Transcription-dependent domain-scale 3D genome organization in dinoflagellates. Response to reviewer comments

Reviewer 1:

> In this manuscript, Marinov et al. used Hi-C to examine the three-dimensional organization of the Breviolum genome. The remarkable divergence of dinoflagellate chromosomes makes this a pertinent topic not only for understanding dinoflagellate biology but concepts in genome organization more broadly. The discovery of transcription-dependent topological domains containing convergent gene arrays is striking and the development of a chromosome scale assembly of the Breviolum genome will be an invaluable resource.

> The manuscript was concise, well written, and well presented. The brevity of the article was appreciated but I think the text could be expanded for clarity and thoroughness in some instances. Nonetheless, my comments are minor and should be straightforward to address.

We thank the reviewer for the thoughtful comments.

> 1. The mention of polycistronic arrays in the abstract and introduction may be misleading because there is little evidence for nuclear polycistronic mRNAs in dinoflagellates (see Beauchemin et al. 2012. PNAS 109:15793-15798). This could be replaced with "tandem gene arrays" to be more accurate.

We thank the reviewer for the comment. We have replaced the term "polycistronic gene arrays" with "tandem gene arrays" throughout the manuscript, though we would like to point out that dinoflagellate transcripts are certainly *trans*-spliced (even if that can happen without them being polycistronic) and that there is little evidence for independent promoter elements for each gene in an array. This is a subject we expect to see dissected in detail in future studies.

> 2. It would be worthwhile to add a brief explanation of what Hi-C is at the start of paragraph four as this technique may not be widely known outside of genomics/chromatin fields.

We have added such a clarification to the text.

> 3. For Figure 1A,B,C,F legends and bin sizes should be added (bin sizes could be added to the figure legend).

We are unsure as to what the reviewer is referring to as the bin size (5-kb) is indeed already specified in the figure legend. We now also mention it in the text.

> 4. In Figure 1C, what does the dotted circle represent? It is not mentioned in the figure legend.

We have added a clarification in the figure legend.

> 5. In Figure 1A and 1C, species names should be italicized.

We have corrected this.

> 6. In Figure 1G, a y-axis should be added. Its a bit confusing that the positive values on the y-axis are above the negative values. Also for the legend, is TPM the correct unit here since the T refers to transcript and the legend indicates the plot is showing Hi-C contacts? A variation of this may be more appropriate (FPM?).

Hi-C maps are generally understood to be 2D matrices with the same axes on the x and the y. We had not added a second labeled axis in order to avoid cluttering the figure with unnecessary text.

Regarding TPMs (we assume the reviewer is referring to 1H, not 1G in that case): the RNA-seq datasets we worked with were not strand-specific, thus simple alignment and calculation of stranded tracks in RPM/FPM units is not possible. As explained in the Methods section, we instead quantified newly assembled (in genome space) transcripts, which are stranded thanks to the presence of splice junctions, then used the TPM units of the transcript quantification to assign strand-specific expression values for each exon of each transcript.

> 7. At the beginning of paragraph five, the topological domains are referred to as strong. What is meant by strong? I assume it is meant to reflect the well defined TAD boundaries/insulation but this should be clarified. It also may be worth citing Figure 1G here.

They are "strong" in the sense that there are few contacts between them and a lot of contacts within them, comparable to TADs in mammals. This domain strength is not seen in other species where supercoiling domains have been observed. The comparison with *S. pombe* and *Caulobacter* that we provide in the supplement (Supplementary Figures S15 and S16) visualizes the difference. We have added a clarification in the text to better explain that point.

> 8. In its present form, the analysis of topoisomerases is not very convincing nor a robust point supporting the plectoneme model. The main issue is that, in general, dinoflagellates have expanded gene repertoires (reflected in their tandem gene arrays) so topoisomerase may not be a special case. This is even demonstrated in Supplemental Figure 17 where the majority of non-topoisomerase genes are also increased in dinoflagellates. To make this claim, it would be important to show that topoisomerases have expanded more relative to other genes, otherwise this could be an argument for almost anything. To do this, I imagine you would need to compare copy number for all orthologues between dinoflagellates and other eukaryotes and then inspect where topoisomerases fall relative to the others.

Dinoflagellates indeed have frequently expanded gene families, but they have neither universally nor uniformly expanded gene repertoires. See for example Marinov & Lynch 2015[1] for an analysis of chromatin-related genes. The analysis in the supplement (Supplementary Figure 18) demonstrates that too (for example, RPA2 and RPA3 are not expanded), as does the analysis in the main text (Figure 2H) – TOP1 topoisomerases are not expanded, while TOP2 genes are greatly amplified, and TOP3 genes are somewhat amplified. Furthermore, expansion is specifically observed for TOP2 and not TOP1 topoisomerase, and it is TOP2 topoisomerases that are expected to deal with transcription-induced supercoiling.

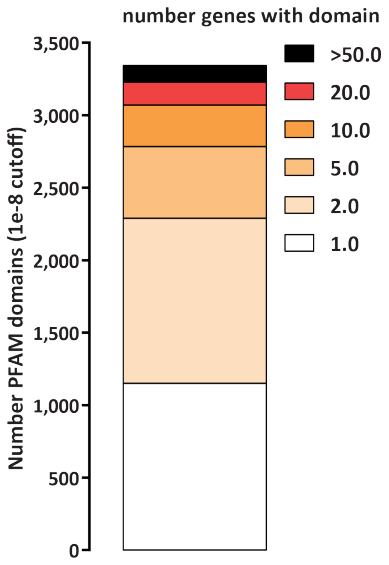
To further clarify this point, we analyzed the number of genes containing a given PFAM domains in the *B. minutum* genome as a means to measure the distribution of gene amplifications (Response To Reviewers Figure 1). This analysis demonstrates that gene family amplifications are highly non-uniform, and that most gene families are in fact not amplified.

We include analysis of PCNA and MCM genes in the supplement because there are in fact putative reasons why those genes are specifically amplified in dinoflagellates that were previously discussed in the literature (see PMID: 30468510 for more details); we have now added a reference to make that connection more explicit.

> 9. In paragraph 15, another likely rationalization for the differences between domain insulation in dinoflagellates compared the bacteria and yeast may be the structure of the chromosomes themselves (i.e., rod shaped instead of traditional interphase chromosomes). A more comparable system may be TADs in mitotic chromosomes.

The relationship between the large-scale shape of dinoflagellate chromosomes and the local domains seen in Hi-C maps is at present difficult to establish. One possibility is that the coil-like structures within the rods that are observed in high-resolution EM images correspond to the domains seen in Hi-C; this would imply one-to-one correspondence between the number and identity of dinoTADs and the number and identity of coil-like structures. However, we suspect that in reality the picture is much more complicated, as existing EM images of *Symbiodinium* cells show examples of rods with many more coils than the number of domains in any of the available pseudochromosomes. It is hard to envisage a possible mechanism through which global compaction into rods could lead to the local domains seen in Hi-C. We hope techniques like superresolution imaging could resolve these questions in the future, but this is beyond the scope of the current manuscript.

[1] Marinov GK, Lynch M. 2015. Diversity and Divergence of Dinoflagellate Histone Proteins. *G3 (Bethesda)* **6**(2):397–422.



Response to Reviewers Figure 1: Dinoflagellate genes are not uniformly massively expanded and such a general non-specific expansion does not explain the observed TOP2 amplification. Shown is the number of genes per PFAM domain (at a cutoff of 1e-8) for the *B. minitum* genome. A majority of domains are found in only one or two genes.

> 10. It would be worthwhile to discuss previous observations of plectoneme like structures in dinoflagellate chromosome spreads (see Livolant and Bouligand 1980. Chromosoma. 80, 97-118). Models have also already been proposed describing a role for plectoneme domains in dinoflagellate chromosomes (see Wong 2019 Microorganisms 7, 27; doi:10.3390/microorganisms7020027). It would be interesting to discuss these models (which were not based on Hi-C data) and observations in the context of the new data. This would also further support the presence of plectonemes given that they don't seem to be very apparent in the Hi-C heatmaps.

We thank the reviewer for the suggestions. We had in fact included a section on previous models of dinoflagellate chromosomes, but eliminated it from the original submission due to length constraints. We have now reincorporated these references and discussion into the text.

> 11. In paragraph 16, additional discussion on the role of dinoflagellate chromatin proteins, specifically DVNP and HLPs, in defining chromosome structure would be interesting. For example, the acquisition of HLPs coincides with the emergence of the unique liquid crystal chromosomes suggesting they may play a role in chromosome structure. Some in vitro work has been done on these proteins and this literature may be worth reviewing and discussing if relevant (possibly in boundary formation?) (see Chan and Wong, 2007, Nucleic Acids Research, 35:2573-2583).

We thank the reviewer for the suggestion. In our view it is more likely that the histone proteins that dinoflagellates still retain, although in a very divergent form, are involved in setting boundaries rather than the DVNPs and the HLPs; this would fit better with the fact that histones are not very abundant (and thus are more likely to associate with specific regions of the genome) while the DVNPs are the main packaging component. Although this is speculation, such a putative role is also supported by the evidence that histones play an inhibitory role on plectoneme formation, as suggested by Hi-C data from dinoflagellates and kinetoplastids. These questions remain to be resolved by future studies that directly map histone occupancy over dinoflagellate genomes; this is at present not possible due to the lack of antibodies that recognize *Symbiodinium* histones, and due to the great difficulty of successfully introducing epitope-tagged transgenes into dinoflagellate cells.

> 12. In the supplemental methods, in the last sentence of Transcription inhibition experiments, "fat" should be "at".

We thank the reviewer for spotting this typo. We have corrected it.

13. The supplemental methods seem to switch between Breviolum and Symbiodinium, this should be standardized throughout.

We thank the reviewer for pointing this out. We have standardized the species name to *B. minutum* throughout the supplement.

Reviewer 2:

The manuscript describes an analysis of the 3D organization of the B. minitum genome. This dinoflagellate is of special interest because its DNA is not packed into nucleosomes, making this organism an interesting system to study the mechanisms underlying the formation of contact domains. Authors conclude that supercoiling plays a critical role in the formation of contact domains in dinoflagellates.

The manuscript could be, in principle, interesting if the authors are able to derive general principles of 3D organization from the study of this organism. Major issues with the paper in its current form include the low resolution of the Hi-C data, which limits the ability of the authors to make meaningful conclusions, and the lack of RNA-seq data in the different conditions used to analyze the possible role of transcription in 3D organization, which makes the results difficult to interpret. The following are specific issues that I hope will help the authors improve the significance and validity of the conclusions:

We thank the reviewer for the comments and suggestions.

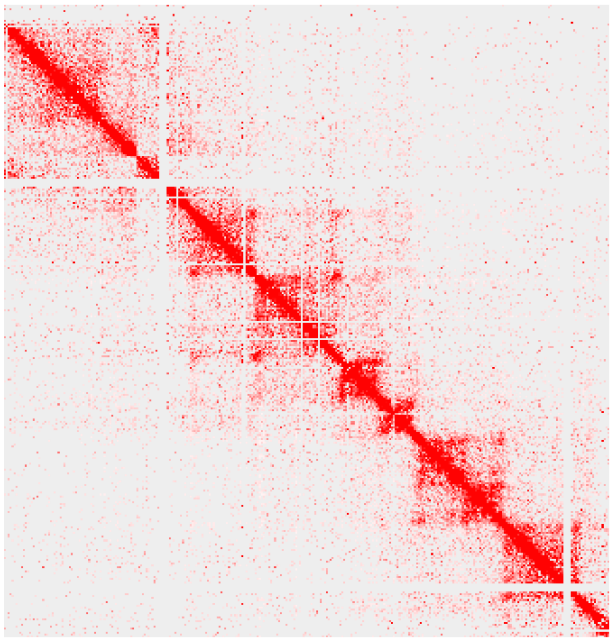
1. One critical issue in the interpretation of Hi-C data is the resolution. Since all analyses described in the manuscript involve intra-chromosomal interactions, authors should include additional columns in Table S1 indicating the fraction of intra- and inter-chromosomal contacts. Each Hi-C heatmap in the figures should indicate the resolution of the data, and scale bars should be included. The easiest way to do this would be to process the data using Juicer and to show plots from Juicebox. Each experiment should have biological, not technical, replicates and information on each replicate should be included. The number of contacts for some of the samples, as low as 15 million, is woefully inadequate. It is difficult to judge the resolution of the data because the size of the B. minitum genome to which the reads are being mapped (excluding repeated sequences) is not mentioned. Any description of the data as high-resolution should be deleted from the text unless it is really high-resolution. In mammals, Hi-C data from Rao et al Cell 2014 containing 4.5 billion contacts is the standard for high-resolution, which in this case is 1 kb.

We are afraid that there is some confusion regarding the resolution of our Hi-C data sets. For the purposes of scaffolding the genome and also for making the claim that loops are not observed, reads were pooled across the initial set of libraries (from samples not treated with transcription inhibitors), resulting in resolution of 5-kbp. The density of reads in a HiC matrix scales as slightly slower than the square of the genome size. Because the *B. miniutum* assembly is approximately 1/5th the size of the human haploid genome (∼613Mbp vs ∼3,000 Mbp; we have now changed Figure 1 to show the assembly size, and we apologize for this omission), we need far fewer contacts (perhaps 1/10th) than the 4.5 billion in Rao et al. We comfortably exceed that threshold, with ≥550M contacts.

Furthermore, the presence of loop and stripe features is readily apparent in an unrelated Hi-C dataset that we have generated for a vertebrate organism with a genome size of 1.2 Gbp, for which only 75M contacts were obtained (Response To Reviewers Figure 2). Therefore we are confident these features do not exist to the degree they do in other vertebrate organisms.

When low-depth sequencing datasets were used, they were only used for the analysis of very large-scale features, for which their depth is sufficient. To further justify this decision, we downsampled libraries to a uniform depth, and found that our conclusions were unchanged (see Response To Reviewers Figure 4 further below). This was in fact an analysis carried out prior to submitting the original manuscript.

Our results are also concordant with those of the related paper by Nand et al.[2], which mapped 3D organization in a different Symbiodiniaceae species, *Symbiodinium microadriaticum*.



Response to Reviewers Figure 2: Sufficiency of the Hi-C coverage obtained in this manuscript for the detection of loop and stripe features. Shown is a Hi-C map for a vertebrate genome whose size is ∼1.2 Gbp, and for which only 75M Hi-C contacts were obtained. Loop and stripe features are readily apparent. In contrast, the analysis presented in the main text is based on ≥550M contacts for a ∼613 Mbp genome assembly.

> 2. Figure 1F. The overlap between the TAD boundaries and the location of divergent transcription is difficult to judge visually. In fact, from this figure, the two do not appear to overlap at many boundaries. Figures S3-13 show many domain boundaries that do not correspond sites of convergent transcription and many sites of convergent transcription that do not correspond to domain boundaries. This conclusion is an important aspect of the manuscript and cannot be reached based on visual observations only. Its unclear whether Figure 1H really addresses this concern. Are all RNA-seq data plotted in this figure or on the data for the convergent genes at the boundaries? Shouldnt transcript levels be zero at the boundary? Authors should map the location of the TAD boundaries and then analyze the overlap with the location of boundaries between divergent gene arrays. Please note that the ability to properly map the location of TAD boundaries will depend on the resolution of the data.

We are confused by why the reviewer thinks that TAD boundaries do not overlap with the location of divergent transcription. There are cases where domains have a more complex structure, i.e. a rather large more diffuse domain includes several smaller weaker domains, which do consist of pairs of divergent arrays (the middle part of Supplementary Figure 4 features a prominent such example). We included those examples in the supplement to better represent the diversity of features observed in the data.

Because Symbiodiniaceae gene arrays are much longer than the Hi-C resolution we are working with, we do not expect that the resolution is problematic with regards to our overall conclusion.

As the reviewer points out, visual examples, though helpful, are insufficient to capture a particular architecture genome-wide. For this reason, we included 1H. However, RNA-seq data is not expected to be zero across TAD boundaries, as the boundaries between transcription units are much finer than the resolution at which TADs

[2]Chromosome-scale assembly of the coral endosymbiont *Symbiodinium microadriaticum* genome provides insight into the unique biology of dinoflagellate chromosomes. Ankita Nand, Ye Zhan, Octavio R. Salazar, Manuel Aranda, Christian R. Voolstra, Job Dekker. *bioRxiv* 2020.07.01.182477

are defined by TAD-calling algorithms. This is why we observe some "bleedthrough" on both sides across TAD boundaries in the metaplots. To avoid this confusion, we could remove the points $\sim \pm 20$ kbp around the slightly "fuzzy" tad boundary to avoid this confusion – we are happy to do this at the editor's discretion.

> 3. Figure 3C. Please include a distance scale and locations of divergent transcription in this panel. Visually, it appears that this TAD is composed of smaller domains that authors may be able to identify with higher-resolution Hi-C data. This is also obvious in supplementary figures 3-13.

We assume the reviewer is actually referring to Figure 1C (there is no Figure 3 in the main text). The scale is in fact already included, as Figure 1C is the inset from Figure 1B. The divergent units are also included, as Figure 1F features the exact same pseudochromosome as Figure 1C, which is also stated in the figure legend.

> 4. Page 3, top left. The nomenclature used to name "contact domains" in the mammalian literature is confusing and following the mammalian convention for dinoflagellates is even more confusing. Here authors use TAD to designate CTCF loops in mammals and suggest a different mechanism for the formation of dinoTADs because these domains lack "corner dots" and extrusion "stripes". It may not be appropriate to make this comparison. Mammalian cells depleted of CTCF lack CTCF loops but still have contact domains similar to the dinoTADs. Therefore, dinoTADs are likely the same as domains present in mammalian cells different from CTCF loops. Please see Rao et al Cell 2017 for a discussion of the nature of these domains.

We thank the reviewer for initiating this discussion. However, we do indeed think that dinoTAD represent a distinct class of features than contact domains in mammals, thus leading us to define a new term to separate them from what has previously been observed. This was done in order to avoid further confusion. Below, we discuss some of the distinctions.

It is true that compartments are preserved after CTCF depletion, and that some contact domain structure remains (Nora et al. 2017[3]). However, those contact domains are not suggested to coincide with divergent transcriptional units the way they do in dinoflagellates. Instead, they are hypothesized to arise from the preferential association of similarly modified nucleosomes.

Moreover, if the loss of histones in dinoflagellates had nothing to do with the appearance of dinoTADs in Hi-C maps, the expectation would be that the same features would be observed in kinetoplastids, which do possess the same kind of long divergent gene arrays but have retained a conventional set of histones. However, that is not the case – in fact, the available *Trypanosoma* Hi-C maps show no domain structure at all. It is also not the case that A/B compartments are observed in dinoflagellates. Thus, we think that for the time being it is best to refer to dinoflagellate domains with a distinct term.

> 5. If I understand correctly, the treatments with HS or Pol II inhibitors was not followed by RNA-seq experiments. It may be important to know how much transcription is affected in each condition in order to properly interpret the Hi-C differences. This is especially important because, unless the authors know this, it is not known whether B. minitum undergoes a heat shock response at the temperatures used and whether triptolide and a-amanitin can get inside of the cells or affect Pol II. For example, a-amanitin does not affect yeast cells because they are not permeable to this compound. Please cite the appropriate references if this is known. One additional concern is the long treatment times used here, which could affect transcription of proteins necessary for interactions detected by Hi-C and thus the observed effects could be indirect.

We fully agree with the reviewer that it is important to show that transcription was indeed inhibited. We would like, however, to point out that it is highly unlikely that α-amanitin and triptolide are not taken up by *Symbiodinium* cells – if indeed these cells were not permeable to them, we expect to see no difference between untreated and treated cells in our chromatin contact maps, and would likewise not expect a dose-dependent response to the drugs, both of which we do observe.

Dinoflagellate RNA Polymerase is indeed known to be sensitive to α-amanitin based on previous studies dating back to the 1970s (Rizzo 1979[4]), although those studies also noted some residual activity, which probably explains

[3]Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG. 2017. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**(5):930–944.e22.

[4]Rizzo PJ. 1979. RNA synthesis in isolated nuclei of the dinoflagellate *Crypthecodinium cohnii*. *J Protozool* **26**(2):290–294.

the dose dependence that we observe in our experiments. Nevertheless, we do agree that inhibition ought to be demonstrated directly within the context of our experiments.

However, RNA-seq itself at the level of gene expression is likely not informative about the extent of inhibition, as it is generally thought that there is very little regulation at the level of transcription in dinoflagellates, and substantial regulation achieved through modulation of transcript stability and translation. Thus transcript levels tell us little about actual transcriptional activity.

The optimal experiment for assessing transcriptional activity is metabolic labeling using a nucleotide analog such as 4SU/4TU and then conversion of newly incorporated thiol-uraciles into cytosines following the SLAM-seq or TUC-seq protocols. Such an experiment has been carried out once in a dinoflagellate previously, in *Karenia brevis*[5]. In that case it was found that *Karenia* does not incorporate 4SU, likely due to poor intake, but does incorporate 4TU, and that was used to determine the half-lives of mRNAs for that species. The mean half life was ∼40 hours and the median half life was 33 hours, and examination of that dataset showed that none of the chromatin-related genes (histones, DVNPs, RNA polymerase) exhibits half lives shorter than the median. If we are to take these results as a guidance, it is unlikely that effects at the level of transcript levels for particular genes had any significant influence over the results we observe at the level of Hi-C domain.

However, these results were in *K. brevis*, not a *Symbiodinium* species. To measure nascent transcripts, we have carried out numerous attempts to metabolically label transcripts in *B. minutum*. However, it appears that *B. minutum* cells are not permeable not just to 4SU, but also to other nucleotide/nucleoside analogs, such as 4TU, which precludes the possibility of using SLAM-seq/TUC-seq for this purpose.

In the absence of a SLAM-seq/TUC-seq experiment, the next best available way of assessing transcriptional activity is to look for nascent RNAs by measuring the fraction of unspliced intronic reads in RNA-seq datasets, although this provides more of a qualitative than quantitative measure of inhibition (as we know very little about the stability of unspliced transcripts in dinoflagellate cells). Therefore we carried out another time course of transcriptional inhibition and performed RNA-seq experiments to measure nascent RNA by calculating the fraction of unspliced intronic reads as a proxy for nascent transcription. We observe reduction of nascent RNA in both α-amanitin- and triptolide-treated cells (Supplementary Figure 30 in the revised submission).

Regarding secondary effects of inhibition in the nucleus, we do not observe large-scale change in transcript levels between treated and untreated cells (Supplementary Figure 31 in the revised submission), making it unlikely that the effects we observe are due to changes in gene expression of particular sets of genes (with all caveats already mentioned above about interpreting the significance of transcript levels in dinoflagellates). Moreover, we again cite the long half-life of dinoflagellate mRNA and that no chromatin-related genes (specifically, histones, DVNPs, or polymerases) exhibited half-lives shorter than the median in the *Karenia* study. We believe this point from the literature, combined with our observation of very few and modest transcript level changes after inhibition, is evidence that decompaction of dinoTADs after transcription inhibition is most likely due directly to loss of transcription, and not due to secondary effects.

> 6. It is unclear what results led the authors to conclude that supercoiling must be involved in the formation of contact domains in B. minitum. Although the authors name the contact domains observed in Hi-C as TADs, there is no evidence that these domains are topological in any organism. The positive/negative supercoiling generated during transcription is released as Pol II advances and it does not accumulate at the 3 of genes. Domains observed in Hi-C data are composed of interactions mediated by proteins that are sufficiently closed to be crosslinked by formaldehyde, which is a short-range (2-3 A) cross-linker. Therefore, proteins mediating interactions in Hi-C are probably interacting biochemically with each other. A contact domain is formed because of changes in the directionality of interactions, which is probably determined by the nature of the proteins present in different regions of the genome rather than by some physical force caused by supercoiling. In other organisms there is no correlation between the location of supercoils determined by psoralen experiments and location of domain boundaries. In addition to inhibiting transcription, triptolide and a-amanitin result in degradation of Poll II, which may also affect the recruitment of other components of the transcription process. Performing ChIP-seq with antibodies to Pol II phosphorylated in Ser2 or ser5, if the polymerase of B. minitum has CTDs, may be a way to explore this question in more detail. In yeast cohesin accumulates at the 3 ends of genes as a consequence of transcription. If this happens in B. minitum and cohesin makes a loop

[5]Morey JS, Van Dolah FM. 2013. Global analysis of mRNA half-lives and de novo transcription in a dinoflagellate, *Karenia brevis*. *PLoS One* **8**(6):e66347.

between the 5 and 3 while it happens (some evidence for this is presented in Zheng et al Nature 2019 PMID 30778195), it may explain the formation of domains in dinoflagellates. This possibility is also supported by results in Drosophila where single DpnII fragment resolution (250 bp on average) Hi-C data suggest the formation of loops between the 5 and 3 of individual genes. Higher resolution Hi-C maps for B. minitum may allow the authors to map subdomains within the dinoTADs and analyze their correlation with genomic and transcription feature. A more accurate map of boundaries may allow the authors to gain insights into these different possibilities.

While dinoflagellate have lost nucleosomes as their main packaging component their DNA is not naked – they have DVNPs and HLPs. Whether these proteins form specific interactions with each other is not known at present as neither they nor their post-translational modifications have been studied in detail, but strong interactions are not necessary to explain the crosslinking to each other of distal regions of the genome in a highly supercoiled state. First, such supercoiling-driven domains are well known in prokaryotes as already discussed in the manuscript, second, polymer simulations show that supercoiling can in fact bring very distal regions of DNA in very close proximity to each other (see for example Krajina & Spakowitz 2016[6]), in which context we can reasonably expect DVNPs to be crosslinked both to DNA and to each other.

To address some of these critiques, we carried out an additional Hi-C experiment in which the denaturation step was omitted (which would in theory better preserve protein-protein contacts), but while we observe accentuation of the strength of dinoTADs in that dataset, we do not find signs of specific contacts.

As noted above, the map generated by pooling contacts across multiple libraries is of sufficient depth to detect loops for the size of the *B. minutum* assembly, but we have carefully manually scanned the whole genome and observe no evidence for any such features.

Unfortunately, ChIP-seq for Pol II in dinoflagellates is currently unobtainable, because their CTD tails lack the phosphorylation motifs typical to conventional eukaryotes[7] and there is no antibody that reliably recognizes any of the dinoflagellate Pol II subunits.

7. I wonder if the authors could learn something about mechanisms of 3D organization in B. minitum by comparing their data to that of S. cerevisiae, where genes are also closely packed. Do domain boundaries in cerevisiae form at each gene boundary, independent of the orientation, or only at sites of convergent genes?

As noted above and in the original submission, kinetoplastids are a much better "natural experiment" than *S. cerevisiae* because they, as a result of convergent evolution, share almost all of the highly divergent features of dinoflagellates – long gene arrays, *trans*-slicing, loss of transcriptional regulation, shift to post-transcriptional regulation as the main mechanism for regulation gene expression, etc. Expect for one – they have conventional nucleosomal chromatin while dinoflagellates have lost it.

Such a collection of characteristics represents as close to a controlled natural experiment as one can hope for, and kinetoplastids show no signs of domain structure associated with gene arrays (Supplementary Figure 16 in the original manuscript). We feel that this is further strong evidence for the model that we propose, in which the absence of histones allows for supercoiling domains to form while their presence and/or interactions with each other override supercoiling's manifestation in Hi-C maps in most eukaryotes. In short, we feel comparison to kinetoplastids is a more highly controlled and relevant consideration than a comparison to *S. cerevisiae*.

Reviewer 3:

In eukaryotes, genome folding is known to be driven by cohesin-mediated loop extrusion, together with CTCF in vertebrates, to establish topologically associating domains (TADs). The role of transcription in genome folding remains elusive. Besides, replication has also been linked to the establishment of genome structure.

[6]Krajina BA, Spakowitz AJ. 2016. *Biophys J* **111**(7):1339–1349. Large-Scale Conformational Transitions in Supercoiled DNA Revealed by Coarse-Grained Simulation.

[7]Marinov GK, Lynch M. 2015. Diversity and Divergence of Dinoflagellate Histone Proteins. *G3 (Bethesda)* **6**(2):397–422.

In this paper, Marinov, Trevino, Xiang and colleagues carry out Hi-C analysis on a species of dinoflagellate, Breviolum minutem, generating 91 major pseudochromosomes, and reveal dinoTADs, appearing as squares on heatmaps, similar to TADs in metazoan genomes. But these dinoTADs are much larger than mammalian TADs. Each dinoTAD corresponds to a pair of divergent gene arrays with a boundary coinciding with convergence between gene blocks.

After transcription inhibition by different methods, these dinoTADs are weakened to different extents. The primary conclusion of this paper is that transcription-induced supercoiling is the primary force driving dinoTAD formation. Though it is plausible, there are a few concerns that require further clarification from the authors.

We thank the reviewer for the comments and suggestions.

Major comments:

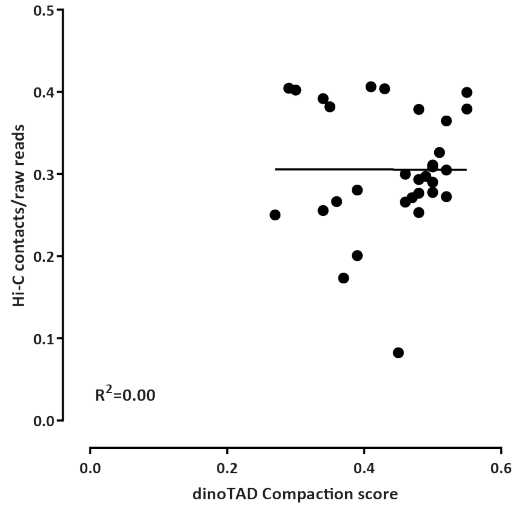
1. The quality of Hi-C experiments may cause unexpected differences in heatmaps, and sometimes may complicate the interpretation. In Supplementary Table 1, the percentage of the Hi-C contacts over the raw read pairs differs for different experimental conditions. For control Hi-C, 27 and 34 degree cultures, the average percentage of Hi-C contact is 33% (27% to 41%). For samples treated with low concentration of alpha amanitin, the average percentage is 31% (26 to 40%), similar to control. Under this condition, the dinoTADs are not obviously weakened to my eyes (Supplementary Figures 21 to 23). However, when the alpha amanitin concentration is increased, the average Hi-C contact percentage drops to averagely 22% (17% to 27%), which, I think, is significantly lower than Hi-C of controls, heat shocked, and low alpha amanitin concentration treated conditions. For triptolide-treated samples, the average percentage of Hi-C contacts is 29% (25% to 33%), similar to Hi-C of controls, heat shocked and low alpha aminitin concentration treated conditions. And accordingly, the dinoTADs seem to be better maintained. Whether the percentage of Hi-C contacts in raw total paired reads can be used as an indicator of Hi-C quality can be debated. However, the correlation of low Hi-C contact percentage and weakened dinoTAD partition may not totally be a coincidence.

We appreciate the reviewer's concern and agree that controlling for Hi-C library metrics is important for interpreting results. We note that the fraction of reads that result in contacts is a function of a myriad factors having to do with DNA isolation, library prep, the details of how the sequencing run went, and many others. Thus is it is challenging to draw conclusions about the quality of the Hi-C itself from that metric alone. We plotted the ratio between Hi-C contacts and raw reads versus the strength of compaction of dinoTADs observed in the Hi-C maps, and found no positive relationship between the two (Response To Reviewers Figure 3). We also note that none of our datasets show other signs of Hi-C failure or suspect quality, such as non-uniform distribution of the four possible orientations of ligation events.

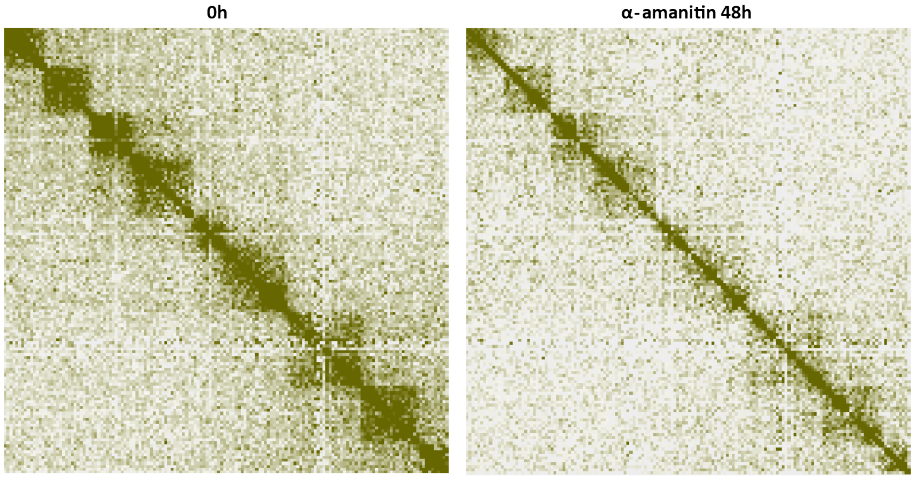
To further assuage this concern, we also reanalyzed the α-amanitin comparison by subsampling libraries down to the same number of Hi-C contacts (Response To Reviewers Figure 4); this did not affect results. We had in fact carried out this analysis before the original submission, since we too had shared these concerns.

2. If transcription inhibition does degrade dinoTADs (Figure 2), we may well expect that after the removal of the inhibiting drugs, transcription will resume and dinoTADs will reappear. The authors may consider to do some Hi-C on cells recovering from transcription inhibition and to show if dinoTADs will reemerge.

We have now carried out such an experiment, in which we inhibited transcription for 48 hours with α-amanitin and triptolide, then removed the drug, and carried out Hi-C after another 48 hours. We do observe partial restoration of domains at that point, but it also appears that transcription may not have fully restarted by that time (based on our estimates of nascent RNA in RNA-seq datasets). We would continue the time course, but we are concerned that at much further past that point, cells dividing and potentially restoring domains through other mechanisms could become a confounding factor.



Response to Reviewers Figure 3: No relationship between the Hi-C contacts/raw reads ratio and observed dinoTAD compaction. The compaction scores were calculated as the ratio of the average contacts within a dinoTAD at a distance of 250 kbp vs. the average contacts off-the diagonal (at 50,000 kb).



Response to Reviewers Figure 4: Subsampling Hi-C libraries to the same number of contacts (down to the number of contacts in the less deeply sequenced one) reveals the same pattern of dinoTAD compaction as observed with all reads. Shown is pseudochromosome 10 as in the main figures (the libraries shown are the "α-amanitin 0h" and "α-amanitin 48h" from the top row in Supplementary Figure 22).

> 3. In Figure 2H, topological domain formation is proposed to be established by transcription. If unidirectional genes are in a block twisted together and separated from genes of different orientation in the adjacent block, which is still in the same dinoTAD, we would expect to see two sub-TADs within a large dinoTAD. Actually, if we examine the Hi-C heatmap plotted at 5kb, indeed, sub-TADs corresponding to single unidirectional gene blocks can be observed (Supplementary Figures 3-13). The authors may reexamine the heatmaps of Hi-C at high resolution (5kb) to reveal to what extent these substructures exist in the genome.

We thank the reviewer for pointing out this trend, which we have observed ourselves too. Indeed, this trend is not only observable in the Supplementary Figures but also in Figure 1B and 1C (all of these figures are already at 5-kb resolution). However, global aggregate analyses obscure this trend because of the wide variation in local gene

array structure. We have revised text to more directly point to that relationship.

> 4. Following point #3, florescence in situ hybridization (FISH) may be carried out to show if genes within the same block are closer than genes in adjacent blocks that are in the same dinoTAD.

We expect direct imaging approaches to resolve a lot of remaining questions about dinoflagellate chromosome organization, but these experiments are part of future efforts and are beyond the scope of the current study.

> 5. Another force that the authors overlooked is replication. How long is the cell cycle? The time length of transcription inhibition may just be enough to reduce gene expression to a level that arrests cells at a specific phase of cell cycle. The authors may need to consider inhibiting replication as well to see if the dinoTADs will be weakened.

We agree with the reviewer that replication could be a driving force for topological strain. However, we note that *B. minutum* cells grow slowly (with doubling times of several days), meanwhile the effects we observe with triptolide (a relatively fast acting inhibitor) are already visible at 8 hours, and those with α-amanitin (which takes at least 12 hours to begin acting, likely due to slow uptake) are very strong at 24 hours. Due to this separation of timescales, we believe replication is unlikely to be confounding the effects of decompaction following transcription inhibition. The question how dinoflagellate chromosome structure changes during cell cycle progression is an important one, but addressing these questions will require establishing protocols for synchronizing and/or isolating specific subpopulation of cells. Given the discrepancy in timing of cell growth versus the relatively fast acting effects following transcription inhibition, we believe this type of experiment to be beyond the scope of the current study.

> Minor comments:
>
> 1. Convergent gene transcription at borders was reported in Drosophila genome by Hou and colleagues in 2012 (Mol. Cell). Despite the huge differences between the genomes of insects and dinoflagellates, it would be interesting to discuss if convergent genes frequently found at boundaries could be evolutionarily conserved.

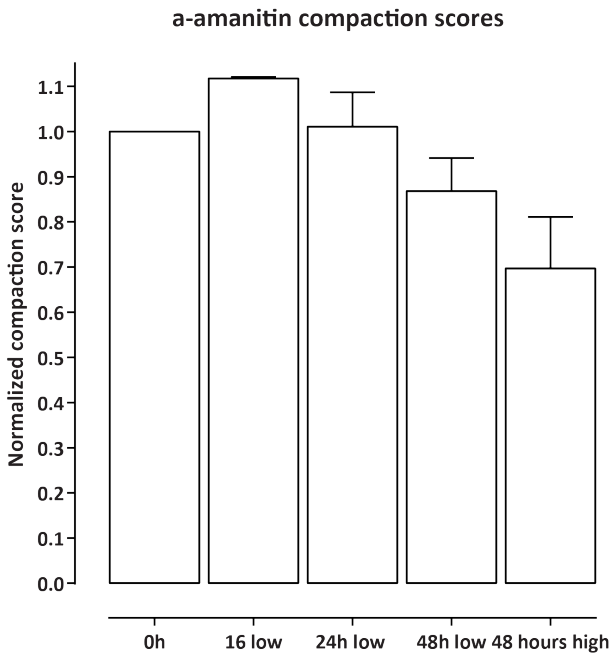
We agree that investigating the relationship between genome architecture and evolutionary conservation is a fascinating area of investigation. However, metazoans and dinoflagellates are separated by about a billion and a half years of evolution, and dinoflagellates also represent a radical departure from conventional genome organization as their genes have become organized into arrays. Essentially no conservation of synteny exists across such vast evolutionary time scales and radically divergent genomic architectures. This does not preclude a further analysis of the phenomenon of synteny at topological domain boundaries in metazoans, which we expect might yield more fruitful insights in future studies.

> 2. Could the authors use any metrics to quantitatively measure the differences, for example, in the border separation, or directionality index, or the dinoTADs internal strength?

We thank the reviewer for this suggestion, which sharpens our existing qualitative and quantitative descriptions of the topological phenomena observed.

We have found a metric that we believe captures the behavior of dinoTADs called a "compaction score". This is calculated as the ratio of the average contacts within a dinoTAD at a certain distance relative to the average contacts just off-the diagonal (also used in the Response To Reviewers Figure 3). How this metric capture decompaction in an α-amanitin time course is shown in Response To Reviewers Figure 5.

As this metric differs slightly from the "insulation score" metrics that have been used previously in the Hi-C literature, for example, to quantify TAD strength in mammalian cells, we feel it necessary to note how the mechanisms of formation of dinoTADs differ from those of mammalian TADs. They are not quite "insulation" domains in the way that mammalian TADs are, i.e., in having a boundary that "blocks" interactions across it. The formation of these potentially supercoiled domains in fact makes very long-range contacts across dinoTAD boundaries more likely, when regions of neighboring supercoiled domains come in close contact to each other, which they would not do as often if they were not supercoiled (see the polymer simulation studies referred to above), and vice versa, the dissolution of the domains also decreases the absolute level of contacts across dinoTADs boundaries. The standard TAD insulation scores thus often do not fully capture the quite clearly observable disappearance of dinoTADs upon transcriptional inhibition.



Response to Reviewers Figure 5: Quantifying dinoTAD compaction and decompaction using compaction scores. The compaction scores were calculated as the ratio of the average contacts within a dinoTAD at a distance of 250 kbp vs. the average contacts off-the diagonal (at 50,000 kb), then normalized to the matching untreated sample.

3. Color bars are missing for nearly all Hi-C heatmaps in both Figures and supplementary materials.

We have now added color bars to all metaplots; we apologize for this omission.

For context, local Hi-C maps have often not featured color bars in the Hi-C literature as the normalizations applied make the units in Hi-C matrices not directly interpretable in the same way that RPM/TPM/RPKM/FPKM units are in ChIP-seq/ATAC-seq/RNA-seq/etc. datasets.