Transcription-dependent domain-scale 3D genome organization in dinoflagellates

Georgi K. Marinov[1,*,#], Alexandro E. Trevino[3,7,*], Tingting Xiang[2,8,*], Anshul Kundaje[1,6], Arthur R. Grossman[2], and William J. Greenleaf[1,3,4,5,#]

[1] *Department of Genetics, Stanford University, Stanford, California 94305, USA*
[2] *Carnegie Institution for Science, Department of Plant Biology, Stanford, California 94305, USA*
[3] *Center for Personal Dynamic Regulomes, Stanford University, Stanford, California 94305, USA*
[4] *Department of Applied Physics, Stanford University, Stanford, California 94305, USA*
[5] *Chan Zuckerberg Biohub, San Francisco, California, USA*
[6] *Department of Computer Science, Stanford University, Stanford, California 94305, USA*
[7] *Department of Bioengineering, Stanford University, Stanford, California 94305, USA*
[8] *Department of Biological Sciences, University of North Carolina at Charlotte, Charlotte, NC 28223, USA*
[*] *These authors contributed equally to this work*
[#] *Corresponding author*

Abstract

Dinoflagellate chromosomes represent a unique evolutionary experiment, as they exist in a permanently condensed, liquid crystalline state, are not packaged by histones, and contain genes organized into tandem gene arrays, with minimal transcriptional regulation. We analyze the 3D genome of *Breviolum minutum*, and find large topological domains without chromatin loops, demarcated by convergent gene array boundaries ("dinoTADs). Transcriptional inhibition degrades dinoTADs, implicating transcription-induced supercoiling as the primary topological force in dinoflagellates.

The three-dimensional (3D) genome architecture of cells has functional consequences for gene regulation, organismal development, replication, and mutational processes. Mechanisms known to drive genome folding in eukaryotes include constraints on cohesin-mediated loop extrusion – imposed by CTCF in vertebrates – that generate topologically associating domains (TADs), and self-associations between similar chromatin states that form compartments[1]. However, the extent to which genome function itself may influence genome folding, for example through transcriptional activity, is poorly understood. There has also been little exploration of 3D organization across eukaryotes, even though major deviations from conventional norms are known to exist, presenting natural experiments that may reveal deeper underlying organizational principles masked in other lineages.

Dinoflagellates are the most radical such departure. They are a diverse, widespread clade playing major roles in aquatic ecosystems, for example, as symbiionts of corals, providing the metabolic basis for reef ecosystems. Dinoflagellates possess numerous highly divergent molecular features[2], including, uniquely among eukaryotes, the loss of nucleosomal packaging of chromatin. Histones are extremely conserved across eukaryotes, were present in their current form already in the Last Eukaryotic Common Ancestor[3], and they and their posttranslational modifications are pivotal to all biochemical processes involving chromatin.

Dinoflagellates are the sole known exception. Their chromosomes exist in a liquid crystalline state, are permanently condensed throughout the cell cycle, and, although highly divergent histone genes are retained in their genomes[4], a combination of virus-derived nucleoproteins and bacterial-derived histone-like proteins have taken over as main packaging components[5]. Dinoflagellate genomes are often huge (up to ≥200 Gbp), genes are organized into tandem gene arrays, individual mRNAs are generated through *trans*-splicing, and transcriptional regulation is largely absent[5]. These fascinating features simultaneously pose intriguing questions regarding the adaptation of transcriptional and regulatory mechanisms to the absence of nucleosomes, and provide a unique opportunity to explore the biophysical forces underlying genomic organization in the context of a large eukaryotic genome nearly devoid of nucleosomes.

To explore these questions, we performed applied chromosome conformation capture using Hi-C on the coral sym-

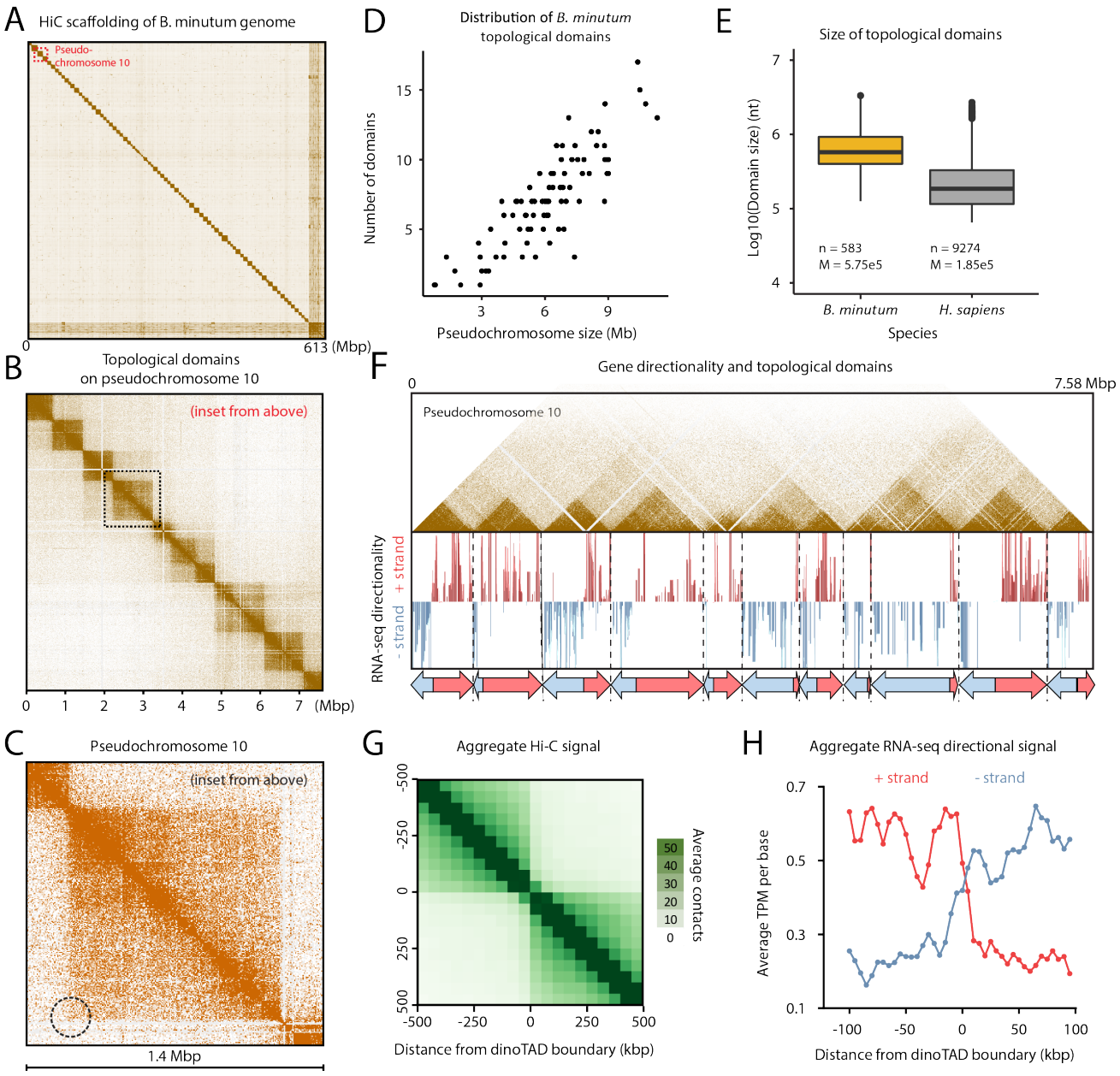

Figure 1: *B. minutum* genome is physically partitioned into dinoTADs defined by tandem gene arrays. (A) Hi-C scaffolding of the *B. minutum* draft genome assembly. (B) Inset from (A). KR-normalized 5-kb resolution Hi-C map for pseudochromosome 10. (C) Inset from (B). Hi-C loops and stripes are not observed in dinoTADs (dotted circle notes where a loop would be). (D) Scales of chromosome size with dinoTAD number. (E) Comparison of human and *B. minutum* topological domain sizes. (F) Hi-C map (5-kb resolution) for pseudochromosome 10 together with forward- and reverse-strand transcript levels and gene arrays. (G) Average Hi-C contacts across dinoTAD boundaries. (H) Average forward- and reverse-strand RNA-seq levels across dinoTAD boundaries.

biont *Breviolum minutum*. We generated multiple libraries under standard growth conditions and for cells grown at elevated temperature, obtaining ∼150–220 million Hi-C contacts for each (Supplementary Table 1). We pooled these libraries to generate a chromosome-level scaffolding of the previously fragmented *B. minutum* assembly[6]. We identified 91 major pseudochromosomes (≥ 500 kbp), encompassing ∼94% of the total sequence (Fig. 1A-B; Supplementary Fig. 1A), the longest being ∼11 Mbp in size, with a median length of 6.7 Mbp (Supplementary Fig. 1A). At 1-Mbp resolution, they exhibit a bipartite (occasionally tripartite) structure (Supplementary Fig. 2).

High-resolution (5-kbp) maps revealed well-defined (comparably so to those observed in mammals) topological

domains, ≤200 kbp–≥2 Mbp in size (Fig. 1B-E; Supplementary Fig. 3–12). In mammals, TAD boundaries are demarcated by CTCF sites blocking loop extrusion, reflected in Hi-C maps by chromatin loops and "stripes". We observe no loop or stripe features in *B. minutum* (Fig. 1C), suggesting a different mechanism for the formation of dinoflagellate TADs, which we term "dinoTADs". Omitting the denaturation step in the Hi-C protocol, which should better preserve protein-protein contacts strongly accentuated dinoTADs, but still did not reveal signs of loops or loop extrusion domains; Supplementary Fig. 14). DinoTAD number correlates with chromosome size (Fig. 1D), and they are considerably larger than mammalian TADs (Fig. 1E).

We next compared Hi-C maps to available annotation features. Remarkably, we found that each dinoTAD corresponds to a pair of divergent gene arrays (Fig. 1F), and dinoTAD boundaries coincide with convergence between gene arrays (Fig. 1G-H).

Numerous models for dinoflagellate chromosome organization have been suggested since the 1960s, primarily based on electron microscopy. These include proposals that chromosomes are organized as "toroidal chromonemas"[8], "stacks of discs"[9], "cored pineapples"[10], or around "central core fibers"[11]. Most of these models imply specific topological constraints maintaining the proposed shapes and are not directly reconcilable with our Hi-C observations.

Instead, the correspondence between dinoTADs and gene arrays suggested a role for transcription in their formation. Although TADs form independently of transcription in metazoan cells, transcription-induced self-interacting domains have been previously demonstrated in bacteria[12], and similar mechanisms have been proposed to explain some topological features in fission yeast[13]. We also note that a handful of models of dinoflagellate chromosome structure have suggested the presence of coil/plectoneme-like features[14,15], but without relating them to gene arrays and transcription. This model is also supported by the observation that frequently each dinoTAD can be divided into more diffuse "sub-dinoTADs" corresponding to the two individual gene arrays in a pair (Fig. 1C; Supplementary Fig. 3–12), which could be the result of torsion generated in each direction of transcription.

The model makes a clear prediction – inhibition of transcription should result in dinoTADs decompaction. To test this relationship, we first compared Hi-C maps for cells grown at 34 °C versus 27 °C, as heat stress could result in general transcription reduction[16]. We observed mild decompaction of dinoTADs at 34 °C, though domains remained intact (Supplementary Fig. 19–21).

We next carried out chemical transcription inhibition experiments. Since transcription inhibition conditions for *B. minutum* are not well established, we chose two inhibitors – triptolide and α-amanitin – with distinct mechanisms of action, and assayed multiple time points and doses (Fig. 2A-B). Amanitin directly inhibits RNA Polymerase II and is slow acting, while triptolide quickly blocks initiation by targeting the TFIIH XPB subunit[17]. While dinoflagellate RNA polymerase II has been reported to be sensitive to α-amanitin, it is possible that the sensitivity is somewhat partial[18]; in addition, the *B. minutum* XPB homolog is highly divergent[6], thus a moderate inhibition effect is not unexpected. We therefore carried out several experiments to directly estimate the extent of transcription inhibition. Direct metabolic labeling approaches[19] were unsuccessful as it appears that Symbiodiniaceae cells are impermeable to nucleotide and nucleoside analogs such as 4SU and 4TU. We were, however, able to qualitatively assess inhibition using the proxy of nascent RNA as measured by the proportion of unspliced reads in PolyA+ RNA-seq datasets (Supplementary Fig. 30); we observe more than 50% reduction in unspliced reads in both α-amanitin- and triptolide-cells after 48 hours suggesting that transcription has indeed been inhibited. We also did not observe large-scale changes in the levels of individual transcripts (Supplementary Fig. 31). Finally, even at high doses, α-amanitin treatment did not detectably affect photosynthetic efficiency or cell viability relative to controls (Fig. 2C), excluding cell death as a confounding factor.

Strikingly, α-amanitin treatment resulted in a dose-dependent, progressive dinoTAD decompaction (Fig. 2D,F; Supplementary Fig. 22–25). These effects were observed in both technical and biological replicates (Supplementary Fig. 22–25). We also observed clear dose-dependent blurring of dinoTAD boundaries after triptolide treatment, though broad dinoTAD-like structures remained visible to a greater extent than in α-amanitin-treated cells (Fig. 2E-F; Supplementary Fig. 26–29).

These experiments support a transcription-induced supercoiling model for dinoTAD formation. Torque generated by active polymerases produces positive/negative supercoiling ahead of/behind the transcription bubble. This can alter the twist of the double helix or induce superhelical writhe, which in turn can be accommodated through nucleosome remodeling, local alterations in DNA secondary structure, or formation of writhed structures such as plectonemes[20], from which we would expect strong Hi-C signals comprising our observed domains.

Although other topological constraints might also be involved, supercoiling-induced plectoneme formation over gene arrays is an intuitive mechanistic explanation for the presence of dinoTADs. An examination of dinoflagellate gene repertoires also corroborates this model, revealing a striking, dinoflagellate-specific expansion of topoisomerase II- and topoisomerase III-like genes (Fig. 1D; Supplementary Fig. 18; Supplementary Table 2), further suggestive of contending with increased levels of writhed forms of helical twist.

Comparison with self-interacting domains in bacteria or *S. pombe* shows much stronger topological insulation for dinoTADs (Supplementary Fig. 15) and 16)). Remarkably, no TAD domains are observed in kinetoplastids, the other

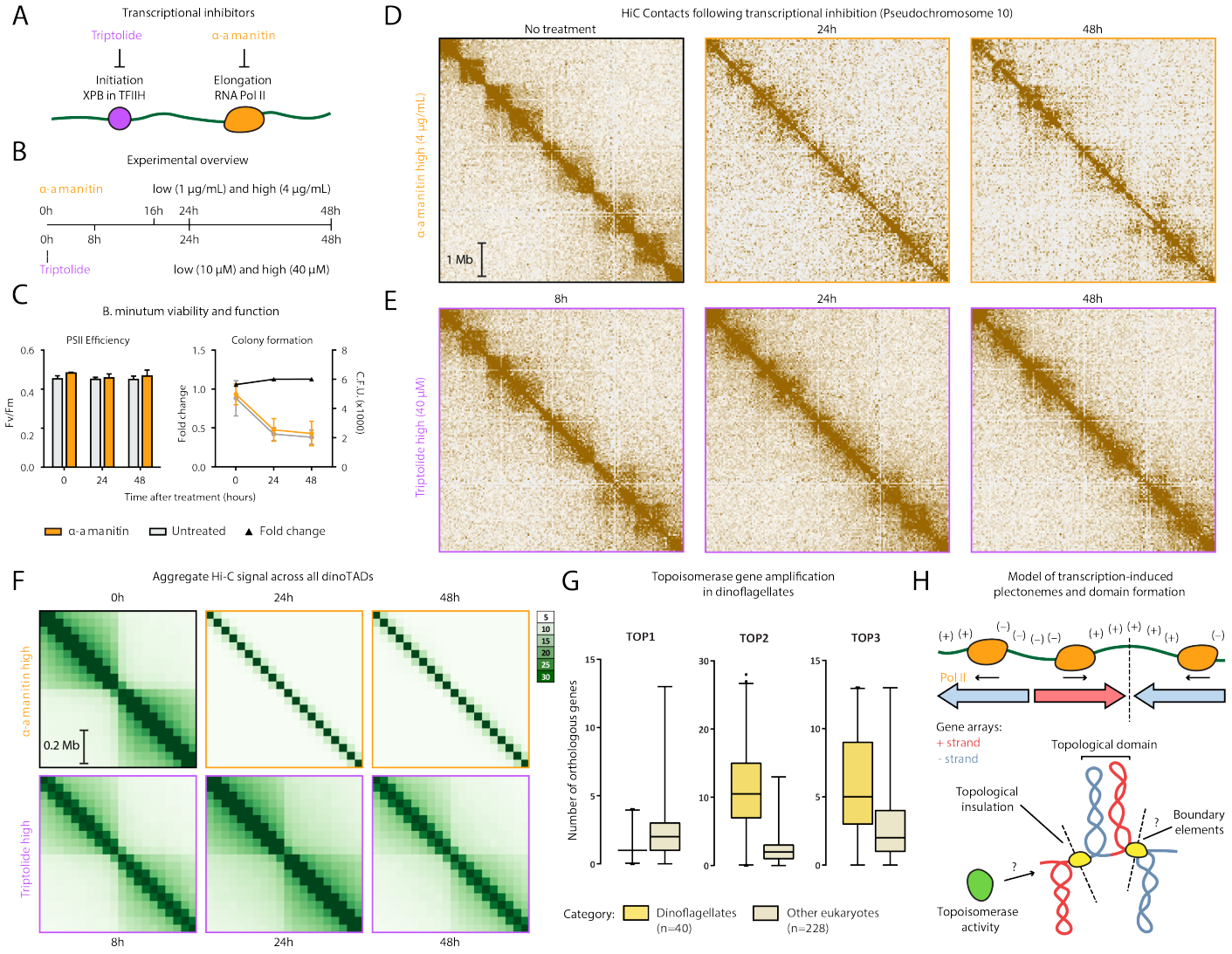

Figure 2: Decompaction of dinoTADs upon application of transcriptional inhibitors and the transcription-induced supercoiling model for their formation. Shown is pseudochromosome 10 as in Fig. 1. (A-B) Outline of transcription inhibition time course experiments. (C) Comparison of cell function, measured by PSII photosythetic efficiency, and cell viability, measured by colony formation (right), between α-amanitin-treated and untreated cells. (D) KR-normalized Hi-C maps (50-kb resolution) show marked loss of dinoTADs after α-amanitin treatment. (E) Hi-C maps show reduction of insulation at dinoTAD boundaries after triptolide treatment. (F) Metaplots of Hi-C signal around domain boundaries (50-kb resolution). (G) Amplification of *TOP2* and *TOP3* topoisomerases in dinoflagellates (based on MMETSP[7] transcriptome assemblies). (H) Transcription-induced supercoiling as driver of dinoflagellate chromatin folding. Transcribing polymerases introduce negative/positive DNA supercoiling behind/ahead of the transcription machinery. Interactions within supercoiled domains could explain the physical association of divergently-oriented arrays. Topological insulation could be driven by supercoiling-related effects, or by specific boundary elements.

lineage with long gene arrays and no transcriptional regulation (Supplementary Fig. 17).

These differences can be rationalized by the unusual dinoflagellate properties. First, neither bacteria nor yeast possess comparably long gene arrays and transcription in those species is highly nonuniform; less transcription-induced torsional stress is therefore expected. Nucleosome loss is the second, and most salient difference. Single mammalian genes as long as dinoTADs are quite common, yet contact domains aligning with gene boundaries is not apparent in mammalian Hi-C maps, nor is it seen in kinetoplastids, which have gene arrays but also have conventional chromatin. We therefore hypothesize that plectonemic structures form due to transcription-induced supercoiling in the nucleosome-depleted genomes of dinoflagellates, while in other eukaryotes, a combination of the wrapping

of DNA around nucleosomes, interactions between nucleosomes, and accumulation of DNA twist, prevent their formation (Fig. 2H).

These results generate a number of open questions. How exactly are boundaries between dinoTADs formed mechanistically? Specific boundary elements of markedly different chromatin state could exist; alternatively, these boundaries may self-organize purely through torsion-related mechanisms. The roles that dinoflagellates' divergent histone genes play is also not clear. Finally, the relationship between Hi-C features and the "toroidal chromonemas" [8] observed by electron microscopy remains unknown. Answers to these questions, together with the dissection of specific roles different topoisomerase classes, will help fully elucidate the interplay between packaging proteins, transcription-induced torsional stress, and genome folding in dinoflagellates.

These observations also identify transcription-induced torsional stress as a key direction of future studies in eukaryotes generally. The strength of dinoTADs underlines the potency of this fundamental biological process for generating topological structure. The precise manner by which torsion is accommodated as twist and writhe, as well as its consequences for regulatory protein occupancy, transcriptional activity, and other chromatin processes, such as the behavior of ATP-dependent chromatin remodelers, are exciting questions remaining to be unraveled.

Author contributions

G.K.M. performed Hi-C experiments. G.K.M and A.E.T. analyzed the data. A.E.T. and T.X. designed and carried out transcription inhibition experiments and cell viability experiments. T.X. carried out *S. minutum* culture and heat stress treatment. W.J.G., A.R.G.. A.K. and J.R.P. supervised the study. G.K.M., A.E.T. and T.X. interpreted data and wrote manuscript with input from all authors.

Acknowledgments

This work was supported by NIH grants (P50HG007735, RO1 HG008140, U19AI057266 and UM1HG009442 to W.J.G., 1UM1HG009436 to W.J.G. and A.K., 1DP2OD022870-01 and 1U01HG009431 to A.K.), the Rita Allen Foundation (to W.J.G.), the Baxter Foundation Faculty Scholar Grant, and the Human Frontiers Science Program grant RGY006S (to W.J.G). W.J.G is a Chan Zuckerberg Biohub investigator and acknowledges grants 2017-174468 and 2018-182817 from the Chan Zuckerberg Initiative. Fellowship support provided by the Stanford School of Medicine Dean's Fellowship (G.K.M.), the Siebel Scholars, the Enhancing Diversity in Graduate Education Program and the Weiland Family Fellowship (A.E.T.). This work is also supported by NSF-IOS EDGE Award 1645164 to A.R.G. and Carnegie Venture grant 10907 (to T.X. and G.K.M.).

The authors would like to thank Andrew J. Spakowitz, Zohar Shipony, Sandy Klemm, Olga Dudchenko, Erez Lieberman Aiden, John R. Pringle, Philip Cleves, and members of the Greenleaf, Kundaje, Pringle and Grossman labs for helpful discussion and suggestions regarding this work.

References

1. Szabo Q, Bantignies F, Cavalli G. 2019. Principles of genome folding into topologically associating domains. *Sci Adv* **5**(4):eaaw1668.
2. Hackett JD, Anderson DM, Erdner DL, Bhattacharya D. 2004. Dinoflagellates: a remarkable evolutionary experiment. *Am J Bot* **91**:1523–1534.
3. Postberg J, Forcob S, Chang WJ, Lipps HJ. 2010. The evolutionary history of histone H3 suggests a deep eukaryotic root of chromatin modifying mechanisms. *BMC Evol Biol* **10**:259.
4. Marinov GK, Lynch M. 2015. Diversity and Divergence of Dinoflagellate Histone Proteins. *G3 (Bethesda)* **6**(2):397–422.
5. Janouškovec J, Gavelis GS, Burki F, Dinh D, Bachvaroff TR, Gornik SG, Bright KJ, Imanian B, Strom SL, Delwiche CF, Waller RF, Fensome RA, Leander BS, Rohwer FL, Saldarriaga JF. 2017. Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. *Proc Natl Acad Sci U S A* **114**(2):E171–E180.
6. Shoguchi E, Shinzato C, Kawashima T, Gyoja F, Mungpakdee S, Koyanagi R, Takeuchi T, Hisata K, Tanaka M, Fujiwara M, Hamada M, Seidi A, Fujie M, Usami T, Goto H, Yamasaki S, Arakaki N, Suzuki Y, Sugano S, Toyoda A, Kuroki Y, Fujiyama A, Medina M, Coffroth MA, Bhattacharya D, Satoh N. 2013. Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Curr Biol* **23**(15):1399–1408.
7. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV, Archibald JM, Bharti AK, Bell CJ, Beszteri B, Bidle KD, Cameron CT, Campbell L, Caron DA, Cattolico RA, Collier JL, Coyne K, Davy SK, Deschamps P, Dyhrman ST, Edvardsen B, Gates RD, Gobler CJ, Greenwood SJ, Guida SM, Jacobi JL, Jakobsen KS, James ER, Jenkins B, John U, Johnson MD, Juhl AR, Kamp A, Katz LA, Kiene R, Kudryavtsev A, Leander BS, Lin S, Lovejoy C, Lynn D, Marchetti A, McManus G, Nedelcu AM, Menden-Deuer S, Miceli C, Mock T, Montresor M, Moran MA, Murray S, Nadathur G, Nagai S, Ngam PB, Palenik B, Pawlowski J, Petroni G, Piganeau G, Posewitz MC, Rengefors K, Romano G, Rumpho ME, Rynearson T, Schilling KB, Schroeder DC, Simpson AG, Slamovits CH, Smith DR, Smith GJ, Smith SR, Sosik HM, Stief P, Theriot E, Twary SN, Umale PE, Vaulot D, Wawrik B, Wheeler GL, Wilson WH, Xu Y, Zingone A, Worden AZ. 2014. The

Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol* **12**(6):e1001889.
8. Oakley BR. Dodge JD. 1979. Evidence for a double-helically coiled toroidal chromonema in the dinoflagellate chromosome. *Chromosoma* **70**:277–291.
9. Livolant F, Bouligand, Y. 1978. New observations on the twisted arrangement of dinoflagellate chromosomes. *Chromosoma* **68**:21–44.
10. Levi-Setti R, Gavrilov KL, Rizzo PJ. 2008. Divalent cation distribution in dinoflagellate chromosomes imaged by high-resolution ion probe mass spectrometry. *Eur J Cell Biol* **87**(12):963–976.
11. Spector DL, Triemer RE. 1981. Chromosome structure and mitosis in the dinoflagellates: an ultrastructural approach to an evolutionary problem. *Biosystems* **14**(3–4):289–298.
12. Le TB, Imakaev MV, Mirny LA, Laub MT. 2013. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* **342**(6159):731–734.
13. Benedetti F, Racko D, Dorier J, Burnier Y, Stasiak A. 2017. Transcription-induced supercoiling explains formation of self-interacting chromatin domains in *S. pombe*. *Nucleic Acids Res* **45**(17):9850–9859.
14. Livolant F, Bouligand Y. 1980. Double helical arrangement of spread dinoflagellate chromosomes. *Chromosoma* **80**:97–118.
15. Wong JTY. 2019. Architectural Organization of Dinoflagellate Liquid Crystalline Chromosomes. *Microorganisms* **7**(2):27.
16. Levin RA, Beltran VH, Hill R, Kjelleberg S, McDougald D, Steinberg PD, van Oppen MJ. 2016. Sex, Scavengers, and Chaperones: Transcriptome Secrets of Divergent Symbiodinium Thermal Tolerances. *Mol Biol Evol* **33**(9):2201–2215.
17. Bensaude O. 2011. Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription* **2**(3):103–108.
18. Rizzo PJ. 1979. RNA synthesis in isolated nuclei of the dinoflagellate *Crypthecodinium cohnii*. *J Protozool* **26**(2):290–294.
19. Herzog VA, Reichholf B, Neumann T, Rescheneder P, Bhat P, Burkard TR, Wlotzka W, von Haeseler A, Zuber J, Ameres SL. 2017. Thiol-linked alkylation of RNA to assess expression dynamics. *Nat Methods* **14**(12):1198–1204.
20. Teves SS, Henikoff S. 2014. DNA torsion as a feedback mediator of transcription and chromatin dynamics. *Nucleus* **5**(3):211–218.
21. Xiang T, Hambleton EA, DeNofrio JC, Pringle JR, Grossman AR. 2013. Isolation of clonal axenic strains of the symbiotic dinoflagellate *Symbiodinium* and their growth and host specificity. *J Phycol* **49**(3):447–458.
22. Xiang T, Nelson W, Rodriguez J, Tolleter D, Grossman AR. 2015. *Symbiodinium* transcriptome and global responses of cells to immediate changes in light intensity when grown under autotrophic or mixotrophic conditions. *Plant J* **82**(1):67–80.
23. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**(7):1665–1680.
24. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**(1):95–98.
25. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**(6333):92–95.
26. Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T. 2018. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun* **9**(1):189.
27. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**(1):15–21.
28. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**(3):R25.
29. Roberts A, Pachter L. 2013. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat Methods* **10**(1):71-73.
30. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**(12):550.
31. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**(3):290-295.
32. Eddy SR. 2011. Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**(10):e1002195.
33. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. 2014. Pfam: the protein families database. *Nucleic Acids Res* **42**(Database issue):D222–230.
34. Knight P, Ruiz D. 2013. A fast algorithm for matrix balancing. *IMA J Num Anal* **33**(3):1029–1047.
35. Hou Y, Ji N, Zhang H, Shi X, Han H, Lin S. 2019. Genome size-dependent pcna gene copy number in dinoflagellates and molecular evidence of retroposition as a major evolutionary mechanism. *J Phycol* **55**(1):37–46.