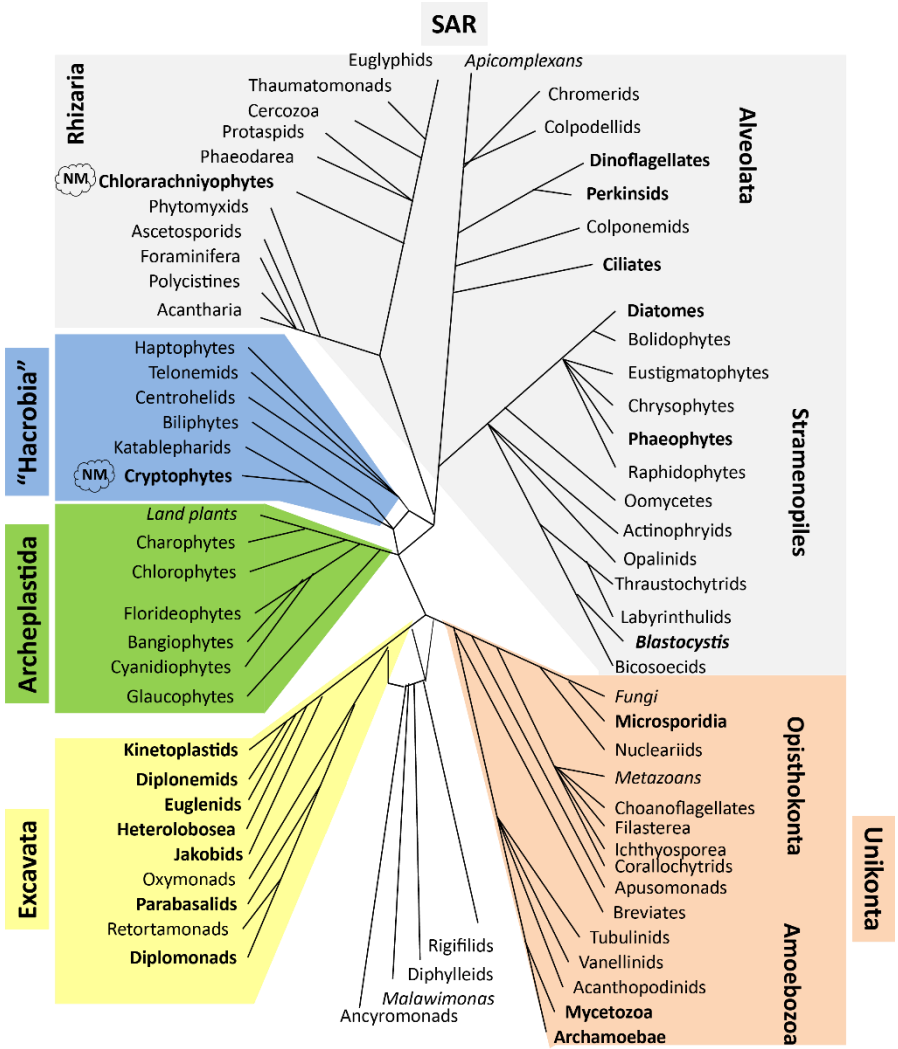
THE PHYSICAL GENOME ACRROSS EVOLUTION

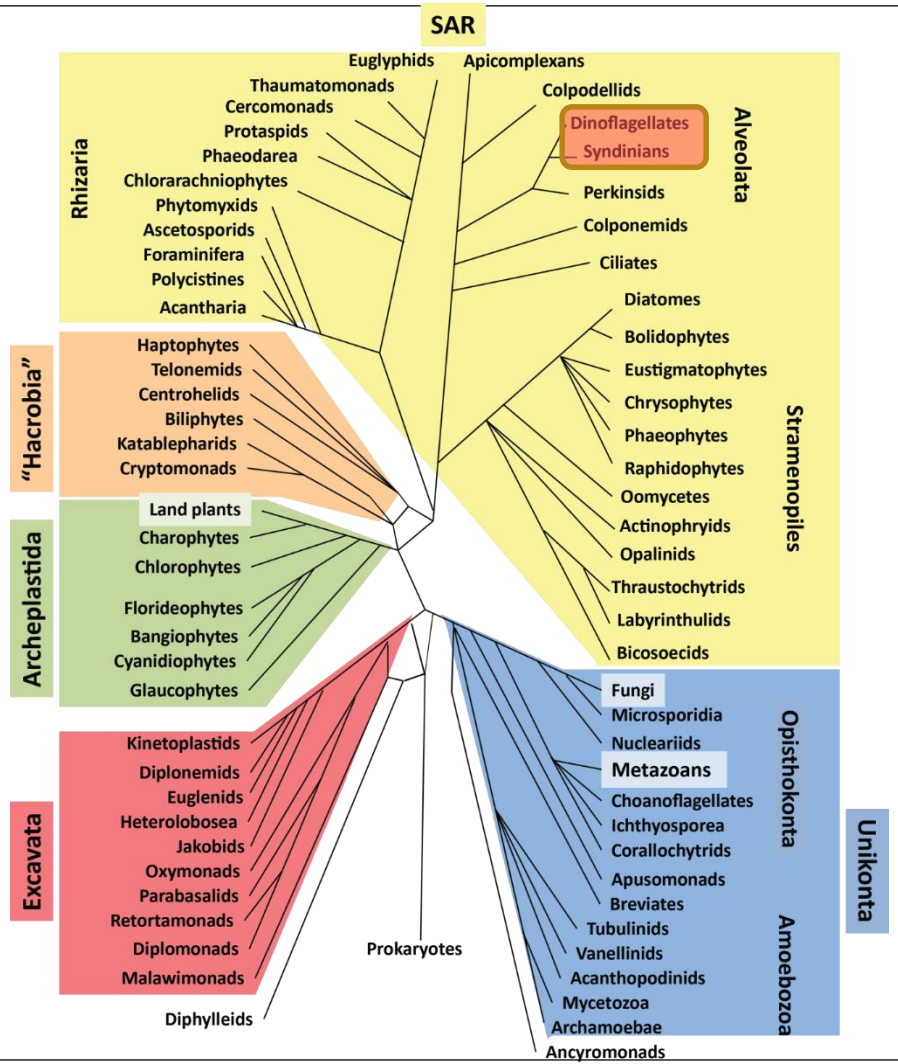
EUKARYOTE TREE



GENERAL QUESTIONS:

- What are the deepest principles of chromatin organization and gene expression? Studying its extremes can make apparent previously obscured features.
- Such deviations from the norm are known, but have generally not been studied at all with modern tools
- How did the regulatory apparatus and mechanisms evolve across the eukaryotic tree of life
- How many times did distal enhancers originate? Why and how?
- What are the conserved and derived chromatin states across different lineages?
- What is the relationship between genome organization and organismal complexity?
- What does all of that tells us about mammalian genomes?
- Finally, some things are just too cool on their own to not be studied.

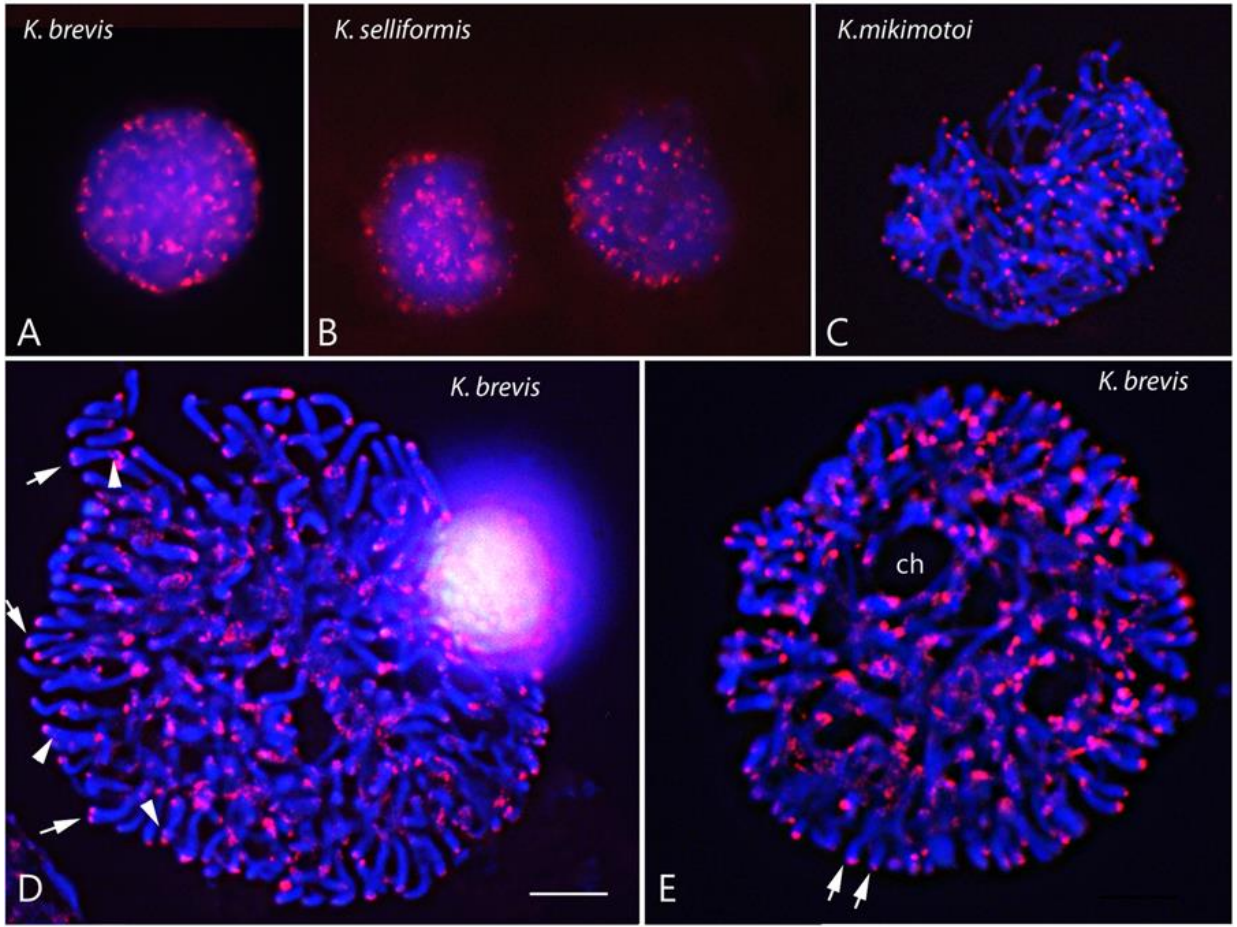
DINOFLAGELLATES



DINOFLAGELLATES AND THEIR SPECIAL FEATURES

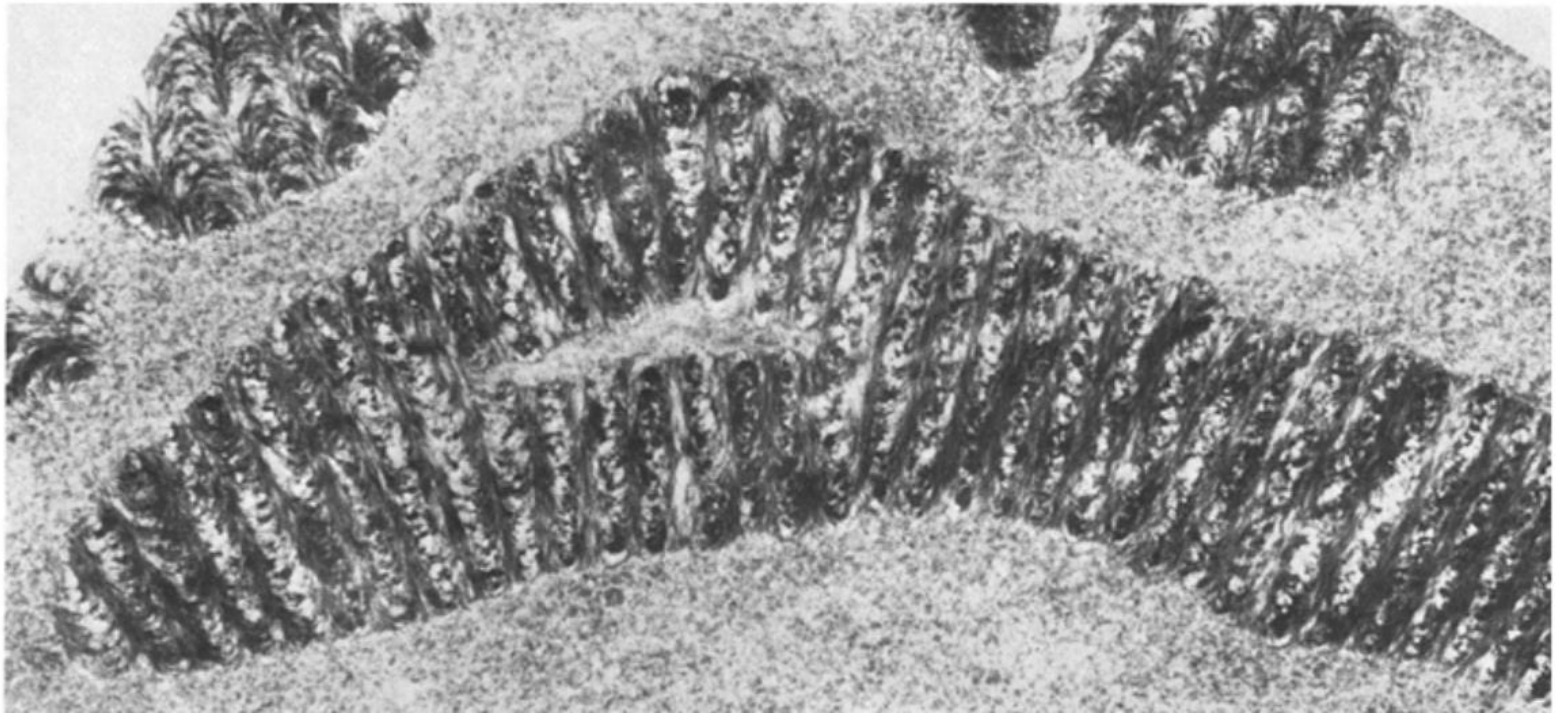
- Permanently condensed fibrilar chromosomes
- Very low protein-to-DNA ratio (~1/10th of the usual)
- Histones are in low abundance, long thought to be completely absent
- High percentage of 5-hydroxymethyluracile (up to 40%)
- Huge genomes, often with tandem arrays of the same gene
- Extremely intron rich (19 introns per gene on average); unique splice sites
- Few transcription factors
- Gene regulation is hypothesized to happen either at the posttrascriptional level or at the level of the control of chromatin looping

DINOFLAGELLATE CHROMOSOMES



- Note: these are interphase chromosomes!

Cuadrado et al. 2019



Livoland & Bouligand, 1978

GENERAL DINOFLAGELLATE QUESTIONS:

- How is chromatin organized in 3D space in permanently condensed chromosomes mostly without histones?
- How is transcriptional regulation (if it exists) accomplished in such an environment?
- How has the transcriptional machinery adapted to transcribing through DVNPs
- What is the role of the very divergent histones?
- What is the role of dhmU?
- and many others

SYMBIODINIUM HI-C

CORAL SYMBIOSIS

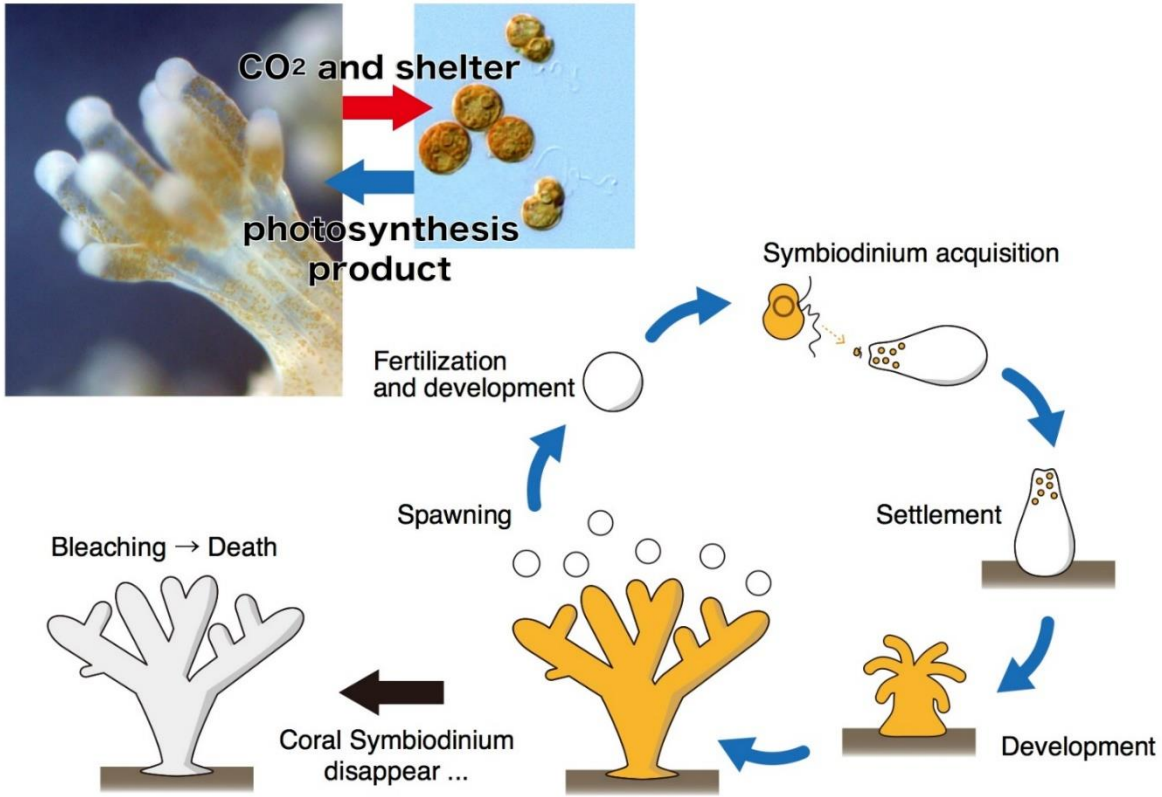


Figure 2. A symbiotic relationship between corals and Symbiodinium

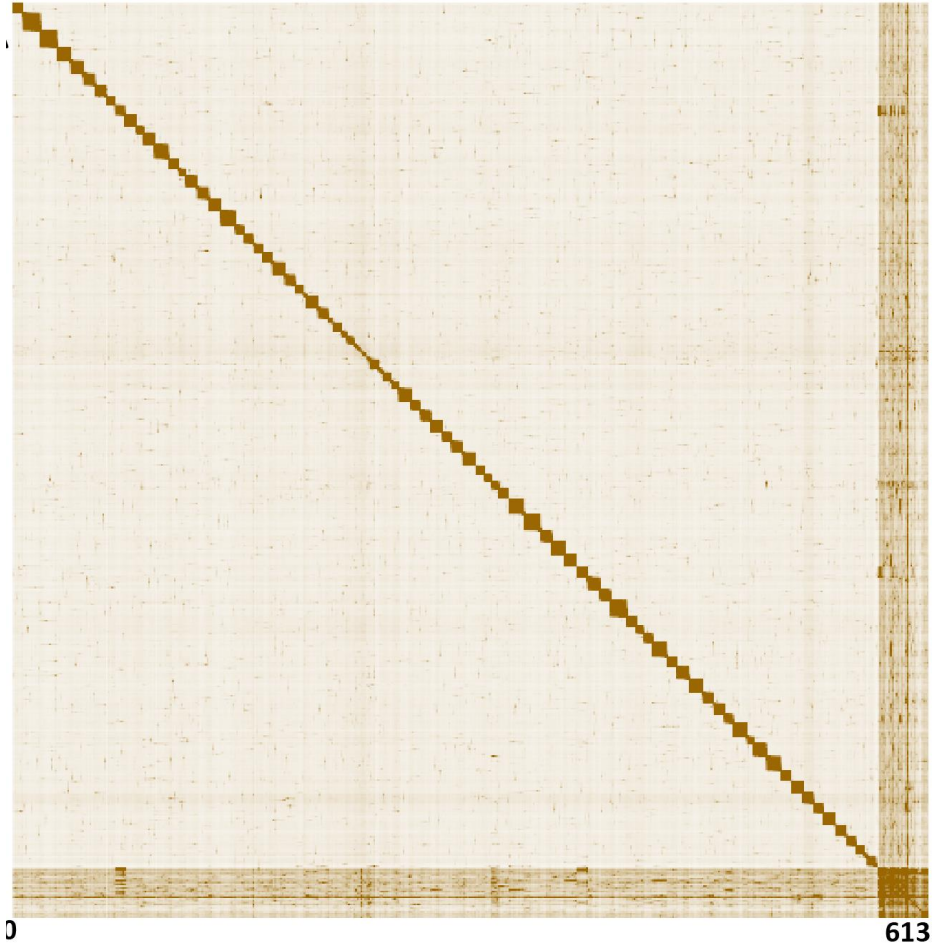
CORAL BLEACHING

SYMBIODINIUM GENOME ASSEMBLIES

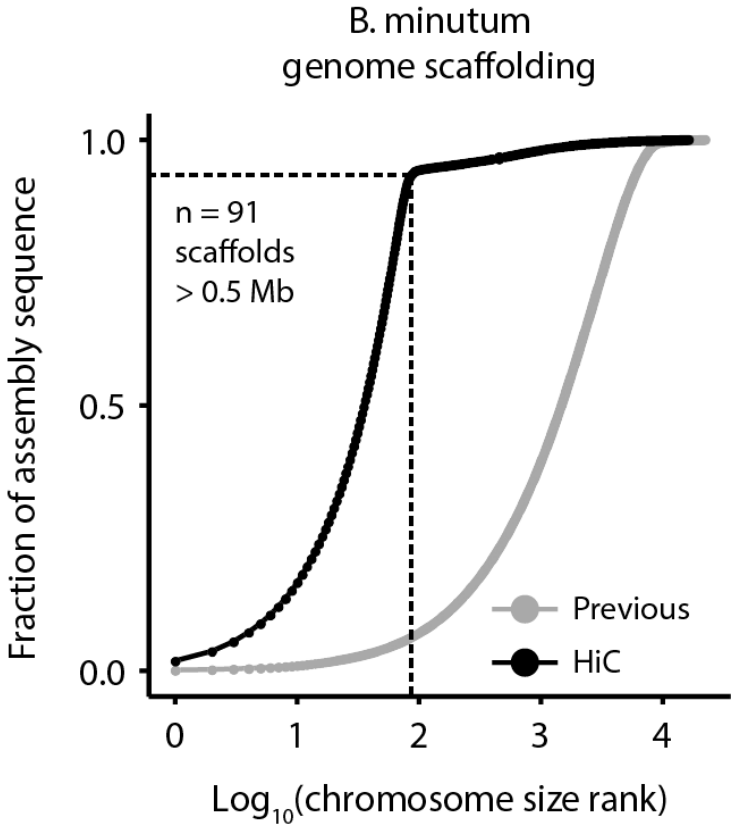
#Contig length	S. kawagutii	S. microadriaticum	S. minutum_Clade_B1
0	23412	136	12051
1,000	2611	6868	2305
10,000	1434	1039	5587
100,000	2534	1516	1956
1,000,000	49	136	0
10,000,000	0	0	0
100,000,000	0	0	0

	S. kawagutii	S. microadriaticum	S. minutum_Clade_B1
N50	125,226	573,512	380,908
N90	31,482	145,806	109,232

HI-C SCAFFOLDING

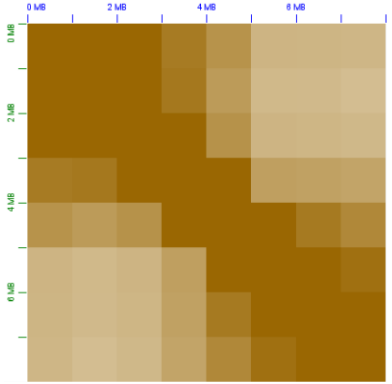


HI-C SCAFFOLDING

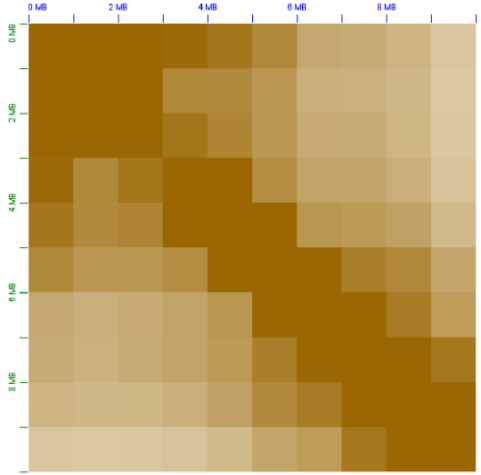


BROAD STRUCTURE:

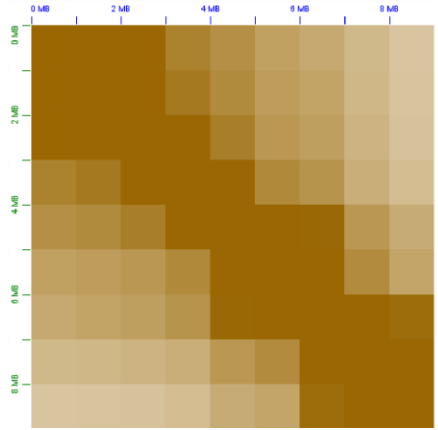
pseudochromosome 10



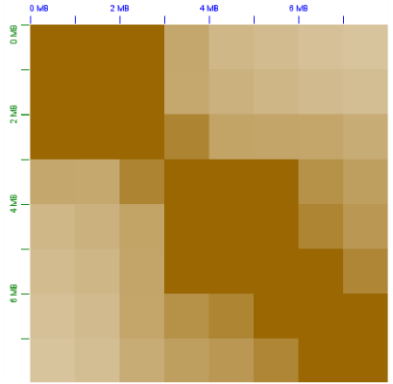
pseudochromosome 4



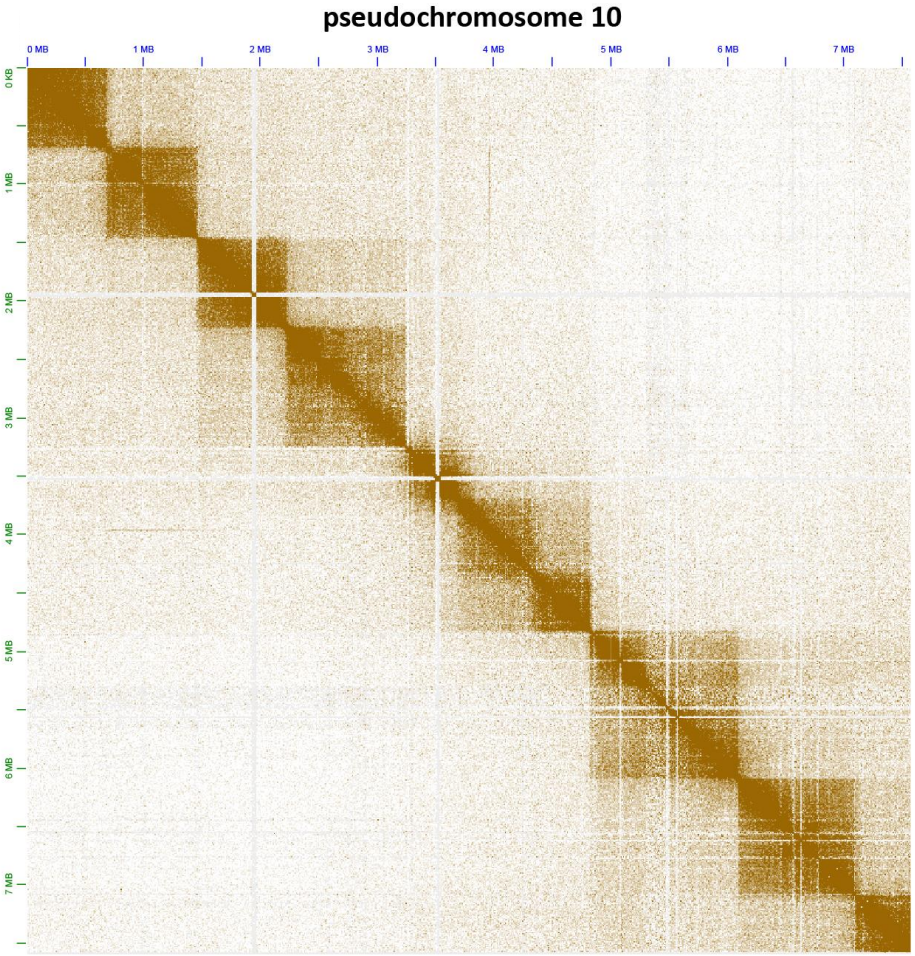
pseudochromosome 5

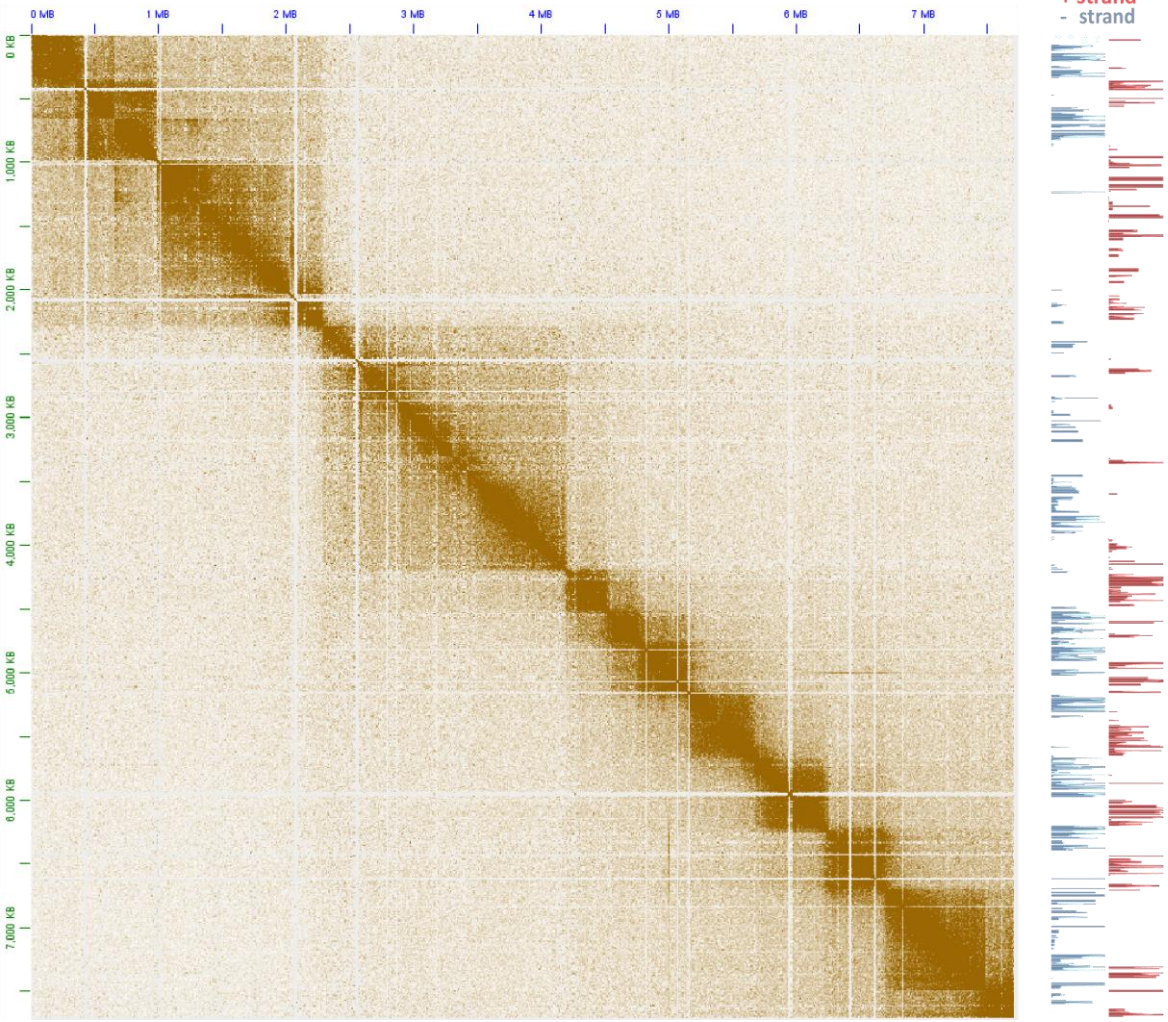


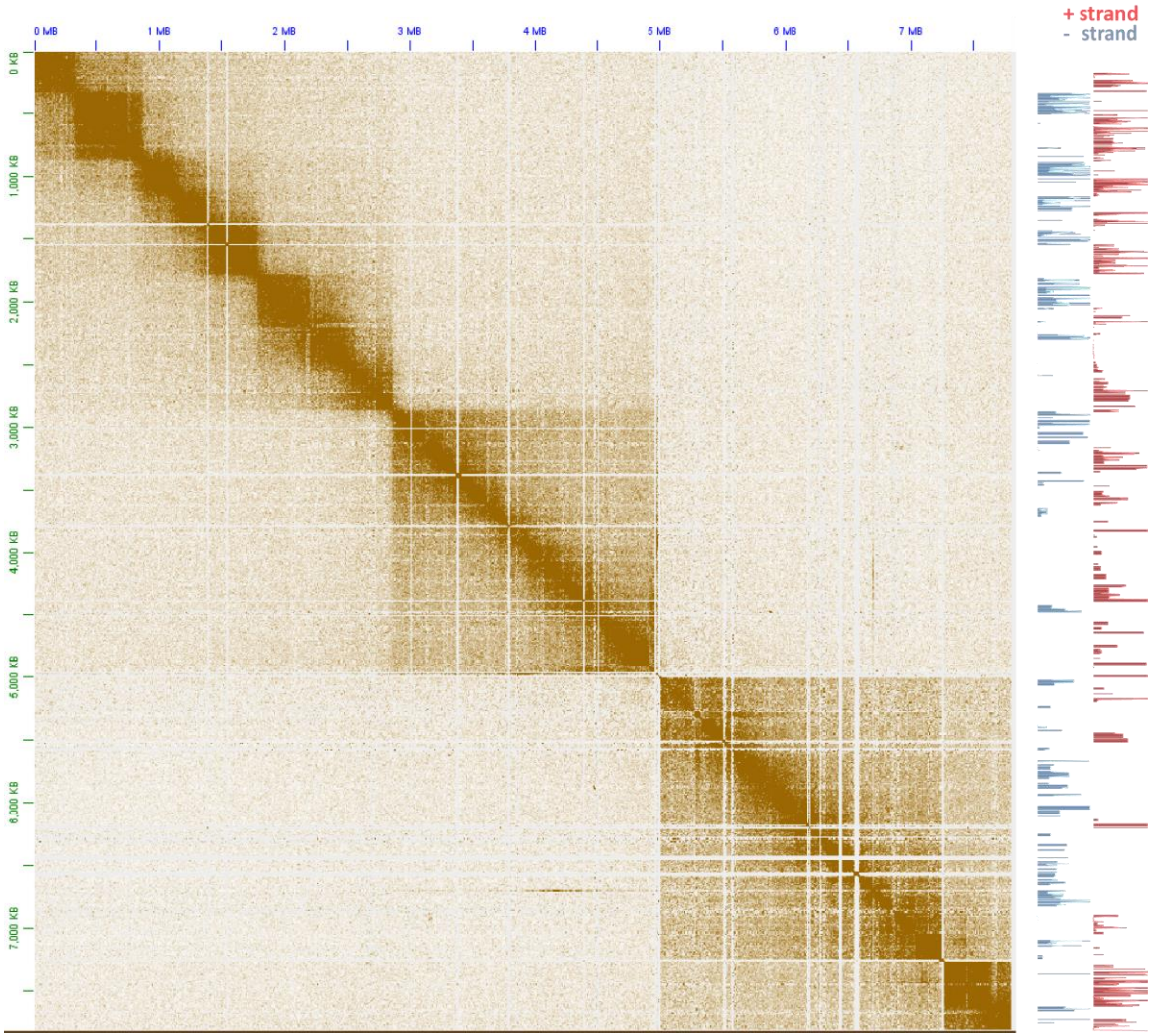
pseudochromosome 6



DOMAIN STRUCTURE:

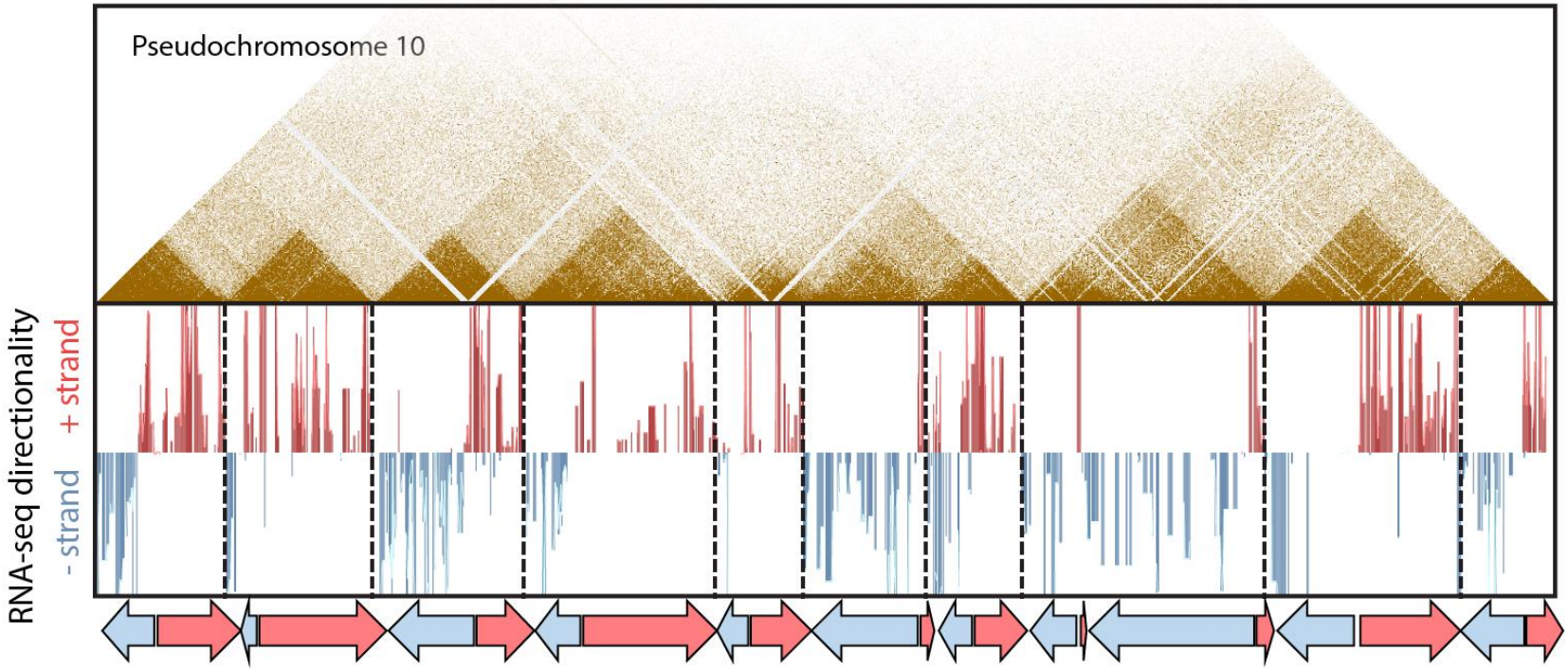



0 MB
1 MB
2 MB
3 MB
4 MB
5 MB
6 MB
7 MB
0 KB
1,000 KB
2,000 KB
3,000 KB
4,000 KB
5,000 KB
6,000 KB
7,000 KB
+ strand
- strand


0 MB
1 MB
2 MB
3 MB
4 MB
5 MB
6 MB
7 MB
0 KB
1,000 KB
2,000 KB
3,000 KB
4,000 KB
5,000 KB
6,000 KB
7,000 KB
+ strand
- strand

F

Gene directionality and topological domains



DOMAIN BOUNDARIES

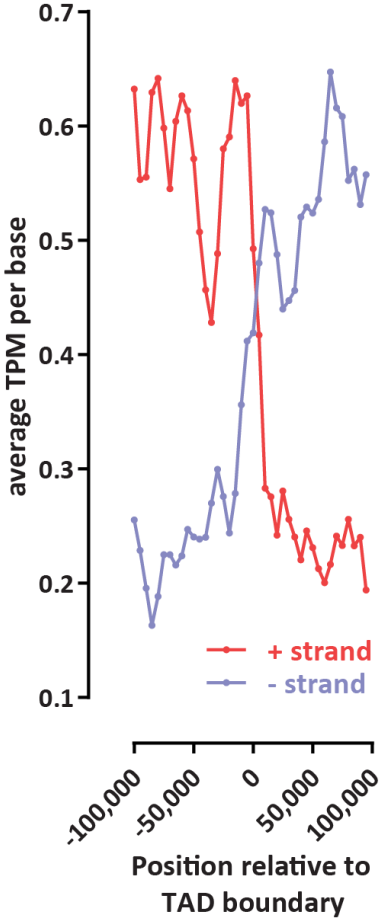
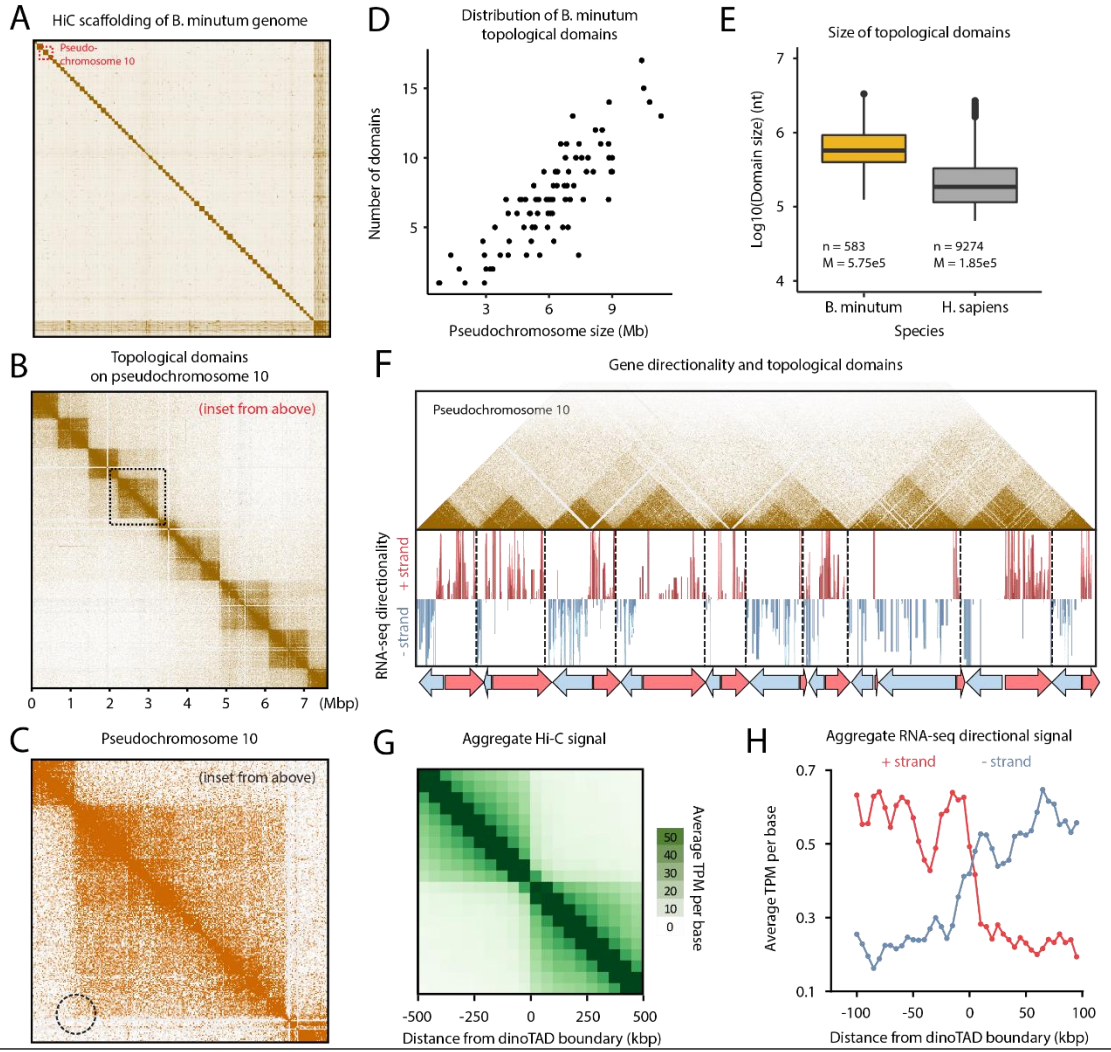
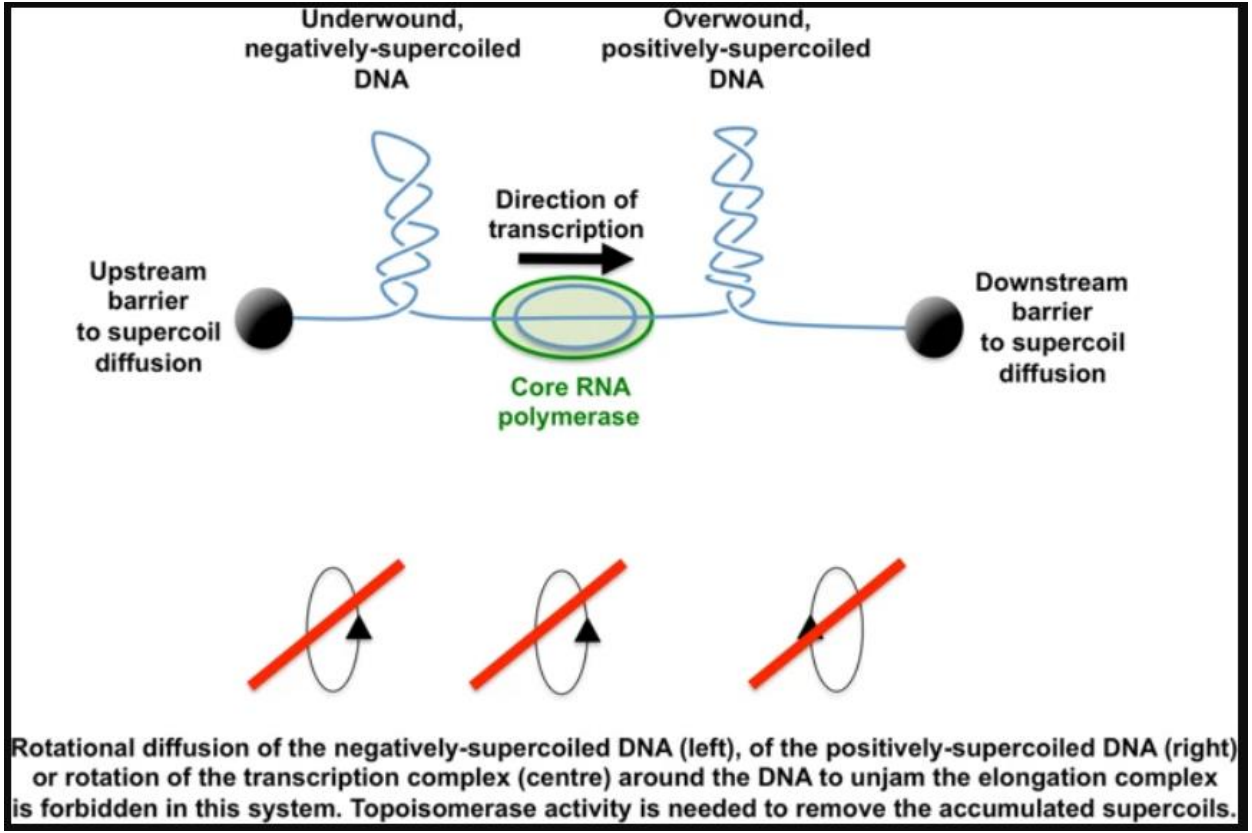


FIGURE 1


A
HiC scaffolding of B. minutum genome
Pseudo-chromosome 10
B
Topological domains on pseudochromosome 10
(inset from above)
0 1 2 3 4 5 6 7 (Mbp)
C
Pseudochromosome 10
(inset from above)
D
Distribution of B. minutum topological domains
Number of domains
Pseudochromosome size (Mb)
E
Size of topological domains
Log10(Domain size) (nt)
n = 583
M = 5.75e5
n = 9274
M = 1.85e5
B. minutum
H. sapiens
Species
F
Gene directionality and topological domains
Pseudochromosome 10
RNA-seq directionality
+ strand
- strand
G
Aggregate Hi-C signal
Average TPM per base
Distance from dinoTAD boundary (kbp)
H
Aggregate RNA-seq directional signal
+ strand
- strand
Average TPM per base
Distance from dinoTAD boundary (kbp)

WHAT IS THE MECHANISM?

- Transcription-induced supercoiling is one possibility



Rotational diffusion of the negatively-supercoiled DNA (left), of the positively-supercoiled DNA (right) or rotation of the transcription complex (centre) around the DNA to unjam the elongation complex is forbidden in this system. Topoisomerase activity is needed to remove the accumulated supercoils.

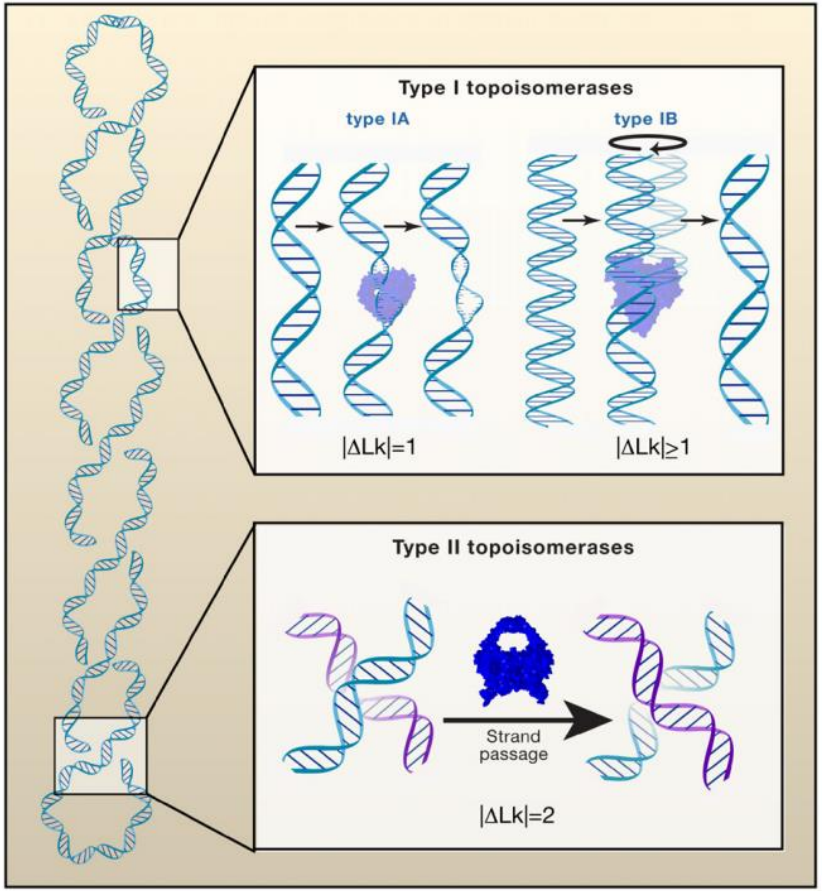


Figure 4. Types of Topoisomerases

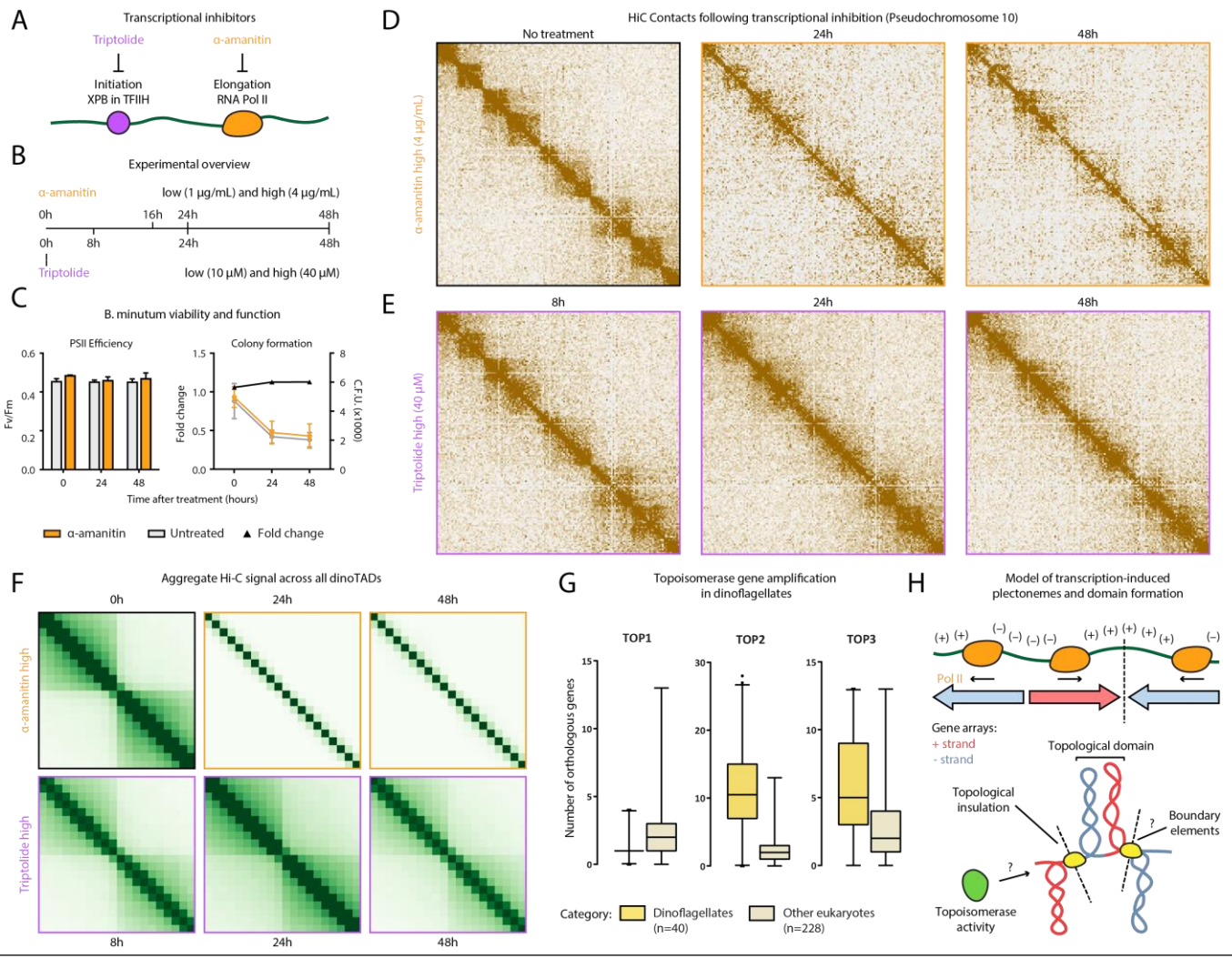
(Top) Type I topoisomerases cleave a single strand of DNA and relax a supercoil either by passing the other strand through an enzyme-DNA linked intermediate (type IA enzymes) or by a strand-swivel mechanism (type IB enzymes).

(Bottom) Type II topoisomerases cleave duplex DNA and then relax the supercoil by passing a second duplex DNA through the transient enzyme-DNA linked intermediate.

PREDICTION:

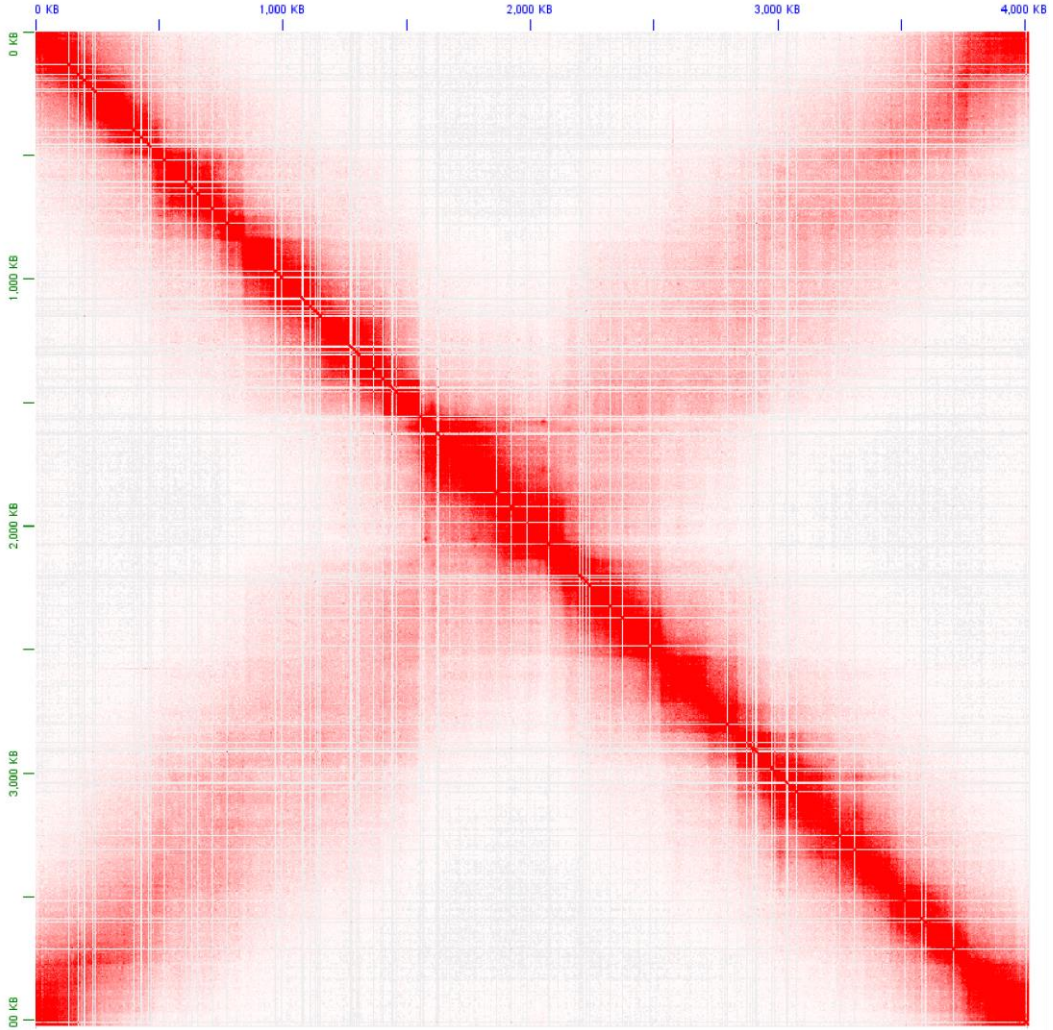
- If transcription is blocked, domains should disappear

FIGURE 2


A
Transcriptional inhibitors
Triptolide
α-amanitin
Initiation
XPB in TFIIH
Elongation
RNA Pol II
B
Experimental overview
α-amanitin
low (1 µg/mL) and high (4 µg/mL)
0h
16h
24h
48h
0h
8h
24h
48h
Triptolide
low (10 µM) and high (40 µM)
C
B. minutum viability and function
PSII Efficiency
Fv/Fm
Colony formation
Fold change
C.F.U. (x1000)
Time after treatment (hours)
α-amanitin
Untreated
Fold change
D
HiC Contacts following transcriptional inhibition (Pseudochromosome 10)
No treatment
24h
48h
α-amanitin high (4 µg/mL)
E
8h
24h
48h
Triptolide high (40 µM)
F
Aggregate Hi-C signal across all dinoTADs
0h
24h
48h
α-amanitin high
Triptolide high
8h
24h
48h
G
Topoisomerase gene amplification in dinoflagellates
TOP1
TOP2
TOP3
Number of orthologous genes
Category:
Dinoflagellates (n=40)
Other eukaryotes (n=228)
H
Model of transcription-induced plectonemes and domain formation
Pol II
Gene arrays:
+ strand
- strand
Topological domain
Topological insulation
Boundary elements
Topoisomerase activity

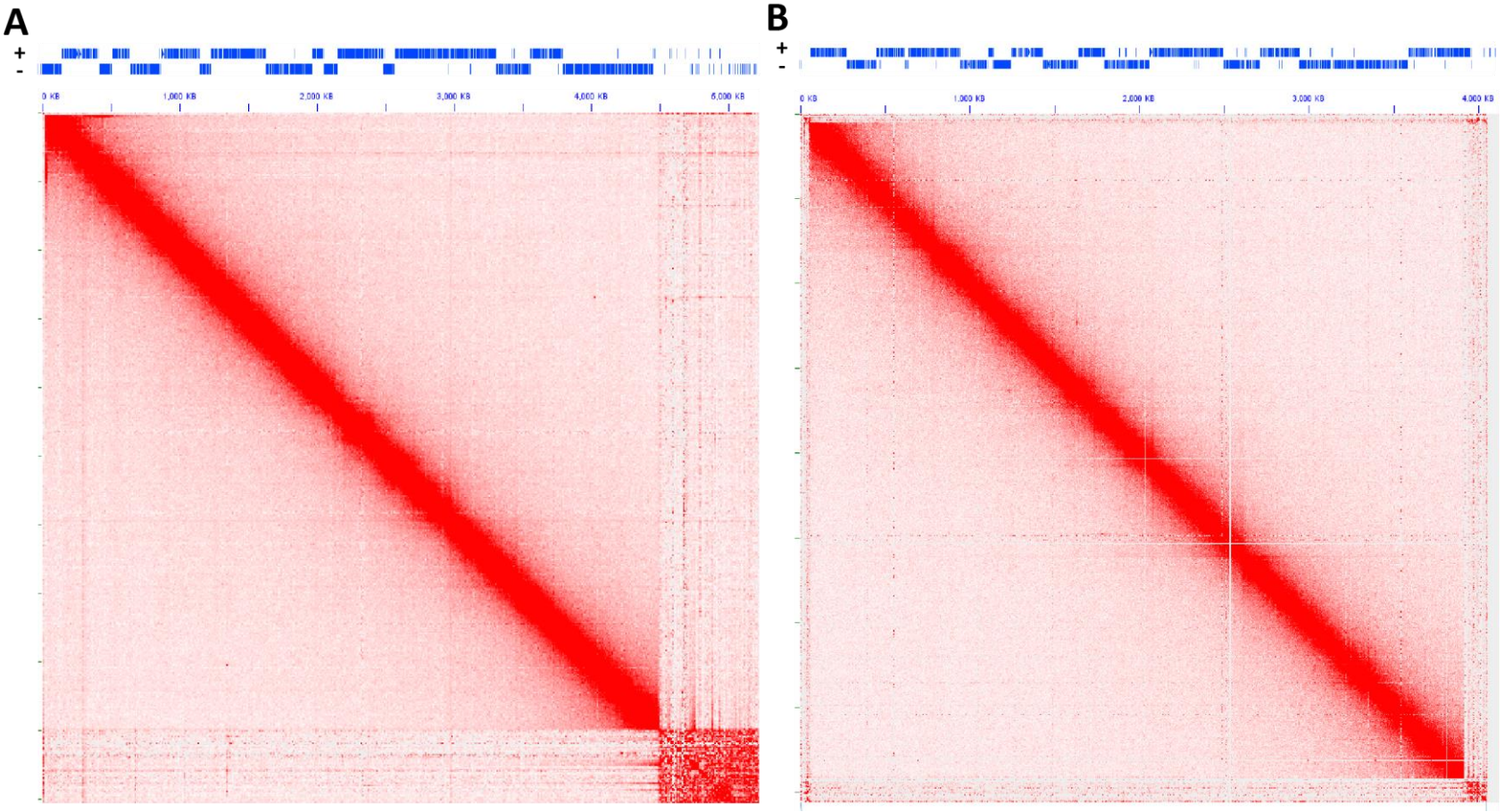
Caulobacter

- prokaryote



Trypanosoma

- eukaryote with gene arrays but also with histones



BRIEF REVIEW OF FOLDING MECHANISMS IN EUKARYOTES:

- There are two main mechanisms driving folding in eukaryotes
- The main one appears to be compartments, driven by associating between similar chromatin states
- Topological insulation on loop extrusion of the CTCF kind operates as an orthogonal mechanism
- After the *Symbiodinium* Hi-C it appears that supercoiling is also a fundamental topological force, but its effects are largely masked in other clades

CONCLUSIONS

- Strong insulation domains are observed in the *Symbiodinium* genome
- These appear to correspond to pairs of divergent multicistronic gene arrays
- Clear loop contacts are not immediately obvious
- Transcription-induced supercoiling appears to be the mechanism driving their formation
- Supercoiling emerges as another fundamental topological force shaping genome folding
- Its effects are masked in most eukaryotes by the presence of nucleosomes – interactions between histones block the formation of plectonemes

OPEN QUESTIONS

- What is the mechanism establishing boundaries?

- It is not immediately obvious why a collision of positive supercoiling trends should form a sharp boundary

- It is possible that there are specific boundary elements with a distinct chromatin state, e.g. perhaps that is where the histones are

- As it is at present not possible to answer that question without extensive additional experiments involving raising antibodies, ChIP-seq, mass-spectrometry, etc., we hope to use computational modelling to gain insights into the possible mechanisms