3D Reconstruction of Dinoflagellate Chromosomes from Hi-C Data Challenges the Cholesteric Liquid Crystal Hypothesis

Authors: Lucas Philipp[1], Georgi K. Marinov[2], Stephen Todd[3,4], Stephanie C. Weber[1,5,*]

Affiliation: [1]Quantitative Life Sciences Program, McGill University, Montreal, Canada, [2]Department of Genetics, Stanford University, Palo Alto, USA, [3]Department of Computing, Goldsmiths, University of London, London, United Kingdom, [4]London Geometry Ltd., London, United Kingdom, [5]Departments of Biology and Physics, McGill University, Montreal, Canada

To whom correspondence should be addressed. E-mail: steph.weber@mcgill.ca

ORCID IDs:

Lucas Philipp: https://orcid.org/0000-0001-6454-4275

Georgi K. Marinov: https://orcid.org/0000-0003-1822-7273

Stephen Todd: N/A

Stephanie C. Weber: https://orcid.org/0000-0002-7297-2875

Keywords

dinoflagellate | chromosome conformation capture | transcription | topologically associating domain | cholesteric liquid crystal

Abstract

Dinoflagellates have permanently condensed chromosomes that are often described as liquid crystalline. Specifically, a Cholesteric Liquid Crystal (CLC) model was proposed in which DNA is organized into parallel fibers within stacked discs, such that the fiber orientation rotates by a constant angle between adjacent discs. Extrachromosomal loops extending from the discs were hypothesized to be more accessible and thus to contain transcriptionally active genes. Although the CLC model

captures some features of dinoflagellate chromosome structure, its validity has not been rigorously tested against modern genomic data. Here, we use chromatin conformation capture (Hi-C) data to simulate 3D conformations of chromosome scaffolds for three dinoflagellate species: *Fugacium kawagutii*, *Symbiodinium microadriaticum*, and *Breviolum minutum*. Consensus and population-based modeling generate diverse polymer conformations with moderate orientational and nematic order. However, we find no evidence of cholesteric discs. Moreover, contact probability curves from empirical Hi-C data are inconsistent with the CLC model. Nevertheless, we show that introducing locus-specific boundaries into the CLC model can produce simulated Hi-C contact maps with topologically associating domains (TADs), which are observed in experimental Hi-C contact maps for these species. Finally, by mapping RNA-seq data onto our simulated conformations, we show that actively transcribed genes are present throughout entire chromosomes, and not exclusively on extrachromosomal loops or at the surface. Our results challenge the long-standing CLC model and suggest that dinoflagellate chromosomes are organized into condensed but non-crystalline structures that do not impede transcription.

Introduction

Dinoflagellates are a diverse group of unicellular algae whose chromosomes remain highly condensed throughout the cell cycle [1]. How this unusual chromosome structure impacts gene expression is a long-standing question, since compact chromatin domains typically exclude RNA polymerase [2, 3] and suppress transcription [4] in other eukaryotes. The leading hypothesis to explain both the unique structure of dinoflagellate chromosomes and its consequences for transcription is the cholesteric liquid crystal (CLC) model. Originally proposed over fifty years ago, the CLC model posits that DNA is packaged into a series of flat discs that stack to form the long axis of the chromosome [5]. Within each disc, chromatin fibers are arranged parallel to a director that rotates by a constant angle between discs [6-9]. The CLC geometry is consistent with banding

47 patterns observed in 2D transmission electron micrographs [5, 6]. In addition, extrachromosomal
48 loops, described previously as 3D protrusions from the chromosome core, were hypothesized to
49 connect both adjacent and non-adjacent discs and to contain transcriptionally-active genes, due to
50 their presumed greater accessibility compared to the condensed chromosome core [5-8, 10-15].
51 However, the validity of the CLC model and its predictions for the spatial organization of
52 transcription have not been rigorously tested against modern genomic data.

53 Chromosome conformation capture (Hi-C) is an experimental sequencing technique that
54 measures which regions of primary DNA sequence contact each other in three-dimensional space. Hi-
55 C analysis of several dinoflagellate species [16-18] revealed topologically associating domains
56 (TADs), regions of primary sequence that interact with each other at much higher probabilities than
57 with other chromosomal regions. These "dinoTADs" correlate with another common feature of
58 dinoflagellate genomes, tandem arrays of co-oriented genes [19-22]. Indeed, the insulation boundaries
59 that separate dinoTADs occur at locations where two oppositely-oriented gene arrays converge [16,
60 17]. Moreover, these boundaries are disrupted by transcription inhibitors, suggesting that active
61 transcription may contribute to dinoTAD formation. Nevertheless, it remains unclear whether
62 dinoTADs correspond to cholesteric discs and whether the CLC model is consistent with Hi-C data
63 more generally [23].

64 Various computational methods have been developed to reconstruct 3D polymer conformations
65 from Hi-C data [24, 25]. In the consensus modelling approach, all Hi-C-derived constraints are
66 applied to a single conformation and contact probability is inversely related to spatial distance
67 between a particular pair of monomers [25-29]. By contrast, in population-based modelling, contacts
68 are distributed across many individual conformations in a population [30-32]. In this case, contact
69 probability corresponds to the fraction of conformations in which a particular monomer pair is
70 constrained to be in close proximity. Consensus modelling is fast and efficient, and therefore suitable

71 for surveying hundreds of genomic regions across multiple species. Population-based modelling is
72 computationally more expensive, but explicitly considers conformational heterogeneity which may
73 result in polymer conformations that more closely reflect those found in individual cells [33, 34].
74 In this article, we use previously published Hi-C data [16-18] and both consensus [26] and
75 population-based [32] modeling tools to simulate 3D conformations of chromosome scaffolds from
76 three dinoflagellate species: *Fugacium kawagutii* (formerly known as *Symbiodinium kawagutii* [35]),
77 *Symbiodinium microadriaticum* and *Breviolum minutum*. Surprisingly, our results indicate that
78 dinoflagellate chromosomes are not organized into a stack of cholesteric discs as predicted by the
79 CLC model. Rather, they contain multiple compact domains, and exhibit more orientational and
80 nematic order than expected by chance, but less than CLCs. By mapping RNA-seq reads [36-38] onto
81 our simulated polymer conformations, we find that active genes are located throughout chromosomes
82 and not solely at their surfaces. This distribution suggests that although dinoflagellate chromosomes
83 are highly condensed, they remain accessible to RNA polymerase. Our findings clarify important
84 aspects of dinoflagellate chromosome structure and resolve a long-standing question as to how this
85 structure relates to transcriptional activity.

86 # Results

87

88 **The CLC model does not produce TADs**

89 The CLC model consists of a stack of cholesteric discs (Fig. 1 A) [5-9]. Within each disc, the
90 chromatin fiber weaves back and forth in a parallel fashion. The director, along which the fiber is
91 oriented, rotates by a constant angle, θ, between adjacent discs, thereby defining the cholesteric pitch.
92 To compare the CLC model to empirical Hi-C data [16-18], we generated a population of 10^6 CLC
93 structures to approximate the number of cells sequenced in a typical bulk Hi-C experiment [39]
94 (Methods). We assumed that discs were arranged consecutively in the stack, in the same order as in
95 primary sequence (Fig. 1 A). To introduce conformational heterogeneity, we randomly incorporated

96 extrachromosomal loops that extend outward from the stack of discs and re-enter either the same disc
97 (intra-disc loops) or an adjacent disc (inter-disc loops) (Fig. 1 B). The length of each
98 extrachromosomal loop was drawn from an exponential distribution. The variable positions and
99 lengths of extrachromosomal loops resulted in variable locations of cholesteric discs along the
100 primary sequence of individual CLC structures in the population (Supplementary Fig. S1 A).

101 We simulated a Hi-C contact map from our population of CLC structures and found a grid-like
102 pattern, with stripes of high contact probability that are parallel to each other and perpendicular to the
103 central diagonal (Fig. 1 C). The number and spacing of stripes are determined by the number and
104 diameter of cholesteric discs in the input structures (Supplementary Fig. S1 B, S2 A, B). The grid-like
105 pattern is robust to changes in CLC model parameters, such as the percentage of total sequence in
106 extrachromosomal loops, the cholesteric pitch, and the tapering of disc diameter at the top and bottom
107 of the stack (Supplementary Fig. S2 C, D, E). However, variation of the number of cholesteric discs
108 per structure within the same population results in a blurring of the stripes (Supplementary Fig. S2 F).

109 The simulated Hi-C contact maps from CLC populations (Fig. 1 C, Supplementary Fig. S2) are
110 qualitatively different from empirical Hi-C contact maps from the dinoflagellates *F. kawagutii* and *S.*
111 *microadriaticum* [16-18] (Fig. 1 F, top right triangles). The most striking difference is the lack of
112 TADs. Indeed, quantitative TAD metrics such as the insulation score [40] and directionality index
113 [31] smoothly oscillate with a constant period for CLCs, while they vary irregularly and abruptly for
114 dinoflagellate chromosomes (Supplementary Fig. S3). Overall, Hi-C contact maps from the CLC
115 model are only modestly predictive of experimental data, with Pearson correlation coefficients r =
116 0.62 for *F. kawagutii* and r = 0.59 *S. microadriaticum*. These results indicate that the CLC model does
117 not produce TADs, which are the dominant structural feature of dinoflagellates chromosomes [16-18].

118 **Dinoflagellate chromosomes lack cholesteric discs**

119 Next, we computed the 3D conformation(s) of dinoflagellate chromosomes that best reproduce
120 empirical Hi-C contact probabilities. To mitigate algorithm bias [42], we used both consensus and
121 population-based modeling approaches. First, we applied the CSynth software package [26], which
122 incorporates Hi-C contacts as spring-like forces between monomers that serve to constrain their
123 positions and outputs a single consensus conformation (Methods). After validation (see SI Text), we
124 used CSynth to simulate conformations for all *F. kawagutii* (Supplementary Fig. S4, Supplementary
125 Table S1) and *S. microadriaticum* (Supplementary Fig. S5, Supplementary Table S2) chromosome
126 scaffolds at 5 kb per monomer resolution. Second, we implemented the Integrative Genome Modeling
127 (IGM) software package [32], which uses maximum likelihood estimation to generate a population of
128 conformations that is statistically consistent with bulk Hi-C data. We used IGM to simulate a
129 population of 100 conformations for the first (largest) chromosome scaffold from *F. kawagutii*
130 (Supplementary Fig. S6) and *S. microadriaticum* (Supplementary Fig. S7), also at 5 kb per monomer
131 resolution.

132 In contrast to the CLC model (Fig. 1 A, B), we found that 3D conformations of dinoflagellate
133 chromosomes lack cholesteric discs (Fig. 1 D, E). Instead, consensus conformations are compact and
134 ellipsoidal (Fig. 1 D, Supplementary Figs. S4, S5), while population-based conformations can be
135 elongated or spherical, with multiple compact domains and long extrachromosomal loops (Fig. 1 E,
136 Supplementary Figs. S6, S7). We observed few extrachromosomal loops in consensus conformations
137 – 0.63 per chromosome in *F. kawagutii* (Supplementary Fig. S4) and 1.8 per chromosome in *S.*
138 *microadriaticum* (Supplementary Fig. S5) – but many more in the population-based conformations –
139 2.5 per conformation in *F. kawagutii* (Supplementary Fig. S6) and 2.9 per conformation in *S.*
140 *microadriaticum* (Supplementary Fig. S7). Extrachromosomal loops were previously reported to be
141 30-60 kb in length [43], corresponding to 6-12 monomers per loop in our simulations. Therefore, the

142 apparent sparseness of extrachromosomal loops in our consensus conformations is not due to limited
143 resolution.

144 We next simulated Hi-C contact maps from our population-based (IGM) conformations of
145 chromosome 1 for both species and found squares along the central diagonal, rather than a grid (Fig. 1
146 F, bottom left triangles). This pattern is consistent with dinoTADs [16-18]. Moreover, quantitative
147 comparison of these contact maps with the experimental input data showed strong correlation (r =
148 0.91 for *F. kawagutii*, r = 0.84 for *S. microadriaticum*) (Fig. 1 F).

149 Finally, we simulated chromosome conformations for a third dinoflagellate species, *B. minutum*.
150 While Hi-C data are available [17], they only cover an estimated 39% of the genome, compared to
151 79% for *F. kawagutii* and 64% for *S. microadriaticum* (Supplementary Table S3). Since *B. minutum*
152 scaffolds may not correspond to full chromosomes, we selected the 10 largest Hi-C scaffolds for
153 reconstruction (Supplementary Table S4). Consensus modeling generated spherical conformations
154 with numerous extrachromosomal loops (5.3 loops per scaffold) (Supplementary Fig. S8 A), in
155 contrast to the ellipsoidal conformations with few loops seen for *F. kawagutii* and *S. microadriaticum*
156 (Fig. 1 D, Supplementary Fig. S4, S5). These conformational differences between species may be due
157 to incomplete scaffolds, rather than true structural differences in *B. minutum* chromosomes.
158 Interestingly, in this species, population-based modeling generated similar conformations, with
159 spherical shapes and many extrachromosomal loops (3.7 loops per conformation) (Supplementary
160 Fig. S8 B). As for the other two species, we found strong correlation (r = 0.93) between Hi-C contact
161 maps simulated from our IGM population and observed for *B. minutum* [17] (Supplementary Fig. S8
162 C), which both exhibit TAD-like squares rather than CLC-like stripes. Overall, our polymer
163 simulations suggest that dinoflagellate chromosomes are not organized into stacks of cholesteric
164 discs.

165 **Locus-specific boundaries can recapitulate TADs in CLCs**

166 In mammalian cells, TADs arise from loop extrusion — which leads to conformational
167 heterogeneity among cells (and over time) — with locus-specific boundaries [34, 44-46]. To
168 determine whether the CLC model is compatible with TADs, we grouped cholesteric discs according
169 to observed dinoTAD sizes [16] and introduced locus-specific boundaries in between groups. The
170 order of discs was randomly permuted within each group but not between groups (Fig. 2 A). The
171 simulated Hi-C contact map for CLCs with disc order permuted within groups shows squares of high
172 contact probability along the central diagonal (Fig. 2 B), which are characteristic of TADs. Indeed, the
173 insulation score and directionality index vary more irregularly and abruptly than for CLCs with
174 consecutive disc order (Supplementary Fig. S3). Therefore, the presence of locus-specific boundaries,
175 together with conformational heterogeneity across the population, can recapitulate TADs in CLCs.
176 Nevertheless, the correlation with the experimental *S. microadriaticum* contact map is only slightly
177 better for CLCs with groups of permuted disc order ($r = 0.67$) than for CLCs with consecutive discs (r
178 $= 0.59$, Fig. 1 C, F).
179 To visualize dinoTADs in our simulated chromosome conformations, we developed a manual
180 TAD annotation tool within the CSynth graphical user interface. This tool allows the user to select the
181 apex of a TAD with a mouse click and then automatically writes the left and right TAD boundary
182 base-pair coordinates to a text file (Fig. 2 C, Methods). As with the rest of CSynth, this tool runs in a
183 web browser and does not require any software installation [26]. Moreover, it can be used to annotate
184 TADs in hundreds of chromosomes in less than a few hours. These TAD annotations can be used to
185 visualize multiple TADs in distinct colors on the same 3D conformation (Fig. 2 D, Methods). Our
186 manual TAD annotation tool has many advantages over existing automatic TAD annotators, which
187 can be sensitive to parameter choice, involve lengthy computation times, require installation, or may
188 not support integrated visualization of their output [47].

189 We used this new tool to annotate dinoTADs in Hi-C datasets [16-18]. A majority of the genome
190 — 84% in *F. kawagutii* and 71% in *S. microadriaticum*, comparable to other eukaryotes [48, 49] — is
191 organized into dinoTADs, which are located throughout our CSynth conformations but not in a
192 regular arrangement or stack (Fig. 2 D). Similarly, dinoTADs in IGM conformations are not groups of
193 cholesteric discs and exhibit diverse shapes depending on the conformation (Fig. 2 E). Therefore,
194 although the CLC model can be made consistent with TADs, our 3D polymer conformations suggest
195 that dinoTADs are not groups of cholesteric discs.

196 **Dinoflagellate contact probability curves are inconsistent with CLC model predictions**

197 Contact probability curves are often used to describe chromosome organizational features [50].
198 This metric quantifies how frequently any two DNA loci, separated by a fixed amount of primary
199 sequence, make contact in 3D space. Mathematically, contact probability can be written as
200 $P(s = |i - j| \times 5\text{ kb}) \propto \sum_{|i-j| \times 5\text{ kb} = s} \frac{CP_{ij}}{L-s} \propto s^{\gamma}$ where $\sum_{|i-j| \times 5\text{ kb} = s} CP_{ij}$ is the sum of Hi-C
201 contact probabilities for all monomer pairs (i, j) separated by $s = |i - j| \times 5$ kb, L is the chromosome
202 length, and γ is the contact probability decay exponent. The exact proportionality depends on the Hi-
203 C normalization procedure. The decay exponent can be compared across species and to predictions
204 from various models [50]. One such model is the equilibrium globule, constructed by first performing
205 a 3D random walk within the confines of a sphere, and then turning on excluded volume interactions
206 and thermally equilibrating the polymer. This procedure generates a knotted, space-filling polymer
207 conformation [50]. Equilibrium globules have a decay exponent of $\gamma = -1.5$, while human interphase
208 chromosomes have a decay exponent of $\gamma = -1$ [50].

209 To determine the decay exponent expected from the CLC model, we calculated the contact
210 probability curve for a population of CLC structures following the methods outlined in [51]. For
211 intermediate genomic separations, s between 10^4 and 10^6 bp, $P(s)$ is very shallow with a decay
212 exponent of $\gamma \approx -0.2$ (Fig. 3 A). This exponent is robust across a range of CLC parameters

(Supplementary Fig. S2), reflecting the high degree of compaction in this polymer model. At larger genomic separations, $s > 10^6$ bp, the curve decays more rapidly, consistent with the linear stacking of cholesteric discs [16]. Furthermore, contact probability curves for CLCs with consecutive disc order have two prominent local maxima at genomic separations corresponding to intra-disc ($\sim 10^5$ bp) and inter-disc ($\sim 10^6$ bp) contacts (vertical dashed lines: Fig. 3 A, Supplementary Fig. S2). CLCs with permuted disc order within groups exhibit the same very shallow decay exponent of $\gamma \approx -0.2$ as CLCs with consecutive disc order (Fig. 3 A). However, the second peak corresponding to inter-disc contacts vanishes. This is expected as discs that are adjacent in primary sequence are no longer adjacent in 3D space and therefore do not regularly contact each other.

Next, we calculated contact probability curves for *F. kawagutii* and *B. minutum*, as was done previously for *S. microadriaticum* [16], using the cooltools package [52]. We selected the largest 75 chromosome scaffolds for each species (Supplementary Tables S1, S2, S4), as they likely represent completely assembled chromosomes, and excluded shorter scaffolds from our analysis. $P(s)$ curves for dinoflagellate chromosome scaffolds are smooth, with a decay exponent of $\gamma \approx -0.5$ (Fig. 3 B-D). Contact probability curves from simulated IGM conformations produce the same $\gamma \approx -0.5$ decay exponent (Fig. 3 B-D). This power-law scaling is steeper than what we observed for the CLC model (Fig. 3 A, Supplementary Fig. S2), even when groups of permuted discs produce TADs in the CLC contact map (Fig. 2 B), suggesting that dinoflagellate chromosomes are less condensed than CLCs. Moreover, the experimental and IGM contact probability curves have no peaks at intermediate genomic separations, as would be expected for cholesteric discs. Overall, we find no evidence of cholesteric organization in the contact probability curves of dinoflagellate chromosome scaffolds. Although the CLC model can be made consistent with TADs, these discrepancies indicate that a stacked organization of cholesteric discs is fundamentally incompatible with dinoflagellate Hi-C data.

Dinoflagellate chromosomes exhibit moderate orientational and nematic order

Despite their apparent lack of cholesteric discs, we next sought to quantify the degree of order in our simulated chromosome conformations. We first quantified the orientational order within our consensus and population-based conformations by calculating the orientational order parameter (OOP) [53] and its Fourier transform. The OOP can be defined as $OOP(|i-j| \times 5\text{ kb}) = < \vec{r}_i \cdot \vec{r}_j >$, where $\vec{r}_i$ is the unit tangent vector to the ith monomer, and the brackets $<\cdot>$ denote the average dot product over all possible pairs of monomers (i, j) separated by $s = |i-j| \times 5$ kb. For helical organization, the OOP is a sinusoid as a function of primary sequence separation and its Fourier transform is a narrow peak corresponding to the period of the helix (Fig. 4 A). For a particular genomic separation s, an OOP of 1 indicates perfect parallel alignment, an OOP of -1 indicates perfect anti-parallel alignment, and an OOP of 0 indicates random orientation or no alignment. Intermediate positive/negative OOP values indicate partial alignment/antialignment.

We compared the OOPs of CLCs as a positive control; equilibrium globules as a negative control; and simulated conformations of dinoflagellate chromosome scaffolds (Fig. 4 B, Supplementary Fig. 8 D). The OOP of CLCs with consecutive disc order exhibits a damped oscillation (Fig. 4 B, Row i). Two peaks are present in its Fourier transform: one peak corresponds to orientational ordering between adjacent cholesteric discs ($\sim 10^{-6}$/bp), and the other corresponds to orientational ordering within a cholesteric disc ($\sim 10^{-5}$/bp). The peak locations depend on the average number of monomers in a cholesteric disc, and not on the percentage of sequence in extrachromosomal loops, as expected (Supplementary Fig. S9). The OOP of CLCs with permuted disc order within groups also has a damped oscillation (Fig. 4 B, Row ii) but lacks the peak corresponding to inter-disc orientational alignment. This is expected because permutation of disc order results in discs that are adjacent in primary sequence but not in 3D space having directors that differ by random multiples of θ (Fig. 1 A).

Equilibrium globules are confined 3D random walks with no orientational order. We calculated the OOP of equilibrium globules using simulated data from [50]. As expected, the OOP monotonically decays at short genomic separations and plateaus to zero. There are no peaks in its Fourier transform, consistent with its lack of any periodic structure (Fig. 4 B, Row iii).

In contrast, the OOPs for consensus conformations of *F. kawagutii, S. microadriaticum* and *B. breviolum* chromosomes lack both the oscillatory signature of the CLC model and the monotonic decay of the equilibrium globule. Instead, they exhibit a negative peak at $\sim 10^4$ bp (Fig. 4 B, Rows iv & v, Supplementary Fig. S8 D), indicating partial anti-parallel orientation at this length scale. Since this corresponds to the average length of gene arrays, these peaks suggest that gene arrays may fold in an anti-parallel orientation in dinoflagellate chromosomes. Interestingly, we detect no orientational order on the scale of dinoTADs ($\sim 10^6$ bp), likely due to their structural heterogeneity (Fig. 2 D, E). Fourier transforms of OOP curves for population-based conformations have smaller peaks at $\sim 10^{-5}$/bp (Fig. 4 B, Rows vi & vii, Supplementary Fig. S8 D), indicating less orientational order in IGM conformations compared to CSynth conformations.

Ordering of the chromosome could also occur independently of primary sequence separation and instead depend on the 3D spatial proximity of DNA loci. To measure this, we calculated the local nematic order parameter (NOP) at two different length scales following the methods outlined in [54]. A NOP of 1 indicates that all tangent vectors within a voxel are co-aligned in either a parallel or antiparallel orientation. A NOP of 0 indicates that tangent vectors within a voxel are randomly oriented with no directional correlation. Across all polymer conformations, the NOP decreases as voxel size increases because tangent vectors are averaged over a larger volume (Fig. 4 C). For CLCs, NOP also decreases as the number of discs increases because more discs with directors that are oriented differently are located within the same voxel (Fig. 4 C). For equilibrium globules, we found a negligible degree of nematic ordering, as expected (Fig. 4 C). Finally, we observe a moderate degree

284 of nematic ordering in dinoflagellate chromosome conformations, intermediate between that of CLCs
285 and equilibrium globules (Fig. 4 C, Supplementary Fig. S8 E). Altogether, both OOP and NOP
286 metrics indicate that dinoflagellate chromosomes, simulated by either modeling approach, are more
287 ordered than equilibrium globules, but less ordered than CLCs.

288 **Active genes occur throughout dinoflagellate chromosomes**

289 Finally, we visualized the 3D spatial distribution of active genes in dinoflagellate chromosomes
290 by mapping RNA-seq reads [36-38] to the genome assembly, and in turn, onto their corresponding 3D
291 locations as predicted by CSynth and IGM. We defined a cylindrical coordinate system where the
292 cylindrical axis is the chromosome's long axis and the radial axis is all perpendicular directions from
293 the cylindrical axis to the chromosome's surface (Fig. 5 A). The CLC model — with or without locus-
294 specific boundaries — is cylindrical in shape and predicts that active transcription occurs on
295 extrachromosomal loops, which are located at the chromosome surface, as opposed to the core, which
296 is hypothesized to be inaccessible to RNA polymerases [5-8, 10-15, 55-58]. Thus, we would expect to
297 see the highest expression levels at the surface and low or no expression in the core (Fig. 5 A).

298 To determine whether transcription levels correlate with proximity to the surface, we used this
299 coordinate system to calculate the average gene expression throughout *F. kawagutii* and *S.*
300 *microadriaticum* chromosome conformations, controlling for variation in DNA density across voxels
301 (Fig. 5 B-C, Methods). In contrast to the CLC model's prediction, we observe active genes throughout
302 dinoflagellate chromosomes modeled using both CSynth (Fig. 5 B) and IGM (Fig. 5 C). Transcription
303 heatmaps of population-based conformations have more diffuse peripheral boundaries compared to
304 consensus conformations, consistent with their more variable shapes and numerous extrachromosomal
305 loops (Fig. 1 E, Supplementary Figs. S6, S7). Interestingly, we find somewhat increased transcription
306 levels away from the center of the cylindrical axis for *S. microadriaticum* chromosomes (Fig. 5 B-C,
307 blue). Telomeres are also preferentially located away from the center of the cylindrical axis

308 (Supplementary Fig. S10), which is consistent with the previous observation that *S. microadriaticum*
309 chromosomes have higher gene density at their telomeres [16]. Overall, we find no correlation
310 between gene expression and radial position in dinoflagellate chromosomes, indicating that active
311 transcription is not restricted to extrachromosomal loops.

312 # Discussion

313 We have systematically and quantitatively compared genomic data for three species of
314 dinoflagellates to predictions of the CLC model. Unexpectedly, our results challenge the long-
315 standing hypothesis that dinoflagellate chromosomes are organized into CLCs. First, empirical Hi-C
316 contact maps have TADs [16-18] instead of a grid-like pattern that we show is characteristic of
317 cholesteric discs (Fig. 1). Second, the scaling of contact probability curves for dinoflagellate
318 chromosome scaffolds are inconsistent with CLCs (Fig. 3). Third, our polymer simulations predict 3D
319 conformations that are less compact and less ordered than CLCs (Fig. 4). Finally, we find active genes
320 located throughout these 3D conformations, rather than on the surface, as expected for CLCs (Fig. 5).

321 The CLC model was proposed to explain the unique structure of dinoflagellate chromosomes [5-
322 9]. It successfully predicted a highly compact twisted chromosome, with significant nematic order,
323 and became the accepted view [5-15, 64-66]. However, our systematic analysis of reconstructed
324 chromosome conformations – as well as experimental data [16-18] – reveals several important
325 disagreements between dinoflagellate chromosomes and the CLC model. Therefore, we conclude that
326 dinoflagellate chromosomes (at least in these three species) are not cholesteric liquid crystals.

327 Different computational modeling approaches can generate different outputs from the same input
328 data, creating uncertainties for interpretation [46]. To mitigate this risk, and to enhance the robustness
329 of our conclusions, we used two different software packages to reconstruct dinoflagellate
330 chromosomes. While both CSynth and IGM produce 3D polymer conformations that disagree with
331 the CLC hypothesis, their predictions differ in some notable ways. For example, the ellipsoidal shape

332 of CSynth conformations is reminiscent of dinoflagellate chromosomes observed by transmission
333 electron microscopy [1, 5-8, 13, 59-63], while IGM conformations sometimes adopt dumbbells and
334 other elongated or irregular shapes (Supplementary Figs. S6, S7) that are less common in the
335 microscopy literature [1]. Nevertheless, IGM conformations give rise to simulated Hi-C contact maps
336 (Fig. 1F) and contact probability curves (Fig. 3) that are highly consistent with empirical data [16-18].
337 Interestingly, CSynth can reproduce CLC-like conformations while IGM cannot (SI text,
338 Supplementary Fig. S15). Overall, our work highlights both the strengths and limitations of these
339 computational models. Finally, IGM is capable of simultaneously integrating Hi-C data together with
340 data from orthogonal experimental methods including fluorescence in situ hybridization (FISH), DNA
341 Adenine Methyltransferase Identification (Dam ID), and single-cell split-pool recognition of
342 interactions by tag extension (SPRITE) [32]. While such data types are not yet available for
343 dinoflagellates, optimizing these methods and applying them to these organisms should be a priority
344 for future research.

345 The 3D conformations produced by CSynth and IGM are static. However, chromosomes are
346 highly dynamic in living cells [46]. Analysis of temporal variation in chromosome structure requires
347 Hi-C data from a synchronized population of cells sampled at different time points throughout the cell
348 cycle [67, 68]. Comparison of Hi-C contact maps of *S. microadriaticum* cultures enriched for either
349 flagellated mastigote cells (G1/S) or immobile coccoid cells (G2/M) showed no apparent differences,
350 indicating that this species' chromosomes are permanently condensed and undergo minimal
351 rearrangements during the cell cycle [16]. Nevertheless, our study does not rule out the possibility of
352 dynamic rearrangement of extrachromosomal loops to the chromosome exterior where they are
353 hypothesized to be more transcriptionally accessible [69]. If this were the case, then the expression of
354 nearly all genes would need to be regulated via this mechanism as nearly all chromosomal regions are
355 transcriptionally active (Fig. 5 B, C). Alternatively, the dense core of dinoflagellate chromosomes

356 may not exclude the transcriptional machinery after all. Indeed, recent Assay for Transposase-
357 Accessible Chromatin using sequencing (ATAC-seq) data on *B. minutum* indicate that while overall
358 accessibility is low, it is more uniform than in other eukaryotic genomes, consistent with the lack of
359 nucleosome-based chromatin [70]. Tn5 transposase (53.3 kD) is much smaller than RNA Polymerase
360 II (Pol II, 514 kD), so it is possible that Tn5 can access regions from which Pol II is excluded.
361 However, bacterial RNA polymerase (457 kD) is able to penetrate the highly condensed nucleoid in
362 stationary-phase *E. coli* cells [71], which has also been described as liquid crystalline [72]. Together,
363 these findings suggest that dinoflagellate chromosomes are broadly accessible, and could explain why
364 active genes are located throughout the chromosome rather than preferentially at the surface (Fig. 5).
365 In addition to the CLC model, many other models for dinoflagellate chromosome organization
366 have been proposed. One model consists of two independent helices coiled around each other [73].
367 Another model consists of a toroidal chromosome twisted into a double helix, with rounded hairpin
368 loops at the chromosome ends rather than exposed telomeres [74]. A third model proposes that
369 dinoflagellate chromosomes are polytene, i.e. they undergo many rounds of complete and/or partial
370 replication without segregation, maintaining close alignment of sister chromatids [75]. A fourth model
371 consists of a hierarchical helix with six degrees of hierarchy [76, 77]. A fifth model consists of a
372 helical condensed core, with abundant extrachromosomal loops [78].
373 Our 3D conformations of dinoflagellate chromosomes are not consistent with any of these
374 models. They do not consist of multiple interwoven helices, are not twisted toroids, are not helices of
375 helices, and do not contain abundant extrachromosomal loops. Furthermore, dinoflagellate Hi-C
376 contact maps do not contain an X-shape pattern that is characteristic of circular chromosomes [79], do
377 not have variable contact probability along the main diagonal as is the case for chromosomes with
378 polytene bands [80], and lack significant trans-chromosomal contacts that would be expected from
379 interwoven helices involving multiple chromosomes [16-18].

380 In addition to dinoflagellates, condensed, twisted, and ordered interphase chromosome structures
381 have been observed in at least three other biological contexts: mosquito chromosomes [81], cricket
382 sperm chromosomes [54], and bacterial nucleoids [82]. The latter two examples are particularly
383 interesting because, like dinoflagellates, they also lack nucleosomes. Indeed, dinoflagellate viral
384 nuclear protein (DVNPs) and histone-like proteins (HLPs) replace histones [56, 57, 83] as the
385 dominant DNA-binding protein in dinoflagellates. Histones are replaced by SNBPs (sperm-specific
386 nuclear basic proteins) in sperm cells [54] and bacteria use a variety of nucleoid-associated proteins,
387 including HU (heat unstable) proteins [84], to package their DNA. Together, these examples suggest
388 that highly condensed, ordered chromosome structures may be more common in the absence of
389 histones.

390 Beyond DNA structure, the CLC model also makes functional predictions, including that active
391 genes are located on the chromosome surface in extrachromosomal loops and not in the condensed
392 chromosome core [5-8, 10-15, 55-58]. These predictions are mainly based on radiolabeling studies,
393 which show incorporation of ^{3}H-adenine or ^{3}H-uridine at the nucleolus [10, 11, 55], but not
394 overlapping with condensed chromosomes. Our results, which leverage recent RNA-seq datasets,
395 suggest that transcription is spatially ubiquitous. We find active genes throughout dinoflagellate
396 chromosomes and not exclusively at the chromosome surface (Fig. 5 B, C). Our findings are
397 consistent with other modern genomic assays such as ATAC-seq, which indicate that dinoflagellate
398 genomes are uniformly accessible to transposases [70]. Moreover, transcription still occurs in highly
399 condensed human mitotic chromosomes [4]. Thus, dinoflagellate chromosomes are likely accessible
400 to RNA polymerase and condensation is not the primary mechanism for regulation of gene
401 expression. Rather, dinoflagellate genes can be actively transcribed regardless of their chromosomal
402 location.

403 Although our results indicate that chromosome conformation does not influence gene expression
404 in dinoflagellates, transcription may contribute to chromosome organization. For example, we observe
405 orientational order at length scales corresponding to gene arrays (Fig. 4 B). The negative peak in the
406 OOP curve for dinoflagellate chromosomes is similar to that observed in model human chromosomes
407 perturbed by extensible force dipoles that represent transcriptional activity [85]. Thus, transcription of
408 co-oriented genes in divergent gene arrays may result in an anti-parallel arrangement. Transcription
409 has also been implicated in dinoTAD formation. Indeed, the boundaries of dinoTADs in *S.*
410 *microadriaticum* [16] and *B. minutum* [17] correspond to positions at which oppositely-oriented gene
411 arrays converge and are sensitive to transcriptional inhibitors. We mapped RNA-seq reads [36-38]
412 onto the *F. kawagutii* genome assembly and also found that dinoTAD boundaries occur between
413 convergent gene arrays (Supplementary Fig. S11), consistent with these previous studies [16, 17].
414 Transmission electron microscopy (TEM) has been used frequently to generate 2D cross-
415 sectional images of dinoflagellate chromosomes [1, 5-8, 13, 59-63]. To date, no study has imaged
416 dinoflagellate chromatin fiber orientation in 3D. Instead, interpretation of the banding patterns
417 observed in 2D cross-sectional images has relied on mathematically projecting 3D CLC models onto a
418 2D imaging plane [5]. The limitation of 2D data makes it difficult to extrapolate ordering observed in
419 a single plane to volumes within a whole chromosome, and to track the contours of individual
420 chromatin fibers as a function of primary sequence. For example, dinoflagellate chromosomes that
421 produce banding patterns in TEM cross-sections, which are typically attributed to CLC organization,
422 appear helical when viewed after a whole-mounting procedure [86]. In addition, a recent TEM study
423 [62] found that while oblique and longitudinal sections of *Prorocentrum minimum* chromosomes
424 agree qualitatively with CLC model predictions, chromatin fibers were not parallel within transverse
425 sections — an observation that is inconsistent with cholesteric discs. The advantage of our Hi-C-

based polymer modeling approach is that the entire dinoflagellate chromosome is modeled in 3D with the added context of primary sequence information.

Overall, this work advances our understanding of dinoflagellate chromosome organization in 3D, providing evidence that dinoflagellate chromosomes are not cholesteric liquid crystals. Future work should focus on 3D imaging and chromatin fiber tracing within dinoflagellate chromosomes, as well as 3D imaging of nascent RNA and transcription [87].

Methods

Hi-C data pre-processing

Hi-C-assisted genome assemblies and contact probabilities were accessed from the Gene Expression Omnibus (GEO) with accession numbers: GSE152150 [16] and GSE153950 [17], or from the Sequence Read Archive (SRA) with accession numbers: SRR25948349, SRR25948348 [88], and further processed following [18].

F. kawagutii Hi-C contact maps were converted from .hic format to .cool format using the hic2cool python package https://github.com/4dn-dcic/hic2cool.

When un-ligated and self-ligated Hi-C artifacts are abundant, Hi-C normalization can greatly reduce off-diagonal contact probabilities [89]. This results in weak CSynth Hi-C-based attraction forces and slow equilibration. Un-ligated and self-ligated Hi-C artifacts do not contain information about the 3D organization of the chromosome; they are short range contacts that map entirely within a single Hi-C bin (main diagonal), or across one Hi-C bin boundary (1st off-diagonal) [16]. To accelerate CSynth conformation equilibration, the ratio of average main diagonal contact probabilities $\overline{CP}_{i,j=i}$ to average off-diagonal contact probabilities $\overline{CP}_{i,j\geq i+2}$ was set to 100 by dividing all main diagonal contact probabilities by $\frac{\overline{CP}_{i,j=i}}{\overline{CP}_{i,j\geq i+2}}/100$, and the ratio of average 1st off-diagonal contact probabilities $\overline{CP}_{i,j=i+1}$ to average off-diagonal contact probabilities $\overline{CP}_{i,j\geq i+2}$ was set to 10 by dividing

449 all 1st off-diagonal contact probabilities by $\frac{\overline{CP}_{i,j=i+1}}{\overline{CP}_{i,j\geq i+2}}/10$. These modifications were chosen such that

450 after normalization, CSynth Hi-C attraction forces remained high enough to allow conformation

451 equilibration within minutes. Main diagonal and 1st off-diagonal contact probabilities were excluded

452 for all contact probability curve calculations.

453 Individual Hi-C contact files for each chromosome were then normalized by Iterative Correction

454 and Eigenvector decomposition (ICE) [89] and written to a text file in a CSynth compatible format via

455 the cooler python package [90].

456 **Generating 3D chromosome conformations with CSynth**

457 CSynth was accessed online at: https://csynth.github.io/csynth/csynth.html. The maximum

458 number of monomers supported by our system's GPU was determined by typing the command

459 gl.getParameter(gl.MAX_TEXTURE_SIZE) into the console of the web browser's developer tools.

460 The following optimized parameters (SI Text) were used:

461 - Spring power: 0. This sets the relative importance of long and short distance effects.

462 - Contact force: 60. This sets the magnitude of attraction between Hi-C contacts.

463 - Push-apart power: -4. This sets the exponent for a global repulsive force between monomers

464 that decays with distance. This exponent is analogous and has a similar value to the

465 empirically derived exponent relating Hi-C contact probability and mean spatial distance

466 determined by FISH in human IMR90 cells on chromosome 21 [91, 92].

467 After equilibration, the xyz coordinates of all monomers were saved to a text file. A picture of the

468 equilibrated structure was saved separately.

469 **Simulating CLC structures**

470 3D polymer models of CLCs were generated using a custom MATLAB script. Polymer length

471 (20 Mb, 4000 monomers) was chosen to match the largest *S. microadriaticum* and *F. kawagutii* Hi-C

472 scaffolds. Cholesteric discs were made by placing monomers on a square grid inside a circle of a

473 chosen diameter. Monomers were connected in primary sequence by "weaving" the polymer
474 backbone back and forth, parallel to a rotating director. The number and length of extrachromosomal
475 loops were randomized while maintaining the total amount of sequence in loops at a user-defined
476 fraction. After specifying the fraction of total monomers in extrachromosomal loops, and the fraction
477 of loop monomers in inter-disc loops (1 minus the fraction of loop monomers in intra-disc loops), the
478 remaining monomers were divided by the number of cholesteric discs minus 1 to give the mean inter-
479 disc loop length. Inter-disc loop lengths were then drawn from an exponential distribution with a
480 parameter determined by the mean inter-disc loop length. This was done to make loop starts/stops as
481 uniformly distributed in monomer space as possible. Intra-disc extrachromosomal loop lengths were
482 drawn from an analogous exponential distribution.

483 The primary sequence order of discs was either consecutive from top to bottom of the stack or
484 permuted within groups but not across locus-specific boundaries which separate groups adjacent in
485 primary sequence. In the case of CLCs with permuted disc order, inter-disc extrachromosomal loops
486 were assigned to connect disc pairs according to that disc pair's spatial distance separation rank
487 because adjacent discs in primary sequence located further away in 3D space require longer loops to
488 connect them.

489 **Simulating Hi-C contact maps**

490 For CLCs and Csynth, Hi-C contacts were simulated using $CP_{ij} \propto \frac{1}{(D_{ij})^4}$ [91, 92] where CP_{ij}
491 represents the Hi-C contact probability for genomic loci pair (i, j), and D_{ij} represents the Euclidean
492 distance between these two monomers. For population level Hi-C contact maps, the contact
493 probability for each single-conformation Hi-C contact map was averaged per pixel (i, j) across all
494 conformations.

495 For IGM, Hi-C contact maps were simulated as previously described in [32].

496 **Quantitative TAD metrics**

497 We used the FanC python package [93] to calculate insulation score [40] and directionality index
498 [41] using a Hi-C bin size of 5 kb and a window size of 100 kb.

499 **Equilibrium globule structures**

500 Equilibrium globule structures were downloaded from GEO using the accession number:
501 GSE18199 [50].

502 **Analyzing spatial organization of transcription**

503 RNA-seq data from asynchronous cell populations were accessed via the SRA using the
504 following accession numbers: for *S. microadriaticum*, SRR3337493 [36]; for *F. kawagutii*,
505 SRR9417753, SRR9417755, SRR9417756, SRR1300302, SRR1300303, SRR1300304, and
506 SRR1300305 [37, 38].

507 Alignment to the genome assembly was performed following the methods outlined in [17].

508 **Calculating normalized gene expression**

509 The largest 75 chromosome scaffolds, likely representing complete or near-complete
510 chromosome assemblies, for each dinoflagellate species were scaled isotropically to the same size,
511 with each chromosome long axis aligned horizontally. Using a cylindrical coordinate system, the total
512 number of RNA-seq reads were counted in each voxel, integrated azimuthally, and divided by the
513 cylindrical Jacobian which controls for variation in amount of DNA per voxel.

514 **Annotating TADs manually with CSynth**

515 After loading a Hi-C contact file into CSynth, the chromosome was hidden and its dynamics
516 were arrested. Hi-C contacts were displayed on CSynth's matrix viewer. The *matdistfar* parameter
517 was adjusted until TADs were clearly visible and rotation was set to 1.5 for a more upright matrix
518 view. To annotate a TAD, the TAD apex was selected using the mouse, and edges then appeared
519 outlining the TAD boundary. TAD coordinates were saved separately for each TAD using option+Z
520 (Macintosh) or alt+Z (Windows), and collectively exported in BED file format using shift+option+Z

521 (Macintosh) or shift+alt+Z (Windows). Finally, TADs were colored in 3D after un-hiding the
522 chromosome and dragging the .bed file into the CSynth browser window. For a detailed list of
523 commands, see SI Text.

524 Data availability

525 Code is available at: https://github.com/lucasphilipp1/3D-Reconstruction-of-Dinoflagellate-
526 Chromosomes-from-Hi-C-Data. *F. kawagutii* Hi-C data, *F. kawagutii* and *S. microadriaticum* CSynth
527 conformations, CLC structures, RNA-seq .bed files, and TAD .bed files, have been deposited to
528 https://doi.org/10.5281/zenodo.14285613.

529 Supplementary Information Appendix (SI)

530 Supplementary Information accompanying this article including SI Text, Supplementary Figures S1-
531 S15 and Tables S1-S4 can be found here _____ .

532 References

533 1. Bhaud, Y. et al. Morphology and behaviour of dinoflagellate chromosomes during the cell cycle
534 and mitosis. *J. Cell Sci.* 133, 1231-1239 (2000).
535 2. Miron, E., et al. Chromatin arranges in chains of mesoscale domains with nanoscale functional
536 topography independent of cohesin. *Sci. Adv.* 6, eaba8811 (2020).
537 3. Chen, J. K., et al. Nanoscale analysis of human G1 and metaphase chromatin in situ. *The EMBO*
538 *Journal*, *44*(9), 2658-2694 (2025).
539 4. Palozola, K.C., et al., Mitotic transcription and waves of gene reactivation during mitotic exit.
540 *Science* 358, 119-122 (2017).
541 5. Bouligand, Y., Soyer, M., Puiseux-Dao, S. La structure fibrillaire et l'orientation des chromosomes
542 chez les dinoflagellés. *Chromosoma* 24, 251-287 (1968).
543 6. Livolant, F., Bouligand, Y. New observations on the twisted arrangement of dinoflagellate

544 chromosomes. *Chromosoma* 68, 21-44 (1978).

545 7. Gautier, A., Michel-Salamin, L., Tosi-Couture, E., McDowall, E., Dubochet, J. Electron
546 microscopy of the chromosomes of dinoflagellates in situ: confirmation of Bouligand's liquid crystal
547 hypothesis. *J. Ultrastruct. Mol. Struct. Res.* 97, 10-30 (1986).

548 8. Rill, R. L., Livolant, F., Aldrich, H.C., Davidson, M. W. Electron microscopy of liquid crystalline
549 DNA: direct evidence for cholesteric-like organization of DNA in dinoflagellate chromosomes.
550 *Chromosoma* 98, 280-286 (1989).

551 9. Bouligand, Y., Norris, V. Chromosome separation and segregation in dinoflagellates
552 and bacteria may depend on liquid crystalline states. *Biochimie* 83, 187-192 (2001).

553 10. Sigee, D. C. Structural DNA and genetically active DNA in dinoflagellate chromosomes.
554 *Biosystems* 16, 203-210 (1983).

555 11. Sigee, D. C. The dinoflagellate chromosome. *Adv. Bot. Res.* 12, 205-264 (1986).

556 12. Soyer-Gobillard, M. O. The genome of the primitive eukaryote dinoflagellates: organization and
557 functioning. *Zool. Stud.* 35, 78-84 (1996).

558 13. Chow, M. H., Yan, K. T., Bennett, M. J., Wong, J. T. Birefringence and DNA condensation of
559 liquid crystalline chromosomes. *Eukaryot. Cell* 9, 1577-1587 (2010).

560 14. Wisecaver, J. H, Hackett, J.D. Dinoflagellate genome evolution. *Annu. Rev. Microbiol.* 65, 369-
561 387 (2011).

562 15. Riaz, S., Niaz, Z., Khan, S., Liu, Y., Sui, Z. Detection, characterization and expression dynamics
563 of histone proteins in the dinoflagellate *Alexandrium pacificum* during growth regulation. *Harmful*
564 *Algae* 87, 101630 (2019).

565 16. Nand, A., et al. Genetic and spatial organization of the unusual chromosomes of the dinoflagellate
566 *Symbiodinium microadriaticum*. *Nat. Genet.* 53(5), 618-629 (2021).

567 17. Marinov, G. K., et al. Transcription-dependent domain-scale three-dimensional organization in the
568 dinoflagellate *Breviolum minutum*. *Nat. Genet. 53*(5), 613-617 (2021).
569 18. Marinov, G. K., Kundaje, A., Greenleaf, W. J., Grossman, A. R. Protocol for mapping the three-
570 dimensional organization of dinoflagellate genomes. *STAR Protoc.* 5, 102941 (2024).
571 19. Aranda, M. et al., Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations
572 conducive to a symbiotic lifestyle. *Sci. Reports* 6, 39734 (2016).
573 20. Stephens, T. G. et al. Genomes of the dinoflagellate *Polarella glacialis* encode tandemly repeated
574 single-exon genes with adaptive functions. *BMC Biol.* 18, 1-21 (2020).
575 21. Bachvaroff, T. R., Place, A. R. From stop to start: tandem gene arrangement, copy number and
576 trans-splicing sites in the dinoflagellate *Amphidinium carterae*. PLoS One 3, e2929 (2008).
577 22. Beauchemin, M. et al. Dinoflagellate tandem array gene transcripts are highly conserved
578 and not polycistronic. *PNAS,* 109, 15793-15798 (2012).
579 23. Lin, S., Song, B., & Morse, D. Spatial organization of dinoflagellate genomes: novel insights and
580 remaining critical questions. *Journal of Phycology*, *57*(6), 1674-1678 (2021).
581 24. Oluwadare, O., Highsmith, M., & Cheng, J. An overview of methods for reconstructing 3-D
582 chromosome and genome structures from Hi-C data. *Biological procedures online*, *21*(1), 7 (2019).
583 25. Hua, N., Tjong, H., Shin, H., Gong, K., Zhou, X. J., & Alber, F. Producing genome structure
584 populations with the dynamic and automated PGS software. *Nat. Protoc.*, *13*(5), 915-926 (2018).
585 26. Todd, S., et al, CSynth: an interactive modelling and visualization tool for 3D chromatin structure.
586 *Bioinformatics* 37, 951-955 (2021).
587 27. Duan, Z., et al. A three-dimensional model of the yeast genome. *Nature*, 465(7296), 363-367
588 (2010).
589 28. Lesne, A., Riposo, J., Roger, P., Cournac, A., & Mozziconacci, J. 3D genome reconstruction from
590 chromosomal contacts. *Nat. Methods*, 11(11), 1141-1143 (2014).

591 29. Wang, H., Yang, J., Zhang, Y., Qian, J., & Wang, J. Reconstruct high-resolution 3D genome
592 structures for diverse cell-types using FLAMINGO. *Nat. Commun.*, *13*(1), 2645 (2022).
593 30. Giorgetti, L., et al. Predictive polymer modeling reveals coupled fluctuations in chromosome
594 conformation and transcription. *Cell*, *157*(4), 950-963 (2014).
595 31. Tjong, H., et al. Population-based 3D genome structure analysis reveals driving forces in spatial
596 genome organization. *PNAS*, *113*(12), E1663-E1672 (2016).
597 32. Boninsegna, L., et al. (2022). Integrative genome modeling platform reveals essentiality of rare
598 contact events in 3D genome organizations. *Nat. Methods*, *19*(8), 938-949.
599 33. Stevens, T. J., et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C.
600 *Nature* 544, 59-64 (2017).
601 34. Bintu, B., et al. Super-resolution chromatin tracing reveals domains and cooperative interactions
602 in single cells. *Science*, *362*(6413), eaau1783 (2018).
603 35. LaJeunesse, T. C., et al. Systematic revision of Symbiodiniaceae highlights the antiquity and
604 diversity of coral endosymbionts. *Curr. Biol.* 28, 2570-2580 (2018).
605 36. Liew, Y. J., Li, Y., Baumgarten, S., Voolstra, C. R., Aranda, M. Condition-specific RNA editing
606 in the coral symbiont *Symbiodinium microadriaticum*. *PLoS Genet.* 13, e1006619 (2017).
607 37. Li, T., et al. RNA-seq profiling of *Fugacium kawagutii* reveals strong responses in metabolic
608 processes and symbiosis potential to deficiencies of iron and other trace metals. *Sci. Total. Environ.*
609 705, 135767 (2020).
610 38. Keeling, P. J. et al. The marine microbial eukaryote transcriptome sequencing project (MMETSP):
611 illuminating the functional diversity of eukaryotic life in the oceans through transcriptome
612 sequencing. *PLoS Biol.* 12, e1001889 (2014).
613 39. Belton, J. M., et al. Hi–C: a comprehensive technique to capture the conformation of
614 genomes. *Methods*, *58*(3), 268-276 (2012).

615 40. Crane, E., et al. Condensin-driven remodelling of X chromosome topology during dosage
616 compensation. *Nature*, *523*(7559), 240-244 (2015).
617 41. Dixon, J. R., et al. Topological domains in mammalian genomes identified by analysis of
618 chromatin interactions. *Nature*, *485*(7398), 376-380 (2012).
619 42. Kubica, J., Korsak, S., Banecki, K. H., Schirman, D., Yadavalli, A. D., & Plewczynski, D. The
620 challenge of chromatin model comparison and validation: A project from the first international 4D
621 Nucleome Hackathon. *PLOS Computational Biology*, *21*(8), e1013358 (2025).
622 43. Yen, C. S., Steele, R. E., Rae, P. M. M. Chromatin structure in a dinoflagellate as revealed by
623 electron microscopy of DNA cross-linked in vivo with trimethylpsoralen. *J. Cell Biol.* 79, 120a
624 (1978).
625 44. Sanborn, A. L., et al. Chromatin extrusion explains key features of loop and domain formation in
626 wild-type and engineered genomes. *PNAS, 112*(47), E6456-E6465 (2015).
627 45. Fudenberg, G., et al. Formation of chromosomal domains by loop extrusion. *Cell reports*, *15*(9),
628 2038-2049 (2016).
629 46. Sabaté, T., et al. Uniform dynamics of cohesin-mediated loop extrusion in living human cells. *Nat.*
630 *Genet.*, 1-13 (2025).
631 47. Xu, J., et al. A comprehensive benchmarking with interpretation and operational guidance for the
632 hierarchy of topologically associating domains. *Nat. Commun.* 15, 4376 (2024).
633 48. An, L., et al. OnTAD: hierarchical domain structure reveals the divergence of activity among
634 TADs and boundaries. *Genome Biol.*, 20, 1-16 (2019).
635 49. Szabo, Q., Bantignies, F., & Cavalli, G. Principles of genome folding into topologically
636 associating domains. *Sci. Adv.*, 5(4), eaaw1668 (2019).
637 50. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding
638 principles of the human genome. *Science* 326, 289-293 (2009).

639 51. Naumova, N., et al. Organization of the mitotic chromosome. *Science* 342, 948-953 (2013).
640 52. Open2C, et al. Cooltools: enabling high-resolution Hi-C analysis in Python. *PLOS Comput. Biol.*
641 20, e1012067 (2024).
642 53. Zhang, B., Wolynes, P. G., Topology, structures, and energy landscapes of human chromosomes.
643 *PNAS,* 112, 6062-6067 (2015).
644 54. Orsi, G. A., et al. Biophysical ordering transitions underlie genome 3D re-organization during
645 cricket spermiogenesis. *Nat. Commun.* 14, 4187 (2023).
646 55. Babillot, C. Étude de l'incorporation d'uridine-^{3}H dans le noyau chez *l'Amphidinium carteri*,
647 dinoflagellé. C. R. Acad. Sci. (Paris) 271, 828-831 (1970).
648 56. Sala-Rovira, M. et al. Molecular cloning and immunolocalization of two variants of the major
649 basic nuclear protein (HCc) from the histone-less eukaryote *Crypthecodinium cohnii* (Pyrrhophyta).
650 *Chromosoma* 100, 510-518 (1991).
651 57. Géraud, M. L., Sala-Rovira, M., Herzog, M., Soyer-Gobillard, M. O. Immunocytochemical
652 localization of the DNA-binding protein HCc during the cell cycle of the histone-less dinoflagellate
653 protoctista *Crypthecodinium cohnii B. Biol. Cell* 71, 123-134 (1991).
654 58. Anderson, D. M., Grabber, A., Herzog, M. Separation of coding sequences from structural DNA
655 in the dinoflagellate *Crypthecodinium cohnii*. *Mol. Mar. Biol. Biotechno*l. 1, 89-96 (1992).
656 59. Allen, J., Roberts, T. M., Loeblich, A. R., Klotz, L. C. Characterization of the DNA from the
657 dinoflagellate *Crypthecodinium cohnii* and implications for nuclear organization. *Cell* 6, 161-169
658 (1975).
659 60. Udy, J. W., Hinde, R., Vesk, M. Chromosomes and DNA in Symbiodinium from Australian hosts.
660 *J. Phycol.* 29, 314-320 (1993).
661 61. Weynberg, K. D., et al. Prevalent and persistent viral infection in cultures of the coral algal
662 endosymbiont Symbiodinium. *Coral Reefs* 36, 773-784 (2017).

663 62. Sergey, G., Mariia, B., Yana, M., Eugene, S., Sergei, S. Ultrastructural organization of the
664 chromatin elements in chromosomes of the dinoflagellate *Prorocentrum minimum*. *Protistology* 12,
665 163-172 (2018).
666 63. Sunesen, I. et al. Morphological and molecular characterization of *Heterocapsa claromecoensis*
667 *sp.* nov. (Peridiniales, Dinophyceae) from Buenos Aires coastal waters (Argentina). *Eur. J. Phycol.*
668 55, 490-506 (2020).
669 64. Kalvelage, J., & Rabus, R. Multifaceted dinoflagellates and the marine model Prorocentrum
670 cordatum. *Microbial Physiology*, *34*(1), 197-242 (2024).
671 65. Li, C., et al. Characterization of the enzyme for 5-hydroxymethyluridine production and its role in
672 silencing transposable elements in dinoflagellates. *PNAS*, *121*(46), e2400906121 (2024).
673 66. Wang, D. Z., et al. 100 Dinoflagellate Genome Project. *The Innovation*. 0(0), 101120 (2025).
674 67. Gibcus, J. H., et al. A pathway for mitotic chromosome formation. *Science*, *359*(6376), eaao6135
675 (2018).
676 68. Contessoto, V. G., et al. Energy landscape analysis of the development of the chromosome
677 structure across the cell cycle. *PNAS*, *122*(12), e2425225122 (2025).
678 69. Chan, Y. H., Wong, J. T. Concentration-dependent organization of DNA by the dinoflagellate
679 histone-like protein HCc3. *Nucleic acids res.*, *35*(8), 2573-2583 (2007).
680 70. Marinov, G. K. et al. Genome-wide distribution of 5-hydroxymethyluracil and chromatin
681 accessibility in the *Breviolum minutum* genome. *Genome Biol.* 25, 115 (2024).
682 71. Janissen, R., et al. Global DNA compaction in stationary-phase bacteria does not affect
683 transcription. *Cell*, *174*(5), 1188-1199 (2018).
684 72. Frenkiel-Krispin, D., et al. Regulated phase transitions of bacterial chromatin: a non-enzymatic
685 pathway for generic DNA protection. *EMBO J.*, 20, 1184-1191 (2001).

686 73. Spector, D., Vasconcelos, A., Triemer, R. DNA duplication and chromosome structure in the
687 dinoflagellates. *Protoplasma* 105, 185-194 (1981).
688 74. Oakley, B. R., Dodge, J. D. Evidence for a double-helically coiled toroidal chromonema in the
689 dinoflagellate chromosome. *Chromosoma* 70, 277-291 (1979).
690 75. Haapala, O., Soyer, M. O. Structure of dinoflagellate chromosomes. *Nat. New Biol.* 244, 195-197
691 (1973).
692 76. Herzog, M., Soyer, M. The native structure of dinoflagellate chromosomes and their stabilization
693 by Ca2+ and Mg2+ cations. *Eur. J. Cell Biol.* 30, 33-41 (1983).
694 77. Dodge, J. D. The chromosomes of dinoflagellates. *Int. Rev. Cytol*. 94, pp. 5-19 (1985).
695 78. Wong, J. T. Y., Architectural organization of dinoflagellate liquid crystalline chromosomes.
696 *Microorganisms* 7, 27 (2019).
697 79. Le, T. B., Imakaev, M. V., Mirny, L. A., Laub, M. T. High-resolution mapping of the spatial
698 organization of a bacterial chromosome. *Science* 342, 731-734 (2013).
699 80. Eagen, K. P., Hartl, T. A., Kornberg, R. D. Stable chromosome condensation revealed by
700 chromosome conformation capture. *Cell* 163, 934-946 (2015).
701 81. Contessoto, V. G., et al. Interphase chromosomes of the Aedes aegypti mosquito are liquid
702 crystalline and can sense mechanical cues. *Nat. Commun.* 14, 1-12 (2023).
703 82. Yee, B., et al. Electron tomography of the nucleoid of *Gemmata obscuriglobus* reveals complex
704 liquid crystalline cholesteric structure. *Front. Microbiol.* 3, 326 (2012).
705 83. Gornik, S. G. et al. Loss of nucleosomal DNA condensation coincides with appearance of a novel
706 nuclear protein in dinoflagellates. *Curr. Biol.* 22, 2303-2312 (2012).
707 84. Yee, B., Sagulenko, E., Fuerst, J. A. Making heads or tails of the HU proteins in the
708 planctomycete *Gemmata obscuriglobus*. *Microbiology* 157, 2012-2021 (2011).

709 85. Saintillan, D., Shelley, M. J., Zidovska, A. Extensile motor activity drives coherent motions in a
710 model of interphase chromatin. *PNAS,* 115, 11442-11447 (2018).
711 86. Costas, E., Goyanes, V. J. Ultrastructure and division behaviour of dinoflagellate chromosomes.
712 *Chromosom*. (Berl) 95, 435-441 (1987).
713 87. Le, P., Ahmed, N., Yeo, G. W. Illuminating RNA biology through imaging. *Nat. Cell Biol.* 24,
714 815-824 (2022).
715 88. Li, T., et al. Genome improvement and core gene set refinement of *Fugacium kawagutii*.
716 Microorganisms 8, 102 (2020).
717 89. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome
718 organization. *Nat. Methods* 9, 999-1003 (2012).
719 90. Abdennur, N., Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled
720 arrays. Bioinformatics 36, 311-316 (2020).
721 91. Su, J. H., Zheng, P, Kinrot, S. S., Bintu, B., Zhuang, X. Genome-scale imaging of the 3D
722 organization and transcriptional activity of chromatin. *Cell* 182, 1641-1659 (2020).
723 92. Wang, S., et al. Spatial organization of chromatin domains and compartments in single
724 chromosomes. *Science* 353, 598-602 (2016).
725 93. Kruse, K., Hug, C. B., & Vaquerizas, J. M. FAN-C: a feature-rich framework for the analysis and
726 visualisation of chromosome conformation capture data. *Genome biology*, *21*(1), 303 (2020).
727

728 # Acknowledgements

729 We would like to thank Audrey Baguette for helpful discussions regarding Hi-C data pre-processing.
730 L.P. was supported by a Canadian Graduate Scholarship – Master's and a Canadian Graduate
731 Scholarship – Doctoral from the National Science and Engineering Research Council of Canada. The

work was funded by the New Frontiers in Research Fund (NFRFE-2019-00189 to S.C.W.). This research was undertaken, in part, thanks to funding from the Canada Research Chairs Program.

Author contributions

L.P. and S.C.W. designed the research, wrote the paper, and obtained funding; G.K.M. improved the *F. kawagutii* genome assembly using raw Hi-C reads, assembled *F. kawagutii* contact maps, and aligned RNA-seq reads to the genome; S.T. added data visualization and manual TAD annotation functionalities to CSynth; L.P. performed all other research aspects.

Competing interest statement

The authors declare no conflict of interest.

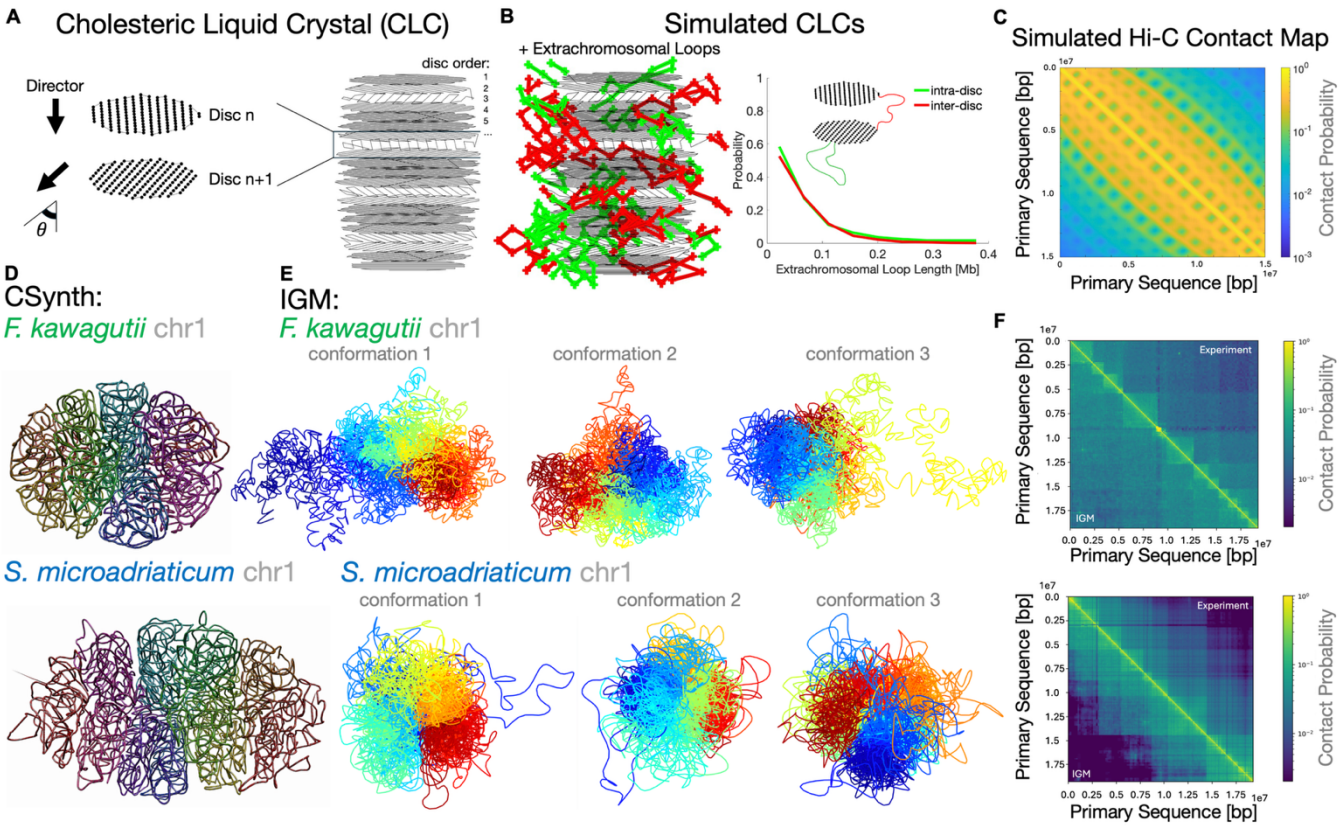

Figure 1. Dinoflagellate chromosomes lack cholesteric discs. A) The Cholesteric Liquid Crystal (CLC) model. The polymer aligns with a director (black arrows), which rotates by a constant angle θ between cholesteric discs. Discs are arranged in a consecutive order. B) CLCs are simulated by adding extrachromosomal loops to the stack of discs. Loops are positioned randomly and have exponentially distributed lengths. Intra-disc loops start and end within the same disc (green) and inter-disc loops connect vertically adjacent discs (red). C) Simulated Hi-C contact map for a population of 10^6 simulated CLC structures. D) CSynth conformations at 5 kb per monomer resolution of chromosome scaffold 1 from *Fugacium kawagutii* (Hi-C data: [18, 88]) and *Symbiodinium microadriaticum* (Hi-C data: [16]). Color indicates primary sequence. For the complete set of consensus conformations of dinoflagellate chromosome scaffolds, see Supplementary Figs. S4, S5. E) IGM conformations at 5 kb per monomer resolution from a population of 100 total conformations of the same chromosomes as in D. For a complete set of population-based conformations, see Supplementary Figs. S6, S7. F) Empirical Hi-C contact maps of representative dinoflagellate chromosomes (data: [16, 18, 88], upper right triangles) and simulated Hi-C contact maps from IGM populations (lower left triangles). Pearson correlation coefficients: $r = 0.91$ for *F. kawagutii*, $r = 0.84$ for *S. microadriaticum*.

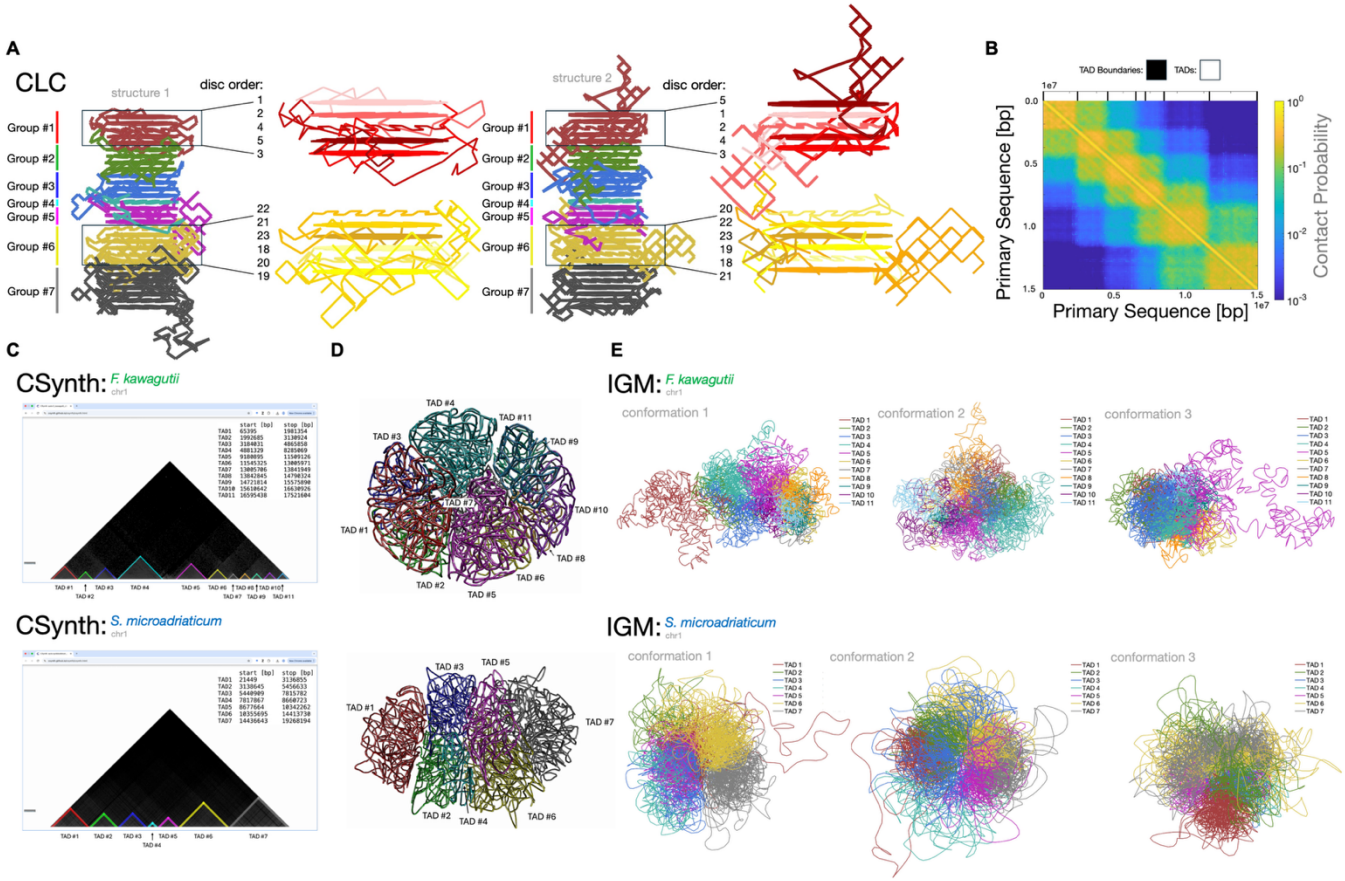

Figure 2. Locus-specific boundaries can recapitulate TADs in CLCs. A) Discs are grouped according to observed TAD size [16]. Insulating boundaries are added in between each group. Within each group, the primary sequence order of discs is permuted in the sub-stack across structures in the population. B) Simulated Hi-C contact map for CLCs with disc order permuted within groups. C) Manual annotation of TADs in CSynth generates a text file of boundary coordinates (inset). TADs can be visualized as distinct colors on the Hi-C contact map (bottom triangle) and in the 3D polymer conformation. D) TADs visualized using CSynth for *F. kawagutii* chr1 (top) and *S. microadriaticum* chr1 (bottom) at 5 kb per monomer resolution. E) TADs visualized in IGM conformations (3 of 100 total) of the same chromosomes.

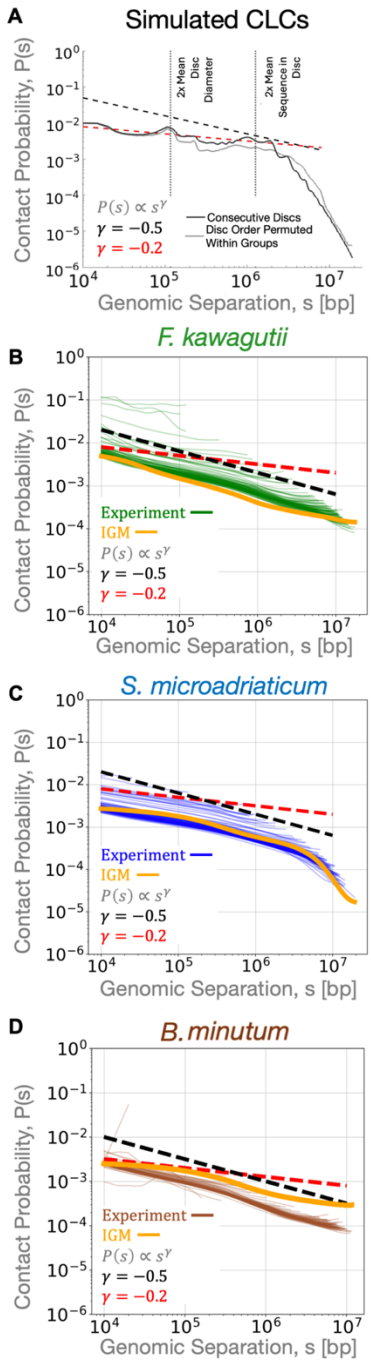

Figure 3. Contact probability curves for dinoflagellate chromosome scaffolds are inconsistent with CLC model predictions. A) Contact probability curve for a population of 10^6 CLC structures with either consecutive disc order or disc order permuted within groups. The genomic separations corresponding to twice the mean cholesteric disc diameter and mean for two cholesteric discs are indicated by vertical dotted lines. Black dashed line represents a power law relationship with an exponent $\gamma \approx -0.5$. Red dashed line represents a power law with exponent $\gamma \approx -0.2$. Contact probability curves for the 75 largest chromosome scaffolds for B) *F. kawagutii* (Hi-C data: [18]); C) *S. microadriaticum* (Hi-C data: [16]); D) *Breviolum minutum* (Hi-C data: [17]). Each curve corresponds to an individual chromosome scaffold. Orange lines correspond to contact probability curves for a population of 100 IGM conformations for chromosome 1.

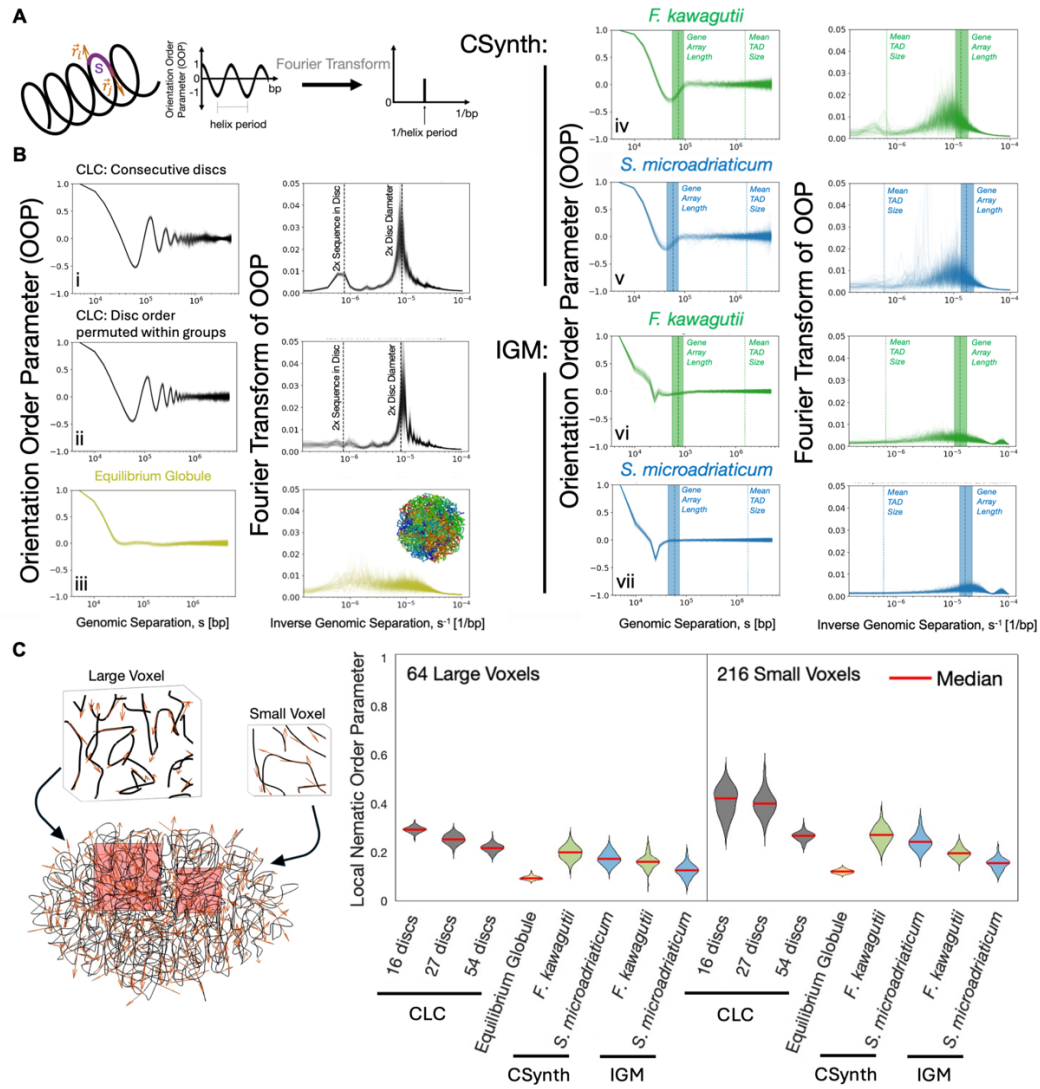

Figure 4. Dinoflagellate chromosomes exhibit moderate orientational and nematic order. A) Orientational order parameter (OOP) and its Fourier transform for a simple helix. Tangent vectors $\vec{r}_i$ and $\vec{r}_j$ (red arrows) at monomers *i* and *j*, located a primary sequence separation *s* (purple) apart. B) Left: OOP. Right: Fourier transform of OOP. Each curve corresponds to one of 75 structures/conformations. Row i: CLCs with consecutive disc order. Row ii: CLCs with disc order permuted within groups. Dashed lines: reciprocal of the average amount of sequence in two cholesteric discs, reciprocal of the amount of sequence in twice the average disc diameter. Row iii: Equilibrium globules (data: [50]). Row iv: *F. kawagutii*, CSynth. Row v: *S. microadriaticum*, CSynth. Each curve corresponds to a consensus conformation of one of the 75 largest chromosomes. Row vi: *F. kawagutii*, IGM. Row vii: *S. microadriaticum*, IGM. Each curve corresponds to one conformation from a population of 100 total conformations of chromosome 1. The mean gene array size (±1 standard deviation) and the mean TAD size are indicated. C) Left: Example CSynth conformation of a dinoflagellate chromosome (black) with tangent vectors (red arrows). Small (edge length = 16.7% chr long axis) and large (edge length = 25% chr long axis) voxels (red boxes) are blown-up as insets and used to estimate short and long range orientational order, respectively. Right: Local nematic order parameter is calculated for all tangent vectors within every voxel and across 75 structures/conformations.

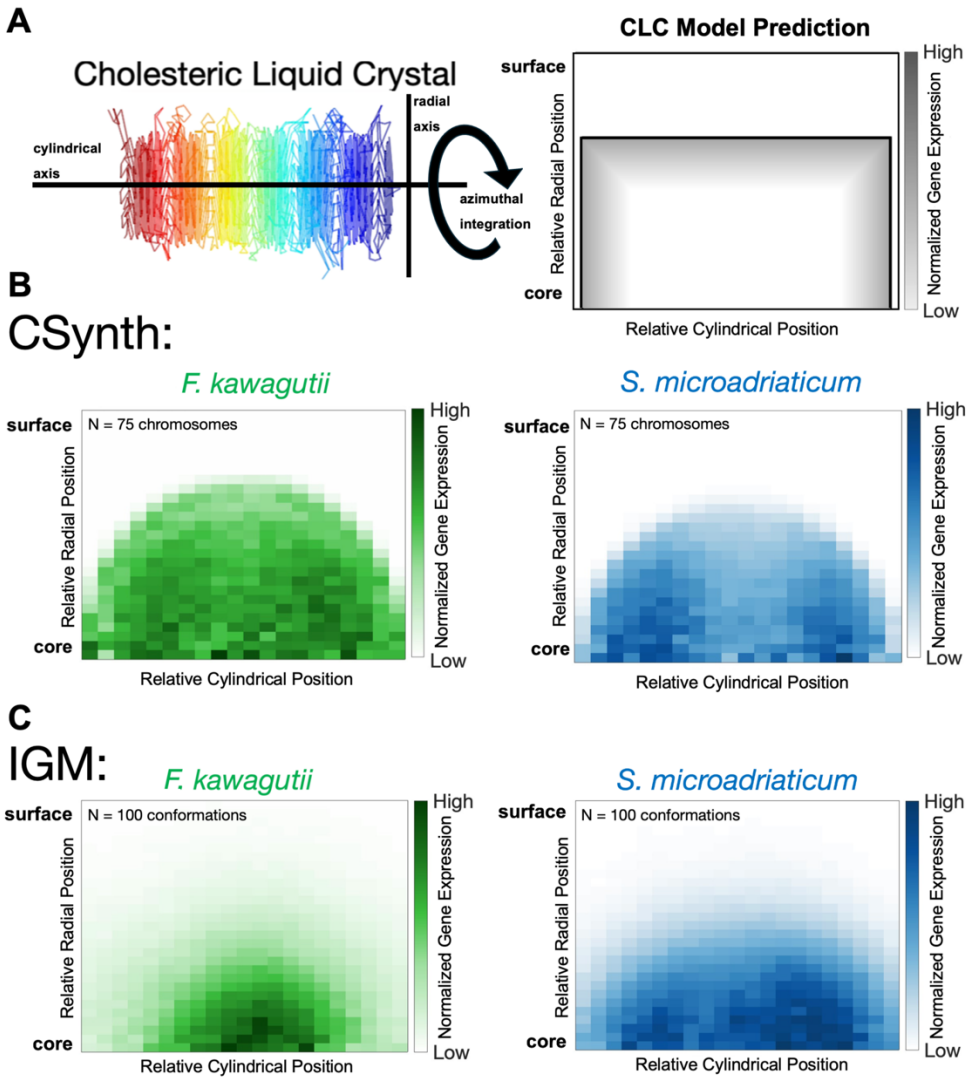

Figure 5. Active genes occur throughout dinoflagellate chromosomes. A) Cylindrical coordinate system with cylindrical axis aligned with the chromosome long axis. The outward radial axis is perpendicular to the cylindrical axis. The CLC model predicts that active genes are near the chromosome surface on extrachromosomal loops, and not in the condensed core. B) Normalized gene expression (data: [36-38]) as a function of distance to core (relative radial position) and position along long axis (relative cylindrical position) of consensus conformations for *F. kawagutti* (green) and *S. microadriaticum* (blue). C) Normalized gene expression of population-based conformations of chromosome 1 for each species using the same cylindrical coordinate system.