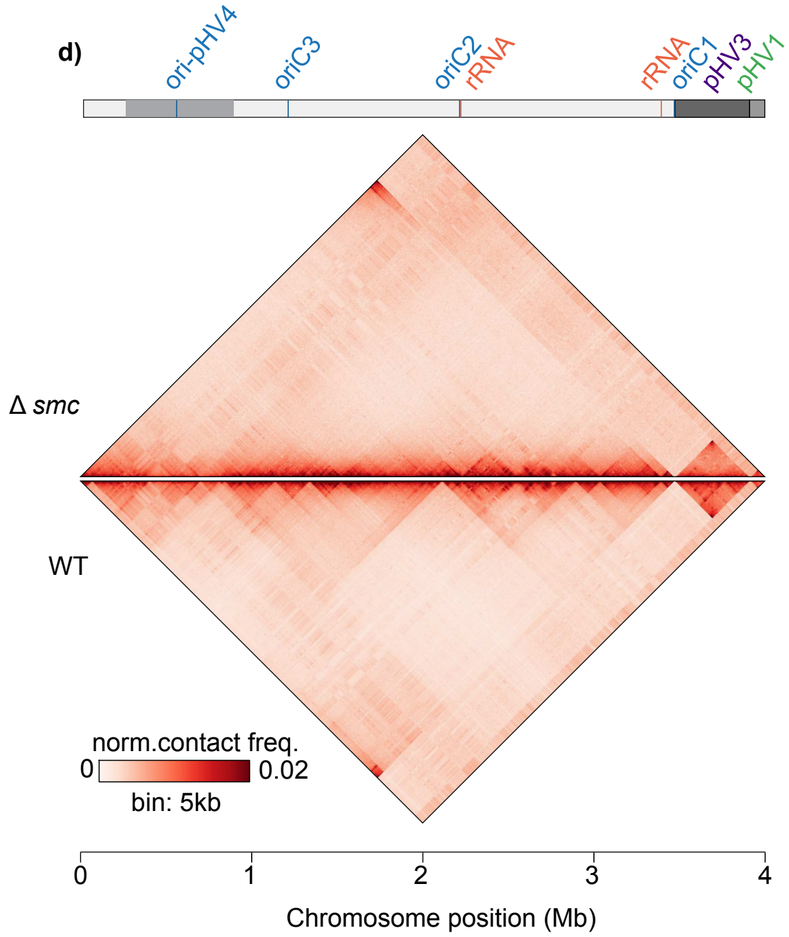
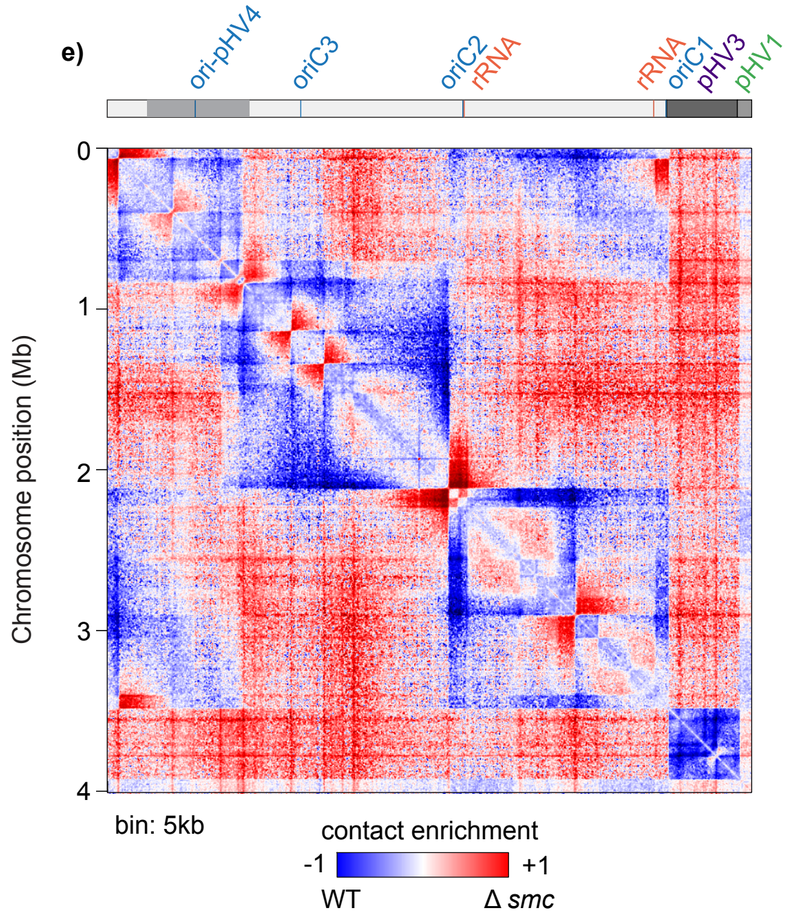
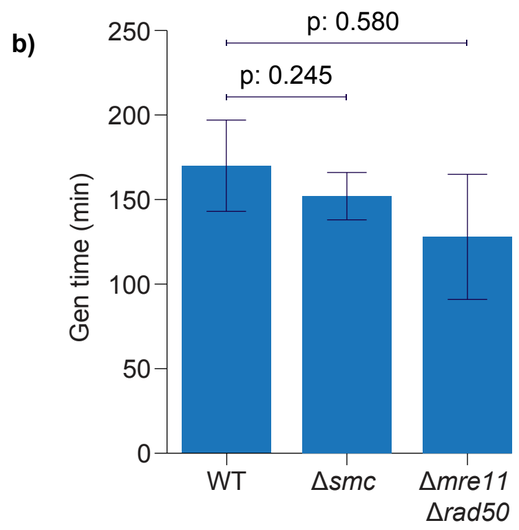

j)
n = 64
μ = 0.331
n = 6
μ = 0.167
n = 58
μ = 0.290
n = 7
μ = 0.147
Loop score
Exponential
Stationary
DMSO
+ Act D

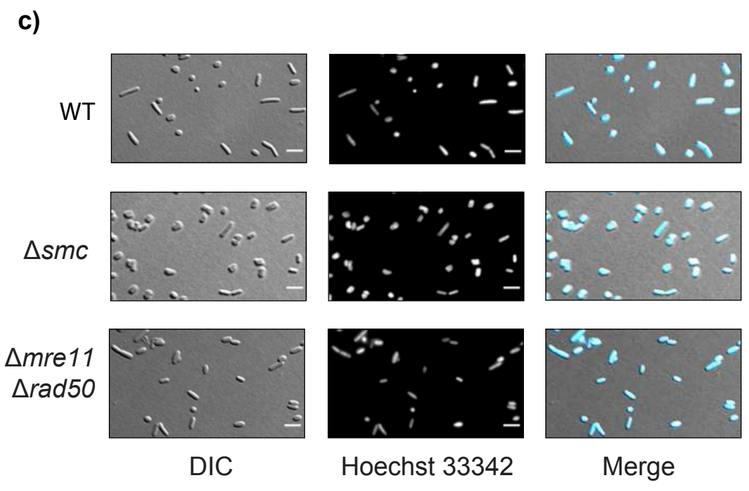
Figure 4

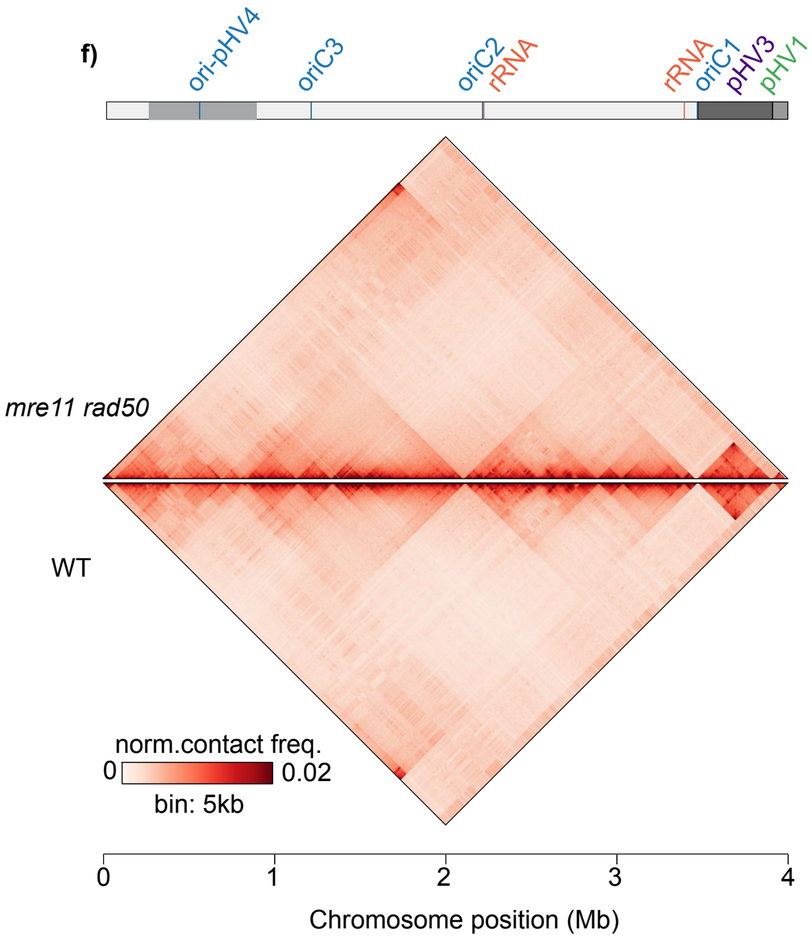

a)
Canonical SMC
Rad50
ClsN
DPANN
Halobacteria
Methanosarcinales
Methanomicrobiales
Methanocellales
Archaeoglobales
Methanomassiliicoccales
Marine group II
Methanopyrales
Methanobacteriales
Thermococcales
Methanococcales
Euryarchaeota
Heimdallarchaeota
Odinarchaeota
Thorarchaeota
Lokiarchaeota
Asgard
Korarchaeota
Thaumarchaeota
Bathyarchaeota
Verstraetearchaeota
Thermoproteales
Desulfurococcales
Fervidicoccales
Sulfolobales
TACK
Absent
Present
Distant homologs
Distant homologs in some species

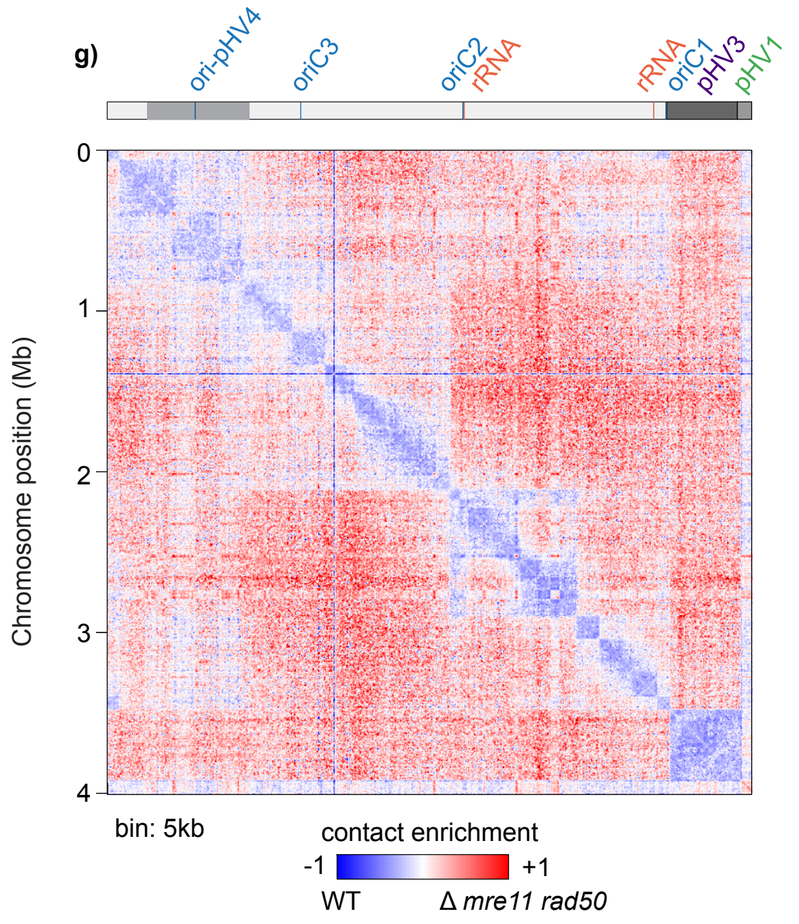

d)
ori-pHV4
oriC3
oriC2
rRNA
rRNA
oriC1
pHV3
pHV1
Δ smc
WT
norm.contact freq.
0
0.02
bin: 5kb
0
1
2
3
4
Chromosome position (Mb)

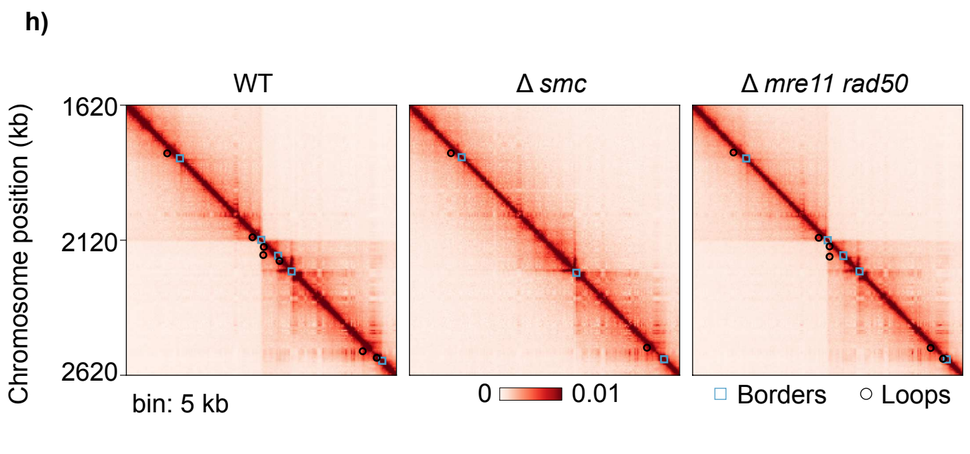

e)
ori-pHV4
oriC3
oriC2
rRNA
rRNA
oriC1
pHV3
pHV1
Chromosome position (Mb)
0
1
2
3
4
bin: 5kb
contact enrichment
-1
+1
WT
Δ smc

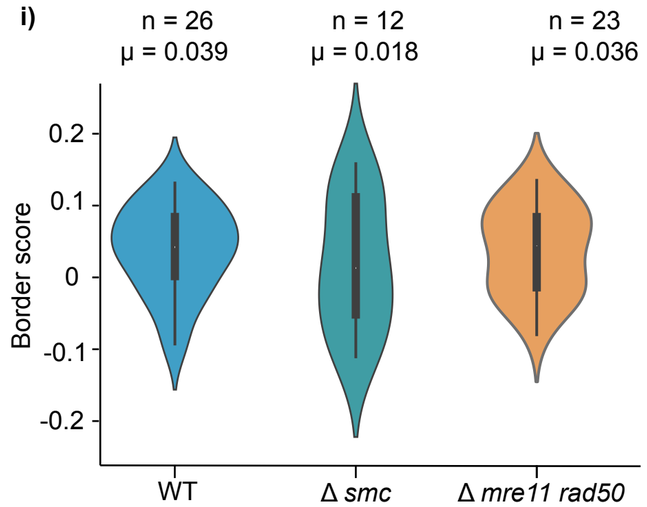

b)
p: 0.580
p: 0.245
Gen time (min)
250
200
150
100
50
0
WT
Δsmc
Δmre11 Δrad50

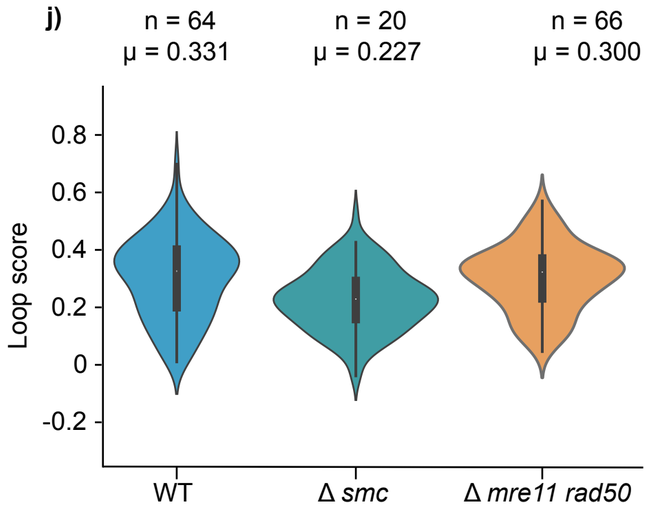

c)
WT
Δsmc
Δmre11 Δrad50
DIC
Hoechst 33342
Merge


f)
ori-pHV4
oriC3
oriC2
rRNA
rRNA
oriC1
pHV3
pHV1
Δ mre11 rad50
WT
norm.contact freq.
0
0.02
bin: 5kb
0
1
2
3
4
Chromosome position (Mb)

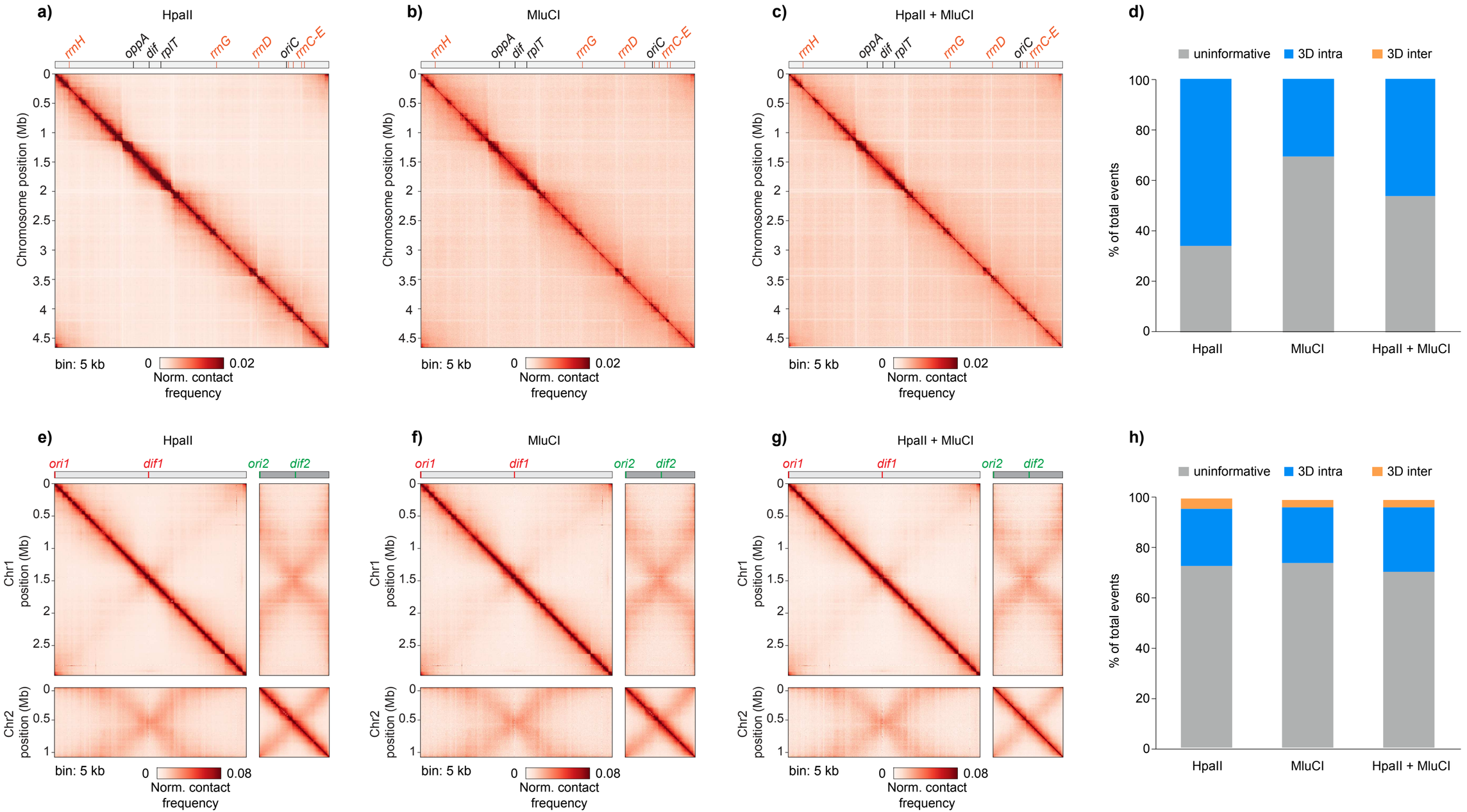

g)
ori-pHV4
oriC3
oriC2
rRNA
rRNA
oriC1
pHV3
pHV1
Chromosome position (Mb)
0
1
2
3
4
bin: 5kb
contact enrichment
-1
+1
WT
Δ mre11 rad50


h)
WT
Δ smc
Δ mre11 rad50
Chromosome position (kb)
1620
2120
2620
bin: 5 kb
0
0.01
Borders
Loops

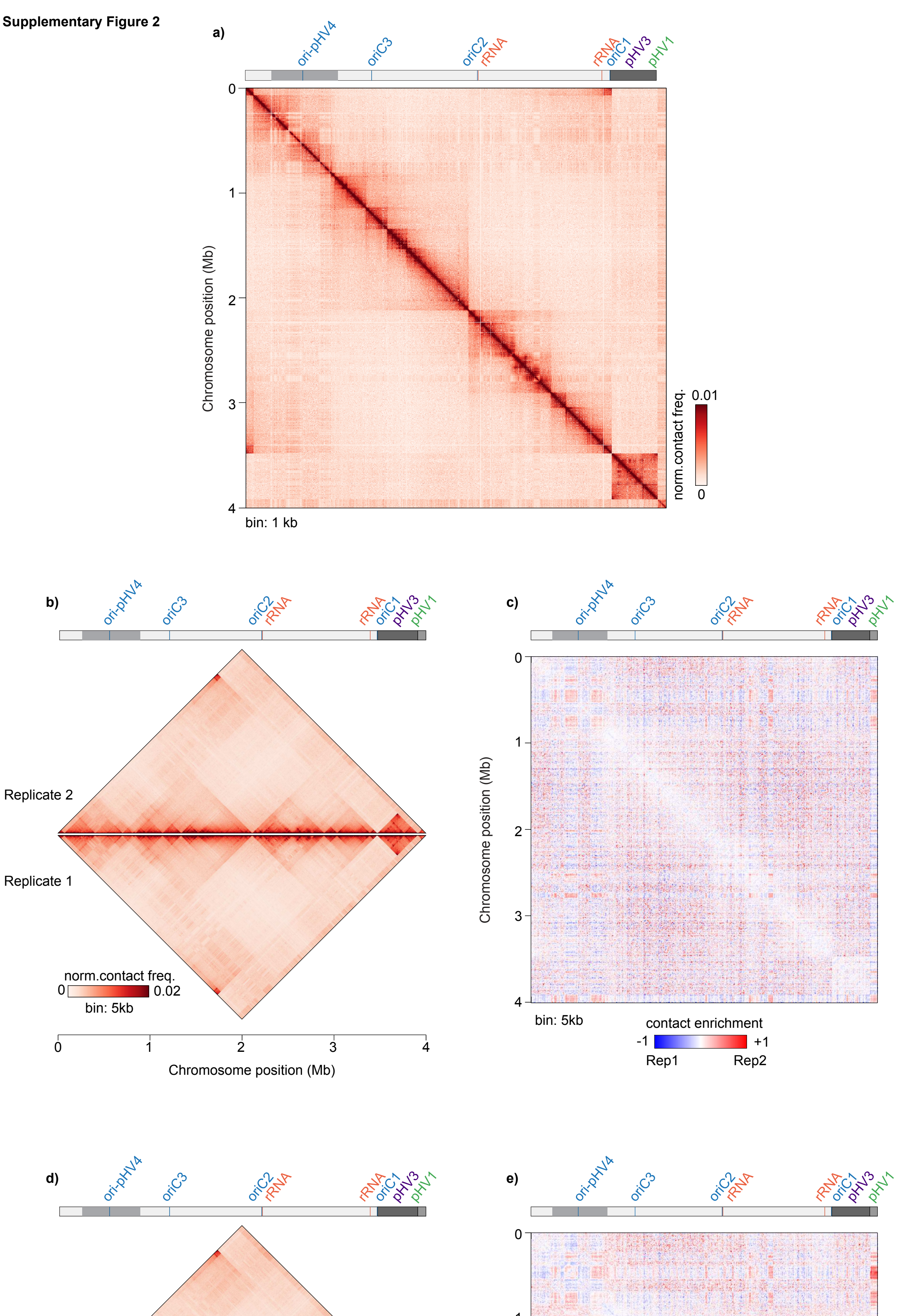

i)
n = 26
μ = 0.039
n = 12
μ = 0.018
n = 23
μ = 0.036
Border score
0.2
0.1
0
-0.1
-0.2
WT
Δ smc
Δ mre11 rad50


j)
n = 64
μ = 0.331
n = 20
μ = 0.227
n = 66
μ = 0.300
Loop score
0.8
0.6
0.4
0.2
0
-0.2
WT
Δ smc
Δ mre11 rad50

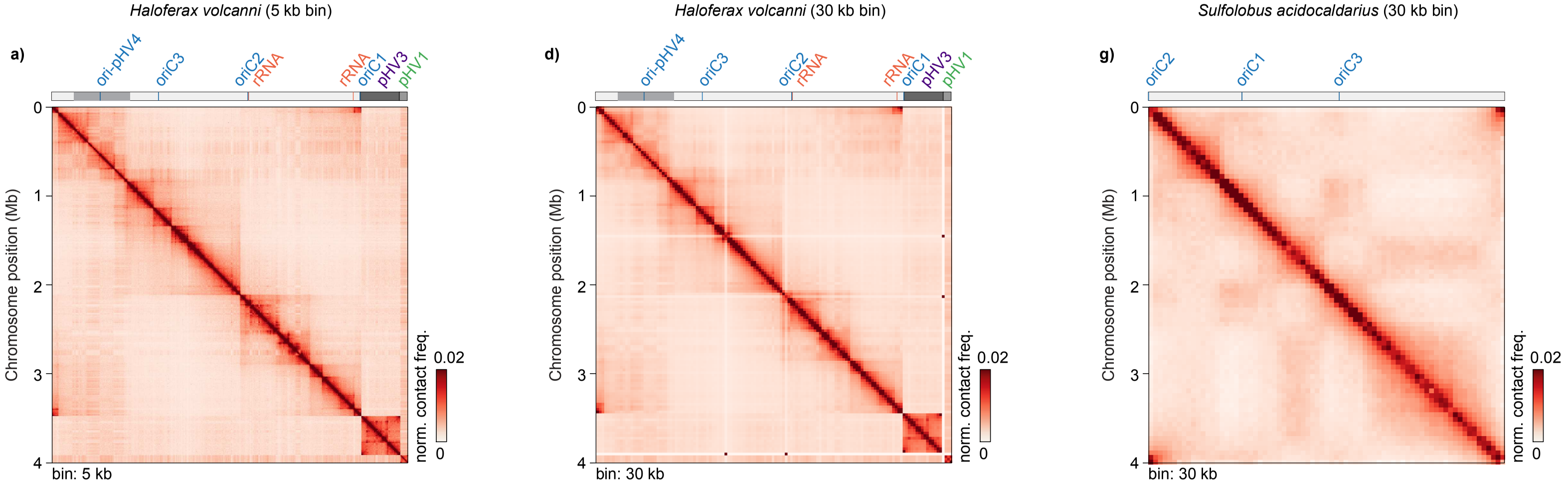
Supplementary Figure 1



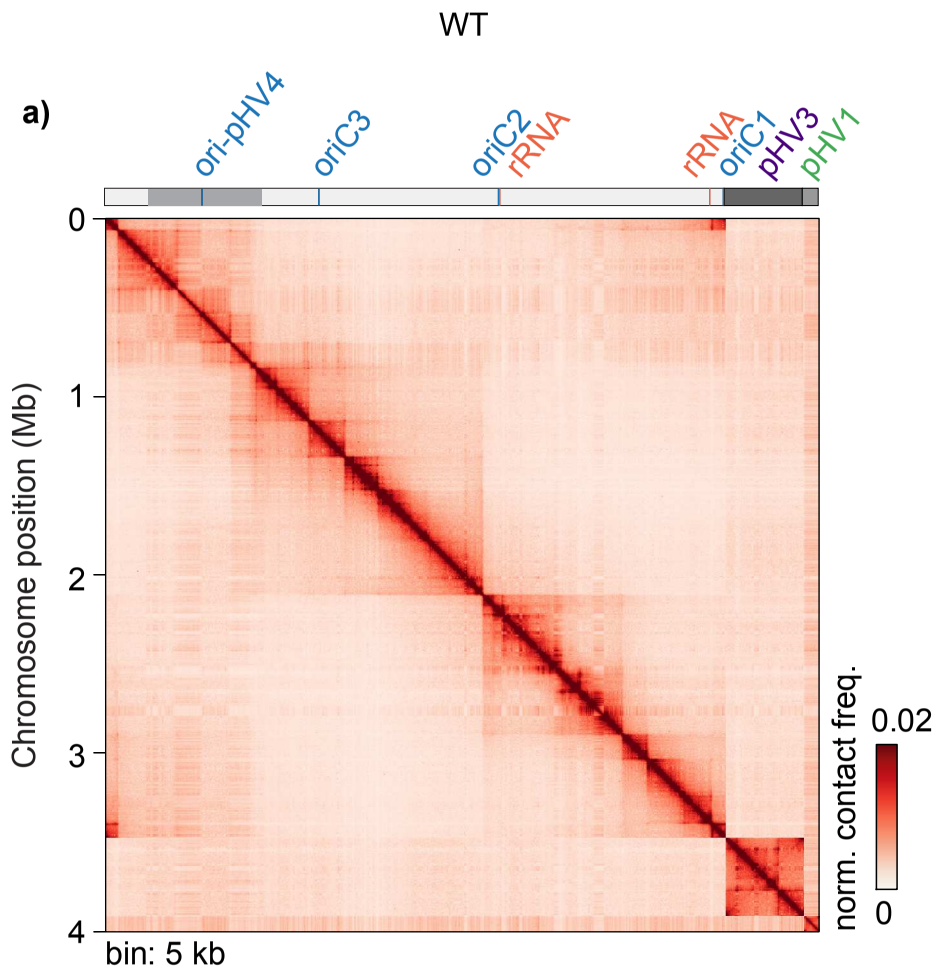
Supplementary Figure 2

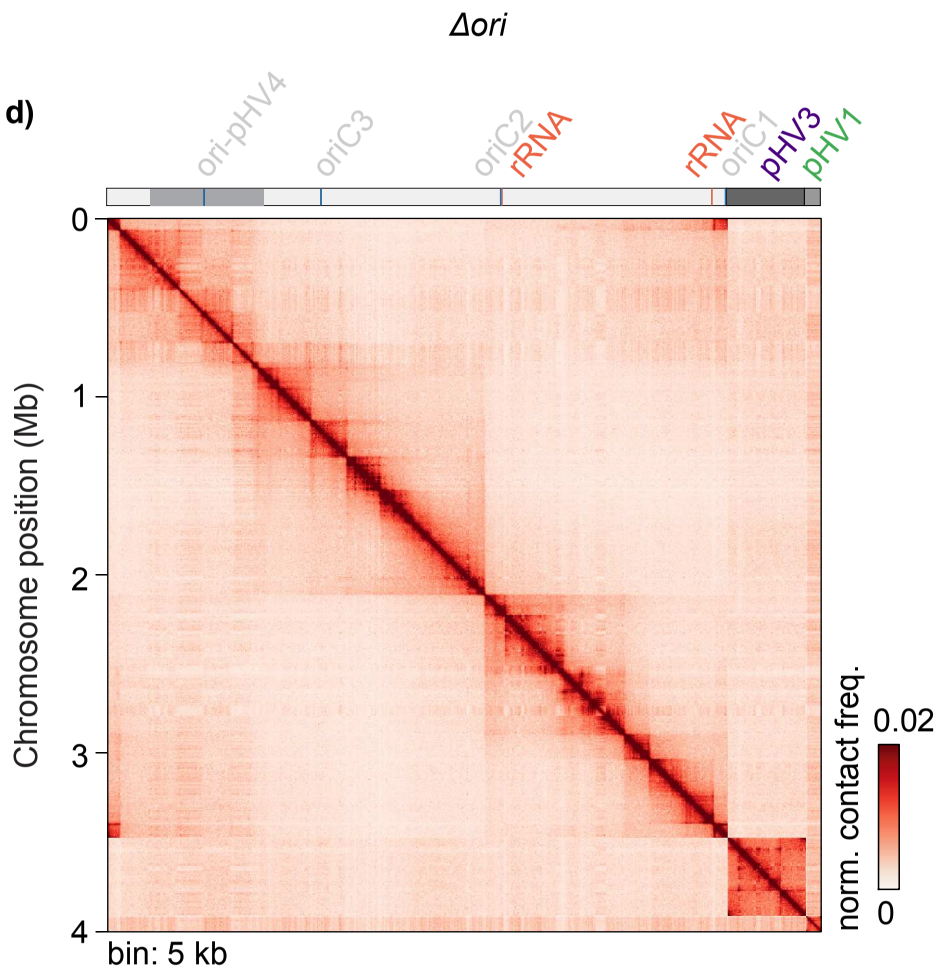


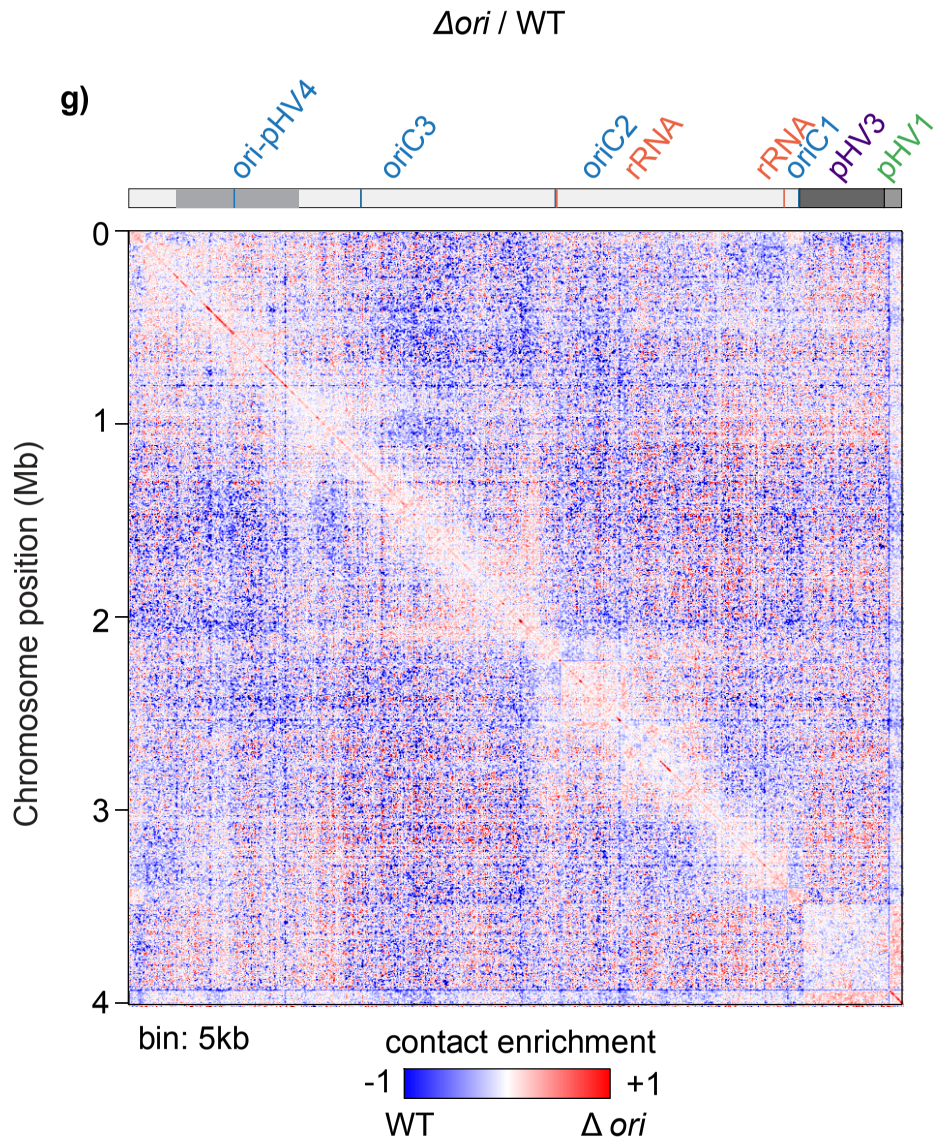
Supplementary Figure 3

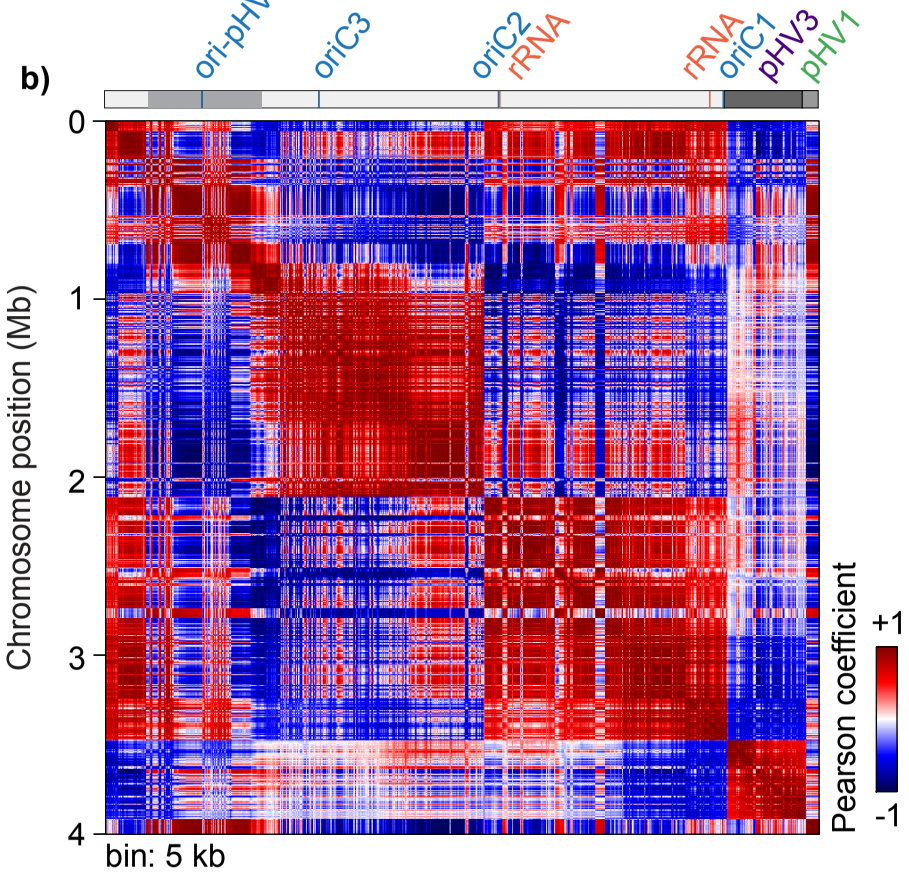


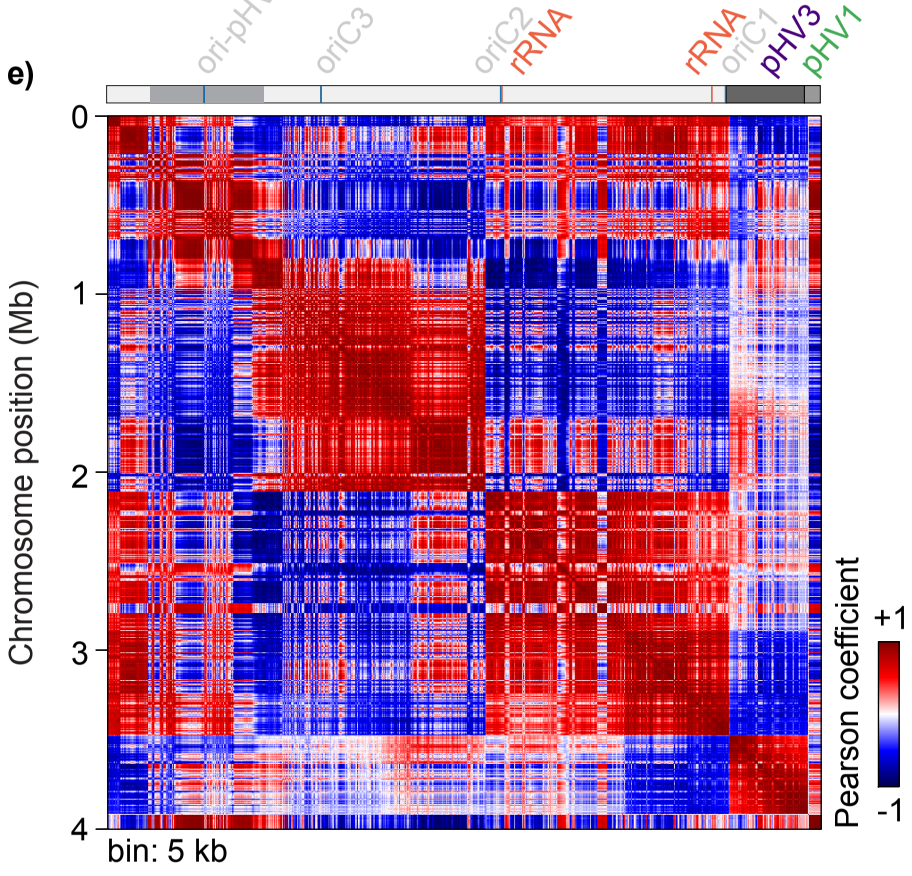
Supplementary Figure 4

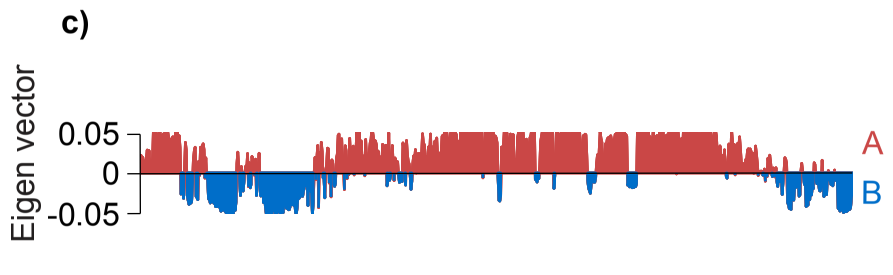


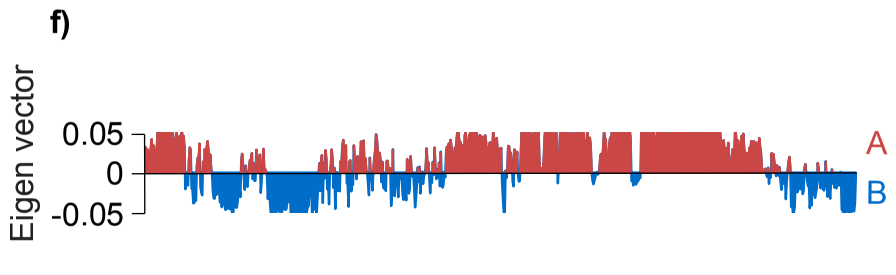




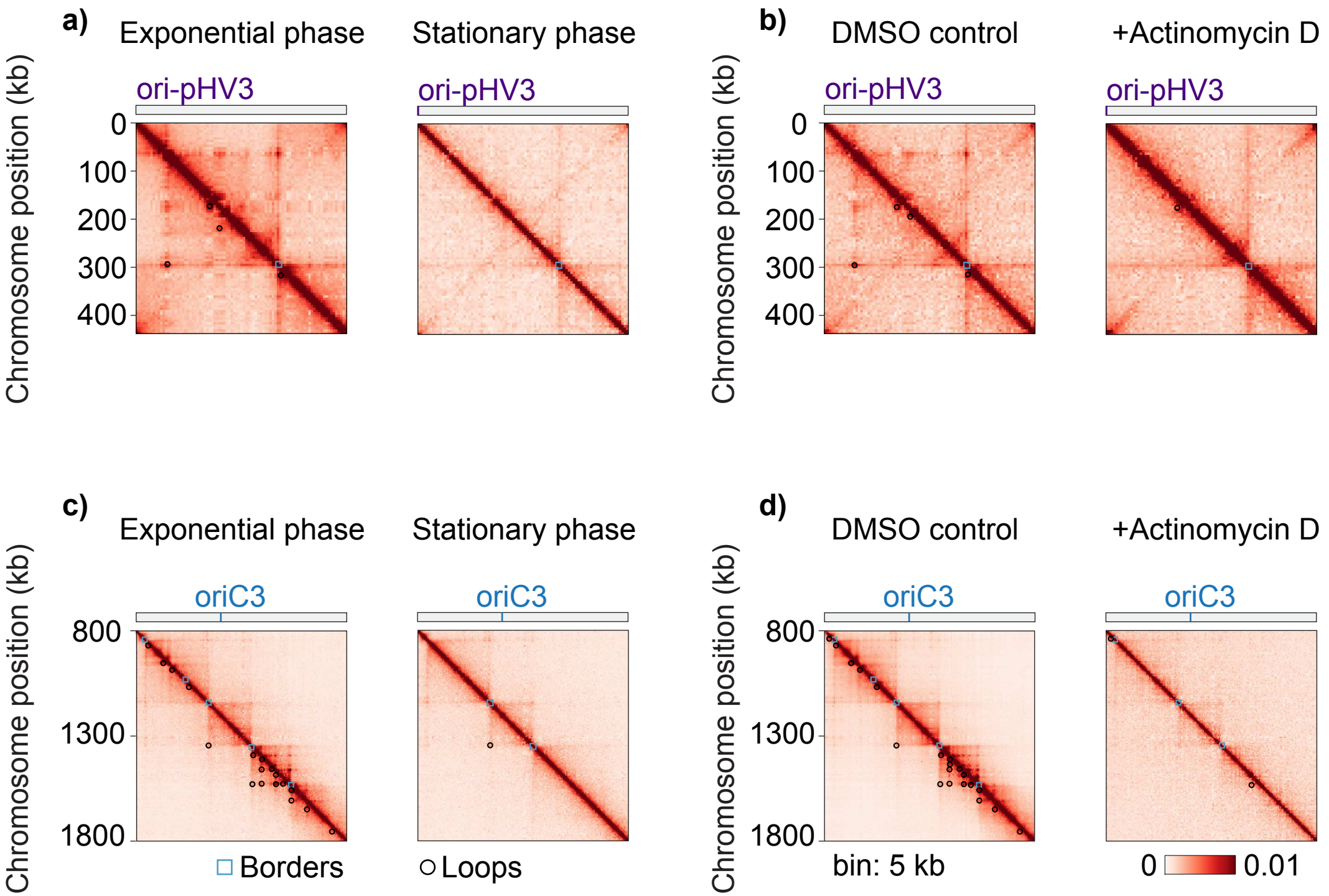


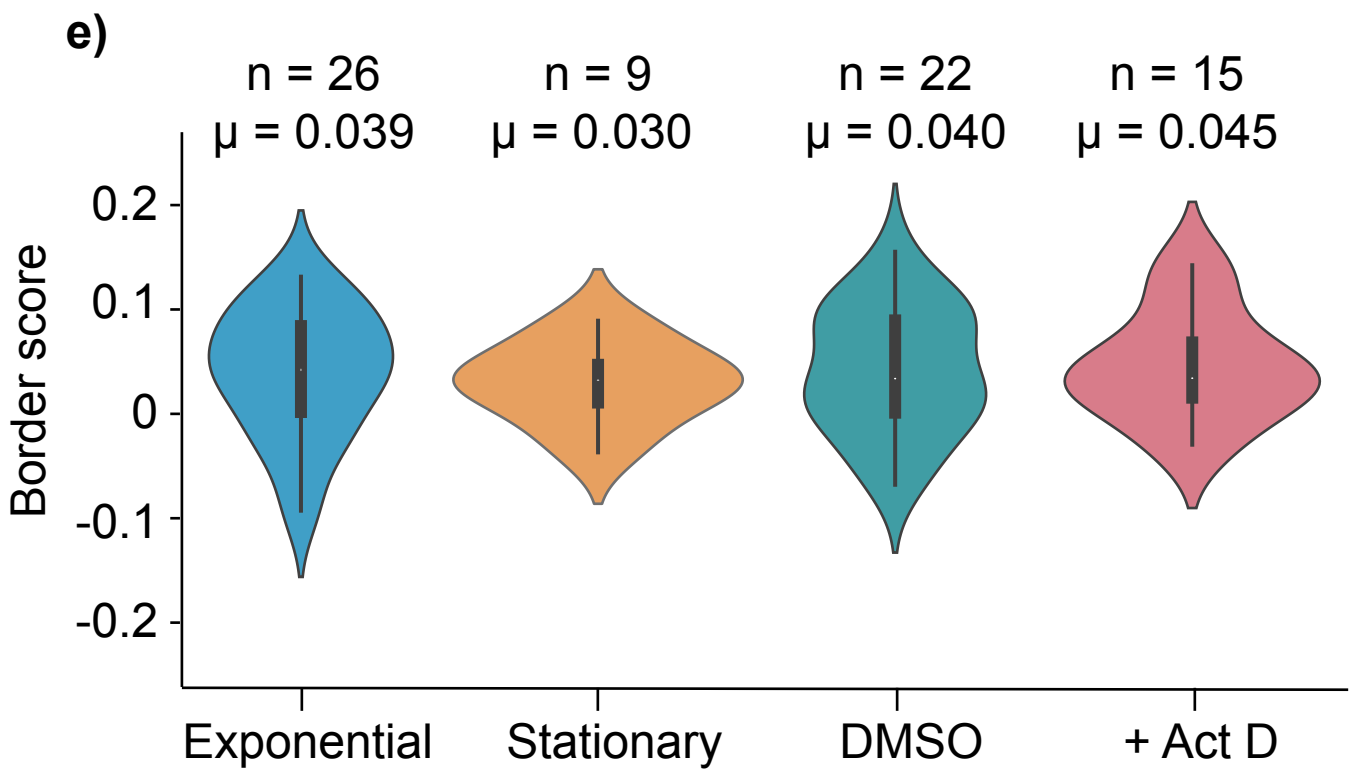




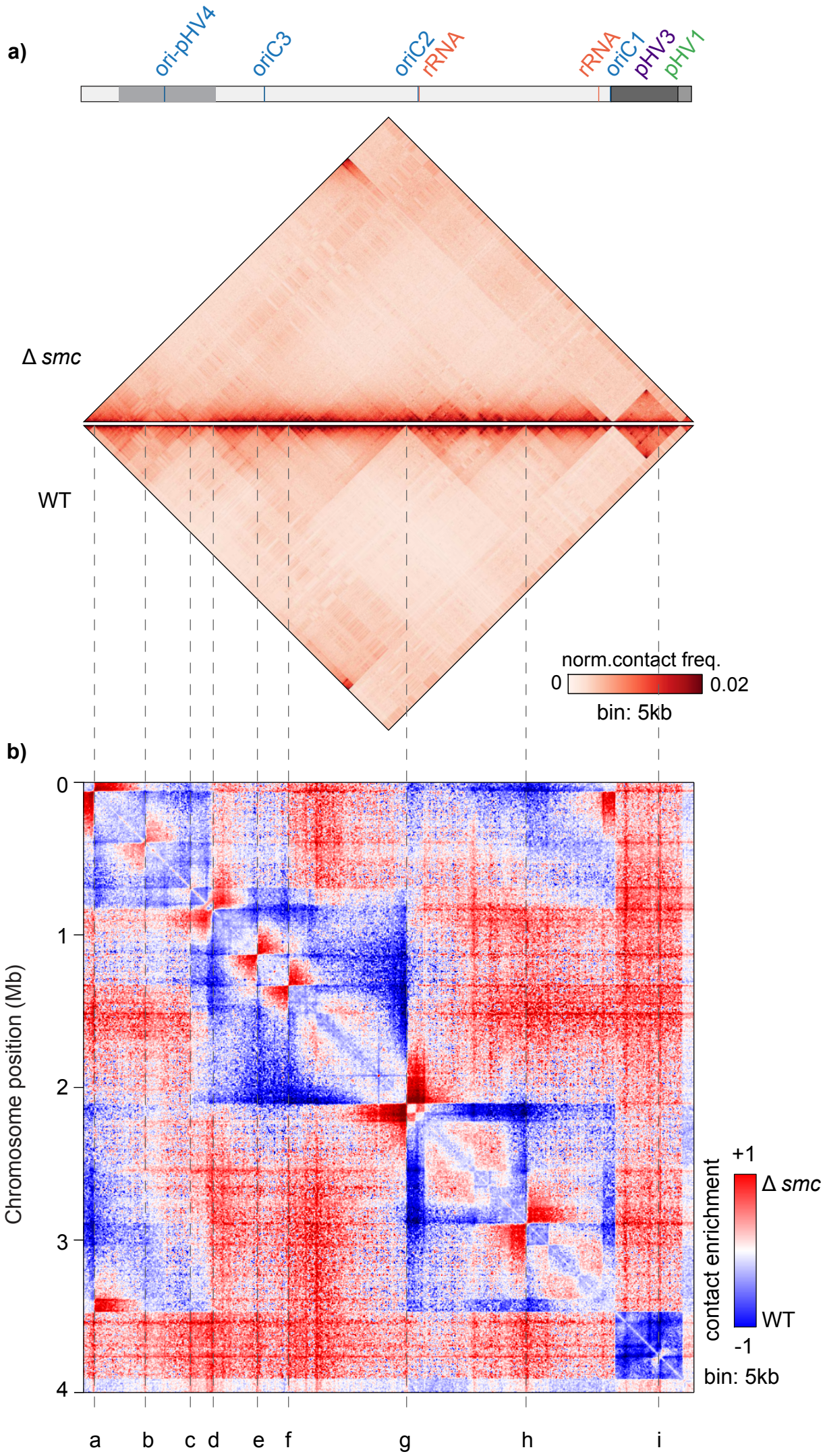


Supplementary Figure 5





Supplementary Figure 6



c)

Border	Chromsome	Approx position (bp)	Closest gene	Proposed funcition
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