Unique territorial and sub-chromosomal organization revealed in the holocentric moth *Bombyx mori*

Gil Jr., J.[1], Rosin, L.F.[2,$], Navarrete, E[3]. Chowdhury, N.[4], Abraham, S.[4], Cornilleau, G.[1], Lei, E.P.[2], Mozziconacci, J.[5], Mirny, L.A.[4,1], Muller, H.[1]*, Drinnenberg, I.A.[1]*

Affiliations:
1 Institut Curie, PSL University, Sorbonne Université, CNRS, Nuclear Dynamics, 75005 Paris, France
2 Nuclear Organization and Gene Expression Section; Laboratory of Biochemistry and Genetics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892 USA
3 Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
4 Institute for Medical Engineering and Science, and Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
5 StrInG Lab, Museum National d'Histoire Naturelle, Paris, France

$ Current address: Unit on Chromosome Dynamics, Division of Developmental Biology, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892 USA

*Correspondence: heloise.muller@curie.fr (H. Muller), ines.drinnenberg@curie.fr (I. Drinnenberg)

Abstract (100-150 words)

The hallmarks of chromosome organization in multicellular eukaryotes are chromosome territories (CT), chromatin compartments, and different types of domains, including topologically associated domains (TADs). Yet, most of these concepts derive from analyses of organisms with monocentric chromosomes. Here we describe the 3D genome architecture of an organism with holocentric chromosomes, the silkworm *Bombyx mori*. At the genome-wide scale, *B. mori* chromosomes form highly separated territories and lack substantial *trans* contacts. As described in other eukaryotes, *B. mori* chromosomes segregate into an active A and an inactive B compartment. Remarkably, we also identify a third compartment, Secluded "S", with a unique contact pattern. Compartment S shows strong enrichment of short-range contacts and depletion of long-range contacts. It hosts a unique combination of genetic and epigenetic features, localizes at the periphery of CTs and shows developmental plasticity. Biophysical modeling shows that formation of such secluded domains requires a new mechanism – a high density of extruded loops within them along with low level of extrusion and compartmentalization of A and B. Together with other evidence of loop extrusion in interphase, this suggests SMC-mediated loop extrusion in this insect. Overall, our analyses highlight the evolutionary plasticity of 3D genome organization driven by a new combination of known processes.

Introduction

The development of the Hi-C technology has enabled the analysis of genome architecture across diverse organisms[1–3]. Based on these studies, principles of chromosome organization have emerged, including the formation of chromosome territories (CTs), affinity-driven segregation of chromosomal regions into active (A) and inactive (B) compartments, and the formation of a variety of Hi-C features, including Topological Associated Domains (TADs), dots and stripes, through structural maintenance of chromosomes (SMC)-mediated loop extrusion [4–13].

Many Hi-C datasets have been generated to aid genome assembly [14–18], but in-depth analyses of 3D genome organization have been performed only for model organisms, including mammals, flies, nematodes, yeasts, bacteria and plants [4–13,19–22]. A recent comparison of diverse organisms provided insights into differences in genome organization. This study described two types of architecture [3]. Type I includes species that have "Rabl-like" features, which include prominent contacts between centromeres or telomeres from different chromosomes or chromosomal arms arranged in parallel. Type II includes species that predominantly exhibit enhanced contact frequencies between loci on the same chromosome reflecting the formation of CT as the main organizational feature. In this study, species with radically different linear chromosome organization were however intentionally excluded from the analyses [3]. This group included species with holocentric chromosomes. Unlike most eukaryotes that have a locally restricted centromeric region, holocentric organisms have centromeric regions distributed along the entire length of chromosomes. Analyses of the spatial organization of holocentric chromosomes could lead to new organizational principles of the genome and help to deduce the role of centromeres in chromatin organization [23,24]. Nevertheless, 3D genome organization of holocentric organisms remains poorly understood. A notable exception is the nematode *Caenorhabditis elegans* [25–27,19]. At a larger scale, *C. elegans* chromosomes spatially segregate into a tripartite structure reflecting its linear organization [28] with preferential contacts between arm and center regions in *cis* and *trans* [26]. At a smaller scale, classical compartmentalization correlated with epigenetic states is also evident along *C. elegans* chromosomes [19] and while TAD-like structures can only be seen on the sex chromosome [25,27], other chromosomes show evidence of loop extrusion by cohesin^{COH-1} [29].

Here, we provide an in-depth characterization of the 3D genome architecture of another holocentric species, the silkworm *Bombyx mori*. We combine Hi-C and genome-wide ChIP-seq datasets with Oligopaint FISH analyses and computational modelling to determine the folding principles of *B. mori* chromosomes. Our analyses reveal a new type of chromatin compartmentalization that results from global affinity-mediated contact and locally concentrated loop extrusion. Our study thus demonstrates that a new combination of classical drivers of spatial genome organization can lead to new principles in genome organization.

Results

Bombyx mori linear genome organization

We generated Hi-C datasets from *B. mori* embryos at three different embryonic post-diapause time points (2h, 24h and 48h hours after diapause release) and one adult stage using the p50 reference strain (Figure 1A). For most of our analysis, we focused on the 24h post-diapause timepoint (PD-D2), for which we confirmed that a proportion of cells have re-entered the cell cycle and are thus no longer arrested in G2 (Figure S1) [30]. Initially, we used these Hi-C datasets to assess the current *B. mori* genome assembly, which is based on long and short read sequence data (hereafter called Bmori_v3) [31]. We were able to confirm the overall continuity of scaffolds assigned to chromosomes, but also identified four scaffold misorientations that are readily visible on the contact maps of Chr11, 19, 22 and 24 (Figure S2). We corrected the assembly and hereby provide version 4 of the *B. mori* genome sequence together with the corresponding gff3 annotation file (Supplementary Files SF1 and SF2). Bmori_v4 was used to remap our embryonic reads and for all downstream analyses described in this paper (Figure S3A).

In this assembly, the Hi-C contact map of Chr11c still reveals a strong insulation point (Figure S3A and B) between the ~5Mb left end and the rest of the chromosome. This location might correspond to the rDNA array, which is known to display strong insulation in other species [5,32]. Although our search of rDNA sequences in the published genome assembly indicates a different position on the same chromosome, our Hi-C contact data support the location of the rDNA array at the insulation point (Figure S3B). We did not correct this error nor other small mis-assemblies that give rise to speckles distant to the diagonal on Hi-C maps, but we provide a full sequence of one unit of rDNA array of *B. mori* in the Supplementary File SF3 as it is incomplete in the reference genome.

Bmori_v4 encompasses 27 autosomes and the Z sex chromosome (Chr01), which vary in length from 8 to 22 Mbp. Genome-wide analysis of GC content revealed a tripartite organization of *B. mori* chromosomes, with a central GC-poor region and two large telomere-proximal GC-rich regions (Figure 1B). This linear organization is reminiscent of observations in the nematode *C. elegans*, another holocentric organism whose chromosomes are divided into centers and arms [33]. We thus used our GC content track along *B. mori* chromosomes to define arms and center regions for each chromosome (see supplemental text for detail, Table S1 and Figure S4). In *C. elegans*, chromosomal arms are enriched in repetitive DNA, while centers are enriched in genes [33] (Figure S4). In *B. mori*, we found a similar organization but with lower local Transposable Element (TE) and gene enrichment patterns (Figure S4).

By performing high quality Hi-C analysis on the p50 *B. mori* reference strain, we were able to improve the current *B. mori* genome assembly. We also delineate the linear organization of *B. mori* chromosome into large center and arm regions. The similarity to the linear chromosomal organization of *C. elegans* might reflect convergent evolution along the two holocentric lineages that evolved independently from different monocentric ancestors [34,35].

Chromosomes in B. mori embryos form highly distinct territories

Next, we explored CT formation in *B. mori*. The Hi-C contact pattern across all 28 chromosomes revealed sharply demarcated contact regions for each chromosome (Figure 1C). Conversely, we found that the inter-chromosomal space is very sparse with no visible

specific contact patterns, even when a low-resolution map was drawn to enhance these contacts (Figure S5A). This is consistent with recent FISH data that revealed that *B. mori* chromosome are tightly folded and occupy distinct CT [36]. The sparseness of inter-chromosomal contacts in *B. mori* is also reflected by the difference between the intra-chromosomal (*cis*) and inter-chromosomal (*trans*) contact frequencies. In contrast to those of other organisms, the *cis* contact frequency decay curve (P(s)) of *B. mori* chromosomes runs at least an order of magnitude higher than the average *trans* contact frequency (Figure 1D, Figure S5B). Despite the similarity in linear genome organization between *B. mori* and *C. elegans*, clustering between centers and arms of different chromosomes, easily seen on *C. elegans* contact maps [25,26], is not evident in *B. mori*. Nevertheless, drawing the average *trans* contact matrix reveals enrichments between large sub-telomeric regions indicating some extent of telomere or arm clustering between chromosomes in *B. mori* (Figure 1E). This observation is also supported by the ratio of *cis* over total (*cis*+*trans*) Hi-C read coverage showing decreasing values toward chromosome ends (Figure S6).

Although inter-chromosomal contacts are sparse, we tested whether known patterns of sub-nuclear positioning of chromosomes seen in other organisms are present in *B. mori*. In humans, small gene-rich chromosomes have been shown to preferentially interact with each other and localize close to each other in the interior of the nuclear space [1,37–39]. A length-dependent contact preference can also be assigned to *B. mori* chromosomes, with a group of small chromosomes (Chr02, 28, 26, 20 and 16) having the highest average inter-chromosomal contact frequency, but the correlation with gene density is mild (Figures 1B, 1F, S7 and Table S2). This might be due to the lower variation in length and gene content among *B. mori* chromosomes compared to human chromosomes (Table S2). We also noticed a positive correlation between inter-chromosomal contacts and GC content (Figure S7 and Table S2), but how these factors contribute to contact preferences remains unclear. Taken together, our data reveal conserved principles in chromosome organization such as the formation of CTs and inter-chromosomal contact preferences in *B. mori*. In contrast to other organisms, however, *B. mori* chromosomes make very limited interchromosomal contacts supporting remarkably strong CTs in this organism.

B. mori chromosomes are organized in three chromatin compartments

We then explored the intra-chromosomal contact maps of *B. mori* for signatures related to the formation of chromatin compartments. The maps show a characteristic checkerboard pattern of alternating regions with high and low contact frequencies (Figures 2A, S3). This is similar to other organisms in which genomic loci spatially segregate into an active A and an inactive B compartment at the sub-chromosomal scale (Figure S5A) [6–8,10–13]. Unexpectedly, our initial visual inspection also revealed large regions that do not conform to this pattern. Instead, these regions make very strong short-range contacts but few longer-range contacts in *cis* with any other chromosomal areas (Figures 2A, S3). Thus, we aimed to identify chromosomal loci that belong to the three observed contact patterns. To do so, we applied a k-means clustering strategy using the three first principal component (PCA) vectors of the observed/expected intra-chromosomal matrices, which together explain over 80 % of the matrix variance ratio for each chromosome (Table S3). This strategy enabled us to define three main clusters for each chromosome (see M&M) (Figures 2B, S8 and S9).

To further characterize the three clusters, we profiled silent (H3K27me3, H3K9me3) and active (H3K36me3, H3K4me3) histone marks by ChIP-seq in the same PD-D2 embryonic

stage of *B. mori*. We also included H2A.Z, as it has been involved in specifying a type of inactive chromatin state in colon cancer cells [24], and H4K20me1, a mark that has been associated with a variety of processes including transcriptional regulation, chromosome replication and segregation, the DNA damage response as well as chromosome compaction [40,41]. One of the clusters is particularly enriched for active chromatin marks. We termed this compartment A because it is reminiscent of the active compartment first described in humans [1,2] (Figure 2A, 2C, S9 and S10A). The other two clusters are depleted of active histone marks and slightly enriched for H3K27me3, a mark associated with facultative heterochromatin [42]. Based on visual correlation with the Hi-C maps, we termed the inactive cluster in the checkerboard pattern compartment B because of its similarity to the inactive compartment described in other organisms [1,2] (Figure 2A, 2C, S9 and S10A). The third cluster correlates with regions of sparse long-range contacts. This has no precedent, and we named this compartment S, to refer to its secluded, spatially segregated, behavior observed on the Hi-C maps (Figure 2A, 2C, S9 and S10A). Like compartment A, S is overall enriched in H4K20me1. This enrichment, however, only applies to about one third of all S domains, which have relatively high levels of this mark (Figure S10B and S10C). The profiles of H3K9me3, a mark associated with constitutive heterochromatin, are not strongly correlated with any of the three compartments (Figure 2C and S10A). Instead, this mark appears to be distributed at low levels along the chromosomes with a slight enrichment over chromosomal arms, particularly in telomere-proximal regions (Figure 2A and S9).

Domains of each compartment are heterogenous in length and cover different fractions of the genome (Figure 2D, 2E, 2F). Overall, A and S domains are smaller, with a median size of 80 kb, while B domains have a broader distribution and are generally larger, with a median size of 120 kb. A and B compartments cover approximately 50 and 30 % of the genome, respectively, while S covers 15 %. The domains are distributed throughout the genome with no preferential clustering towards the chromosomal center or arms (see Figure 2G for S compartments distribution and S9 for all).

To confirm that S constitutes an independent compartment and not a sub-compartment of either A or B, we determined the genome-wide homotypic and heterotypic *cis* off-diagonal P(s) for all combinations of A, B and S (Figure S11). A-A, B-B and A-B curves confirm that homotypic contact frequencies are higher than heterotypic ones at all genomic distances. In addition, A-A contact frequencies are always higher than B-B, arguing for a higher affinity among compartment A domains compared to those among B. The off-diagonal P(s) of S domains reveals an elevated contact preference between S domains at distances lower than 600 kb. At longer distances, the curve drops below A-B contact probabilities and merges with the very low S-A or S-B contact probabilities at around 1 Mbp. These data confirm that at intermediate distances compartment S interacts more frequently with itself and makes only few contacts beyond 1 Mbp, with no preference for any of the three compartments.

These data are also recapitulated by the average off-diagonal (inter) contact plots between compartment types in *cis* (Figure 2H, also see below Figure 3E). The strongest enrichment corresponds to A-A homotypic contacts, while S compartments contacts are depleted with any of the three types. On the other hand, the frequent short-range contacts observed in S are recapitulated by on-diagonal (intra) compartment average plots showing higher contact frequencies than expected within the S compartment (Figure 2I, also see below Figure 3E). To illustrate the distinctiveness of the three compartments using a visual representation, we constructed the average 3D representation of Hi-C maps for each *B. mori* chromosome using the ShRec3D algorithm (Figure S12). These maps recapitulate the spatial

segregation between A and B compartments. Domains of S are isolated from both and protrude as bubble-like structures at the outside of chromosome territories. While these renderings reflect trends in Hi-C data, 3D models of chromosomes were also developed using a different approach (see below).

Taken together, we conclude that *B. mori* chromosomes segregate into three intra-chromosomal compartments. As is common in other organisms, we identified A and B compartments corresponding to epigenetically defined active and inactive regions of the genome. The newly identified compartment S exhibits repressive epigenetic features yet with a specific enrichment of H4K20me1. Uniquely, contiguous S domains have distinct short- and long-range contact profiles with high density within each domain, and depletion of contacts with each other or A and B domains, and collectively includes about one sixth of the genome.

Phenomenological model reveals principles of A/B/S compartmental organization

Given the clear compartmental pattern observed in Hi-C, we first sought to explain the unique behavior of S compartment by modeling affinity-mediated compartmentalization of chromosomes, which has successfully recapitulated compartmentalization in a variety of other organisms [43,44].

We began by reducing the system down to recapitulate S phenomenologically. To do so, we generated a series of polymer models with equal sized A, B, and S domains (Figure 3A). For each model in the series, we applied a different set of homotypic affinities for A, B, and S (*i.e.,* A-A, B-B, and S-S attractions) but kept heterotypic contact neutral. We deemed a parameter set successful if it yielded enrichment within contiguous S domains while maintaining depletions between S and A, S and B, and between different S domains. For each model we measured frequencies of A-S, B-S, S-S_{inter}, and plotted them versus S-S_{intra} (Figure 3A, right). We found that none of the models tested satisfied our criteria. While we could reproduce conventional checkerboard patterning as desired for A and B compartments, we could not reproduce the patterns of S domains, with starkly depleted contacts between S and everything else along the heatmap (Figure 3A right panel and 3B left panel). We therefore concluded that the observed pattern of contact across S domains cannot result from affinity-mediated compartmentalization alone.

Aiming to explain pattern of contacts of S domains we turned to loop extrusion, a mechanism of chromosome organization seen across species, including holocentric nematodes [19,29]. In vertebrates, loop extrusion by the SMC complex cohesin, when occluded by extrusion boundaries, generates domains of local compaction, TADs. One of the documented effects of cohesin's activity is a decrease in compartment contacts [45]. We therefore hypothesized that loop extrusion could ameliorate the two missing features in our phenomenological map: local compaction within S and decreased distal contacts with S and everything else. To test this, we devised a model with active loop extrusion exclusively within S regions (Figure 3B middle panel). In our model, loop extruders are loaded only within S domains and domain boundaries act as barriers to extrusion. We found that this model reproduced dense S domains and successfully prevented contact between S and other domains but failed to produce the compartmentalization of A and B regions.

We then integrated our two sets of models and simulated compartmentalization of A and B and loop extrusion within S together (Figure 3B right panel). This integrated model successfully reproduces the pattern of compartmental checkerboarding of A and B regions as well as dense S domains that are depleted in contacts with the rest of the chromosome. Based

on these observations, we conclude that a model combining compartmentalization and S-specific loop extrusion can reproduce the patterns of genome organization observed in *B. mori*.

Evidence of loop extrusion in the S compartment

To evaluate whether this model was a plausible explanation for S domain organization, we searched for evidence of loop extrusion in the PD-D2 Hi-C. Key indicators of loop extrusion activity can be (i) specific patterns seen in Hi-C maps such as insulated domains, dots and stripes [46]; (ii) shapes of the contact frequency P(s) curves [47,48]. While specific patterns are attributed to cohesin and CTCF in vertebrate systems [49], these signatures of loop extrusion in Hi-C are agnostic to the molecular players involved and reflect structures resulting from a barrier-restricted loop extrusion process [50]. Similarly, shapes of the P(s) curves have been used across biological systems, including yeast and *C. elegans,* to detect loop extrusion and infer sizes of loops, irrespective of presence of extrusion barriers [47,48].

We visually inspected the PD-D2 maps to find examples of extrusion-mediated features such as insulated domains, dots, and stripes. While these features are not present for all S domains, we found multiple domains with structures indicative of extrusion with barriers (Figure 3C). Strong insulation points are also evident both inside and at edges of S domains. The presence of internal structures and insulation points within S domains further substantiates our hypothesis that loop extrusion is active within S domains. Following up on these visual inspections, we examined local insulation genome-wide in PD-D2 using high resolution 5 kb matrices and 25 kb diamond shaped sliding windows (Figure S13). We found that strong boundaries correlate with a local enrichment in H3K4me3 flanked by H3K36me3 enrichments (Figure S14), suggesting that many boundaries are linked to gene expression. To explore potential differences in insulation strength within A, B and S, we generated pileup plots of internal boundaries within large (>200 kb) domains of each type (Figure S15). In A, insulation points are visible with the majority representing boundaries correlated with expressed genes (Figure S14B). In B, most genes are silenced, so boundaries are not associated with gene expression (Figure S14B) and appear to be much weaker compared to A (Figure S15). In S, both types of boundaries, associated or not with gene expression, are present (Figure S14B) and appear stronger on average compared to A and B (Figure S15).

We also examined the curves of contact frequency P(s) as a function of genomic separation s. Loops creates a local contact enrichment seen as shoulder in the P(s) curve and as a peak in the log-derivative of the P(s) curve [47,48]. The maximum of the peak reflects the average loop size, which is approximately 100kb in the case of vertebrate (Figure S5B). We computed these values for the PD-D2 Hi-C both genome-wide (Figure 1D) and separately for each compartment type (aggregated intra domain) (Figure 3D). In both P(s) representations, we indeed observe extrusion-indicative shoulders on P(s) and a peak in its log-derivative. Genome-wide, the average loop size is approximately 40-60 kb. The P(s) shoulder is most prominent for the S domains and indicate an average loop size of ~40-60 kb. Although less pronounced, the curves corresponding to compartment B also suggests loop extrusion there, though with more sparse loops as suggested by the smaller height of the peak.

Together these analyses strongly suggest loop extrusion as a mechanism of genome organization in *B. mori.* All indicators point as high density of extruded loops in S compartment, yet some, lower, level of extrusion across A and B are also suggested.

Integrated model of affinity-based compartmentalization and S-enriched loop extrusion

Using principles uncovered by the phenomenological model and indications of loop extrusion in S and possibly A and B, we next sought to develop a chromosome-size polymer model that can reproduce genome organization of *B. mori* (Figure 3E). We focused on the first 6.5 Mb of Chr15 and simulated three stitched copies of it thus modeling a chromosome-size chain. To best encapsulate the strong territoriality of *B. mori*, we simulated a chromosome as a polymer in spherical confinement. We assigned compartmental identities based on those obtained from PD-D2 Hi-C data. The polymer was subject of A-A and B-B affinities, and loop extrusion both within S and outside S (in A+B regions) albeit with different densities of loops. Assuming that the same extruder is present in S and non-S regions, we set the processivity of the loop extruder to be the same everywhere and equal inferred loop size of 55 kb. Based on the inferences from Hi-C, we set separation between extruders to be much smaller in S than in non-S regions, achieving high density of extruded loops S, as in the phenomenological model. We also tested models with extrusion exclusively occurring in S. We created a simplified barrier landscape, such that S compartment boundaries acted as extrusion boundaries. All together the model had four parameters: A-A affinity, B-B affinity, separation of extruders in S regions, separation of extruders outside of S regions. Sweeping values of these parameters, and equilibrating the polymer, we generated more than 1000 models.

Instead of trying to identify a single set of parameters to definitively encapsulate *B. mori*'s genome architecture, we identified parameter sets that encapsulated its general features such as (i) broad agreement with experimental AB compartmentalization (within 20% of measured frequencies); (ii) agreement of S_{intra} contact density; (iii) depleted S-A, S-B and S_{inter} contacts; and (iv) average S-A and S-B contacts to be within 10% of each other (Figure 3F). We define enrichment and depletion as the average observed value relative to the average value for a given separation/distance along the polymer. After testing over a thousand models, about 50 met our criteria.

First, we find that models with compartmentalization in A and B, and high density of extruders (i.e., separation less than processivity) in S domains give rise to maps that recapitulate remarkable features of *B. mori* (Figure 3E and F). Simulated chromosomes have compact S regions that are depleted in contacts with the rest of the chromosome similar to the experimental data of the same chromosomal region on Chr15 (Figure 3E). Second, such depletion required some degree of extrusion outside S (10-fold sparser loops outside than inside S). Extrusion outside S allowed to achieve depletion of S-A, S-B, S-S contact frequencies while having condensed S regions with contact frequency consistent with Hi-C (Figure 3E and F). In the absence of extrusion outside S, achieving depletion of S-A, S-B, S-S required exceedingly dense S regions. Third, we found that attraction energies among A-type monomers are consistently more than twice those of B-type monomers. This suggests that in contrast to mammals where compartmentalization is driven largely by B-B interactions [43], A-A interactions can play a key role in compartmentalization in *B. mori*. Finally, simulations show that compartmentalized A and B regions do not show preferential central or peripheral location within a CT, while S regions demonstrate a clear tendency to be close to the chromosomal surface (Figure 3G).

Compartment S domains are found preferentially on CT surfaces

To determine the location of A, B and S domains with respect to CTs, we used Oligopaint FISH to label portions of single A, B or S domains as well as the whole CT for Chr04,

17 and 23 in embryonic nuclei (Figure 4A and S16). Shell analysis (see Methods for details) and measuring the distance from the domain center to the CT edge revealed that S domains are more likely to occupy peripheral CT shells and are closer to the CT edge than A or B domains for all chromosomes and loci analyzed (Figure 4B and Figure S16).

Although we did not observe many contacts between chromosomes in the Hi-C data, we wanted to test whether this preferential localization would influence the average contact frequency among S domains in *trans*. Therefore, we built pileup plots showing the average *trans* contacts between each compartment types (Figure 4C). We found that the average value of S-S *trans* contacts is higher compared to any other combination in *trans*, and in particular A-A and B-B. This is consistent with the preferential positioning of S towards the periphery of CT, at a favored location for *trans* contact.

Finally, to evaluate compactness of compartments, we took advantage of the designed compartment-specific FISH probes, which cover equal linear sizes on the chromosome, in order to measure domain volumes relative to CT volumes. We found that on average A compartment probes cover a substantially larger area than B or S compartment probes when compared across all chromosomes (Figure 4D). This observation confirms that compartment A is on average less compact than B and S consistent with our Hi-C data (Figure 3H). Taken together, S domains are preferentially located at the CT peripheries and form dense structures, which is consistent with the biophysical simulation and Hi-C data analyses.

Compartment S domains correspond to regions depleted in genes

The distinct contact pattern and localization of compartment S prompted us to explore whether S is distinct from A and B with respect to genetic features (Figure 5). We found that compartment S has a lower GC content and repeat coverage (TE) compared to the whole genome. Despite these similarities to A compartments, S domains correspond to gene-poor regions and is even more depleted in genes than B compartments. This is consistent with the depletion of histone marks associated with active transcription in both compartments (Figure 2C). However, unlike B, S is not a repressive compartment, and the expression level of the few genes that are located within S is in the range of the genome-wide distribution. Out of the 608 genes within S, about half (300) are expressed at the PD-D2 stage (TPM>10) (Figure 5 and Table S4). Gene ontology analyses revealed an enrichment of DNA-binding and transcription regulation processes (Table 1). We also specifically tested for an enrichment of homeotic and Polycomb group (PcG) response genes because one of the largest S domains in the genome corresponds to the Hox cluster, a bona fide PcG target [51]. Using previously published datasets [51,52], we generated a list of 399 homeotic and PcG response genes in *B. mori* and determined their position within A, B or S (Table S5, see M&M for details). We did not find an enrichment in homeotic and PcG response genes in compartment S but rather in A (Table S5). Taken together, our analyses show that compartment S largely encompasses gene deserts. Nevertheless, about half of the genes that are present are expressed in embryos 2 h after diapause release and are enriched for functions related to DNA-binding and transcription regulation.

Compartment S genomic localization changes during development

Taking advantage of our Hi-C datasets from different developmental stages including three embryonic stages and one adult stage, we next explored the developmental dynamics of compartment S. Based on initial visual inspections, we could identify compartment

switching to or from S on multiple chromosomes between timepoints (see examples in Figure S17A). To test the dynamics of S compartment switching more systematically, we repeated the compartment calling analysis for the adult stage (Adult Heads, AH). We found that a comparable fraction of the AH genome (12%) folds into compartment S compared to the PD-D2 embryonic stage (Figure S18). Nevertheless, several embryonic compartment S domains visually show checkerboard patterning in the AH Hi-C maps, including three of the largest S domains on Chr06 and Chr23 (Figure 6A and S17B). To compare the two datasets further, we restricted our analyses to large (>200 kb) domains to allow for visual confirmation of compartment assignment based on the Hi-C contact pattern. We found that only about 45% of compartment S domains defined in the late embryonic stage are maintained in the AH (referred to S->S) (Figure 6B). This fraction is much lower compared to compartment A or B domains. Large S domains that change compartment assignment in the AH exclusively turn into B domains (referred to S->B). Notably, this change also affects the expression of genes and is associated with a significant reduction in their expression levels (Figure 6C and S19). When zooming into those S domains that change to B, the Hi-C contact maps show a loss of features driven by loop extruding complexes including insulation points and off-diagonal dots (Figure 6A and S17B). The weakening of insulation is also evident when comparing pileup analyses of boundaries called within S domains that are maintained (S->S) or lost (S->B) in the AH (Figure 6D). The loss of extrusion features together with the increased checkerboard patterning of S domains that turn into B thus further supports our model that loop extrusion processes underlie the spatial segregation of compartment S by counteracting affinity-based compartmentalization. Finally, the developmental dynamics of several S domains also argue against a strict genetic specification of compartment S but rather support the presence of epigenetic features involved in its formation.

Discussion

Our investigation of the spatial genome organization in *B. mori* revealed both conserved principles as well as novel folding behaviors. Specifically, we observed the formation of strong CTs and spatial segregation of chromosomes into active A and inactive B compartments, as seen in other eukaryotes. Unexpectedly, we have also discovered a novel compartment structure, which lacks preferential interactions between noncontiguous domains.

The remarkably strong CTs and the low frequency of inter-chromosomal contacts is consistent with recent whole chromosome Oligopaint analyses of six *B. mori* chromosomes, which revealed spatially distinct CTs [36]. In the context of the recent study by Hoencamp et al. [3] categorizing a variety of eukaryotic genomes into two types of architectures, Rabl-like (I) and strong CT (II), *B. mori* represents an extreme case of type II [3]. In this study as well as a study in *D. melanogaster,* folding, volume and intermixing of CTs have been associated with condensin II [3,53]. Condensin II subunits can be identified in the *B. mori* genome, and it would be interesting to evaluate the effects of these proteins on CT formation in the future.

CT strength is also intriguing from an evolutionary point of view. Previous studies in *D. melanogaster* cell lines have described an inverse relationship between strength of CTs and genomic translocation potential of cells [54]. Interestingly, karyotypes and synteny are highly conserved across diverse Lepidoptera including *B. mori* [55]. This suggests that the strongly demarcated CTs in *B. mori* may contribute to low structural variations and a high karyotype conservation in these organisms.

The presence of a new chromatin compartment with domains that strongly interact with themselves at short distances but segregate away from the rest of the genome is remarkable. We emphasize that this contact profile is not due to assembly or mapping issues. Any type of large mis-assembly or recent translocation event would be readily visible as a strong intra- or inter-chromosomal signal far from the main diagonal and this signature is not apparent in our matrices. In addition, a previous study that used both whole-chromosome and sub-chromosomal Oligopaints designed on *B. mori* chromosomes that contain large S domains showed a colinear organization in meiotic prophase, thus providing another line of evidence to argue against mis-assemblies [56]. Regarding mapping concerns, low quality alignments often resulting from repeated regions would result in filtered bins ("white lines") in Hi-C contact maps. This is not the case in S domains which are also repeat poor compared to the full genome.

To our knowledge, there is no known precedent for a compartment with similar contact and epigenetic profiles as compartment S. Nevertheless, it is not the first time that an additional compartment besides compartments A and B has been described. Nor is it the first time that specific DNA loci have been observed at the CT periphery. For example, regions that do not show a checkerboard pattern during cell differentiation or tumorigenesis have been referred to as intermediate or compartment I [57–59]. Although compartment S shares several characteristics with compartment I, such as enrichment of H3K27me3 and developmental plasticity, they differ in their most prominent feature. Compartment I domains show an overall homogeneous contact pattern across the chromosome with either conventional A or B compartments, while S segregates away from both. A recently described that B_0 compartment identified in a colon cancer cell line also shows a neutral contact profile [24]. However, B_0, as well as other compartment types, tend to interact with homotypic regions. Compartment S, on the contrary, is depleted in homotypic contact preferences, a feature that makes S truly unique compared to any previously identified compartment. Finally, the localization of compartment S at the periphery of CTs is reminiscent of the localization of some gene-rich chromatin regions that frequently extend away from CTs in human and mouse cells [60]. Although compartment S is gene-poor, it will nevertheless be interesting to test whether transcription contributes to the peripheral localization of S domains like gene-rich regions in human and mouse.

Our biophysical simulations together with Hi-C data suggest that high levels of loop extrusion, specifically localized within contiguous S regions, underlie formation of compartment S. To evaluate this model, it will be informative to determine whether S domains are enriched in SMC complexes such as cohesin or condensin. Importantly, this mechanism is different from the mechanism of TAD formation in vertebrates. TADs emerge due to extrusion-halting boundaries acting on otherwise largely uniformly distributed cohesin. Modeling shows that the formation of S compartment in *B. mori* requires high density of extruders within S, and almost 10-times lower density outside of S compartment. Such localization of extruders to specific regions has not been seen in other systems and suggest that S regions can constitute SMC loading sites, possible via specific histone marks (see below). Such targeted SMC loading is known to play a key role in dosage compensation in *C. elegans* [25,61] and has been recently reported at vertebrate enhancers [62,63].

Our simulations further imply the presence of barriers that block loop extruders from traversing past the boundaries of S domains. The enrichment of the promoter mark H3K4me3 at a significant number of insulation points is consistent with previous studies that have found transcription sites can form extrusion barriers [64]. This does not preclude the possibility of

additional extrusion barriers such as CTCF, CP190, and Mod(mdg4), which are conserved in *B. mori* [65,66] playing a role as well.

Interestingly, H4K20me1 is the only histone mark we found to be enriched in S. In *C. elegans*, it has been shown that the demethylase activity of a member of the SMC complex that is involved in dosage compensation enriches H4K20me1 on the inactive sex chromosome [41]. H4K20me1 has also been associated with facultative heterochromatin on the mouse X chromosome. In that system, H4K20me1 has been proposed to enhance X chromosome compaction during X inactivation [67]. Together, these studies suggest that H4K20me1 may be a mark of chromatin compaction, which may be associated with loop extrusion, possibly by triggering loading of extruders. In fact, S compartments share similarity to inactivated sex chromosomes by virtue of being (i). locally compacted and isolated from the rest of the genome (like Xi in mammals), and (ii) hosting high density of specific loop extruders (like SMCHD1 in Xi or dosage-compensation in *C. elegans*).

The function of S remains an intriguing question. It might be that the isolated genomic environment of these domains ensures the precise transcriptional regulation of the few genes present in compartment S. The fact that S-located genes are functionally enriched in transcription-related processes might suggest that S represents a developmental transition state to either A or B, as hypothesized for compartments I and B_0 [24,57,58]. Such a model is supported by the observation that these S domains are variable among different developmental stages. We consider it unlikely that S compartments are involved in centromere specification because the genomic distribution of S domains does not correlate with centromere profiles that we generated from asynchronous *B. mori* derived cell line [68].

In summary, our study describes a new chromosome sub-structure, likely formed by localized rather than demarcated loop extrusion, thus expanding our knowledge of eukaryotic 3D genome folding, and highlighting the need of research in non-model systems to reveal the diversity and plasticity of genome organization.

Figure legends

Figure 1: **Hi-C in *Bombyx mori* embryos reveals highly distinct chromosome territories**.

(**A**) Experimental design for Hi-C and ChIP samples (**B**) Schematics of *B. mori* chromosomes drawn to approximate scale indicated in Mbp below. For each chromosome, total GC content in % is noted and local GC content per 100 kb windows is indicated as a color scale from blue (<20 %) to red (>60 %). A suffix "c" added to chromosome name indicates assembly corrections. (**C**) Contact map at 80 kb bin resolution of five selected *B. mori* chromosomes (on top) (**D**) Contact probability P(s) (*cis*) (top) and its derivative (slope) (bottom) for all aggregated *B. mori* chromosomes computed from the contact map at 1 kb bin resolution. Red dashed line represents average interchromosomal contact frequency (*trans*) (**E**) Average interchromosomal (*trans*) contact versus expected matrix for all scaled *B. mori* chromosomes computed at 40 kb bin resolution. (**F**). Heat map illustrating observed vs expected interchromosomal contact frequencies as a color scale from blue to red. Chromosomes have been clustered and ordered to reflect similar contact patterns.

Figure 2: **Hi-C compartmental analysis reveals three genome-wide contact patterns**.

(**A**) Representative Hi-C contact map of Chr09 at 40 kb bin resolution with gene track (black), tracks of the first two PCA components (grey), locations of domains of compartment A (orange), B (turquoise) and S (purple) and ChIP-seq tracks for various histone marks. (**B**) Example of cluster calling along PCA1 and PCA2 of Pearson's correlation matrix for Chr19c. Clusters called using k-means function of scikit-learn are shown in different colors. For each cluster, centroid and the corresponding attribution to A, B and S compartments is indicated. (**C**) Heatmap showing genome-wide enrichment of several histone marks across compartment A, B and S. Enrichment values correspond to median of IP/input ratios at 40 kb resolution normalized to genome-wide median for each mark. (**D**) Box blots showing size distribution of A, B and S domains. Box region corresponds to data between first to third quartile. Line indicates median of distribution while cross corresponds to mean. Whiskers extends to lowest and highest data points, excluding outliers, which have been removed. (**E**) Bar graph showing relative genomic coverage of A, B and S. (**F**) Interpolated Hi-C maps of regions of Chr10, 13 and 23 at 40 kb resolution showing examples of various compartment S domains. (**G**) Distribution of S compartment domains (purple boxes) along *B. mori* chromosomes. (**H**) Average *cis* off-diagonal (inter) contact versus expected plots within and between all scaled A, B and S compartments. (**I**) Rescaled average *cis* on-diagonal (intra) contact frequency compared to expected for the three compartment types.

Figure 3: Simulations of affinity-based compartmentalization and activity-based loop extrusion recapitulate patterns of contact seen in *B. mori*.

(**A**) Summary of the biophysical model of compartmentalization of *B. mori* chromosomes. Each simulation with specific values of A-A, B-B and S-S attractions results in a contact frequency map. Maps are then compared in their A-S, B-S and S-S contact frequency, and intra-S contact frequency. No model was able to come close to experimental values. (**B**) Model with compartmentalization and loop extrusion with S regions. Panels are representative examples of contact maps generated through affinity-based compartmentalization (left), S domain specific loop extrusion (middle), and a combination of both processes (right). The combination was able to capture all essential features. (**C**) Evidence of extrusion in Hi-C data. Examples of maps with domains (indicated with gene track below matrices), dots and stripes within S regions. Matrices are plot at 5 kb bin resolution, gray arrows indicate corner peaks, blue

arrows indicate stripes. (**D**) Contact frequency P(s) as a function of genomic separation (s) curves for A, B and S contiguous regions (c*is* on-diagonal) and their slopes (log-derivatives), computed from a Hi-C matrix at 1 kb bin resolution. Position of the peak on the slope curve indicates the mean size of the loop. (**E**). Polymer model with both compartmentalization and loop extrusion of a region of *B. mori* Chr15 (0 - 6.5Mb). The top right half of the image is the Hi-C contact map and the bottom half is the *in-silico* contact map generated by one of best fitting models. (**F**) Average enrichment of contacts between different types of compartments for Hi-C data (top right half) and shown simulations (bottom left half). S-S enrichments have been further split into enrichments within one S domain (S_{intra}) and between distinct S-domains (S_{inter}). (**G**) Rendering of a single chromosome and its central slice. Notice peripheral locations of S domains.

Figure 4: **Compartment S domains preferentially localize to the periphery of CT**.
(**A**) 4-color Oligopaint FISH labeling single A (red), B (green), and S (purple) compartment domains as well as the corresponding CT (white). White dashed lines indicate CT edge (left column). Gray dashed squares (second column) indicate zoom panels in third column. Fourth column: 3D rendering of zoomed CT from TANGO. Microscope images are Z-projections of 10 Z stacks. Background in the CT channel acts as a proxy for the nuclear edge. (**B**) Shell analysis measuring compartment foci positions within their CTs for Chr04, 17 and 23. The location of Oligopaint FISH probes within the chromosome is shown above each plot. Dots indicate mean of 3 biological replicates (different embryos, n>250 nuclei). Error bars show standard error of mean. P-value and statistical significance by unpaired t-test (KS test) between distributions is indicated at the bottom right of each graph. (**C**) Average *trans* contact versus expected plots within and between all scaled A, B and S compartments. (**D**) Box plot showing foci volume as a fraction of CT volume for all domains across all chromosomes examined. Box region corresponds to data between first to third quartile. Line indicates median of distribution. Whiskers extends to lowest and highest data points, excluding outliers, which have been removed. Statistical significance by Welch's t-test is indicated by asterisks on top of graph (*p = 0.01 - 0.05, **p = 0.001 - 0.005, ***p < 0.001).

Figure 5: **Compartment S domains are GC-poor, repeat-poor and depleted in genes.**
Box plots showing distribution of genetic features including GC percentage, TE coverage and gene coverage per 40 kb windows of whole genome (n=11143) or within A (n=3294), B (n=5772) and S (n=1670), or per gene for gene expression (in TPM) (whole genome n=13869, in A n=7404, in B n=4500, in S n=608). Box region corresponds to data between first to third quartile. Line indicates median of distribution while cross corresponds to mean. Whiskers extends to lowest and highest data points, excluding outliers, which have been removed. All distributions in each compartment for all features are significantly different from each other and from genome-wide distribution by Mann-Whitney (95 %, two-tailed) statistical test.

Figure 6: **Compartment S changes during development.**
(**A**) Hi-C contact maps of the full Chr23 at 40 kb bin resolution and Chr23:3000000..8000000 region at 10 kb bin resolution in PD-D2 embryos on the left and in AH on the right. Below or above each matrix are gene locations, RNA-seq expression aggregated at 1 kb resolution and compartments A, B and S genomic location, for the corresponding stage and region. (**B**) Compartment assignment in AH datasets of large domains (>200 kb) assigned to large compartment A, B or S in PD-D2, in % of the compartments assigned in PD-D2. (**C**) Box plots of

distributions of log_2 ratio of gene expression (TPM) between PD-D2 embryos and AH. In purple is the distribution for genes that are in domains assigned to compartment S in the two stages (S→S) (n=119) and in green is the distribution for genes that are in region assigned to compartment S in PD-D2 and that switched to B in AH (S→B) (n=120). Box region corresponds to data between first to third quartile. Line indicates median of distribution while cross corresponds to mean. Whiskers extends to lowest and highest data points, and dots outside correspond to outliers. Asterisks indicate distributions significantly different by Kolmogorov-Smirnov test (*p = 0.01 - 0.05). (**D**) Pileup plots centered on internal boundaries within all large S compartment domains (>200 kb, excluding 40kb on each side of compartment boundaries) in AH or S→S and S→B categories, as described in the right panel of (B), at 5kb resolution and extending to 30 kb on each side.

Acknowledgement:
We would like to thank the members of the Lei, Mirny and Drinnenberg labs, as well as Axel Cournac for helpful discussion. We thank all members of the Koszul lab and in particular Pierrick Moreau for helping in the initial development of the Hi-C protocol. We thank Stéphane Descorps-Declère, Ilya M. Flyamer and Nicolas Servant for their help in using bioinformatics tools and the 4DN lab which generated some data used in this study. We would also like to thank Life Science Editors for editing services (www.lifescienceeditors.com). EN is funded by NSF graduate research fellowship. LAM is the recipient of a « Chaires Internationales de Recherche Blaise Pascal ». HM and IAD receive salary support from the CNRS. This work is supported by the Labex DEEP ANR-11-LABX-0044 part of the IDEX Idex PSL ANR-10-IDEX-0001-02 PSL, an ATIP-AVENIR Research grant, Institut Curie, the ERC (CENEVO-758757) and INSB «Diversity of biological mechanisms». This work was supported in part by the Intramural Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health (NIDDK; DK015602 to E.P.L.), the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health (NICHD; 1K99HD104851 to L.F.R).

REFERENCES:

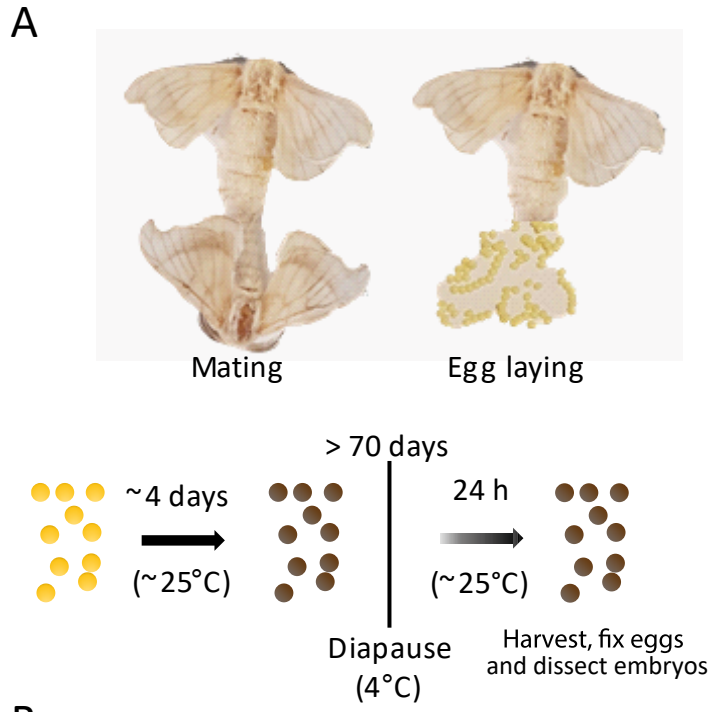
1. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
2. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
3. Hoencamp, C. *et al.* 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. *Science* **372**, 984–989 (2021).
4. Bastié, N. *et al.* Smc3 acetylation, Pds5 and Scc2 control the translocase activity that establishes cohesin-dependent chromatin loops. *Nat Struct Mol Biol* **29**, 575–585 (2022).
5. Lazar-Stefanita, L. *et al.* Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle. *EMBO J.* **36**, 2684–2697 (2017).
6. Mirny, L. & Dekker, J. Mechanisms of Chromosome Folding and Nuclear Organization: Their Interplay and Open Questions. *Cold Spring Harb Perspect Biol* **14**, a040147 (2022).
7. Mirny, L. A., Imakaev, M. & Abdennur, N. Two major mechanisms of chromosome organization. *Curr Opin Cell Biol* **58**, 142–152 (2019).
8. Misteli, T. The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell* **183**, 28–45 (2020).
9. Mizuguchi, T., Barrowman, J. & Grewal, S. I. S. Chromosome domain architecture and dynamic organization of the fission yeast genome. *FEBS Lett.* **589**, 2975–2986 (2015).
10. Oudelaar, A. M. & Higgs, D. R. The relationship between genome structure and function. *Nat Rev Genet* **22**, 154–168 (2021).
11. Pontvianne, F. & Grob, S. Three-dimensional nuclear organization in Arabidopsis thaliana. *J Plant Res* **133**, 479–488 (2020).
12. Rowley, M. J. *et al.* Condensin II Counteracts Cohesin and RNA Polymerase II in the Establishment of 3D Chromatin Organization. *Cell Reports* **26**, 2890-2903.e3 (2019).
13. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19**, 789–800 (2018).
14. Burton, J. N. *et al.* Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat Biotechnol* **31**, 1119–1125 (2013).
15. Dudchenko, O. *et al.* De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
16. Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Systems* **3**, 99–101 (2016).
17. Marie-Nelly, H. *et al.* High-quality genome (re)assembly using chromosomal contact data. *Nat Commun* **5**, 5695 (2014).
18. Wang, X. *et al.* Genome-wide detection of enhancer-hijacking events from chromatin interaction data in rearranged genomes. *Nat Methods* **18**, 661–668 (2021).
19. Das, M. *et al.* Condensin I folds the C. elegans genome. *bioRxiv* 2022.06.14.495661 (2022) doi:10.1101/2022.06.14.495661.
20. Lioy, V. S. *et al.* Multiscale Structuring of the E. coli Chromosome by Nucleoid-Associated and Condensin Proteins. *Cell* **172**, 771-783.e18 (2018).
21. 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).
22. Ouyang, W., Xiong, D., Li, G. & Li, X. Unraveling the 3D Genome Architecture in Plants: Present and Future. *Mol Plant* **13**, 1676–1693 (2020).
23. Muller, H., Gil, J. & Drinnenberg, I. A. The Impact of Centromeres on Spatial Genome Architecture. *Trends Genet.* **35**, 565–578 (2019).

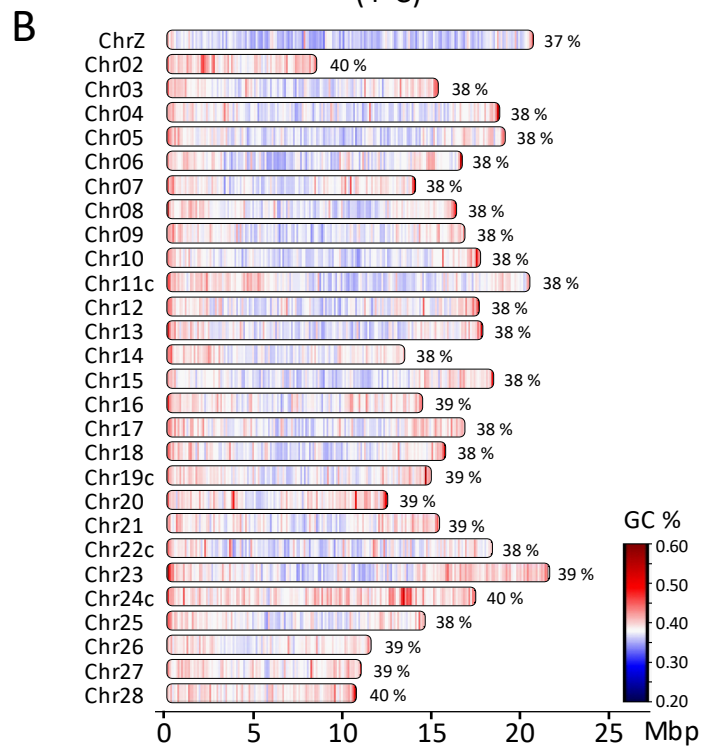
24. Spracklin, G. *et al.* Diverse silent chromatin states modulate genome compartmentalization and loop extrusion barriers. *Nat Struct Mol Biol* (2022) doi:10.1038/s41594-022-00892-7.
25. Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).
26. Gabdank, I., Ramakrishnan, S., Villeneuve, A. M. & Fire, A. Z. A streamlined tethered chromosome conformation capture protocol. *BMC Genomics* **17**, 274 (2016).
27. Rowley, M. J. *et al.* Analysis of Hi-C data using SIP effectively identifies loops in organisms from C. elegans to mammals. *Genome Res* **30**, 447–458 (2020).
28. Liu, T. *et al.* Broad chromosomal domains of histone modification patterns in C. elegans. *Genome Research* **21**, 227–236 (2011).
29. Bolaji N. Isiaka *et al.* Cohesin forms fountains at active enhancers in C. elegans. *bioRxiv* 2023.07.14.549011 (2023) doi:10.1101/2023.07.14.549011.
30. Nakagaki, M., Takei, R., Nagashima, E. & Yaginuma, T. Cell cycles in embryos of the silkworm, Bombyx mori: G2-arrest at diapause stage. *Rouxs Arch Dev Biol* **200**, 223–229 (1991).
31. Kawamoto, M. *et al.* High-quality genome assembly of the silkworm, Bombyx mori. *Insect Biochem. Mol. Biol.* **107**, 53–62 (2019).
32. Mercy, G. *et al.* 3D organization of synthetic and scrambled chromosomes. *Science* **355**, (2017).
33. C. elegans Sequencing Consortium. Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
34. Melters, D. P., Paliulis, L. V., Korf, I. F. & Chan, S. W. L. Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. *Chromosome Res.* **20**, 579–593 (2012).
35. Senaratne, A. P., Cortes-Silva, N. & Drinnenberg, I. A. Evolution of holocentric chromosomes: Drivers, diversity, and deterrents. *Semin Cell Dev Biol* **127**, 90–99 (2022).
36. Rosin, L. F., Chen, D., Chen, Y. & Lei, E. P. Dosage compensation in Bombyx mori is achieved by partial repression of both Z chromosomes in males. *Proc Natl Acad Sci U S A* **119**, e2113374119 (2022).
37. Boyle, S. *et al.* The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet* **10**, 211–219 (2001).
38. Croft, J. A. *et al.* Differences in the Localization and Morphology of Chromosomes in the Human Nucleus. *J Cell Biol* **145**, 1119–1131 (1999).
39. Tanabe, H. *et al.* Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc Natl Acad Sci U S A* **99**, 4424–4429 (2002).
40. Beck, D. B., Oda, H., Shen, S. S. & Reinberg, D. PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. *Genes Dev* **26**, 325–337 (2012).
41. Meyer, B. J. The X chromosome in C. elegans sex determination and dosage compensation. *Curr Opin Genet Dev* **74**, 101912 (2022).
42. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
43. Falk, M. *et al.* Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570**, 395–399 (2019).
44. Abdulla, A. Z., Salari, H., Tortora, M. M. C., Vaillant, C. & Jost, D. 4D epigenomics: deciphering the coupling between genome folding and epigenomic regulation with biophysical modeling. *Curr Opin Genet Dev* **79**, 102033 (2023).

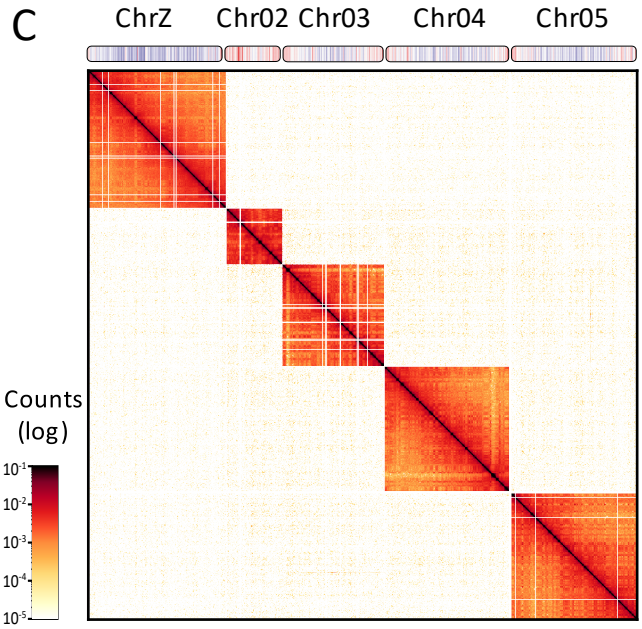
45. Goloborodko, A., Imakaev, M. V., Marko, J. F. & Mirny, L. Compaction and segregation of sister chromatids via active loop extrusion. *Elife* **5**, e14864 (2016).
46. Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A. & Mirny, L. A. Emerging Evidence of Chromosome Folding by Loop Extrusion. *Cold Spring Harb Symp Quant Biol* **82**, 45–55 (2017).
47. Gassler, J. *et al.* A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *EMBO J.* **36**, 3600–3618 (2017).
48. Polovnikov, K., Belan, S., Imakaev, M., Brandão, H. B. & Mirny, L. A. Fractal polymer with loops recapitulates key features of chromosome organization. *bioRxiv* 2022.02.01.478588 (2022) doi:10.1101/2022.02.01.478588.
49. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944.e22 (2017).
50. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* **15**, 2038–2049 (2016).
51. Chai, C.-L. *et al.* A genomewide survey of homeobox genes and identification of novel structure of the Hox cluster in the silkworm, Bombyx mori. *Insect Biochem Mol Biol* **38**, 1111–1120 (2008).
52. Li, Z. *et al.* Genome-wide identification of polycomb target genes reveals a functional association of Pho with Scm in Bombyx mori. *PLoS One* **7**, e34330 (2012).
53. Rosin, L. F., Nguyen, S. C. & Joyce, E. F. Condensin II drives large-scale folding and spatial partitioning of interphase chromosomes in Drosophila nuclei. *PLoS Genet* **14**, e1007393 (2018).
54. Rosin, L. F. *et al.* Chromosome territory formation attenuates the translocation potential of cells. *Elife* **8**, e49553 (2019).
55. d'Alençon, E. *et al.* Extensive synteny conservation of holocentric chromosomes in Lepidoptera despite high rates of local genome rearrangements. *Proc Natl Acad Sci USA* **107**, 7680–7685 (2010).
56. Rosin, L. F., Gil, J., Drinnenberg, I. A. & Lei, E. P. Oligopaint DNA FISH reveals telomere-based meiotic pairing dynamics in the silkworm, Bombyx mori. *PLoS Genet* **17**, e1009700 (2021).
57. Johnstone, S. E. *et al.* Large-Scale Topological Changes Restrain Malignant Progression in Colorectal Cancer. *Cell* **182**, 1474-1489.e23 (2020).
58. Vilarrasa-Blasi, R. *et al.* Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation. *Nat Commun* **12**, 651 (2021).
59. Schwarzer, W. *et al.* Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
60. Mahy, N. L., Perry, P. E. & Bickmore, W. A. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J Cell Biol* **159**, 753–763 (2002).
61. Kim, J. *et al.* Condensin DC loads and spreads from recruitment sites to create loop-anchored TADs in C. elegans. *Elife* **11**, e68745 (2022).
62. Guo, Y. *et al.* Chromatin jets define the properties of cohesin-driven in vivo loop extrusion. *Mol Cell* **82**, 3769-3780.e5 (2022).
63. Galitsyna, A. *et al.* Extrusion fountains are hallmarks of chromosome organization emerging upon zygotic genome activation. *bioRxiv* 2023.07.15.549120 (2023) doi:10.1101/2023.07.15.549120.

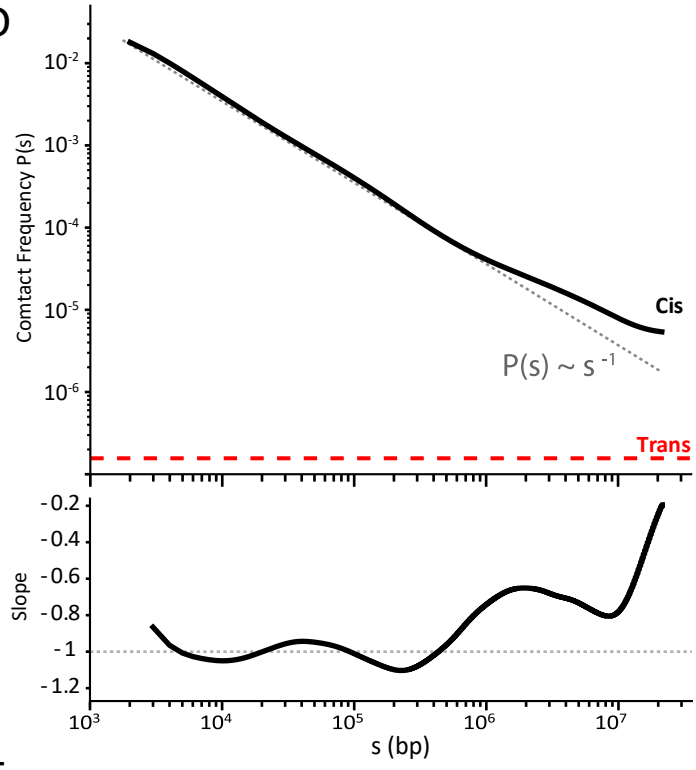
64. Banigan, E. J. *et al.* Transcription shapes 3D chromatin organization by interacting with loop extrusion. *Proc Natl Acad Sci U S A* **120**, e2210480120 (2023).
65. Heger, P., George, R. & Wiehe, T. Successive gain of insulator proteins in arthropod evolution. *Evolution* **67**, 2945–2956 (2013).
66. Pauli, T. *et al.* Transcriptomic data from panarthropods shed new light on the evolution of insulator binding proteins in insects : Insect insulator proteins. *BMC Genomics* **17**, 861 (2016).
67. Tjalsma, S. J. D. *et al.* H4K20me1 and H3K27me3 are concurrently loaded onto the inactive X chromosome but dispensable for inducing gene silencing. *EMBO Rep* **22**, e51989 (2021).
68. Senaratne, A. P. *et al.* Formation of the CenH3-Deficient Holocentromere in Lepidoptera Avoids Active Chromatin. *Curr Biol* **31**, 173-181.e7 (2021).

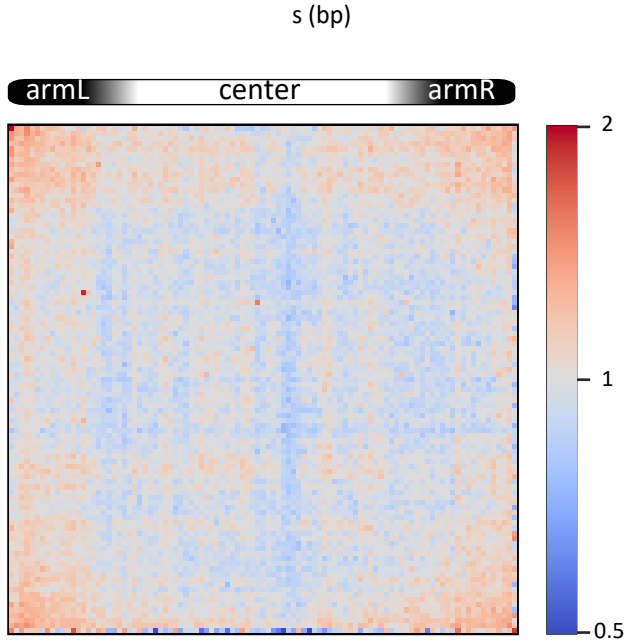
Figure 1


A
Mating
Egg laying
> 70 days
~4 days
(~25°C)
24 h
(~25°C)
Diapause
(4°C)
Harvest, fix eggs
and dissect embryos


B
ChrZ 37 %
Chr02 40 %
Chr03 38 %
Chr04 38 %
Chr05 38 %
Chr06 38 %
Chr07 38 %
Chr08 38 %
Chr09 38 %
Chr10 38 %
Chr11c 38 %
Chr12 38 %
Chr13 38 %
Chr14 38 %
Chr15 38 %
Chr16 39 %
Chr17 38 %
Chr18 38 %
Chr19c 39 %
Chr20 39 %
Chr21 39 %
Chr22c 38 %
Chr23 39 %
Chr24c 40 %
Chr25 38 %
Chr26 39 %
Chr27 39 %
Chr28 40 %
GC %
0.60
0.50
0.40
0.30
0.20
0
5
10
15
20
25
Mbp


C
ChrZ
Chr02
Chr03
Chr04
Chr05
Counts
(log)


D
Comtact Frequency P(s)
Cis
P(s) ~ s -1
Trans
Slope
s (bp)


E
armL
center
armR
2
1
0.5

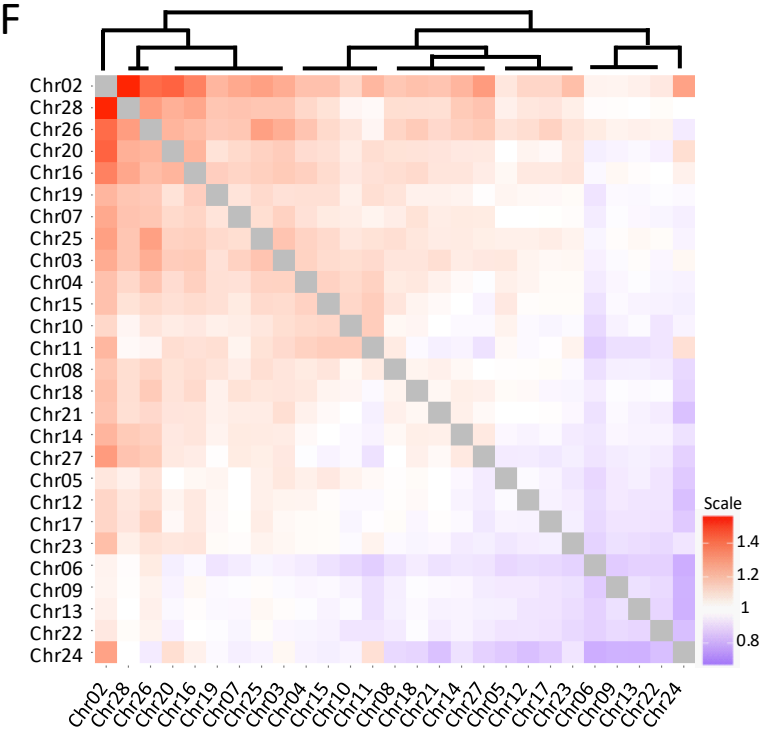

F
Chr02
Chr28
Chr26
Chr20
Chr16
Chr19
Chr07
Chr25
Chr03
Chr04
Chr15
Chr10
Chr11
Chr08
Chr18
Chr21
Chr14
Chr27
Chr05
Chr12
Chr17
Chr23
Chr06
Chr09
Chr13
Chr22
Chr24
Scale
1.4
1.2
1
0.8

Figure 2

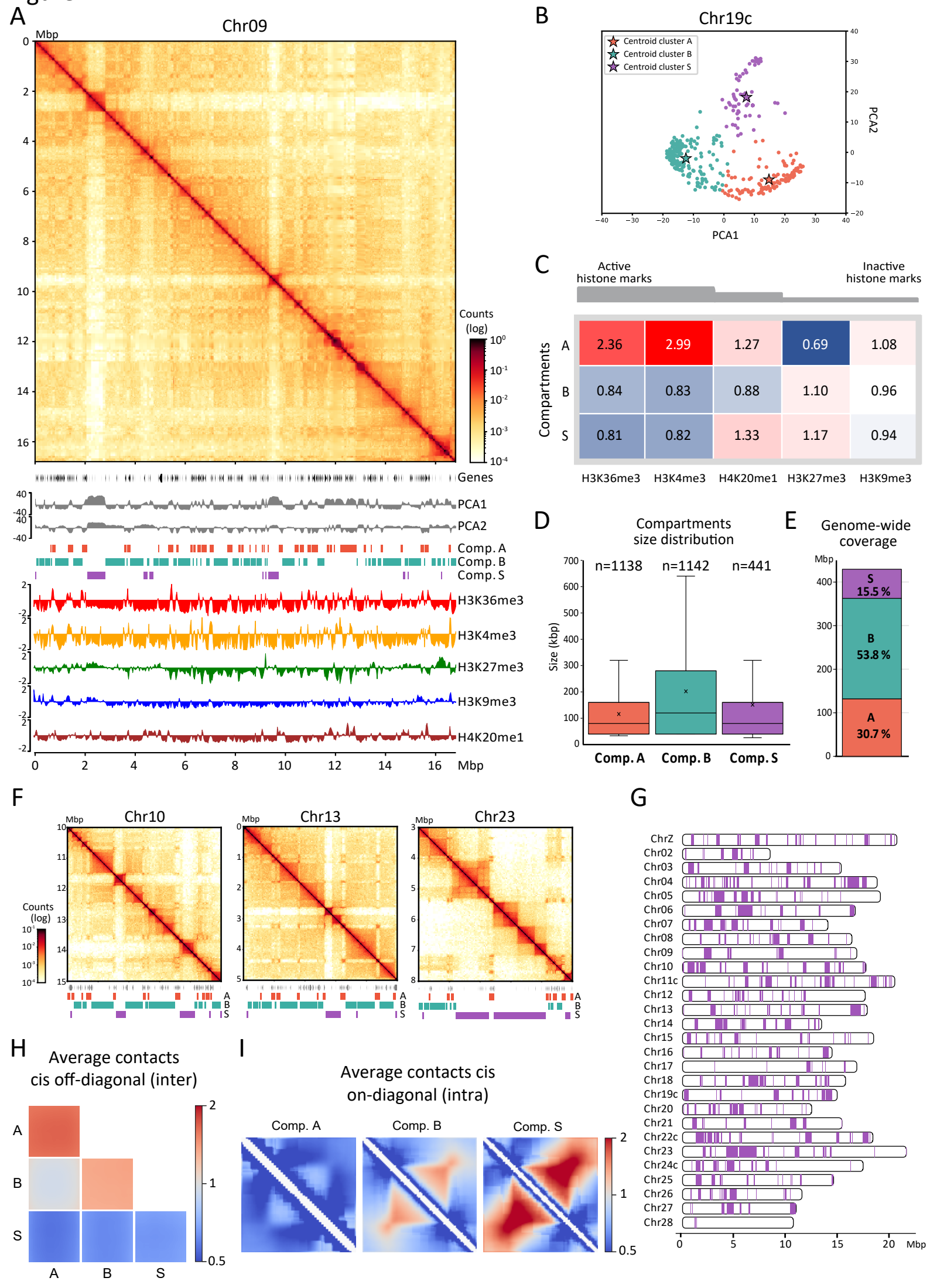

A
Chr09
Mbp
Counts (log)
Genes
PCA1
PCA2
Comp. A
Comp. B
Comp. S
H3K36me3
H3K4me3
H3K27me3
H3K9me3
H4K20me1
Mbp
B
Chr19c
Centroid cluster A
Centroid cluster B
Centroid cluster S
PCA1
PCA2
C
Active histone marks
Inactive histone marks
Compartments
A 2.36 2.99 1.27 0.69 1.08
B 0.84 0.83 0.88 1.10 0.96
S 0.81 0.82 1.33 1.17 0.94
H3K36me3 H3K4me3 H4K20me1 H3K27me3 H3K9me3
D
Compartments size distribution
n=1138 n=1142 n=441
Size (kbp)
Comp. A Comp. B Comp. S
E
Genome-wide coverage
Mbp
S 15.5 %
B 53.8 %
A 30.7 %
F
Chr10
Chr13
Chr23
Mbp
Counts (log)
A
B
S
G
ChrZ
Chr02
Chr03
Chr04
Chr05
Chr06
Chr07
Chr08
Chr09
Chr10
Chr11c
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19c
Chr20
Chr21
Chr22c
Chr23
Chr24c
Chr25
Chr26
Chr27
Chr28
Mbp
H
Average contacts cis off-diagonal (inter)
A
B
S
I
Average contacts cis on-diagonal (intra)
Comp. A
Comp. B
Comp. S

Figure 3

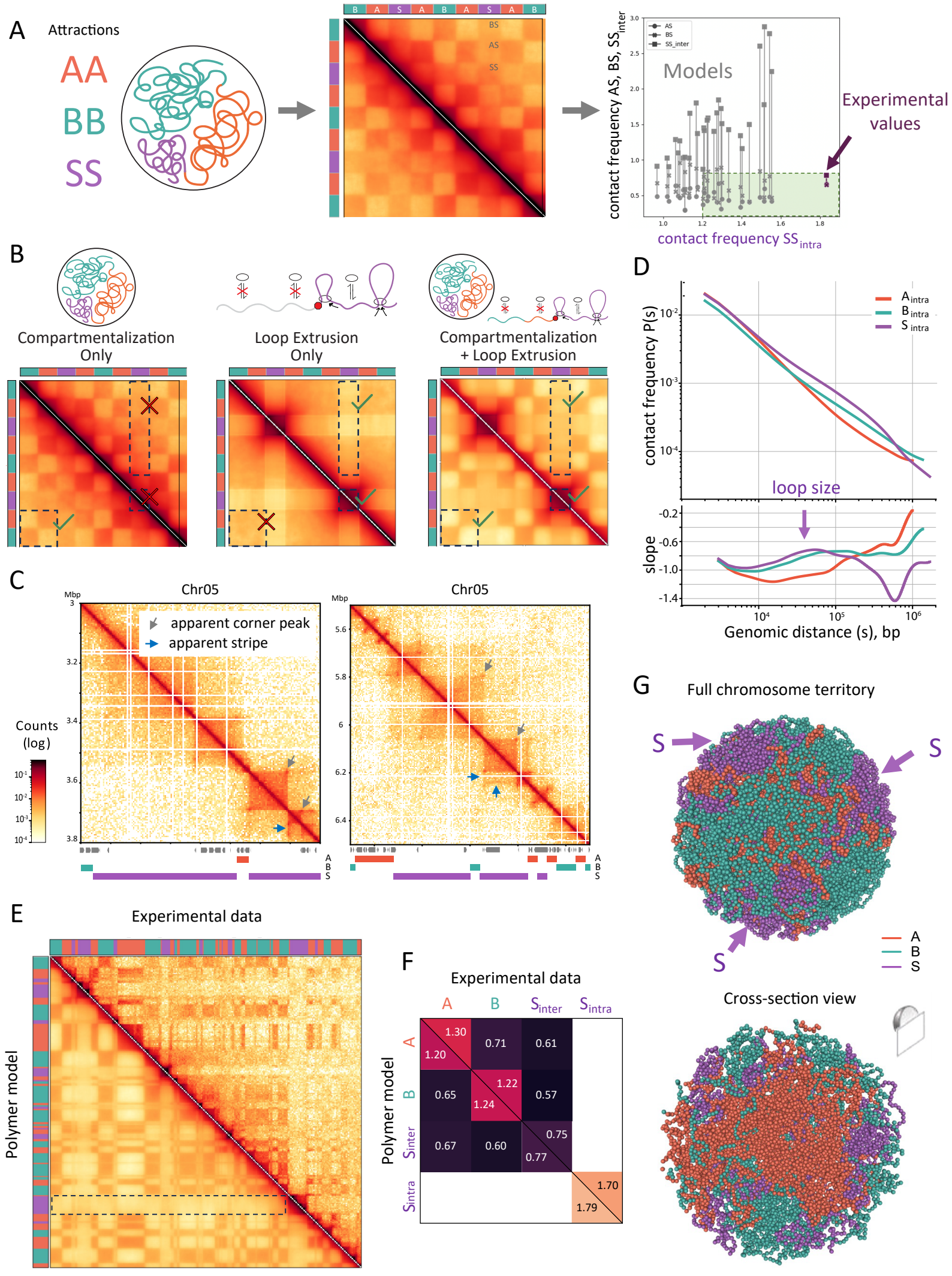

A
Attractions
AA
BB
SS
B A S A B A S A B
BS
AS
SS
Models
AS
BS
SS_inter
contact frequency AS, BS, SS_inter
contact frequency SS_intra
Experimental values
B
Compartmentalization Only
Loop Extrusion Only
Compartmentalization + Loop Extrusion
D
A_intra
B_intra
S_intra
contact frequency P(s)
loop size
slope
Genomic distance (s), bp
C
Chr05
Chr05
Mbp
apparent corner peak
apparent stripe
Counts (log)
A
B
S
G
Full chromosome territory
S
S
S
A
B
S
Cross-section view
E
Experimental data
Polymer model
F
Experimental data
A
B
S_inter
S_intra
Polymer model
1.30
1.20
0.71
0.61
0.65
1.22
1.24
0.57
0.67
0.60
0.75
0.77
1.70
1.79

Figure 4

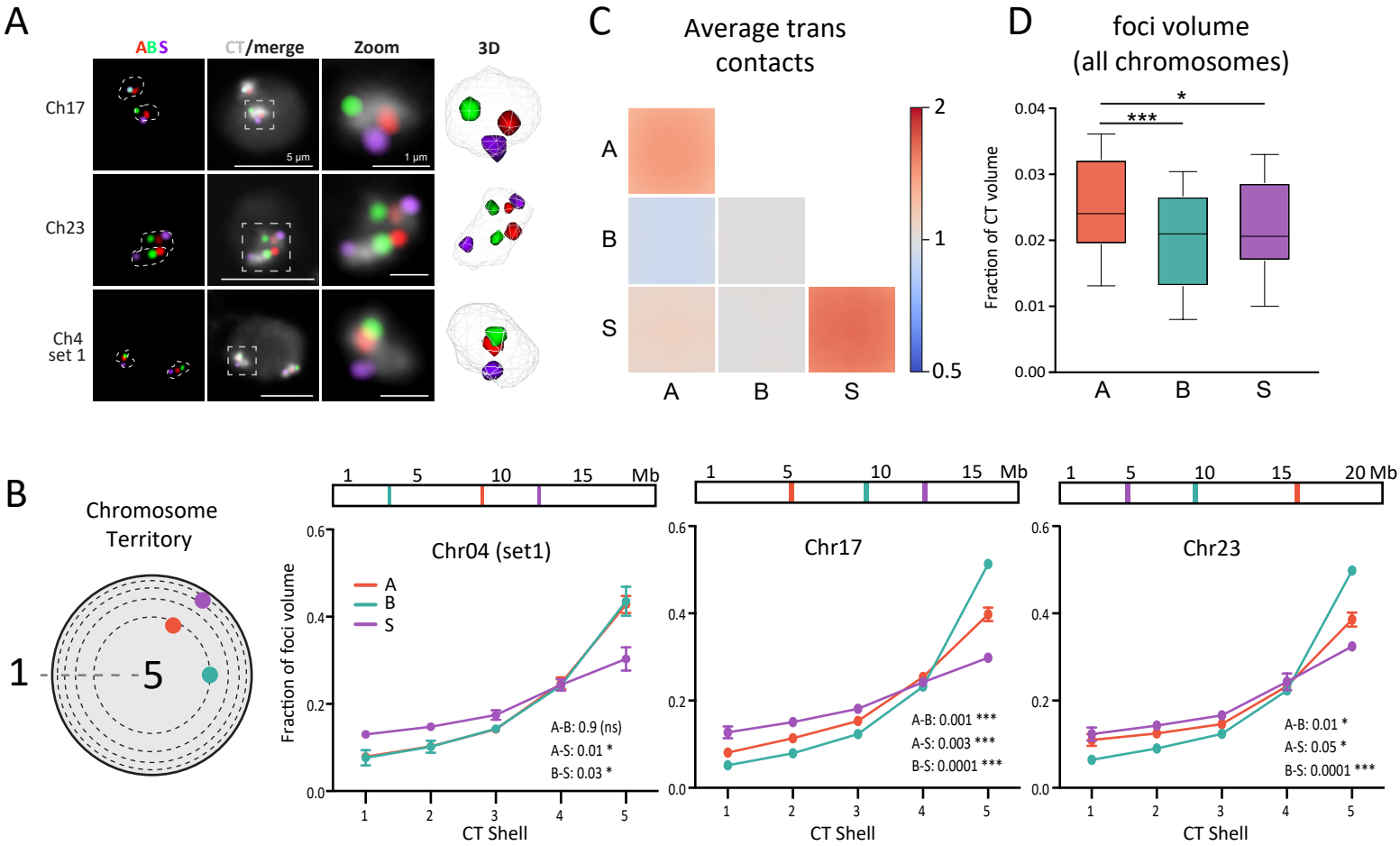

A
ABS
CT/merge
Zoom
3D
Ch17
Ch23
Ch4 set 1
5 µm
1 µm
C
Average trans contacts
A
B
S
2
1
0.5
D
foci volume (all chromosomes)
Fraction of CT volume
0.04
0.03
0.02
0.01
0.00
*
***
A
B
S
B
Chromosome Territory
1
5
Chr04 (set1)
Chr17
Chr23
Fraction of foci volume
CT Shell
Mb
A-B: 0.9 (ns)
A-S: 0.01 *
B-S: 0.03 *
A-B: 0.001 ***
A-S: 0.003 ***
B-S: 0.0001 ***
A-B: 0.01 *
A-S: 0.05 *
B-S: 0.0001 ***

Figure 5

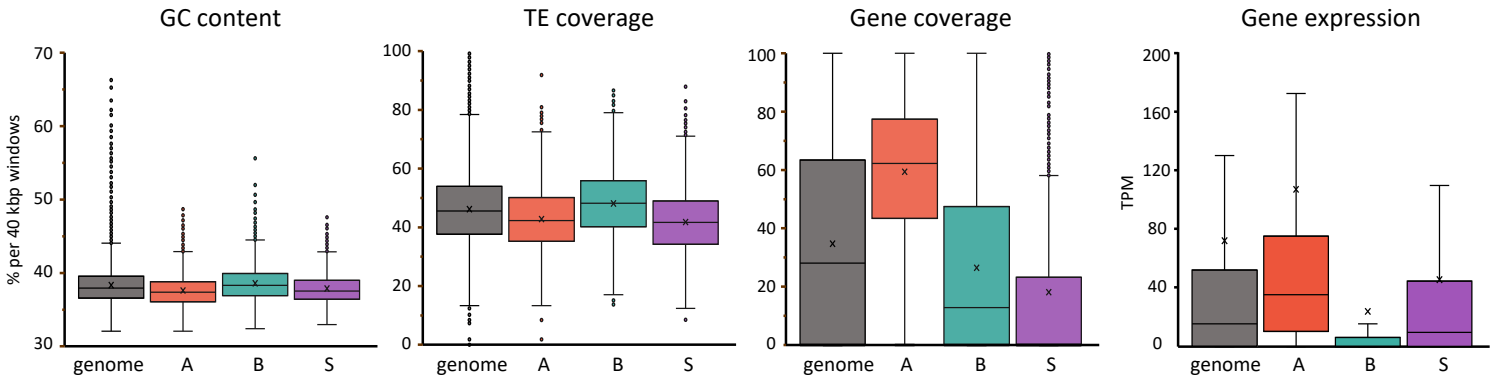
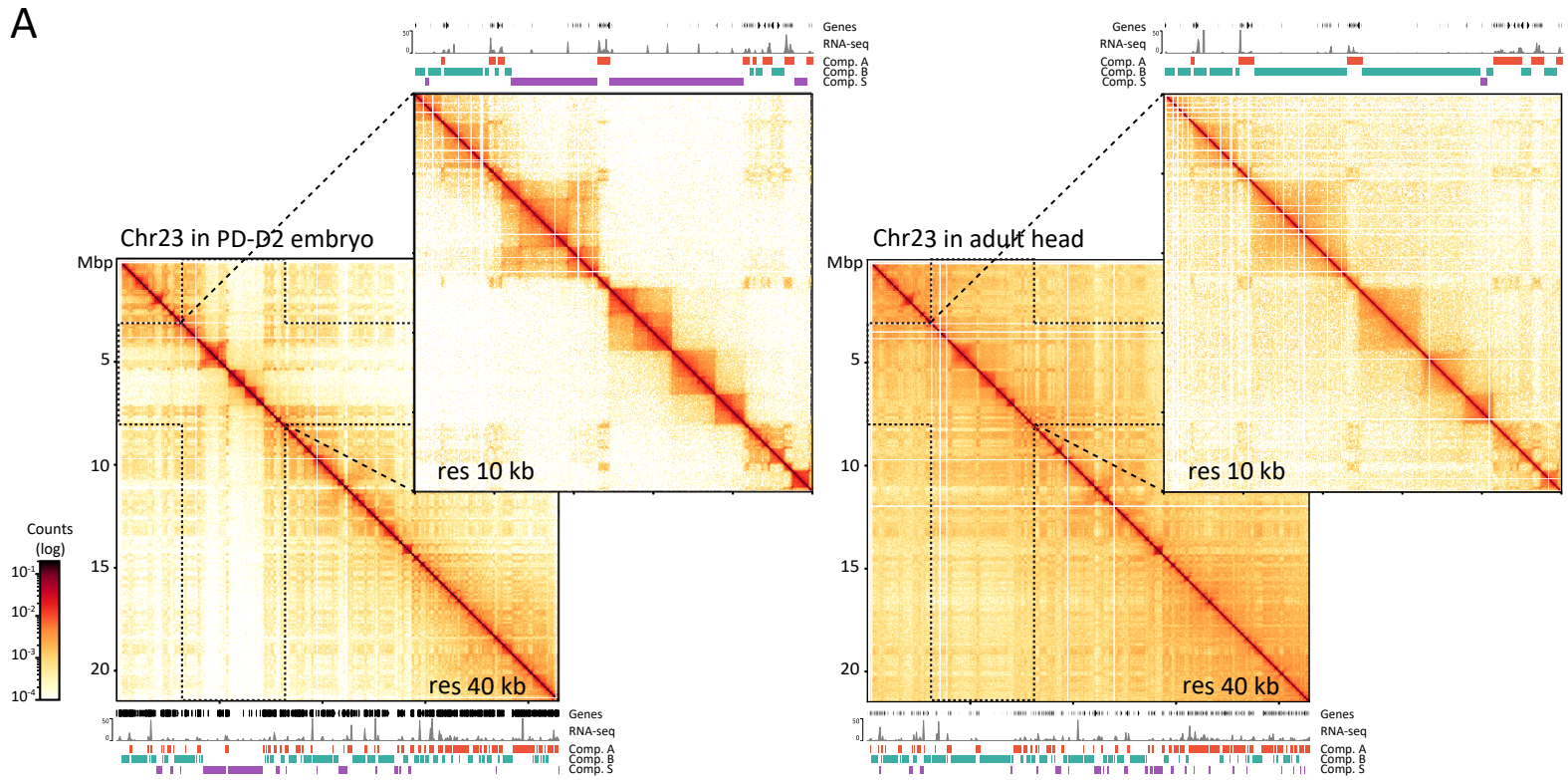
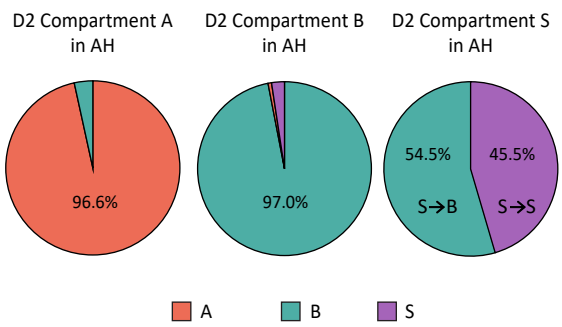

GC content
TE coverage
Gene coverage
Gene expression
% per 40 kbp windows
TPM
genome
A
B
S

Figure 6

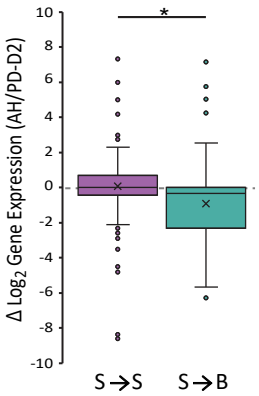
A



B



C



D

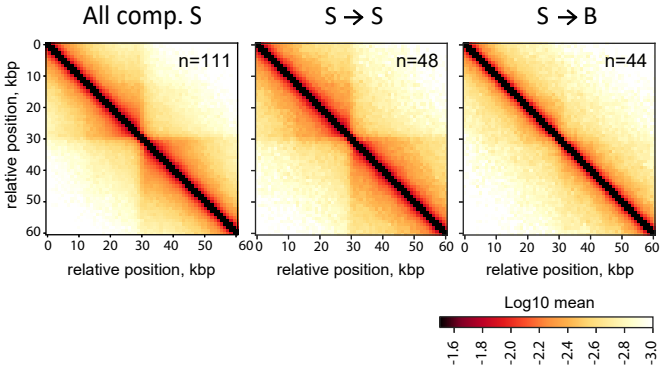


Table 1: ShinyGO Molecular Function enrichment of *B.mori* genes in S.

Pathway	GO ID	Enrichment FDR	Number of genes in category	Total number of pathway genes	Fold Enrichment
DNA binding	GO:0003677	6.10E-23	66	592	4.45
Nucleic acid binding	GO:0003676	7.43E-14	89	1461	2.43
DNA-binding transcription factor activity	GO:0003700	3.18E-13	31	220	5.63
Transcription regulator activity	GO:0140110	5.13E-12	32	262	4.88
Sequence-specific DNA binding	GO:0043565	7.10E-08	19	139	5.46
DNA-binding transcription factor activity, RNA polymerase II-specific	GO:0000981	2.811E-06	13	80	6.49
Nuclear receptor activity	GO:0004879	4.05E-03	4	12	13.32
Ligand-activated transcription factor activity	GO:0098531	4.05E-03	4	12	13.32
Steroid hormone receptor activity	GO:0003707	1.22E-02	4	16	9.99
Signaling receptor binding	GO:0005102	2.77E-02	7	67	4.17