EVOLUTION AND DIVERSITY OF CHROMATIN ARCHITECTURE ACROSS THE DEEP ORGANISMAL DIVERSITY

GEORGI K. MARINOV

January 31st 2024

department of GENETICS
Stanford University
School of Medicine

Greenleaf Lab

I. INTRODUCTION

PERSONAL BACKGROUND

- Wold Lab, Caltech
    - As part of the ENCODE Project
        - Development of methods for optimizing and analyzing ChIP-seq and RNA-seq datasets
        - Development of early scRNA-seq techniques
        - Mapping gene regulatory networks involved in muscle differentiation
    - Role of piRNAs in silencing repetitive elements in *Drosophila* and mammals
    - Mapping the regulatory landscape of mitochondrial genomes

- Michael Lynch Lab (IU Bloomington)
    - Genome evolution in *Paramecium* ciliates
    - Phylogenomics of chromatin proteins in eukaryotes
    - Energetics of cellular organization and its role in evolution

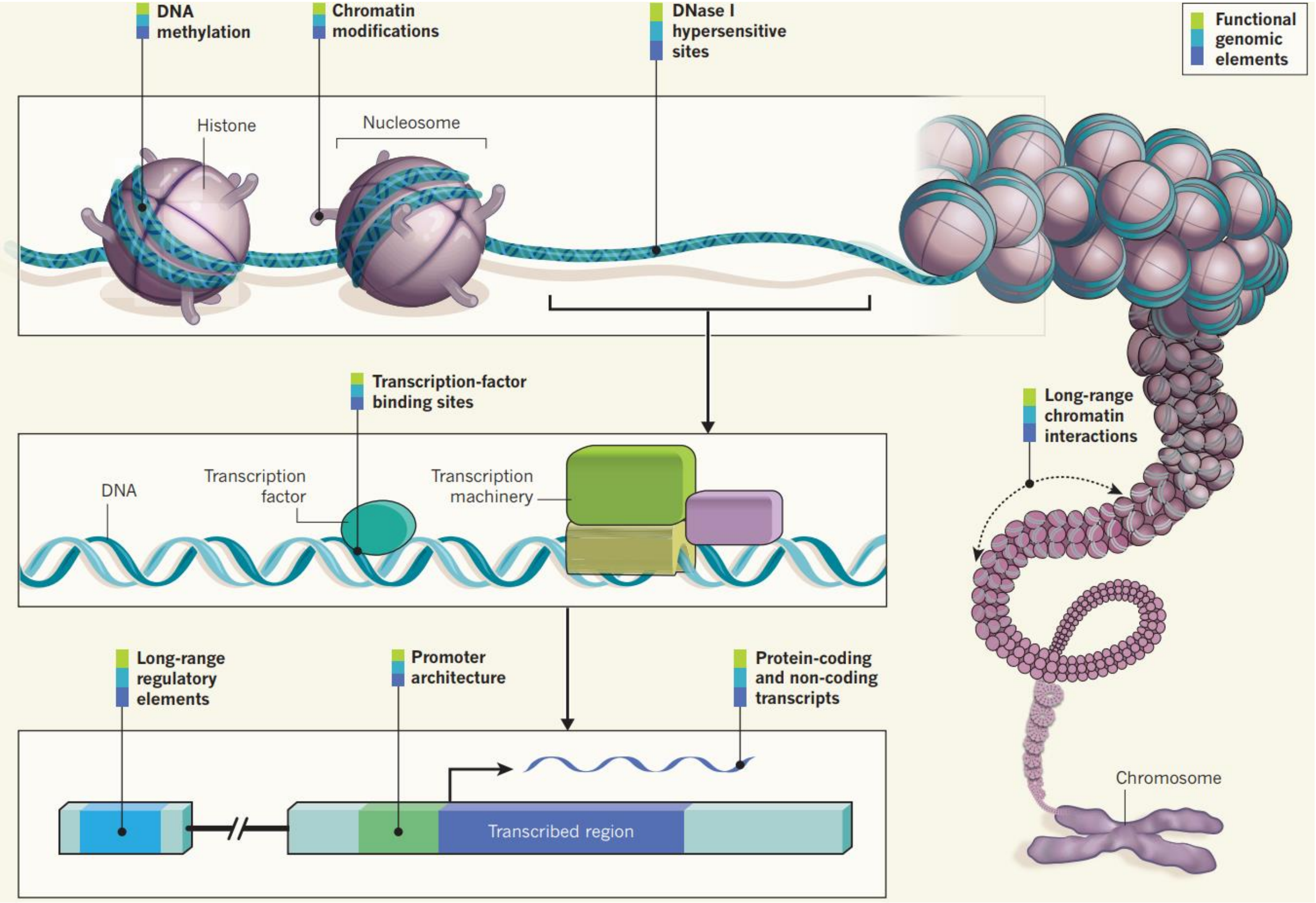
PERSONAL BACKGROUND

- Greenleaf and Kundaje labs (Stanford)

    - Charting the deep evolution of chromatin organization across the tree of life

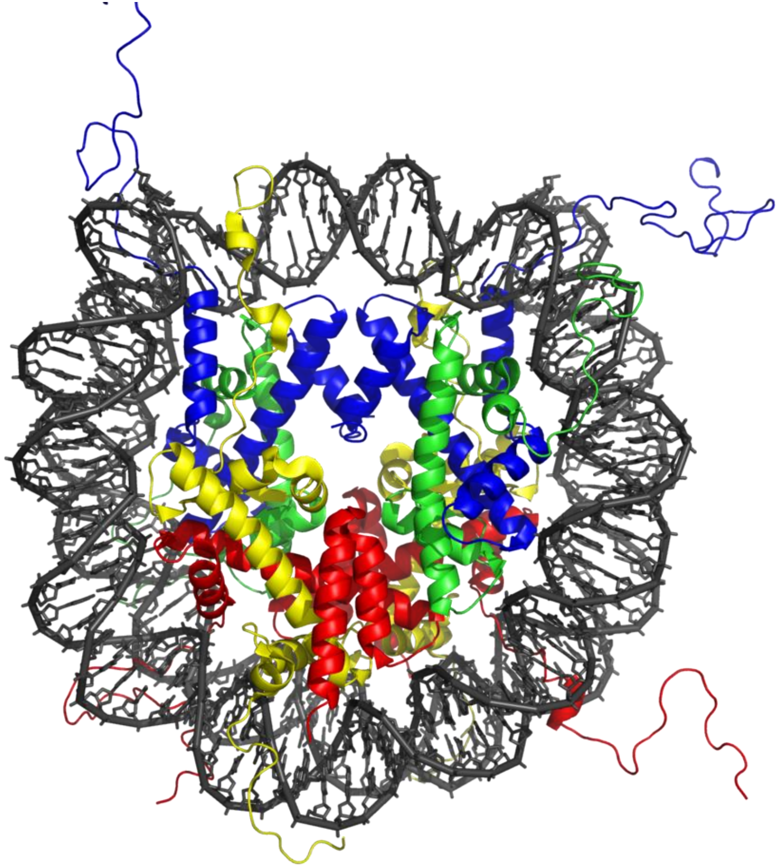
    - Development of new genomics tools for mapping the regulatory genome
        - Mapping functional elements in the genome using CRISPR tools
        - Single-cell methods for mapping chromatin, transcriptomes and perturbations
        - Single-molecule methods for mapping chromatin architecture

CONVENTIONAL VIEW OF EUKARYOTIC CHROMATIN



ENCODE Project 2012

EUKARYOTIC NUCLEOSOME

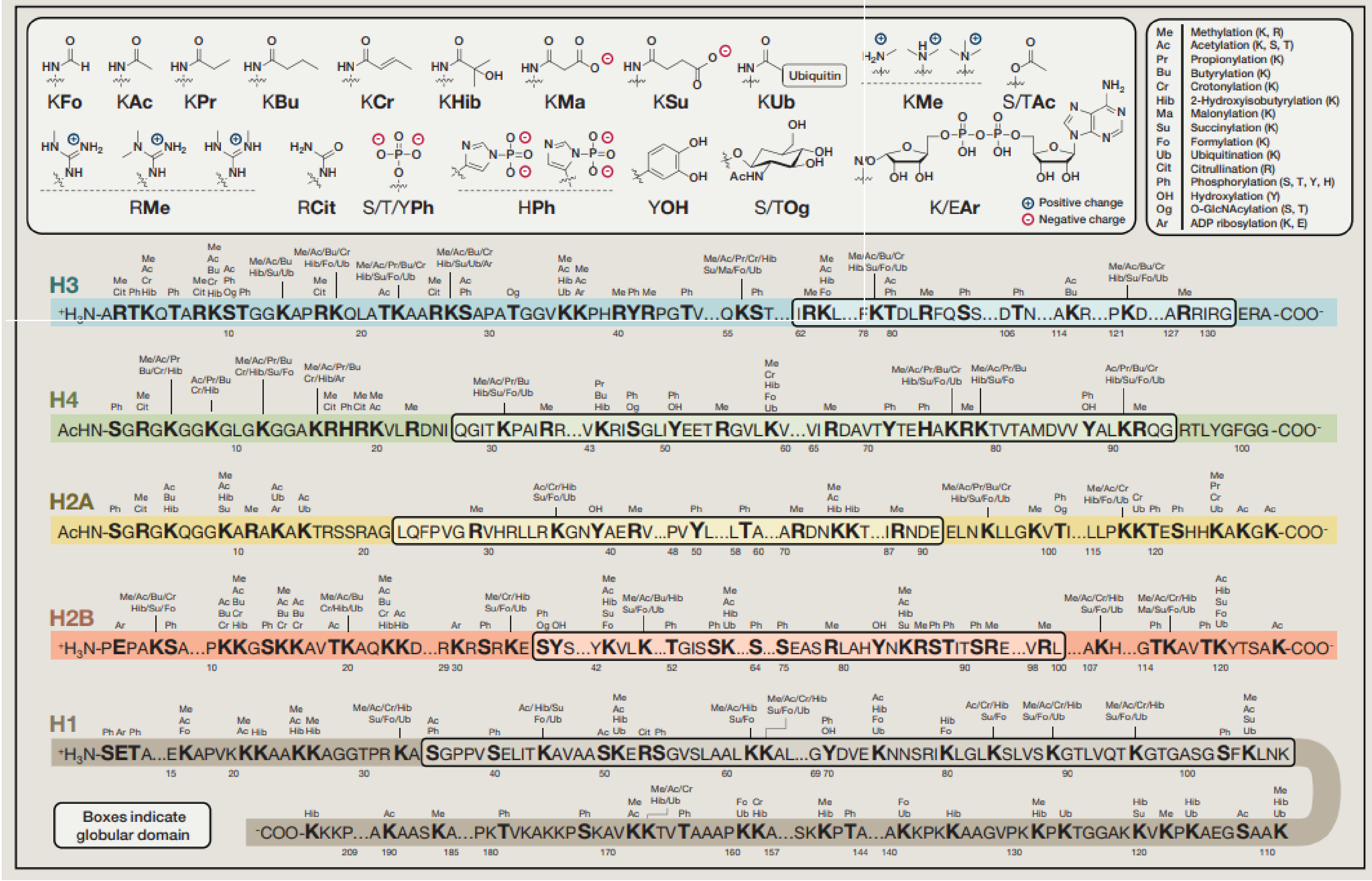


Octamer of two copies each of:

- H2A
- H2B
- H3
- H4

*there’s also linker H1

HISTONE CODE



Huang et al. 2014

HISTONE CODE

- Histone modifications are deposited by histone modifying enzymes

- Histone modifications recruit other proteins
    - chromatin remodelers
    - histone modification writers and erasers
    - other effector proteins

- Every chromatin process involves histone marks in some capacity
    - transcription and its regulation
    - chromosome compaction
    - DNA replication and chromosome segregation
    - DNA repair

- Histone tails, especially H3 and H4 are extremely conserved across nearly all eukaryotes

CONVENTIONAL VIEW OF EUKARYOTIC CHROMATIN

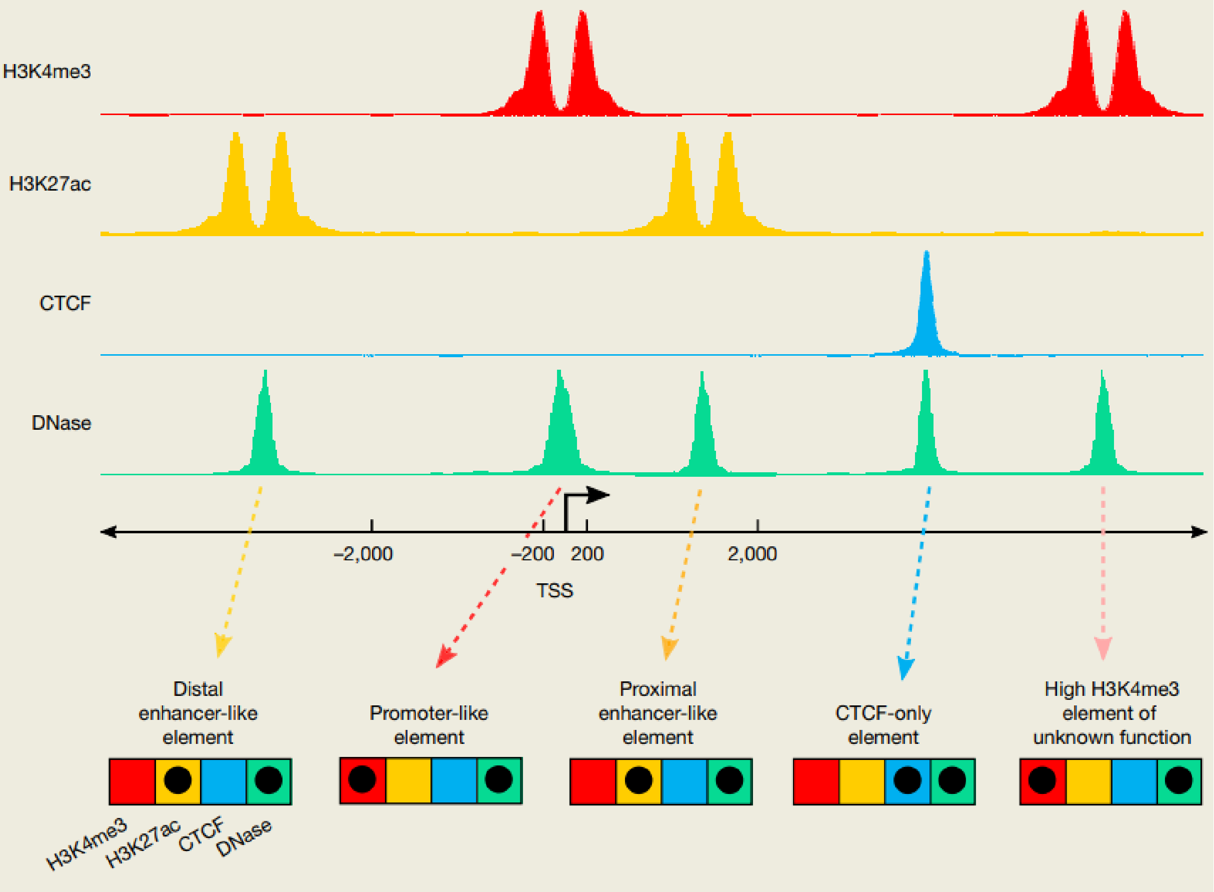
- We now have an unprecedented in its breadth and depth view of the human genome, enabled by genome-wide functional genomic assays

- Methods for mapping the functional genome:
  - ChIP-seq for mapping protein-DNA interactions
  - RNA-seq and derivatives to map the transcription landscape
  - ATAC-seq/DNase-seq for mapping open chromatin
  - Hi-C for mapping 3D genome organization
  - GRO-seq/PRO-seq/KAS-seq for mapping active transcription
  - and many others

- Large collections of datasets across cell types and conditions
  - ENCODE Consortium Project
  - Single-cell atlases across tissues, development, individuals and diseases

CONVENTIONAL VIEW OF EUKARYOTIC CHROMATIN

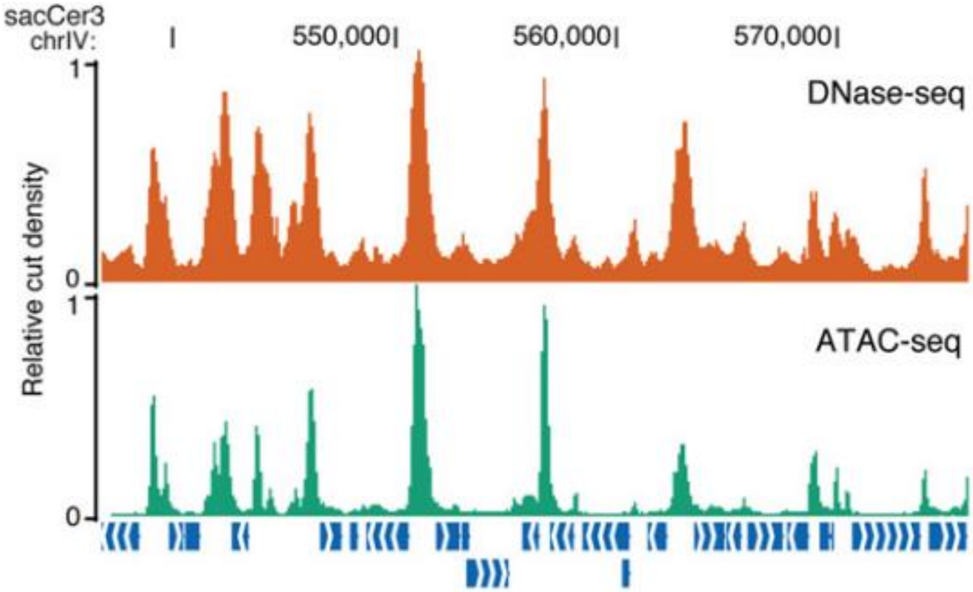
- Human genome
    - large (3Gbp)
    - sparse (mostly introns and intergenic space, lots of repeats)
    - many (>10) distal regulatory elements per gene



ENCODE Project 2020

CONVENTIONAL VIEW OF EUKARYOTIC CHROMATIN

- Yeast genome
  - small (12Mbp)
  - few and short introns
  - Little intergenic space
  - few to no distal regulatory elements



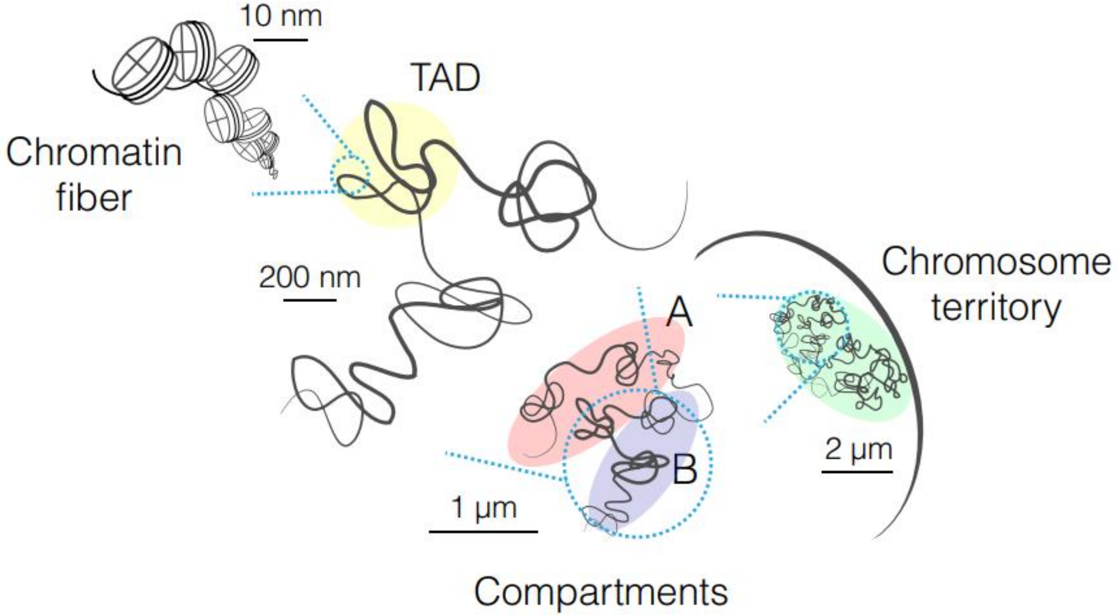
Schep et al. 2015

PRINCIPLES OF GENOME FOLDING:

The mammalian-centric view:

- TAD (topologically associated domains) formed by constraints on loop extrusion (CTCF)

- Compartments formed by interactions between similar chromatin states

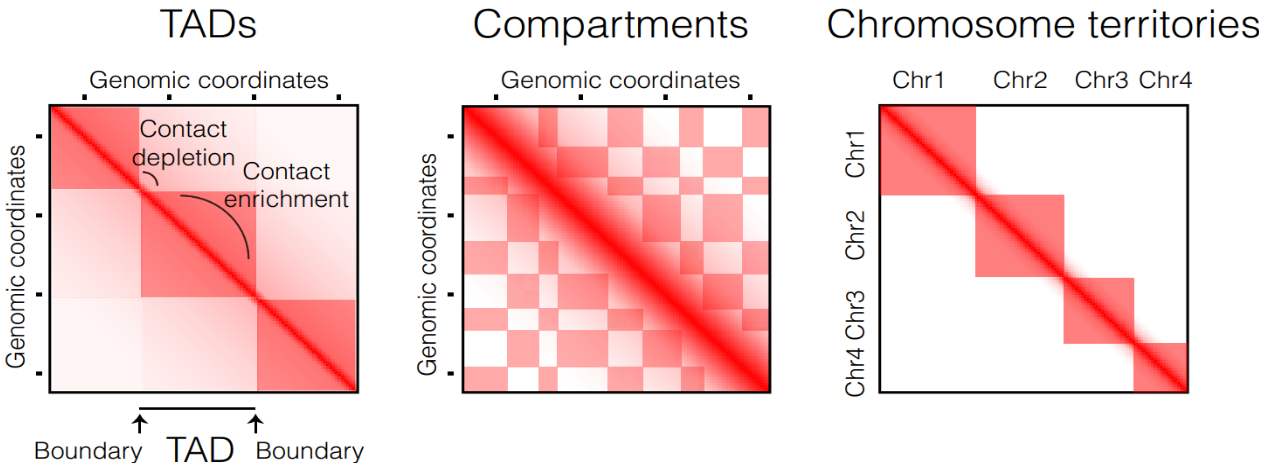


Szabo et al. 2019

PRINCIPLES OF GENOME FOLDING:

The mammalian-centric view:

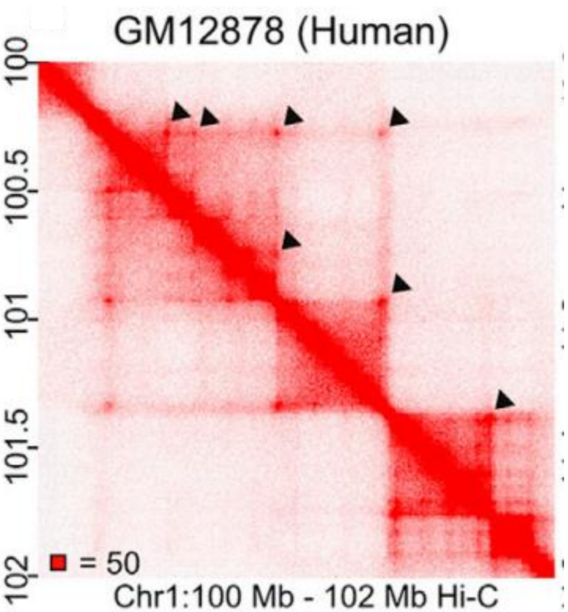
- TAD (topologically associated domains) formed by constraints on loop extrusion
- Compartments formed by interactions between similar chromatin states

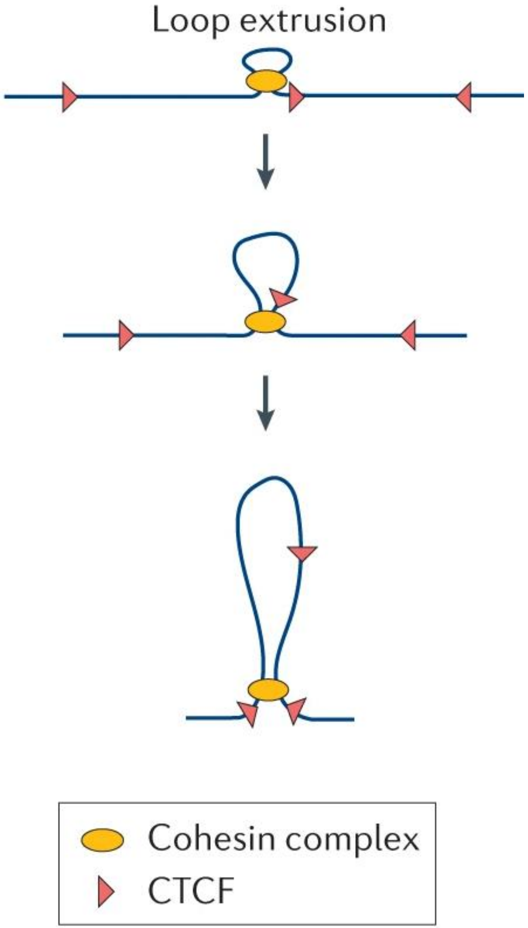


Szabo et al. 2019

LOOP EXTRUSION DOMAIN SIGNATURES

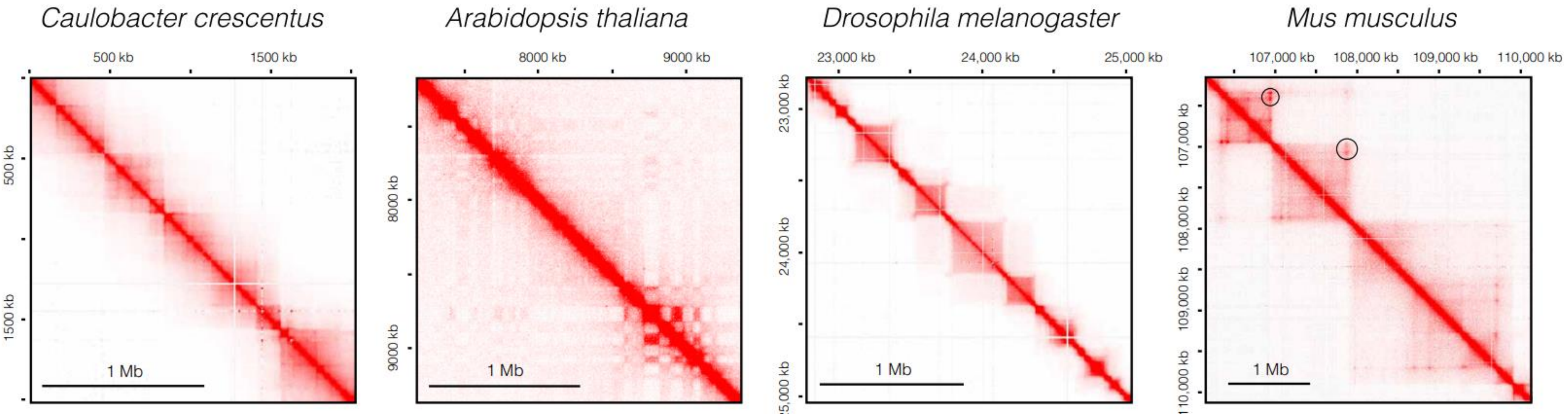
- Clear chromatin loops
- Topological stripes
- CTCF motifs in opposite orientations





Rowley et al. 2017
Davidson et al. 2021

TYPICAL HI-C MAPS


Caulobacter crescentus

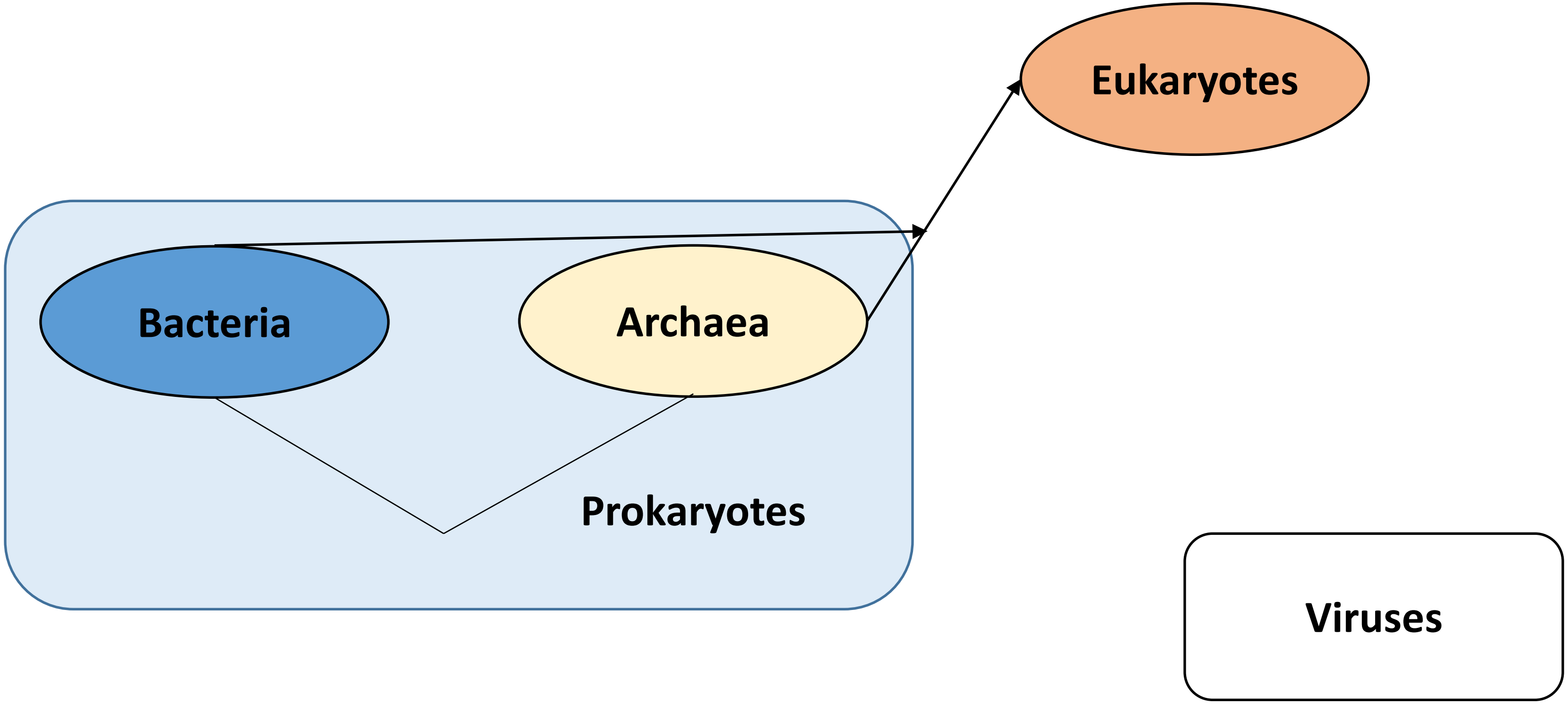
Arabidopsis thaliana

Drosophila melanogaster

Mus musculus

Szabo et al. 2019

LIFE ON EARTH



Eukaryotes

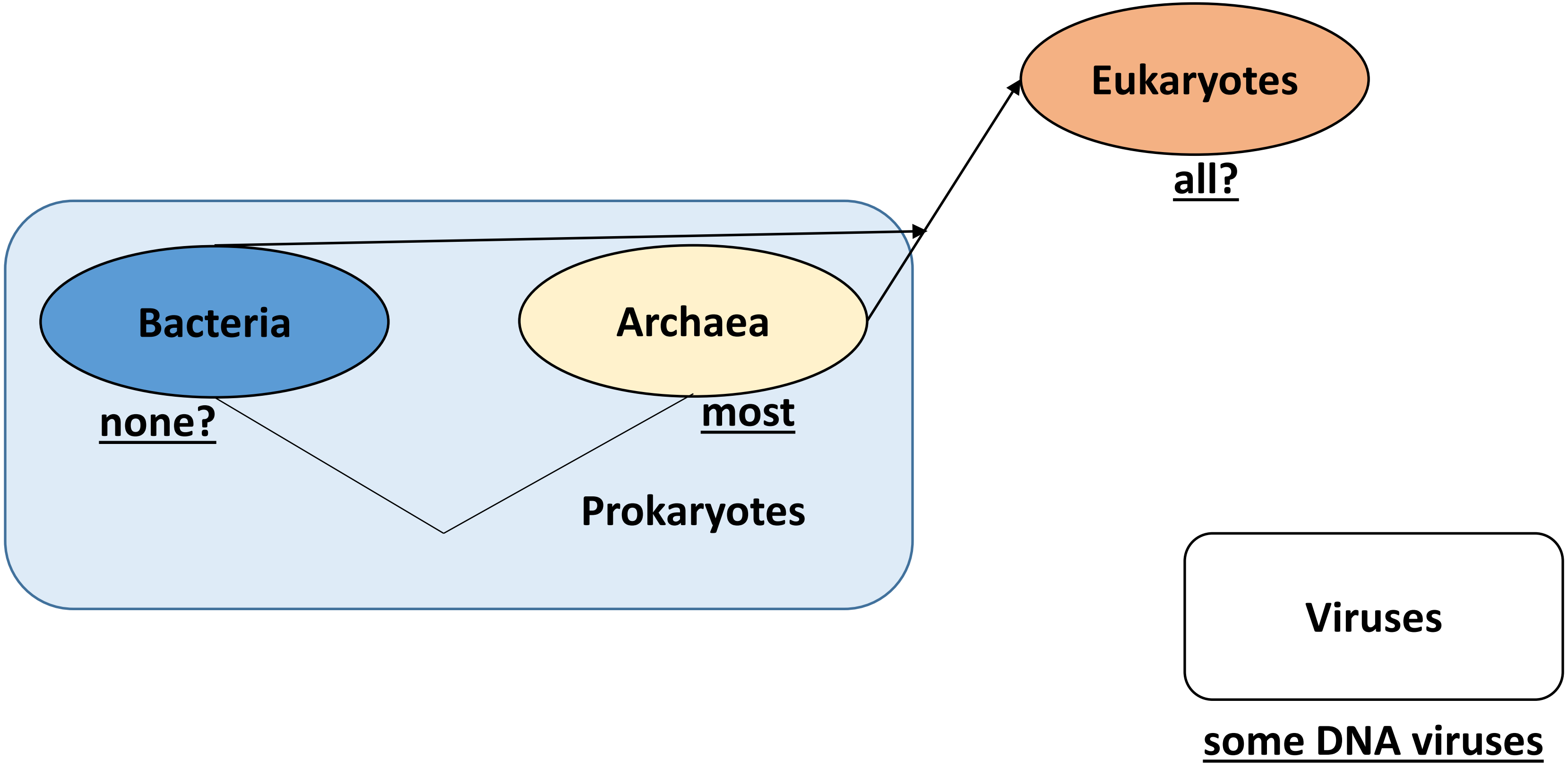
Bacteria

Archaea

Prokaryotes

Viruses

HISTONES AND CHROMATIN



Eukaryotes

all?

Bacteria

none?

Archaea

most

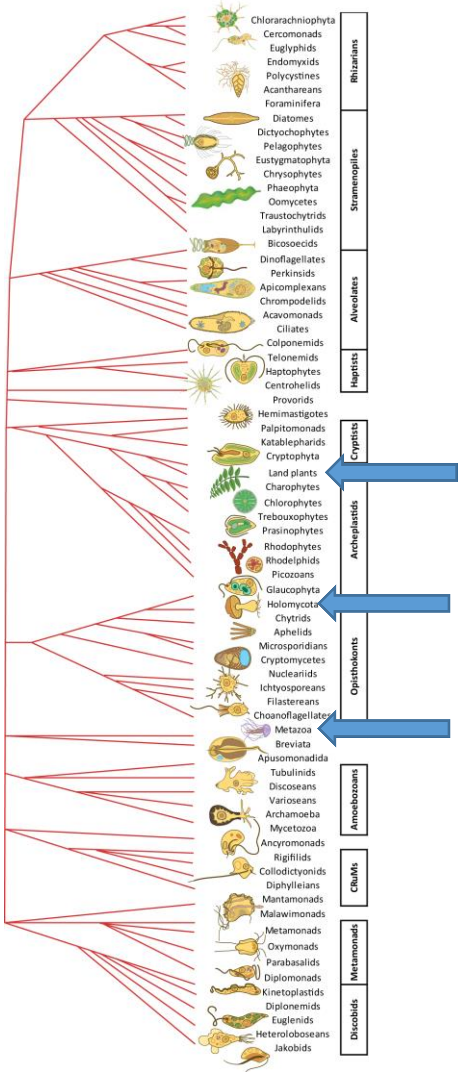
Prokaryotes

Viruses

some DNA viruses

KNOWN EUKARYOTE DIVERSITY

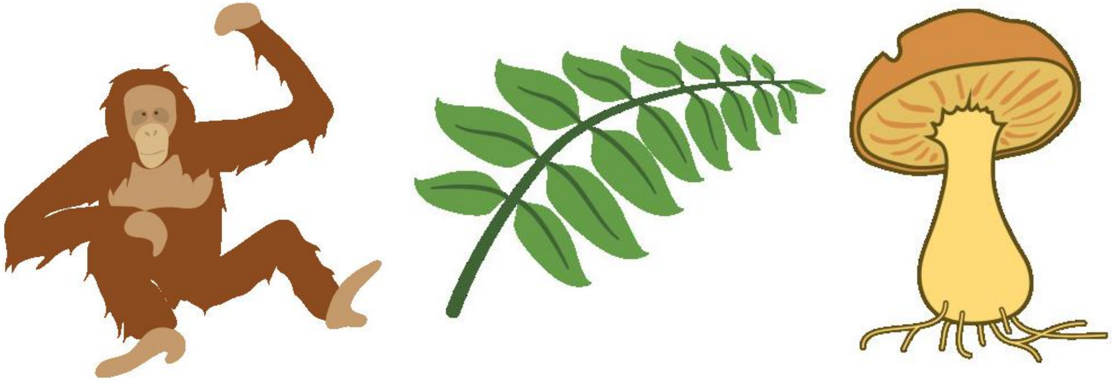
- Much deeper than the well-known macroscopic clades



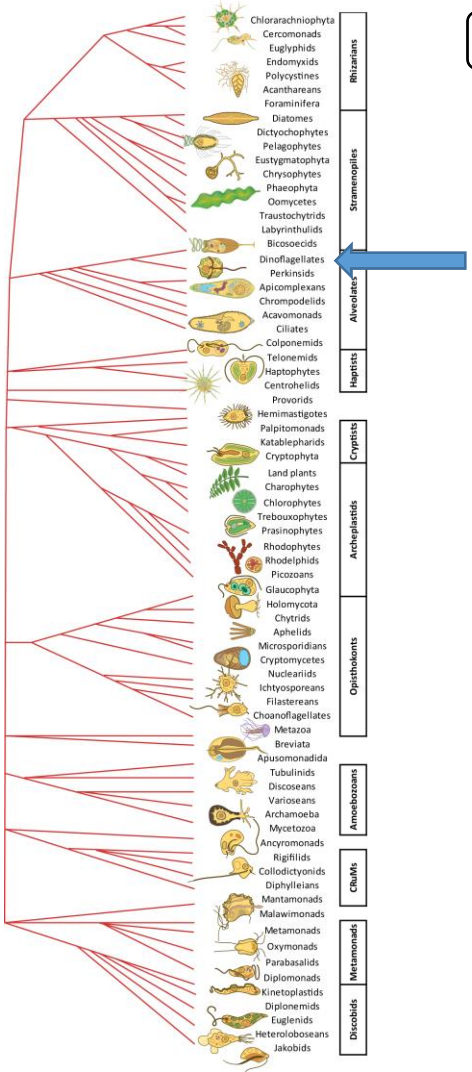
plants

fungi

animals



KNOWN EUKARYOTE DIVERSITY

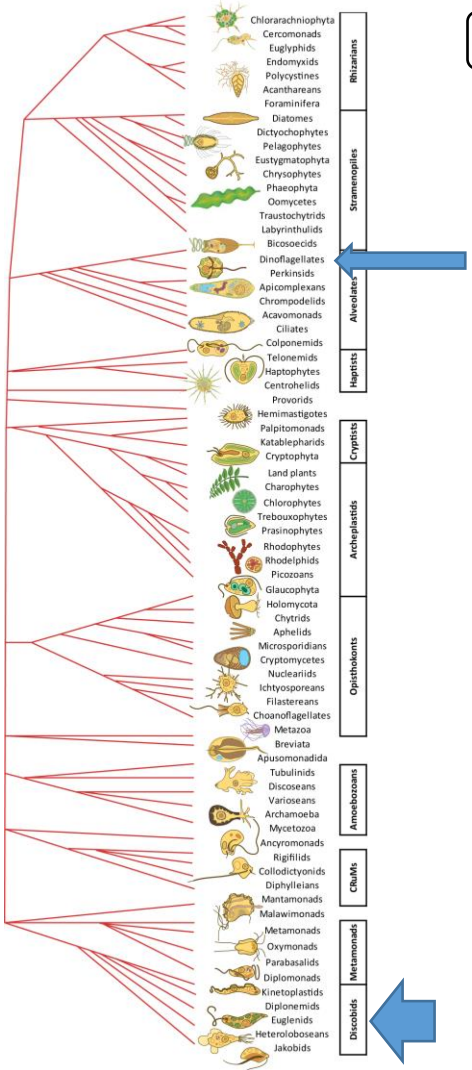


- Much deeper than the well-known macroscopic clades

- Contains numerous known exceptions to the conventional chromatin organization and regulatory logic

  - loss of histones as main packaging component
    - dinoflagellates

KNOWN EUKARYOTE DIVERSITY

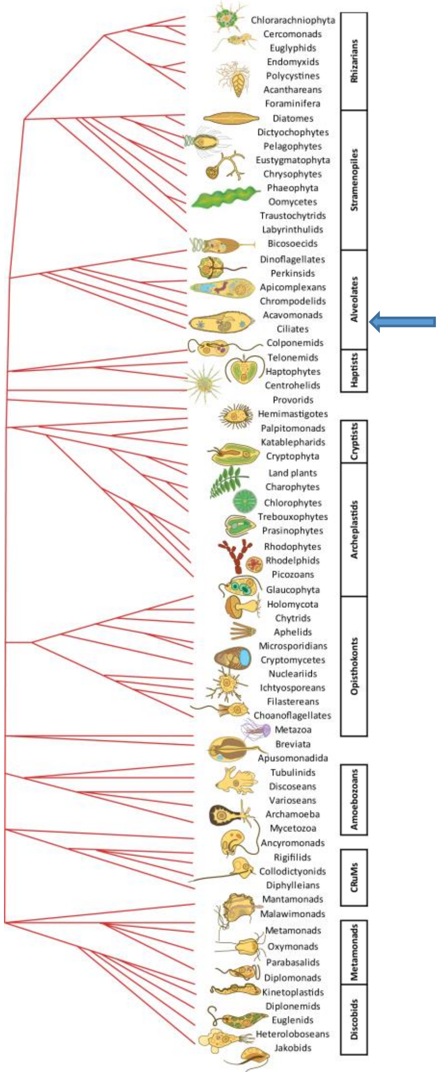


- Much deeper than the well-known macroscopic clades

- Contains numerous known exceptions to the conventional chromatin organization and regulatory logic

  - loss of histones as main packaging component
  - loss of transcriptional regulation
    - dinoflagellates, euglenozoans

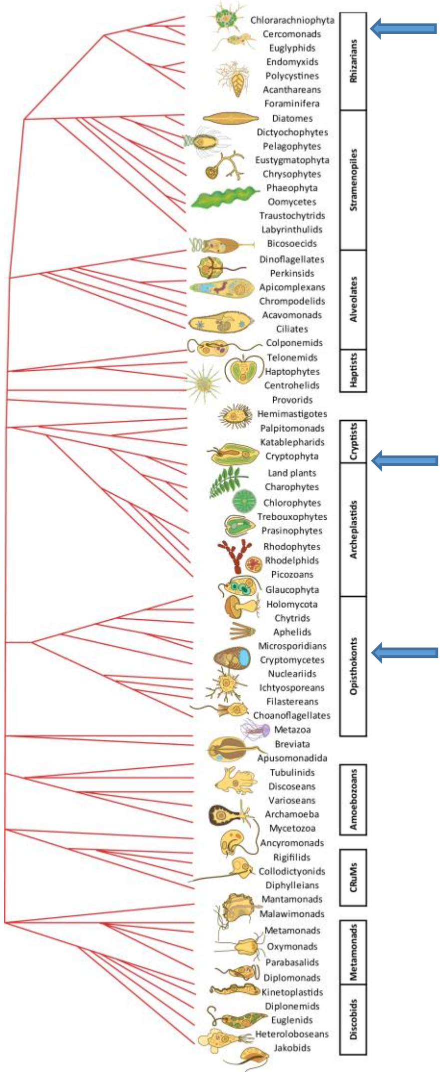
KNOWN EUKARYOTE DIVERSITY



- Much deeper than the well-known macroscopic clades

- Contains numerous <u>known</u> exceptions to the conventional chromatin organization and regulatory logic

  - loss of histones as main packaging component
  - loss of transcriptional regulation
  - <u>drastic fragmentation of the genome in nanochromosomes</u>
    - ciliates

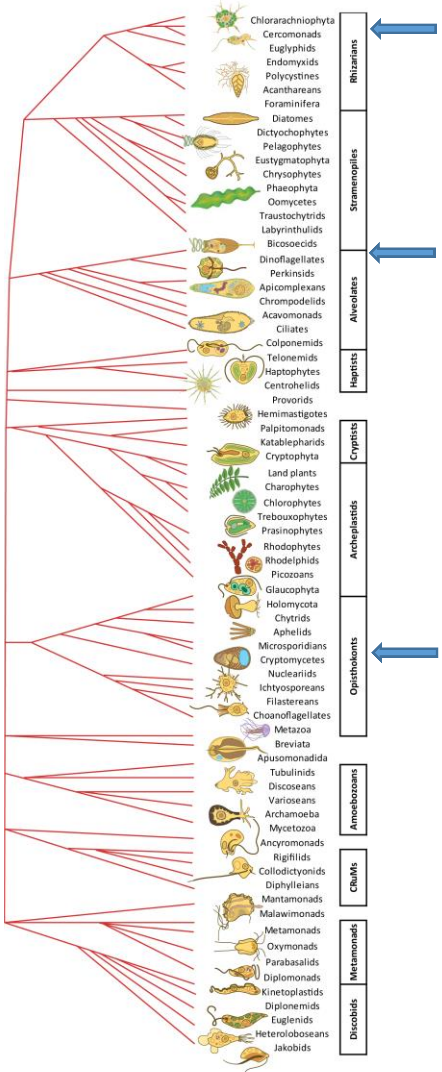


KNOWN EUKARYOTE DIVERSITY

- Much deeper than the well-known macroscopic clades

- Contains numerous <u>known</u> exceptions to the conventional chromatin organization and regulatory logic

  - loss of histones as main packaging component
  - loss of transcriptional regulation
  - drastic fragmentation of the genome in nanochromosomes
  - <u>drastic reduction of the genome</u>
    - nucleomorphs (in cryptophytes and chlorarachnophytes)
    - microsporidians



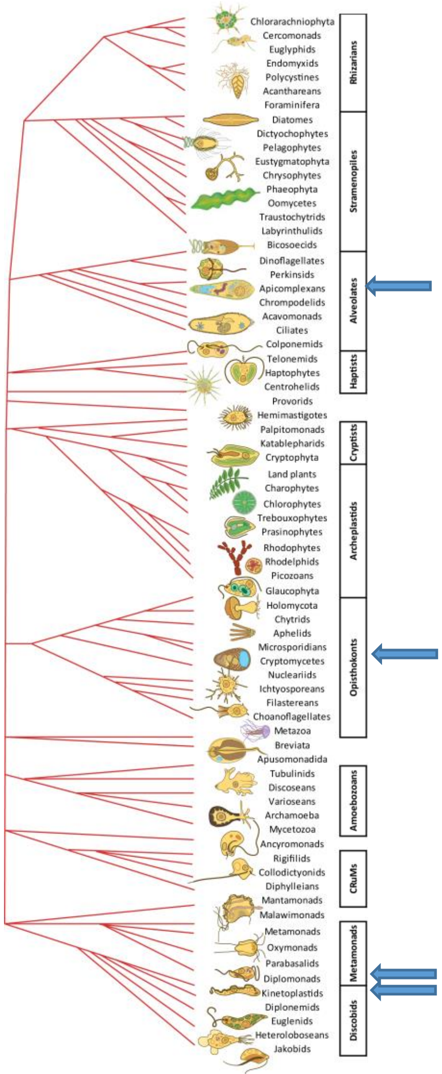
KNOWN EUKARYOTE DIVERSITY

- Much deeper than the well-known macroscopic clades

- Contains numerous known exceptions to the conventional chromatin organization and regulatory logic

  - loss of histones as main packaging component
  - loss of transcriptional regulation
  - drastic fragmentation of the genome in nanochromosomes
  - drastic reduction of the genome
  - extreme divergence of histone proteins
    - dinoflagellates
    - nucleomorphs

KNOWN EUKARYOTE DIVERSITY



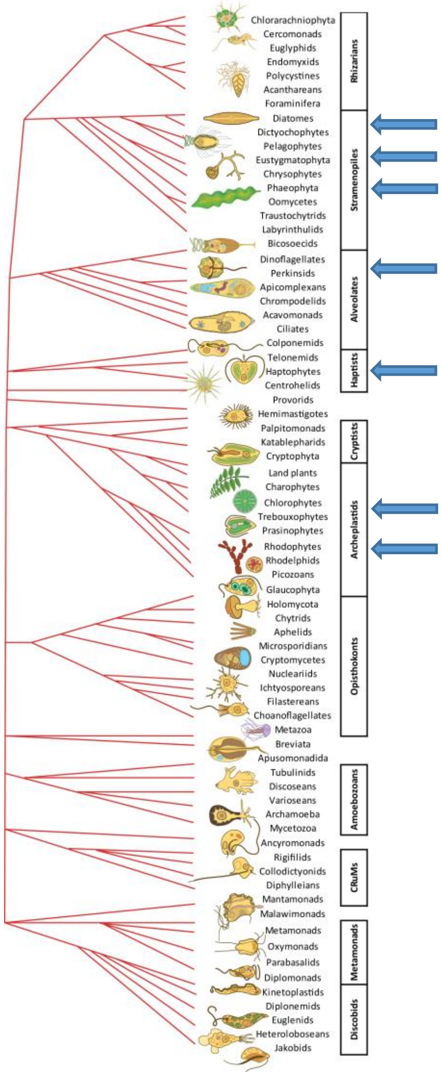
- Much deeper than the well-known macroscopic clades

- Contains numerous <u>known</u> exceptions to the conventional chromatin organization and regulatory logic

  - loss of histones as main packaging component
  - loss of transcriptional regulation
  - drastic fragmentation of the genome in nanochromosomes
  - drastic reduction of the genome
  - extreme divergence of histone proteins

- Importance of protozoans
  - <u>human pathogens</u>
    - apicomplexans (malaria, toxoplasma)
    - kinetoplastids (*Trypanosoma*, *Leishmania*)
    - parabasalids (*Trichomonas*)
    - diplomonads (*Giardia*)
    - heteroloboseans (*Naegleria*)
    - microsporidians

KNOWN EUKARYOTE DIVERSITY



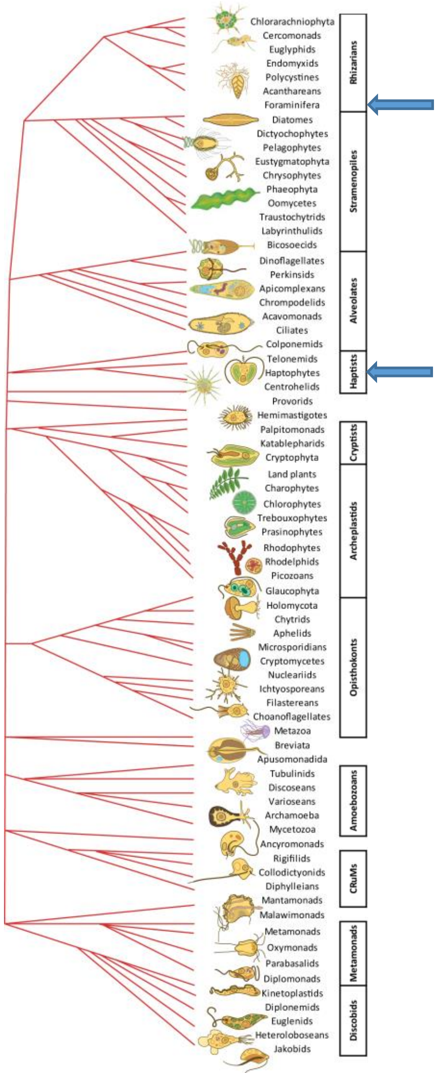
- Much deeper than the well-known macroscopic clades

- Contains numerous <u>known</u> exceptions to the conventional chromatin organization and regulatory logic

    - loss of histones as main packaging component
    - loss of transcriptional regulation
    - drastic fragmentation of the genome in nanochromosomes
    - drastic reduction of the genome
    - extreme divergence of histone proteins

- Importance of protozoans
    - human pathogens
    - <u>key primary producers</u>
        - diatoms
        - haptophytes
        - green algae
        - red algae
        - brown algae
        - dinoflagellates
        - eustygmatophytes

KNOWN EUKARYOTE DIVERSITY

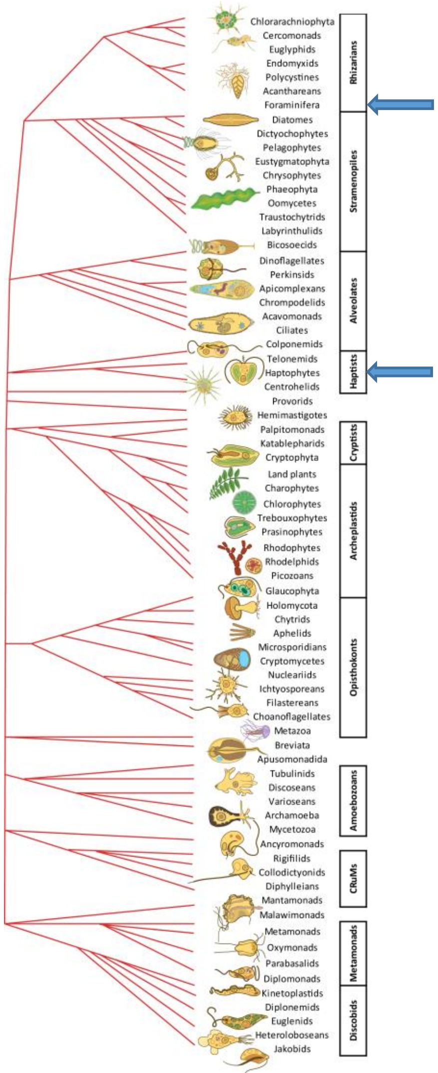


- Much deeper than the well-known macroscopic clades

- Contains numerous known exceptions to the conventional chromatin organization and regulatory logic

  - loss of histones as main packaging component
  - loss of transcriptional regulation
  - drastic fragmentation of the genome in nanochromosomes
  - drastic reduction of the genome
  - extreme divergence of histone proteins

- Importance of protozoans
  - human pathogens
  - key primary producers
  - key role in the carbon cycle
    - haptophytes
    - foraminiferans



KNOWN EUKARYOTE DIVERSITY

- Much deeper than the well-known macroscopic clades

- Contains numerous <u>known</u> exceptions to the conventional chromatin organization and regulatory logic

  - loss of histones as main packaging component
  - loss of transcriptional regulation
  - drastic fragmentation of the genome in nanochromosomes
  - drastic reduction of the genome
  - extreme divergence of histone proteins

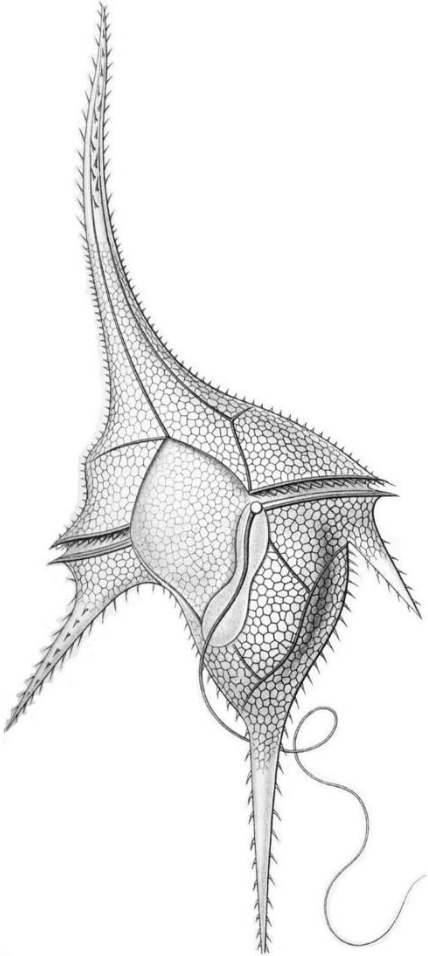
- Importance of protozoans
  - human pathogens
  - key primary producers
  - key role in the carbon cycle

- <u>Little to nothing is known</u> about the biology of most of the deeply branching lineages

II. DINOFLAGELLATE CHROMATIN

DINOFLAGELLATES

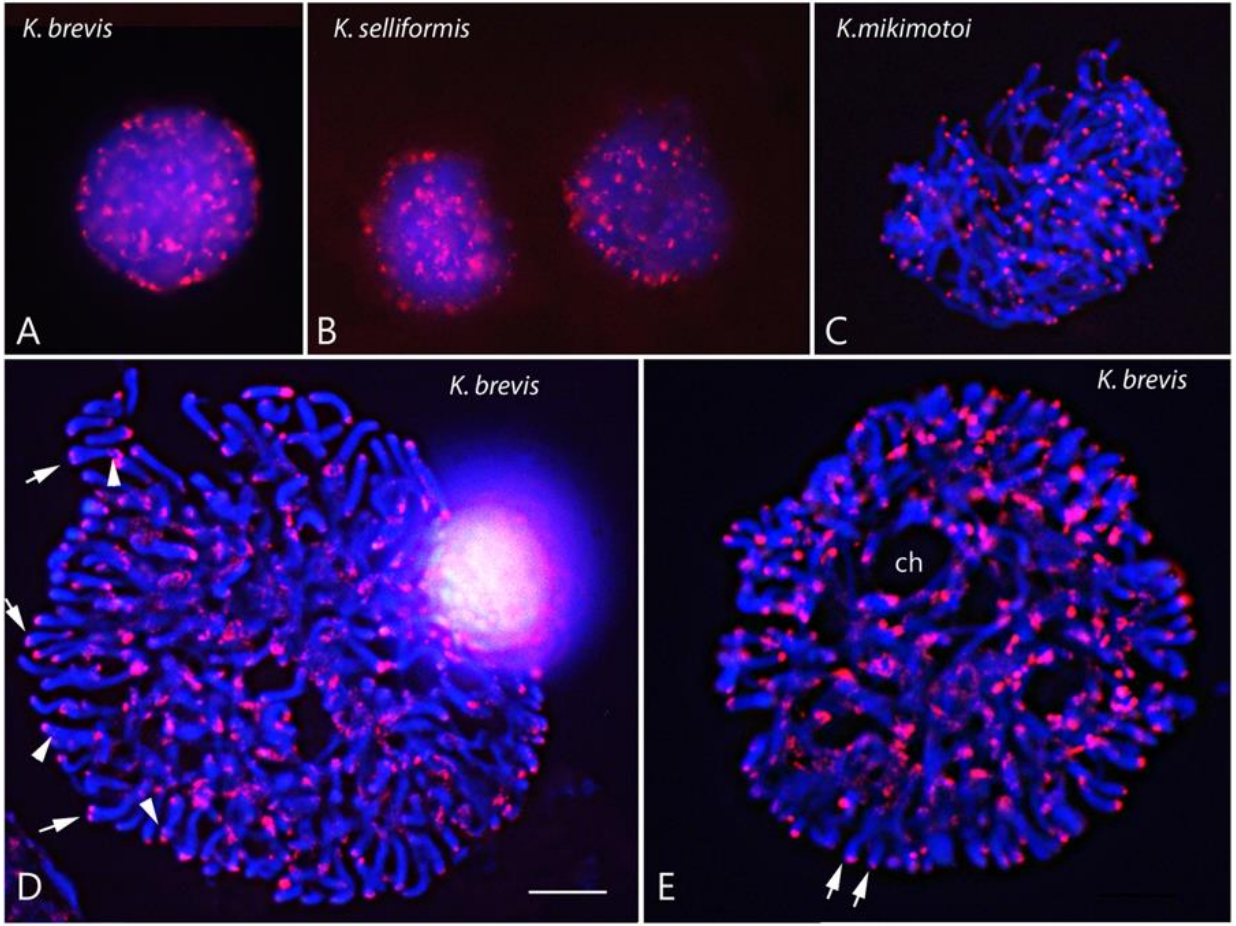
- highly diverse group
- photosynthetic, predatory, and parasitic
- mostly unicellular plankton
- numerous secondary and tertiary symbiotic events
- photosynthetic dinoflagellates are the trophic foundation of coral reefs
- numerous extreme divergences from the conventional eukaryote state
- once proposed to be intermediate between prokaryotes and eukaryotes
    - mesokaryote (Dodge, 1965)



SPECIAL FEATURES OF DINOFLAGELLATES (PARTIAL LIST)

- Permanently condensed fibrilar chromosomes
- Very low protein-to-DNA ratio (~1/10th of the usual)
- Histones in low abundance, long thought to be completely absent.
  - replaced by other proteins of viral and bacterial origin (DVNPs and HLPs)
- High percentage of 5-hydroxymethyluracile (dhmU)
- Huge genomes
- Genes are organized into multigene/polycistronic arrays
- Widespread trans-splicing
- Transcriptional regulation mostly absent
  - few transcription factors;
  - gene regulation happens at the posttranscriptional level

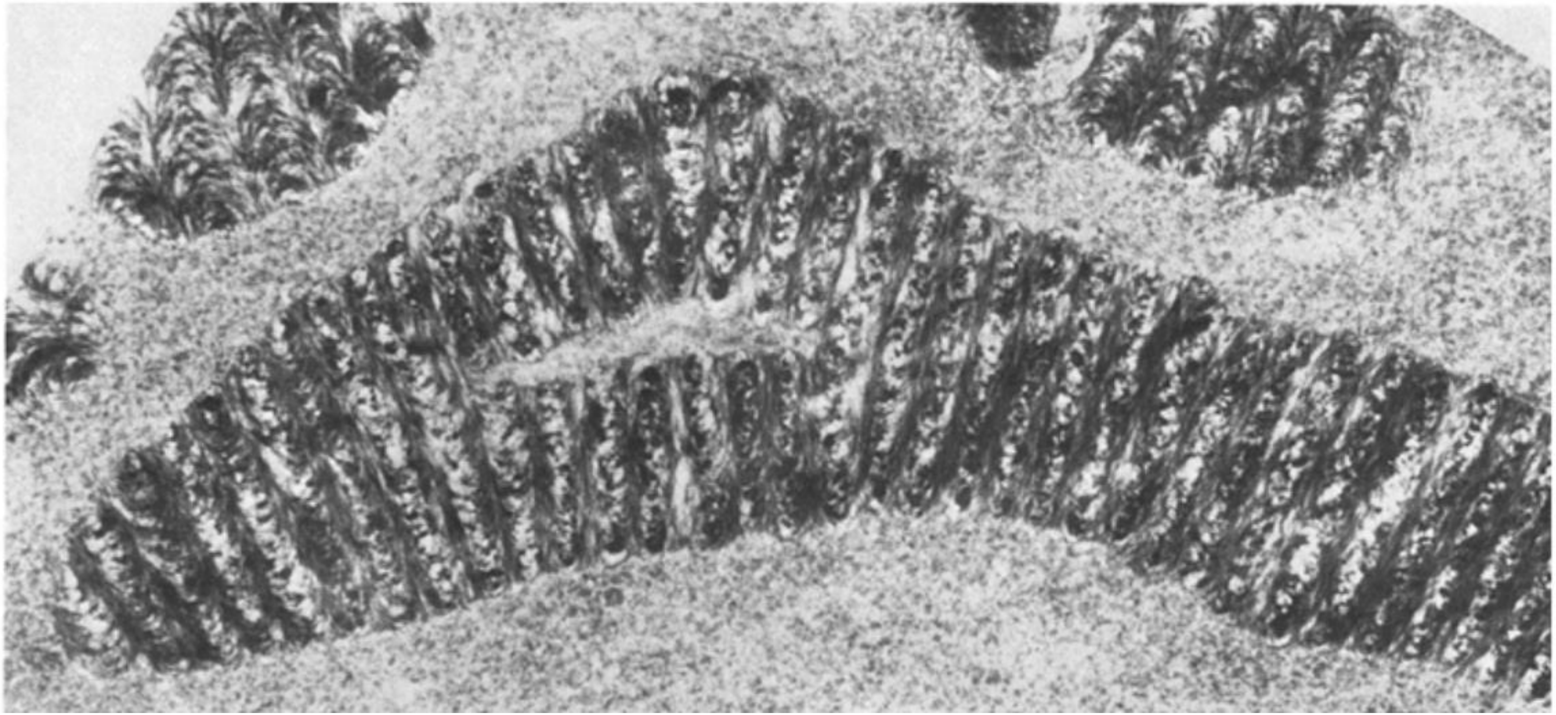
DINOFLAGELLATE CHROMOSOMES



these are interphase chromosomes!

Cuadrado et al. 2019

DINOFLAGELLATE CHROMOSOMES



Livoland & Bouligand, 1978

LONG STANDING QUESTIONS:

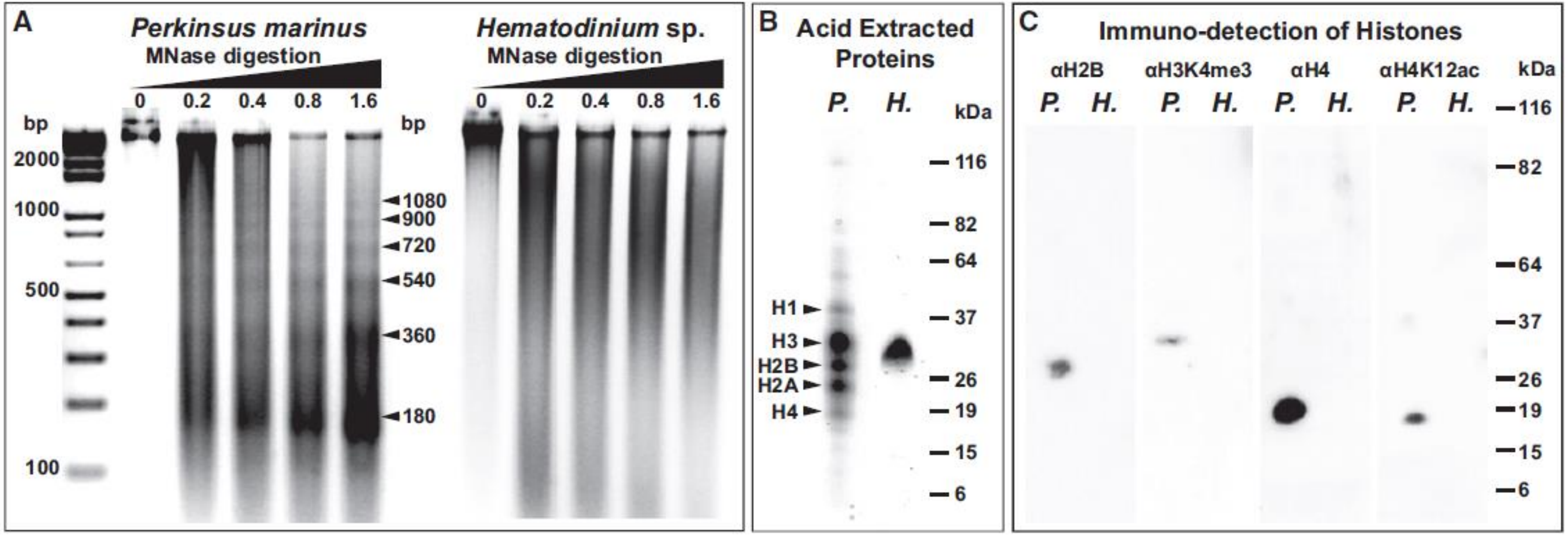
- How is the genome organized in 3D space in permanently condensed chromosomes mostly without histones?

- How is transcription and gene regulation accomplished in such an environment?
  - one decades-old proposal featured regulated looping of genes out of the permanently condensed chromosome for the purpose of transcription

- What is the role of the novel set of chromatin proteins?
  - might there be an analog to the histone code that has evolved?

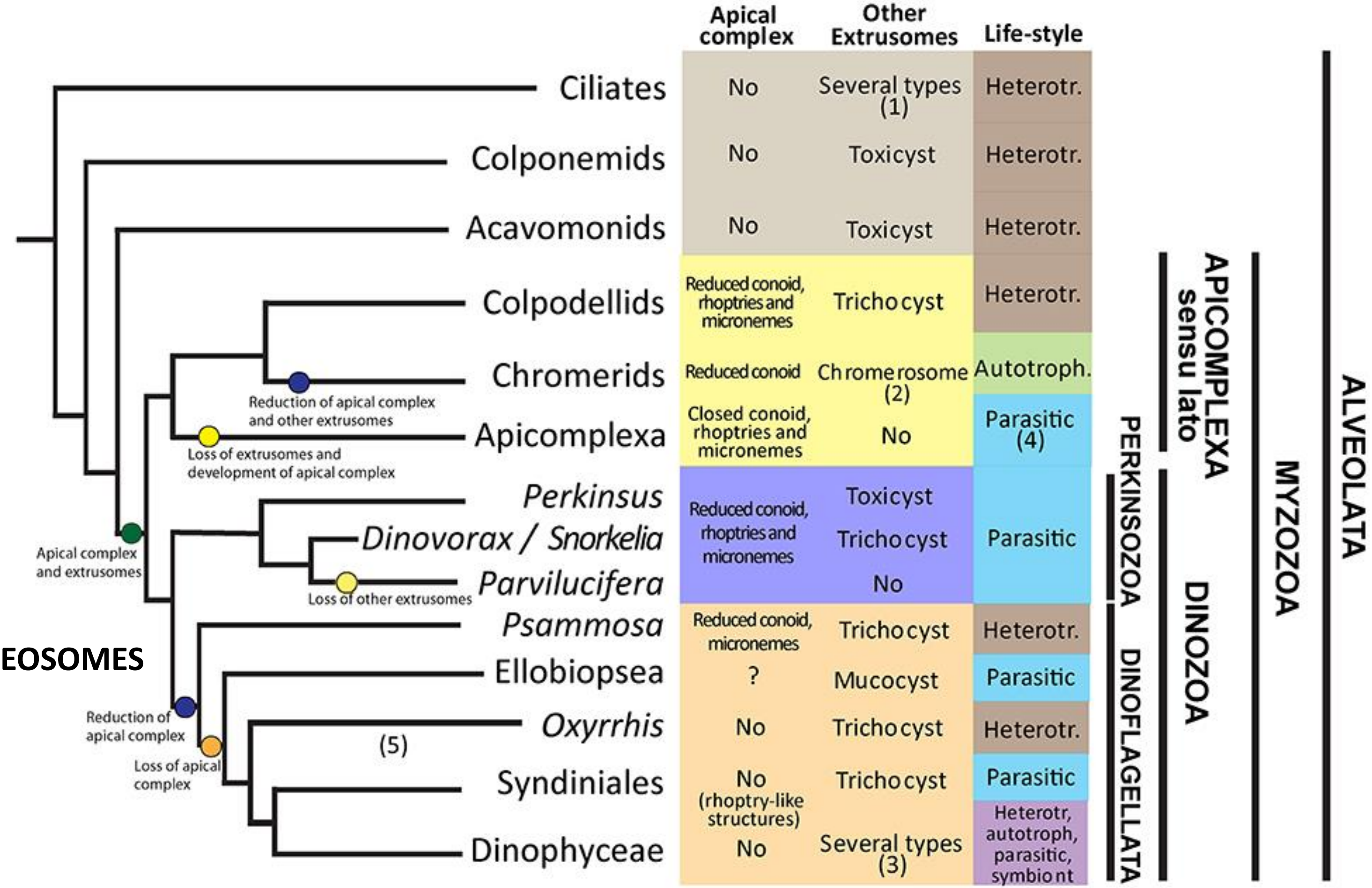
- What is the role of dhmU?

- DVNPs (Dinoflagellate Viral Nucleoproteins, unrelated to histones) and
  HLPs (Histone-Like Proteins of bacterial origin)
  are thought to be the main packaging components



Gornik et al. 2012

LOSS OF NUCLEOSOMES



	Apical complex	Other Extrusomes	Life-style
Ciliates	No	Several types (1)	Heterotr.
Colponemids	No	Toxicyst	Heterotr.
Acavomonids	No	Toxicyst	Heterotr.
Colpodellids	Reduced conoid, rhoptries and micronemes	Trichocyst	Heterotr.
Chromerids	Reduced conoid	Chromerosome (2)	Autotroph.
Apicomplexa	Closed conoid, rhoptries and micronemes	No	Parasitic (4)
Perkinsus	Reduced conoid, rhoptries and micronemes	Toxicyst	Parasitic
Dinovorax / Snorkelia		Trichocyst	
Parvilucifera		No	
Psammosa	Reduced conoid, micronemes	Trichocyst	Heterotr.
Ellobiopsea	?	Mucocyst	Parasitic
Oxyrrhis	No	Trichocyst	Heterotr.
Syndiniales	No (rhoptry-like structures)	Trichocyst	Parasitic
Dinophyceae	No	Several types (3)	Heterotr, autotroph, parasitic, symbiont

Janouskovec et al. 2017

BUT HISTONES ARE IN FACT PRESENT

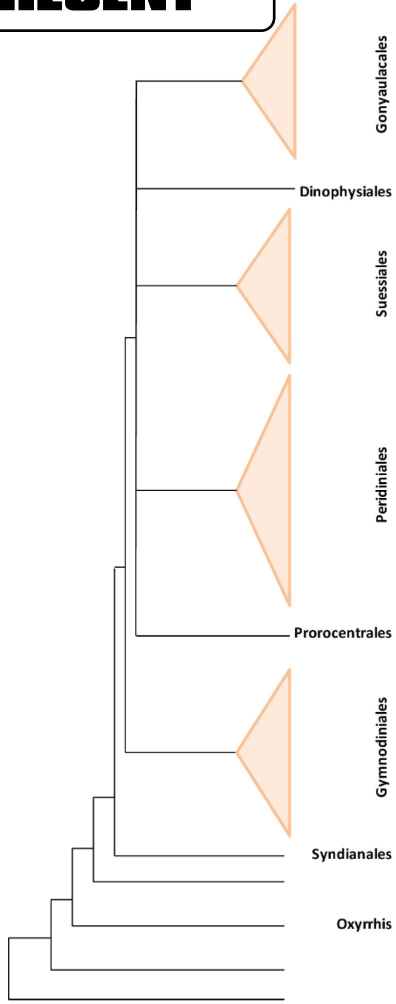
Diversity and Divergence of Dinoflagellate Histone Proteins

Georgi K. Marinov[1] and Michael Lynch
Department of Biology, Indiana University, Bloomington, Indiana 47405
ORCID ID: 0000-0003-1822-7273 (G.K.M.)

- Phylogenomic analysis of the presence and properties of histones in available dinoflagellate transcriptomes

- Also charted the evolution of known chromatin modifiers and remodellers

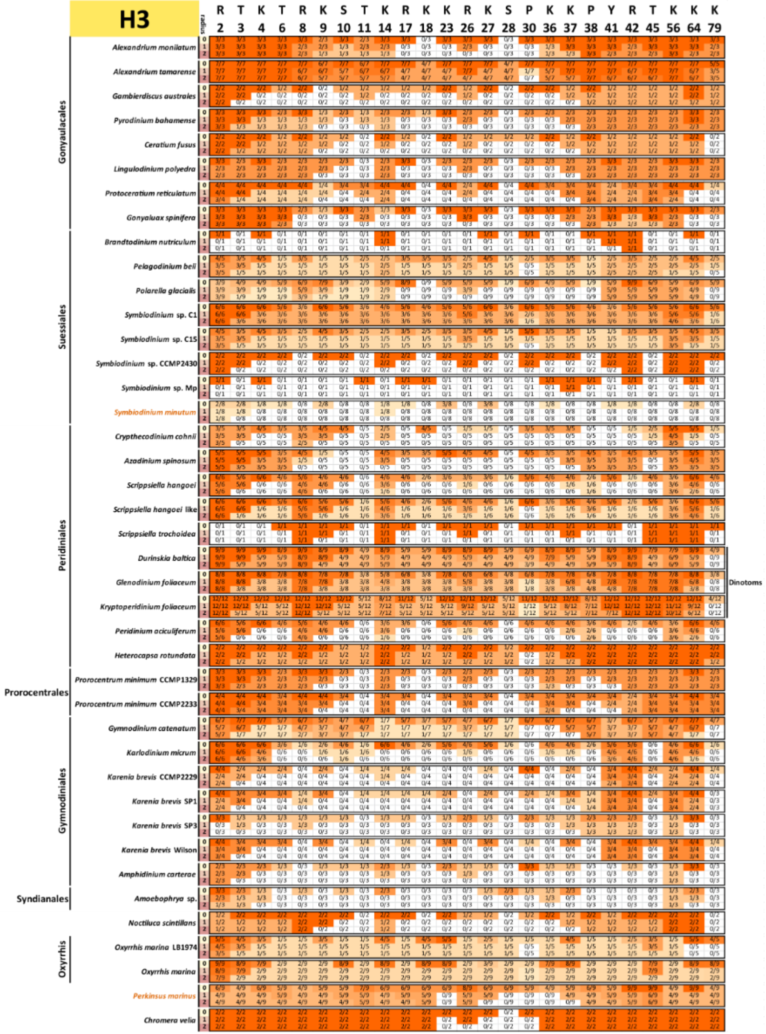
HISTONES ARE PRESENT



	Total						Complete						
	DVNP	H2A	H2B	H3	H4	H1	DVNP	H2A	H2B	H3	H4	H1	
Alexandrium monilatum	38	3	4	5	1	10	12	1	1	-	1	3	
Alexandrium tamarense	16	15	7	9	2	-	7	7	5	3	-	-	
Gambierdiscus australes	17	2	2	5	1	-	4	-	1	-	-	-	
Pyrodinium bahamense	23	4	1	6	3	-	4	-	-	1	1	-	
Ceratium fusus	15	3	1	4	1	2	3	-	-	2	-	-	
Lingulodinium polyedra	23	3	3	5	2	-	2	-	-	-	-	-	
Protoceratium reticulatum	25	5	1	5	1	4	3	-	-	1	-	1	
Gonyaulax spinifera	14	3	2	6	2	-	2	-	-	1	1	-	
Dinophysis acuminata	11	5	3	3	1	2	5	-	-	-	-	-	
Pelagodinium beii	14	4	3	6	2	4	4	2	3	2	1	1	
Polarella glacialis	15	2	3	11	4	5	1	2	1	8	2	-	
Symbiodinium sp C1	13	7	6	10	4	2	4	5	2	4	2	1	
Symbiodinium sp C15	8	5	3	7	4	1	6	3	1	5	3	1	
Symbiodinium sp CCMP2430	10	2	2	6	3	-	4	-	-	1	1	-	
Symbiodinium sp Mp	10	2	2	5	2	-	2	1	2	1	1	-	
Symbiodinium minutum	24	2	2	8	2	19	24	2	2	8	2	19	
Brandtodinium nutriculum	19	7	4	4	1	-	1	-	-	-	-	-	
Crypthecodinium cohnii	17	3	4	7	1	2	11	3	2	5	1	1	
Azadinium spinosum	29	3	1	9	1	-	1	1	1	2	1	-	
Scrippsiella hangoei	23	4	3	7	3	-	8	-	1	1	-	-	
Scrippsiella hangoei -like	23	4	3	9	3	-	7	-	1	1	-	-	
Scrippsiella trochoidea	18	3	5	7	3	1	5	1	1	1	1	-	
Durinskia baltica	19	11	10	12	7	4	6	3	4	5	2	3	Dinotoms
Glenodinium foliaceum	21	10	6	11	5	3	2	3	-	3	1	3	
Kryptoperidinium foliaceum	36	18	8	22	9	5	4	5	3	4	2	3	
Peridinium aciculiferum	17	4	4	7	2	2	5	-	1	1	-	-	
Heterocapsa rotundata	14	3	6	4	2	1	2	-	1	1	-	-	
Prorocentrum minimum CCMP1329	24	3	3	5	3	1	6	-	-	1	2	-	
Prorocentrum minimum CCMP2233	15	3	2	6	4	-	2	-	-	2	1	-	
Togula jolla	9	2	3	5	2	-	1	-	1	-	-	-	
Gymnodinium catenatum	9	8	1	10	6	2	-	2	-	2	2	-	
Karlodinium micrum	26	4	4	6	3	3	9	1	2	1	3	-	
Karenia brevis CCMP2229	13	5	3	6	3	5	5	3	-	4	2	-	
Karenia brevis SP1	29	6	3	8	2	9	13	3	-	2	2	-	
Karenia brevis SP3	20	6	3	6	3	5	12	3	3	2	2	1	
Karenia brevis Wilson	29	6	3	8	2	6	18	3	2	4	2	-	
Amphidinium carterae	7	2	2	6	2	1	4	1	1	3	1	1	
Amoebophrya sp.	3	4	2	4	5	-	2	3	1	3	3	-	
Noctiluca scintillans	11	6	4	8	3	1	-	2	3	2	2	1	
Oxyrrhis marina LB1974	13	4	3	6	1	2	5	1	-	3	1	-	
Oxyrrhis marina	12	8	3	11	3	4	2	1	1	3	2	-	
Perkinsus marinus	-	20	10	9	10	17	-	20	10	9	10	17	
Chromera velia	-	3	4	2	2	1	-	1	1	2	1	1	

…BUT HIGHLY DIVERGENT

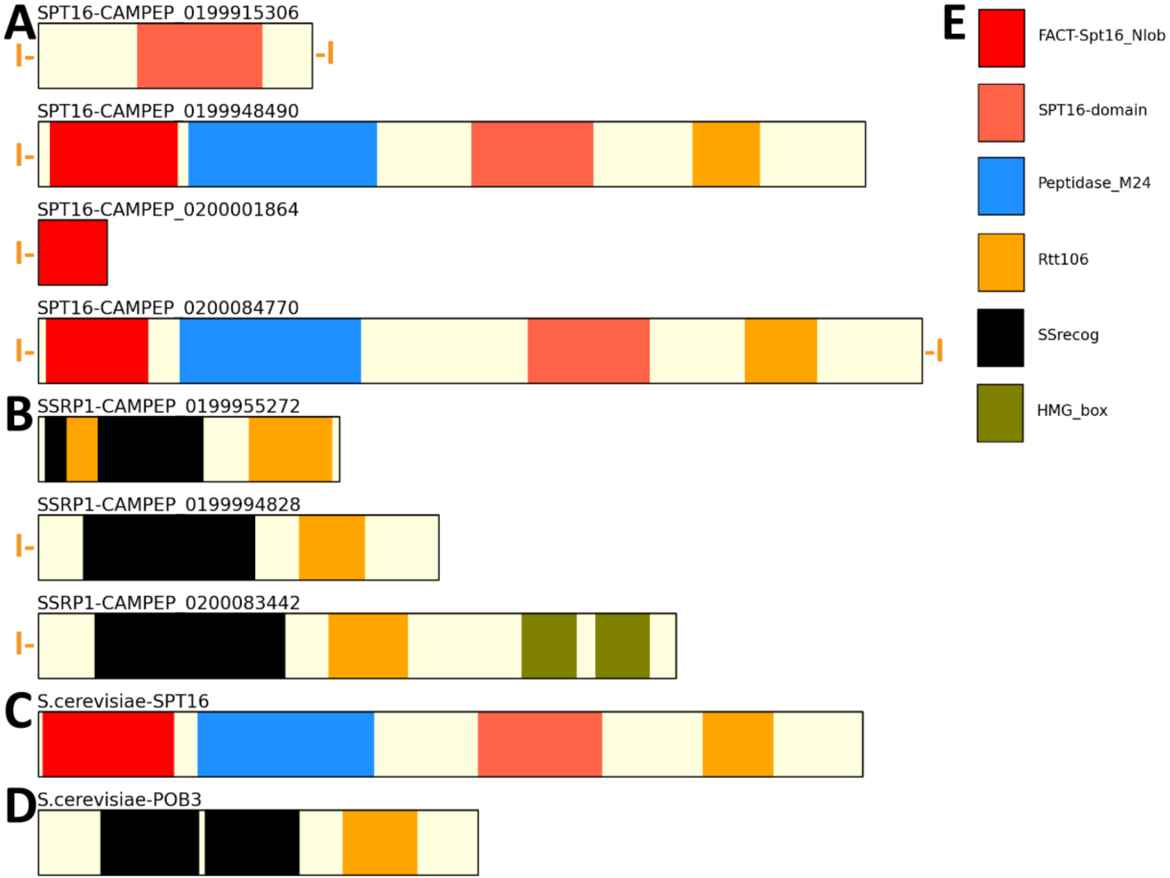
0 - +/- 0aa (1-aa) conservation
1 - +/- 1aa (3-aa) conservation
2 - +/- 2aa (5-aa) conservation



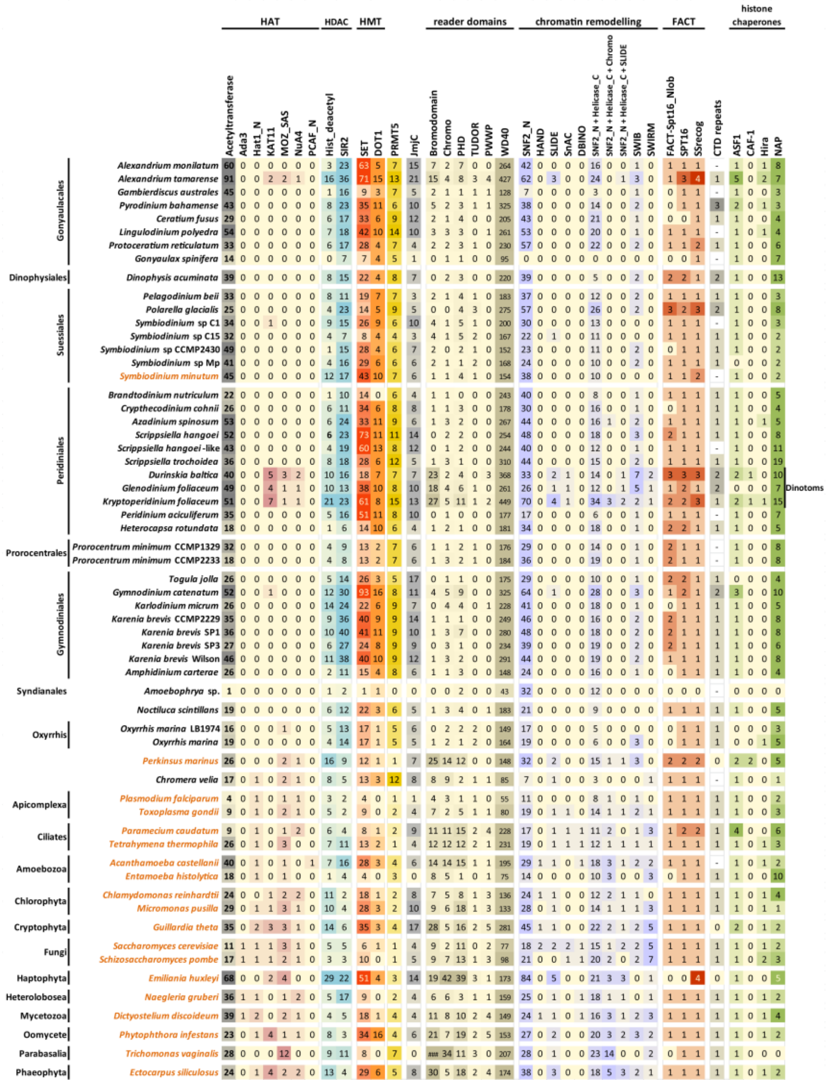
	H3	radius	R 2	T 3	K 4	T 6	R 8	K 9	S 10	T 11	K 14	R 17	K 18	K 23	R 26	K 27	S 28	P 30	K 36	K 37	P 38	Y 41	R 42	T 45	K 56	K 64	K 79	
Gonyaulacales	*Alexandrium monilatum*	0	3/3	3/3	3/3	3/3	3/3	2/3	3/3	2/3	3/3	3/3	0/3	3/3	3/3	3/3	0/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	
		1	3/3	3/3	3/3	3/3	2/3	2/3	1/3	2/3	2/3	0/3	0/3	0/3	2/3	0/3	0/3	0/3	1/3	3/3	3/3	3/3	2/3	3/3	3/3	3/3	2/3	
		2	3/3	3/3	3/3	3/3	2/3	1/3	1/3	1/3	1/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	1/3	1/3	3/3	2/3	2/3	2/3	3/3	2/3	2/3	
	Alexandrium tamarense	0	7/7	7/7	7/7	7/7	7/7	6/7	7/7	6/7	7/7	7/7	4/7	7/7	7/7	7/7	4/7	7/7	7/7	7/7	7/7	7/7	7/7	7/7	7/7	7/7	5/5	
		1	7/7	7/7	7/7	7/7	6/7	6/7	5/7	6/7	6/7	4/7	4/7	4/7	7/7	4/7	4/7	1/7	5/7	7/7	7/7	7/7	6/7	7/7	7/7	6/7	3/5	
		2	7/7	7/7	7/7	7/7	6/7	5/7	5/7	5/7	5/7	4/7	4/7	4/7	4/7	4/7	4/7	0/7	5/7	5/7	7/7	6/7	6/7	6/7	7/7	6/7	3/5	
	Gambierdiscus australes	0	2/2	2/2	2/2	1/2	2/2	0/2	1/2	1/2	1/2	1/2	1/2	2/2	1/2	2/2	0/2	2/2	2/2	1/2	2/2	1/2	1/2	1/2	1/2	2/2	1/2	
		1	2/2	2/2	0/2	0/2	0/2	0/2	0/2	1/2	0/2	0/2	0/2	0/2	1/2	0/2	0/2	0/2	0/2	1/2	1/2	1/2	1/2	1/2	1/2	2/2	1/2	
		2	2/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	
	Pyrodinium bahamense	0	3/3	3/3	3/3	2/3	3/3	1/3	2/3	1/3	2/3	2/3	1/3	3/3	2/3	2/3	0/3	2/3	2/3	2/3	2/3	2/3	2/3	2/3	2/3	3/3	2/3	
		1	3/3	3/3	1/3	1/3	1/3	1/3	0/3	1/3	1/3	0/3	0/3	0/3	2/3	0/3	0/3	0/3	0/3	2/3	2/3	2/3	2/3	2/3	2/3	3/3	2/3	
		2	3/3	1/3	1/3	1/3	1/3	0/3	0/3	0/3	1/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	2/3	2/3	2/3	2/3	2/3	2/3	2/3	
	Ceratium fusus	0	2/2	2/2	2/2	1/2	2/2	1/2	1/2	0/2	2/2	1/2	0/2	2/2	1/2	1/2	1/2	1/2	2/2	1/2	2/2	1/2	1/2	1/2	1/2	2/2	0/2	
		1	2/2	2/2	1/2	1/2	1/2	1/2	0/2	0/2	1/2	0/2	0/2	0/2	1/2	0/2	0/2	0/2	0/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	0/2	
		2	2/2	1/2	1/2	1/2	1/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	1/2	1/2	1/2	1/2	1/2	1/2	0/2	
	Lingulodinium polyedra	0	3/3	2/3	3/3	2/3	2/3	2/3	2/3	0/3	2/3	3/3	0/3	2/3	2/3	2/3	0/3	2/3	2/3	2/3	2/3	3/3	2/3	2/3	3/3	3/3	2/3	
		1	2/3	2/3	2/3	2/3	2/3	2/3	0/3	0/3	2/3	0/3	0/3	0/3	2/3	0/3	0/3	0/3	0/3	2/3	2/3	2/3	2/3	2/3	2/3	2/3	2/3	
		2	2/3	2/3	2/3	2/3	2/3	0/3	0/3	0/3	2/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	2/3	2/3	2/3	2/3	2/3	2/3	2/3	
	Protoceratium reticulatum	0	4/4	4/4	4/4	4/4	4/4	1/4	3/4	3/4	4/4	4/4	0/4	4/4	2/4	4/4	0/4	4/4	3/4	3/4	3/4	2/4	3/4	3/4	4/4	4/4	1/4	
		1	4/4	4/4	1/4	1/4	1/4	1/4	0/4	2/4	2/4	0/4	0/4	0/4	2/4	0/4	0/4	0/4	0/4	3/4	3/4	2/4	2/4	3/4	3/4	0/4	0/4	
		2	3/4	1/4	1/4	1/4	1/4	0/4	0/4	0/4	2/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	3/4	2/4	2/4	3/4	2/4	0/4	0/4	
	Gonyaluax spinifera	0	3/3	3/3	3/3	3/3	2/3	1/3	3/3	2/3	1/3	3/3	0/3	3/3	3/3	3/3	0/3	3/3	3/3	3/3	2/3	3/3	3/3	3/3	3/3	3/3	2/3	
		1	3/3	3/3	3/3	3/3	0/3	0/3	0/3	2/3	0/3	0/3	0/3	0/3	3/3	0/3	0/3	0/3	0/3	2/3	2/3	3/3	1/3	3/3	2/3	0/3	0/3	
		2	3/3	3/3	3/3	2/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	2/3	1/3	1/3	1/3	2/3	0/3	0/3	
Suessiales	*Brandtodinium nutriculum*	0	1/1	0/1	1/1	0/1	0/1	0/1	0/1	0/1	1/1	0/1	0/1	0/1	0/1	1/1	0/1	1/1	0/1	0/1	1/1	1/1	1/1	0/1	0/1	1/1	0/1	
		1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	1/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	1/1	1/1	0/1	0/1	0/1	0/1	
		2	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	1/1	0/1	0/1	0/1	0/1	
	Pelagodinium beii	0	4/5	3/5	4/5	1/5	3/5	1/5	1/5	2/5	2/5	3/5	3/5	3/5	2/5	4/5	1/5	2/5	3/5	3/5	1/5	2/5	3/5	2/5	2/5	4/5	2/5	
		1	3/5	3/5	1/5	1/5	1/5	1/5	1/5	1/5	2/5	1/5	1/5	1/5	2/5	1/5	1/5	0/5	1/5	1/5	1/5	2/5	2/5	2/5	2/5	2/5	1/5	
		2	3/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	2/5	1/5	1/5	1/5	1/5	1/5	1/5	0/5	1/5	1/5	1/5	2/5	2/5	2/5	1/5	1/5	0/5	
	Polarella glacialis	0	3/9	4/9	4/9	5/9	6/9	7/9	3/9	2/9	5/9	8/9	0/9	5/9	5/9	5/9	1/9	6/9	4/9	5/9	1/9	5/9	9/9	6/9	5/9	6/9	5/9	
		1	3/9	3/9	1/9	1/9	5/9	3/9	1/9	1/9	2/9	0/9	0/9	0/9	3/9	0/9	0/9	0/9	0/9	0/9	0/9	5/9	5/9	5/9	5/9	4/9	0/9	
		2	3/9	1/9	1/9	1/9	3/9	1/9	1/9	1/9	2/9	0/9	0/9	0/9	0/9	0/9	0/9	0/9	0/9	0/9	0/9	5/9	5/9	5/9	5/9	4/9	0/9	
	Symbiodinium sp. C1	0	6/6	6/6	6/6	5/6	3/6	6/6	5/6	3/6	4/6	5/6	4/6	5/6	5/6	6/6	3/6	6/6	5/6	5/6	4/6	3/6	5/6	5/6	5/6	6/6	5/6	
		1	6/6	6/6	3/6	3/6	3/6	3/6	3/6	3/6	3/6	3/6	3/6	3/6	5/6	3/6	3/6	2/6	3/6	3/6	3/6	3/6	3/6	3/6	5/6	5/6	1/6	
		2	6/6	3/6	3/6	3/6	3/6	3/6	3/6	3/6	3/6	3/6	3/6	3/6	3/6	3/6	3/6	1/6	3/6	3/6	3/6	3/6	3/6	3/6	4/6	3/6	1/6	
	Symbiodinium sp. C15	0	4/5	3/5	4/5	3/5	2/5	4/5	3/5	2/5	2/5	3/5	2/5	3/5	3/5	4/5	1/5	5/5	3/5	3/5	1/5	1/5	3/5	3/5	3/5	4/5	3/5	
		1	3/5	3/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	3/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	3/5	3/5	1/5	
		2	3/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	0/5	1/5	1/5	1/5	1/5	1/5	1/5	3/5	1/5	1/5	
	Symbiodinium sp. CCMP2430	0	2/2	2/2	2/2	2/2	0/2	2/2	2/2	0/2	2/2	2/2	0/2	2/2	2/2	2/2	0/2	2/2	2/2	2/2	0/2	2/2	2/2	0/2	2/2	2/2	2/2	
		1	2/2	2/2	0/2	0/2	0/2	0/2	0/2	0/2	2/2	0/2	0/2	0/2	2/2	0/2	0/2	2/2	0/2	0/2	0/2	2/2	2/2	0/2	2/2	2/2	0/2	
		2	2/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	2/2	0/2	2/2	2/2	0/2	
	Symbiodinium sp. Mp	0	1/1	0/1	1/1	0/1	0/1	0/1	0/1	1/1	0/1	1/1	1/1	0/1	0/1	0/1	0/1	0/1	1/1	1/1	1/1	0/1	1/1	0/1	0/1	1/1	0/1	
		1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	1/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	
		2	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	
	Symbiodinium minutum	0	2/8	2/8	1/8	1/8	0/8	2/8	0/8	0/8	1/8	1/8	0/8	3/8	0/8	3/8	0/8	0/8	1/8	0/8	0/8	1/8	1/8	0/8	0/8	2/8	0/8	
		1	1/8	1/8	0/8	0/8	0/8	0/8	0/8	0/8	1/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	
		2	1/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8	
Peridiniales	*Crypthecodinium cohnii*	0	3/5	3/5	4/5	3/5	3/5	4/5	4/5	0/5	2/5	0/5	4/5	0/5	1/5	1/5	0/5	3/5	3/5	3/5	0/5	0/5	1/5	2/5	5/5	5/5	1/5	
		1	3/5	3/5	0/5	0/5	3/5	3/5	0/5	0/5	2/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	1/5	4/5	1/5	0/5	
		2	3/5	0/5	0/5	0/5	2/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	3/5	0/5	0/5	
	Azadinium spinosum	0	5/5	5/5	5/5	3/5	4/5	1/5	0/5	0/5	4/5	3/5	3/5	5/5	4/5	4/5	0/5	4/5	3/5	3/5	4/5	3/5	4/5	3/5	3/5	5/5	3/5	
		1	5/5	5/5	3/5	3/5	1/5	0/5	0/5	0/5	3/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	3/5	3/5	3/5	3/5	0/5	3/5	4/5	3/5	
		2	5/5	3/5	3/5	3/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	3/5	3/5	3/5	0/5	3/5	3/5	3/5	
	Scrippsiella hangoei	0	6/6	5/6	6/6	4/6	5/6	4/6	4/6	0/6	4/6	4/6	2/6	3/6	3/6	3/6	1/6	5/6	4/6	4/6	2/6	5/6	1/6	4/6	3/6	6/6	4/6	
		1	5/6	5/6	0/6	0/6	4/6	4/6	0/6	0/6	3/6	0/6	0/6	0/6	1/6	0/6	0/6	0/6	0/6	1/6	1/6	0/6	0/6	0/6	3/6	4/6	0/6	
		2	5/6	0/6	0/6	0/6	4/6	0/6	0/6	0/6	1/6	0/6	0/6	0/6	0/6	0/6	0/6	0/6	0/6	0/6	1/6	0/6	0/6	0/6	3/6	2/6	0/6	
	Scrippsiella hangoei like	0	6/6	6/6	6/6	5/6	6/6	5/6	5/6	1/6	5/6	4/6	2/6	5/6	4/6	4/6	1/6	6/6	3/6	5/6	4/6	5/6	2/6	5/6	3/6	6/6	5/6	
		1	6/6	6/6	1/6	1/6	5/6	5/6	1/6	1/6	4/6	1/6	1/6	1/6	3/6	1/6	1/6	0/6	1/6	1/6	3/6	1/6	1/6	1/6	3/6	5/6	1/6	
		2	6/6	1/6	1/6	1/6	5/6	1/6	1/6	1/6	3/6	1/6	1/6	1/6	1/6	1/6	1/6	0/6	1/6	1/6	1/6	1/6	1/6	1/6	3/6	4/6	1/6	
	Scrippsiella trochoidea	0	0/1	0/1	0/1	1/1	1/1	1/1	1/1	0/1	1/1	1/1	0/1	1/1	1/1	1/1	0/1	1/1	1/1	1/1	1/1	1/1	0/1	1/1	1/1	1/1	1/1	
		1	0/1	0/1	0/1	0/1	1/1	1/1	0/1	0/1	1/1	0/1	0/1	0/1	1/1	0/1	0/1	0/1	0/1	1/1	0/1	0/1	0/1	1/1	1/1	1/1	0/1	
		2	0/1	0/1	0/1	0/1	1/1	0/1	0/1	0/1	1/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	1/1	1/1	1/1	0/1	
	Durinskia baltica	0	9/9	9/9	9/9	8/9	9/9	8/9	8/9	4/9	8/9	5/9	5/9	8/9	8/9	8/9	5/9	6/9	8/9	8/9	5/9	8/9	9/9	7/9	7/9	9/9	4/9	
		1	9/9	9/9	5/9	5/9	8/9	8/9	4/9	4/9	5/9	4/9	4/9	4/9	5/9	4/9	4/9	4/9	7/9	5/9	5/9	8/9	8/9	4/9	6/9	5/9	0/9	
		2	9/9	5/9	5/9	5/9	8/9	4/9	4/9	4/9	5/9	4/9	4/9	4/9	4/9	4/9	4/9	3/9	4/9	4/9	5/9	5/9	8/9	4/9	6/9	5/9	0/9	
	Glenodinium foliaceum	0	8/8	8/8	8/8	7/8	8/8	7/8	7/8	3/8	5/8	6/8	3/8	7/8	6/8	6/8	4/8	6/8	7/8	7/8	6/8	7/8	8/8	7/8	7/8	8/8	3/8	Dinotoms
		1	8/8	8/8	3/8	3/8	7/8	7/8	3/8	3/8	4/8	3/8	3/8	3/8	5/8	3/8	3/8	1/8	3/8	6/8	4/8	7/8	7/8	7/8	7/8	6/8	0/8	
		2	8/8	3/8	3/8	3/8	7/8	3/8	3/8	3/8	3/8	3/8	3/8	3/8	3/8	3/8	3/8	1/8	3/8	3/8	4/8	4/8	7/8	7/8	7/8	3/8	0/8	
	Kryptoperidinium foliaceum	0	12/12	12/12	12/12	12/12	12/12	12/12	12/12	5/12	8/12	11/12	5/12	12/12	12/12	12/12	5/12	11/12	12/12	12/12	8/12	12/12	12/12	12/12	12/12	12/12	4/12	
		1	12/12	12/12	5/12	5/12	12/12	12/12	5/12	5/12	7/12	5/12	5/12	5/12	9/12	5/12	5/12	1/12	5/12	8/12	7/12	12/12	12/12	12/12	11/12	9/12	0/12	
		2	12/12	5/12	5/12	5/12	12/12	5/12	5/12	5/12	7/12	5/12	5/12	5/12	5/12	5/12	5/12	1/12	5/12	5/12	7/12	7/12	12/12	12/12	10/12	6/12	0/12	
	Peridinium aciculiferum	0	6/6	5/6	6/6	4/6	5/6	4/6	4/6	0/6	3/6	3/6	0/6	5/6	4/6	4/6	0/6	4/6	4/6	4/6	3/6	4/6	2/6	4/6	3/6	6/6	4/6	
		1	5/6	5/6	0/6	0/6	4/6	4/6	0/6	0/6	3/6	0/6	0/6	0/6	1/6	0/6	0/6	0/6	0/6	2/6	2/6	0/6	0/6	0/6	3/6	4/6	0/6	
		2	5/6	0/6	0/6	0/6	4/6	0/6	0/6	0/6	1/6	0/6	0/6	0/6	0/6	0/6	0/6	0/6	0/6	0/6	2/6	0/6	0/6	0/6	3/6	2/6	0/6	
	Heterocapsa rotundata	0	2/2	2/2	2/2	2/2	2/2	2/2	1/2	1/2	2/2	2/2	1/2	2/2	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	
		1	2/2	2/2	1/2	1/2	2/2	1/2	1/2	1/2	2/2	1/2	1/2	1/2	2/2	1/2	1/2	0/2	1/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	
		2	2/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	2/2	1/2	1/2	1/2	1/2	1/2	1/2	0/2	1/2	1/2	2/2	2/2	2/2	2/2	2/2	2/2	1/2	
Prorocentrales	*Prorocentrum minimum* CCMP1329	0	3/3	3/3	3/3	2/3	3/3	3/3	2/3	0/3	2/3	2/3	0/3	2/3	2/3	2/3	0/3	2/3	2/3	2/3	2/3	2/3	2/3	2/3	2/3	3/3	2/3	
		1	3/3	3/3	2/3	2/3	2/3	2/3	0/3	0/3	2/3	0/3	0/3	0/3	2/3	0/3	0/3	0/3	2/3	2/3	0/3	2/3	2/3	2/3	2/3	2/3	2/3	
		2	3/3	2/3	2/3	2/3	2/3	0/3	0/3	0/3	2/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	1/3	0/3	0/3	0/3	2/3	2/3	2/3	2/3	2/3	
	Prorocentrum minimum CCMP2233	0	4/4	4/4	4/4	3/4	4/4	4/4	3/4	0/4	3/4	3/4	0/4	3/4	3/4	3/4	0/4	3/4	3/4	3/4	3/4	3/4	2/4	3/4	3/4	4/4	3/4	
		1	4/4	4/4	3/4	3/4	3/4	3/4	0/4	0/4	3/4	0/4	0/4	0/4	3/4	0/4	0/4	0/4	3/4	3/4	0/4	2/4	2/4	3/4	3/4	3/4	3/4	
		2	4/4	3/4	3/4	3/4	3/4	0/4	0/4	0/4	3/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	2/4	0/4	0/4	0/4	2/4	3/4	3/4	3/4	3/4	
Gymnodiniales	*Gymnodinium catenatum*	0	6/7	7/7	7/7	5/7	6/7	5/7	4/7	6/7	1/7	5/7	3/7	5/7	4/7	6/7	1/7	6/7	6/7	6/7	6/7	3/7	6/7	5/7	6/7	7/7	4/7	
		1	5/7	6/7	1/7	1/7	4/7	3/7	4/7	4/7	1/7	1/7	1/7	1/7	3/7	1/7	1/7	0/7	0/7	5/7	5/7	3/7	3/7	5/7	4/7	6/7	0/7	
		2	5/7	1/7	1/7	1/7	2/7	3/7	3/7	1/7	1/7	1/7	1/7	1/7	1/7	1/7	1/7	0/7	0/7	0/7	5/7	3/7	3/7	5/7	4/7	1/7	0/7	
	Karlodinium micrum	0	6/6	6/6	6/6	3/6	1/6	2/6	4/6	1/6	6/6	4/6	2/6	5/6	4/6	5/6	1/6	0/6	4/6	4/6	2/6	5/6	5/6	0/6	4/6	6/6	1/6	
		1	6/6	6/6	4/6	0/6	0/6	1/6	1/6	1/6	0/6	0/6	0/6	0/6	1/6	0/6	0/6	0/6	1/6	1/6	0/6	4/6	4/6	0/6	4/6	4/6	0/6	
		2	6/6	4/6	3/6	0/6	0/6	0/6	1/6	0/6	0/6	0/6	0/6	0/6	0/6	0/6	0/6	0/6	0/6	0/6	0/6	3/6	4/6	0/6	4/6	1/6	0/6	
	Karenia brevis CCMP2229	0	4/4	2/4	2/4	2/4	0/4	2/4	0/4	1/4	1/4	1/4	0/4	0/4	0/4	1/4	0/4	4/4	0/4	2/4	0/4	3/4	4/4	2/4	2/4	4/4	1/4	
		1	2/4	2/4	0/4	0/4	0/4	0/4	0/4	0/4	1/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	3/4	3/4	0/4	2/4	2/4	0/4	
		2	2/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	2/4	3/4	0/4	2/4	2/4	0/4	
	Karenia brevis SP1	0	4/4	3/4	3/4	3/4	1/4	3/4	0/4	1/4	1/4	1/4	1/4	2/4	0/4	2/4	0/4	2/4	1/4	3/4	1/4	3/4	4/4	2/4	3/4	4/4	0/3	
		1	2/4	3/4	0/4	0/4	1/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	1/4	1/4	3/4	3/4	0/4	3/4	2/4	0/3	
		2	2/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	1/4	3/4	3/4	0/4	3/4	2/4	0/3	
	Karenia brevis SP3	0	3/3	1/3	1/3	1/3	1/3	1/3	1/3	0/3	1/3	1/3	0/3	1/3	2/3	1/3	0/3	2/3	1/3	1/3	2/3	2/3	2/3	0/3	1/3	3/3	0/3	
		1	0/3	1/3	0/3	0/3	1/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	1/3	1/3	1/3	1/3	0/3	1/3	0/3	0/3	
		2	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	1/3	1/3	1/3	0/3	1/3	0/3	0/3	
	Karenia brevis Wilson	0	4/4	3/4	3/4	3/4	0/4	3/4	0/4	1/4	0/4	1/4	0/4	3/4	0/4	3/4	0/4	1/4	0/4	3/4	0/4	3/4	4/4	3/4	3/4	4/4	1/4	
		1	3/4	3/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	3/4	3/4	0/4	3/4	3/4	0/4	
		2	3/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	3/4	3/4	0/4	3/4	3/4	0/4	
	Amphidinium carterae	0	2/3	2/3	2/3	1/3	0/3	1/3	1/3	1/3	2/3	1/3	0/3	2/3	1/3	1/3	0/3	3/3	1/3	1/3	0/3	0/3	1/3	0/3	1/3	3/3	0/3	
		1	2/3	2/3	0/3	0/3	0/3	0/3	0/3	0/3	1/3	0/3	0/3	0/3	1/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	1/3	0/3	0/3	
		2	2/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	1/3	0/3	0/3	
Syndianales	*Amoebophrya* sp.	0	3/3	2/3	1/3	0/3	1/3	0/3	1/3	0/3	2/3	0/3	0/3	0/3	0/3	1/3	2/3	1/3	1/3	2/3	0/3	0/3	0/3	0/3	1/3	2/3	0/3	
		1	2/3	1/3	1/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	1/3	0/3	0/3	0/3	0/3	0/3	1/3	1/3	0/3	
		2	1/3	1/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	1/3	0/3	0/3	
	Noctiluca scintillans	0	1/2	2/2	2/2	2/2	2/2	2/2	2/2	0/2	2/2	2/2	0/2	2/2	1/2	1/2	0/2	1/2	2/2	1/2	2/2	1/2	2/2	2/2	2/2	2/2	0/2	
		1	1/2	1/2	1/2	1/2	2/2	2/2	0/2	0/2	1/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	1/2	1/2	1/2	1/2	1/2	2/2	2/2	0/2	
		2	1/2	1/2	1/2	1/2	2/2	0/2	0/2	0/2	1/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	1/2	1/2	1/2	1/2	2/2	2/2	0/2	
Oxyrrhis	*Oxyrrhis marina* LB1974	0	5/5	4/5	3/5	1/5	1/5	3/5	1/5	1/5	4/5	1/5	4/5	5/5	1/5	2/5	1/5	1/5	1/5	4/5	1/5	1/5	2/5	3/5	1/5	5/5	4/5	
		1	4/5	3/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	0/5	1/5	1/5	1/5	1/5	1/5	3/5	1/5	0/5	0/5	
		2	3/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	0/5	1/5	1/5	1/5	1/5	1/5	1/5	1/5	0/5	0/5	
	Oxyrrhis marina	0	9/9	8/9	7/9	2/9	2/9	2/9	8/9	2/9	8/9	2/9	8/9	8/9	2/9	3/9	2/9	2/9	7/9	8/9	2/9	2/9	2/9	7/9	2/9	8/9	8/9	
		1	8/9	7/9	2/9	2/9	2/9	2/9	2/9	2/9	2/9	2/9	2/9	2/9	2/9	2/9	2/9	1/9	2/9	2/9	2/9	2/9	2/9	7/9	2/9	2/9	2/9	
		2	7/9	2/9	2/9	2/9	2/9	2/9	2/9	2/9	2/9	2/9	2/9	2/9	2/9	2/9	2/9	1/9	2/9	2/9	2/9	2/9	2/9	2/9	2/9	2/9	2/9	
	Perkinsus marinus	0	6/9	4/9	6/9	5/9	4/9	5/9	5/9	7/9	6/9	6/9	6/9	6/9	8/9	6/9	6/9	5/9	4/9	5/9	6/9	5/9	9/9	9/9	4/9	9/9	4/9	
		1	4/9	4/9	4/9	5/9	4/9	4/9	5/9	5/9	4/9	5/9	5/9	0/9	5/9	5/9	0/9	0/9	0/9	4/9	4/9	5/9	5/9	6/9	4/9	4/9	4/9	
		2	4/9	4/9	4/9	4/9	4/9	4/9	4/9	5/9	4/9	4/9	5/9	0/9	5/9	0/9	0/9	0/9	0/9	0/9	4/9	4/9	5/9	6/9	4/9	4/9	4/9	
	Chromera velia	0	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	0/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	
		1	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	0/2	2/2	0/2	0/2	0/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	
		2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	0/2	0/2	0/2	0/2	0/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	

THERE IS A FACT COMPLEX

- Thus, presumably transcription through nucleosomes too



HISTONE MODIFYING ENZYMES AND CHROMATIN REMODELLERS



		HAT							HDAC		HMT				reader domains						chromatin remodelling										FACT				histone chaperones				
		Acetyltransferase	Ada3	Hat1_N	KAT11	MOZ_SAS	NuA4	PCAF_N	Hist_deacetyl	SIR2	SET	DOT1	PRMT5	JmjC	Bromodomain	Chromo	PHD	TUDOR	PWWP	WD40	SNF2_N	HAND	SLIDE	SnAC	DBINO	SNF2_N + Helicase_C	SNF2_N + Helicase_C + Chromo	SNF2_N + Helicase_C + SLIDE	SWIB	SWIRM	FACT-Spt16_Nlob	SPT16	SSrecog	CTD repeats	ASF1	CAF-1	Hira	NAP	
Gonyaulacales	*Alexandrium monilatum*	60	0	0	0	0	0	0	3	23	63	5	7	15	7	2	7	0	0	264	42	0	0	0	0	16	0	0	1	0	1	1	1	-	1	0	1	8	
	Alexandrium tamarense	91	0	0	2	2	1	0	16	36	71	15	13	21	15	4	8	3	4	427	62	0	3	0	0	24	0	1	3	0	1	3	4	1	5	0	2	7	
	Gambierdiscus australes	45	0	0	0	0	0	0	1	16	9	3	7	5	2	1	1	1	1	128	7	0	0	0	0	3	0	0	1	0	0	1	1	-	1	0	0	3	
	Pyrodinium bahamense	43	0	0	0	0	0	0	8	23	35	11	6	10	5	2	3	1	1	325	38	0	0	0	0	14	0	0	2	0	0	1	1	3	2	0	1	3	
	Ceratium fusus	29	0	0	0	0	0	0	6	17	33	6	9	12	2	1	4	0	0	205	43	0	0	0	0	21	0	0	1	0	0	0	1	1	1	0	0	4	
	Lingulodinium polyedra	54	0	0	0	0	0	0	7	18	42	10	14	10	3	3	3	0	1	261	53	0	0	0	0	20	0	0	1	0	1	1	1	-	1	0	1	4	
	Protoceratium reticulatum	33	0	0	0	0	0	0	6	17	28	4	7	4	2	2	3	1	0	230	57	0	0	0	0	22	0	0	2	0	1	1	2	1	1	0	0	6	
	Gonyaulax spinifera	14	0	0	0	0	0	0	0	7	7	4	5	1	0	1	1	0	0	95	0	0	0	0	0	0	0	0	0	0	0	0	1	-	1	0	0	7	
Dinophysiales	*Dinophysis acuminata*	39	0	0	0	0	0	0	8	15	22	4	8	7	0	2	3	0	0	220	39	0	0	0	0	5	0	0	2	0	2	2	1	2	1	0	0	13	
Suessiales	*Pelagodinium beii*	33	0	0	0	0	0	0	8	11	19	7	7	3	2	1	4	1	0	183	37	0	0	0	0	12	0	0	2	0	1	1	1	1	1	0	0	3	
	Polarella glacialis	25	0	0	0	0	0	0	4	23	14	5	9	5	0	0	4	3	0	275	57	0	0	0	0	26	0	0	2	0	3	2	3	2	1	0	0	8	
	Symbiodinium sp C1	34	0	0	1	0	0	0	9	15	26	9	6	10	4	1	5	1	0	200	30	0	0	0	0	13	0	0	0	0	1	1	1	-	1	0	0	3	
	Symbiodinium sp C15	32	0	0	0	0	0	0	4	7	8	4	4	3	4	1	5	2	0	167	22	0	1	0	0	11	0	0	0	0	1	1	1	1	1	0	0	2	
	Symbiodinium sp CCMP2430	49	0	0	0	0	0	0	1	15	28	4	6	7	2	0	2	1	0	152	23	0	0	0	0	11	0	0	2	0	0	1	1	1	1	0	0	2	
	Symbiodinium sp Mp	41	0	0	0	0	0	0	4	16	29	6	6	6	2	1	1	2	0	168	24	0	0	0	0	10	0	0	2	0	1	1	1	1	0	0	0	2	
	Symbiodinium minutum	45	0	0	0	0	0	0	12	17	43	10	7	6	1	1	4	1	0	154	38	0	0	0	0	10	0	0	0	0	1	1	2	-	1	0	0	2	
Peridiniales	*Brandtodinium nutriculum*	22	0	0	0	0	0	0	1	10	14	0	6	4	1	1	0	0	0	243	40	0	0	0	0	8	0	0	1	0	1	1	1	1	1	0	0	5	
	Crypthecodinium cohnii	26	0	0	0	0	0	0	6	11	34	6	8	8	1	1	3	0	0	178	30	0	0	0	0	16	0	0	1	0	0	1	1	1	1	0	0	4	
	Azadinium spinosum	53	0	0	0	0	0	0	6	24	33	11	9	6	1	3	2	0	0	267	44	0	0	0	0	16	1	0	2	0	1	1	1	1	1	0	1	5	
	Scrippsiella hangoei	52	0	0	0	0	0	0	6	23	73	11	11	14	0	2	2	0	0	254	48	0	0	0	0	18	0	0	3	0	2	1	1	1	1	0	0	8	
	Scrippsiella hangoei -like	43	0	0	0	0	0	0	4	19	60	13	8	12	0	1	2	0	0	244	40	0	0	0	0	12	0	0	1	0	1	1	1	-	1	0	0	11	
	Scrippsiella trochoidea	36	0	0	0	0	0	0	8	18	28	6	12	5	1	3	1	0	0	310	44	0	0	0	0	15	0	0	2	0	1	1	1	1	1	0	0	19	
	Durinskia baltica	40	0	0	5	3	2	0	10	16	18	7	7	7	23	2	4	0	3	368	33	0	2	1	0	14	0	1	7	2	3	3	3	2	2	1	0	10	
	Glenodinium foliaceum	49	0	0	4	1	1	0	10	13	38	10	8	10	18	4	6	1	0	261	26	0	1	1	0	12	0	1	5	1	1	2	1	2	0	0	0	7	Dinotoms
	Kryptoperidinium foliaceum	51	0	0	7	1	1	0	21	23	61	8	15	13	27	5	11	1	2	449	70	0	4	1	0	34	3	2	2	1	2	2	3	1	2	1	1	15	
	Peridinium aciculiferum	35	0	0	0	0	0	0	5	16	51	11	8	10	0	1	0	0	0	177	17	0	0	0	0	6	0	0	1	0	1	1	1	-	1	0	0	7	
	Heterocapsa rotundata	18	0	0	0	0	0	0	1	6	14	10	6	4	1	2	1	0	0	181	34	0	0	0	0	18	0	0	1	0	2	2	1	1	1	0	0	5	
Prorocentrales	*Prorocentrum minimum* CCMP1329	32	0	0	0	0	0	0	4	9	13	2	7	6	1	1	2	1	0	176	29	0	0	0	0	14	0	0	1	0	2	1	1	-	1	0	0	8	
	Prorocentrum minimum CCMP2233	18	0	0	0	0	0	0	4	8	13	2	7	6	1	3	2	1	0	184	36	0	0	0	0	19	0	0	1	0	2	1	1	-	1	0	0	8	
Gymnodiniales	*Togula jolla*	26	0	0	0	0	0	0	5	14	26	3	5	17	0	1	1	0	0	175	29	0	0	0	0	10	0	0	1	0	2	2	1	1	0	0	0	4	
	Gymnodinium catenatum	52	0	0	1	0	0	0	12	30	93	16	8	11	4	5	9	0	0	325	64	0	1	0	0	28	0	0	3	0	1	2	1	2	3	0	0	10	
	Karlodinium micrum	26	0	0	0	0	0	0	14	24	22	6	9	7	0	4	4	0	1	228	41	0	0	0	0	18	0	0	1	0	0	1	1	1	1	0	0	5	
	Karenia brevis CCMP2229	35	0	0	0	0	0	0	9	36	40	9	9	14	1	1	1	0	0	249	46	0	0	0	0	16	0	0	2	0	2	1	1	1	1	0	0	8	
	Karenia brevis SP1	36	0	0	0	0	0	0	10	40	41	11	9	10	1	3	7	0	0	280	48	0	0	0	0	18	0	0	2	0	2	1	1	1	1	0	0	8	
	Karenia brevis SP3	27	0	0	0	0	0	0	6	27	24	8	9	9	0	3	1	0	0	234	39	0	0	0	0	19	0	0	2	0	2	1	1	1	1	0	0	6	
	Karenia brevis Wilson	46	0	0	0	0	0	0	11	38	40	10	9	12	1	3	2	0	0	291	44	0	0	0	0	19	0	0	2	0	1	1	1	1	1	0	0	8	
	Amphidinium carterae	26	0	0	0	0	0	0	2	11	15	4	8	6	1	1	3	0	0	148	24	0	0	0	0	16	0	0	1	0	0	1	1	1	1	0	0	4	
Syndianales	*Amoebophrya* sp.	1	0	0	0	0	0	0	1	2	1	1	0	0	0	0	2	0	0	43	32	0	0	0	0	12	0	0	0	0	0	0	0	-	0	0	0	0	
	Noctiluca scintillans	19	0	0	0	0	0	0	6	12	22	3	6	5	1	3	4	0	1	183	21	0	0	0	0	9	0	0	0	0	1	1	1	1	1	0	0	5	
Oxyrrhis	*Oxyrrhis marina* LB1974	16	0	0	0	1	0	0	5	13	17	1	5	6	0	2	2	2	0	149	17	0	0	0	0	5	0	0	0	0	0	1	1	1	0	0	0	3	
	Oxyrrhis marina	19	0	0	0	0	0	0	4	14	17	1	5	5	1	2	1	2	0	164	19	0	0	0	0	6	0	0	3	0	0	1	1	1	0	0	1	5	
	Perkinsus marinus	26	0	0	0	2	1	0	16	9	12	1	1	7	25	14	12	0	0	148	32	0	2	0	0	15	1	1	3	0	2	2	2	0	2	2	0	5	
	Chromera velia	17	0	1	0	2	1	0	8	5	13	3	12	8	8	9	2	1	1	85	7	0	1	0	0	3	0	0	0	1	1	1	1	-	1	0	0	1	
Apicomplexa	*Plasmodium falciparum*	4	0	1	0	1	1	0	3	2	4	0	1	1	4	3	1	1	0	55	11	0	0	0	0	8	1	0	1	0	1	1	1	1	1	0	0	2	
	Toxoplasma gondii	9	0	1	0	2	1	0	5	2	9	0	2	4	7	2	5	1	1	80	19	0	1	1	0	14	1	1	2	1	1	1	1	1	1	0	0	3	
Ciliates	*Paramecium caudatum*	9	0	1	0	1	2	0	6	4	8	1	2	9	11	11	15	2	4	228	17	0	1	1	1	11	2	0	1	3	1	2	2	1	4	0	0	6	
	Tetrahymena thermophila	26	0	1	0	3	0	0	7	11	13	2	1	4	12	12	12	2	1	231	19	0	1	1	1	12	1	1	1	1	1	1	1	1	1	0	1	3	
Amoebozoa	*Acanthamoeba castellanii*	40	0	1	0	1	0	1	7	16	28	3	4	6	14	14	15	1	1	195	29	1	1	0	1	18	3	1	2	2	1	1	1	-	1	0	1	2	
	Entamoeba histolytica	18	0	1	0	1	0	0	1	4	4	0	3	0	8	5	1	0	1	75	14	0	0	0	0	10	3	0	0	3	0	1	1	1	1	0	0	10	
Chlorophyta	*Chlamydomonas reinhardtii*	24	0	0	1	2	2	0	11	2	18	1	2	8	7	5	8	1	3	136	24	1	1	0	0	12	2	1	1	0	1	1	1	1	1	0	1	4	
	Micromonas pusilla	29	0	1	1	3	1	0	10	4	28	3	2	10	9	6	18	1	3	133	28	0	1	0	0	14	1	1	1	3	1	1	1	1	1	0	1	1	
Cryptophyta	*Guillardia theta*	35	0	2	3	3	1	0	14	6	35	3	4	17	28	5	16	2	5	281	45	1	1	0	0	22	2	1	2	5	1	1	1	0	2	0	1	2	
Fungi	*Saccharomyces cerevisiae*	11	1	1	1	3	1	0	5	5	6	1	1	4	9	2	11	0	2	77	18	2	2	2	1	15	1	2	2	5	1	1	1	1	1	0	1	2	
	Schizosaccharomyces pombe	17	1	1	1	2	1	0	3	3	10	0	1	5	9	7	13	1	3	98	21	0	0	1	1	20	2	0	2	7	1	1	1	1	1	0	2	3	
Haptophyta	*Emiliania huxleyi*	68	0	0	2	4	0	0	29	22	51	4	3	14	19	42	39	3	1	173	84	0	5	0	0	21	3	3	0	1	0	0	4	0	1	0	0	5	
Heterolobosea	*Naegleria gruberi*	36	1	1	0	1	2	0	5	17	9	0	2	4	6	6	3	1	1	159	25	0	1	0	1	18	1	1	0	1	1	1	1	1	1	0	1	2	
Mycetozoa	*Dictyostelium discoideum*	39	1	2	0	2	1	0	4	5	18	1	4	4	11	8	10	2	4	149	24	1	1	0	1	16	3	1	2	3	1	1	1	1	1	0	1	4	
Oomycete	*Phytophthora infestans*	23	0	1	4	1	1	0	8	3	34	16	4	6	21	7	19	2	5	153	27	0	2	0	0	20	3	2	3	2	1	1	1	1	1	0	1	2	
Parabasalia	*Trichomonas vaginalis*	28	0	0	0	12	0	0	9	11	8	0	7	0	###	34	11	3	0	207	28	0	1	0	0	23	14	0	0	2	0	1	1	1	1	0	0	0	
Phaeophyta	*Ectocarpus siliculosus*	24	0	1	4	2	2	0	13	4	29	6	5	8	30	5	18	2	4	174	38	0	3	0	0	18	5	3	2	1	1	1	1	1	1	0	1	2	

CORAL SYMBIOSIS AS MODEL SYSTEM

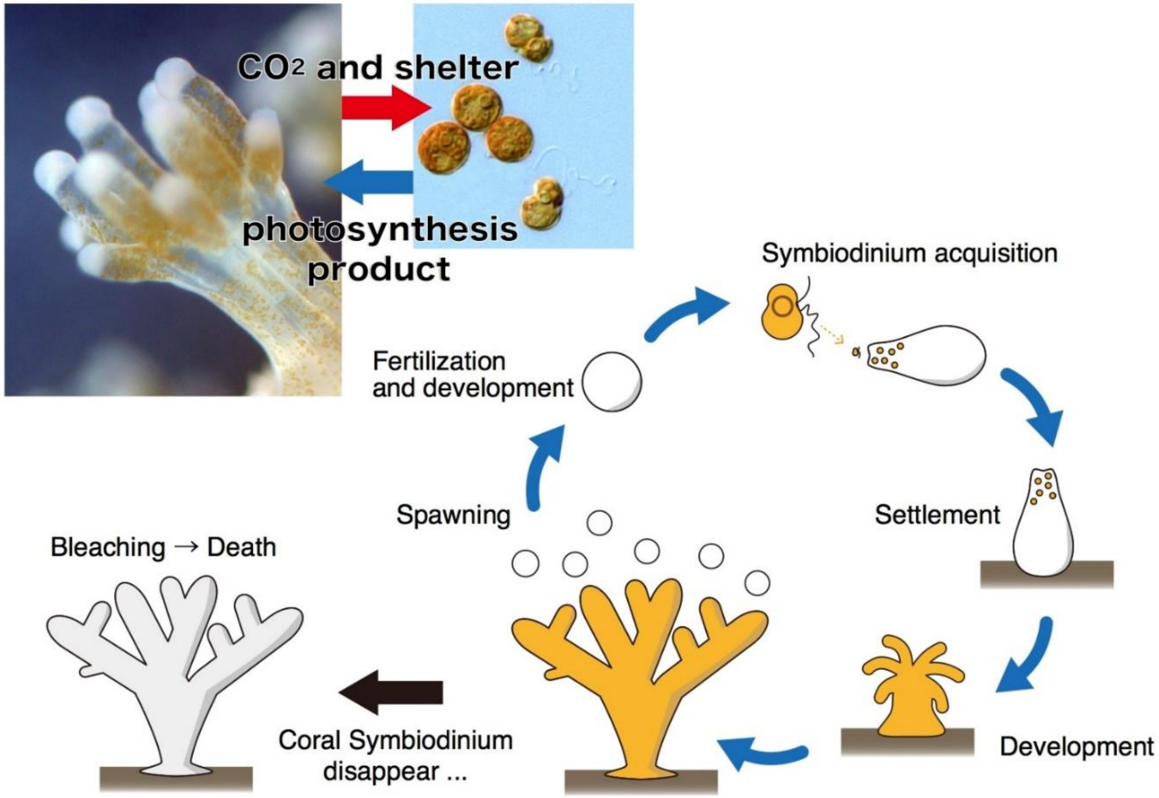
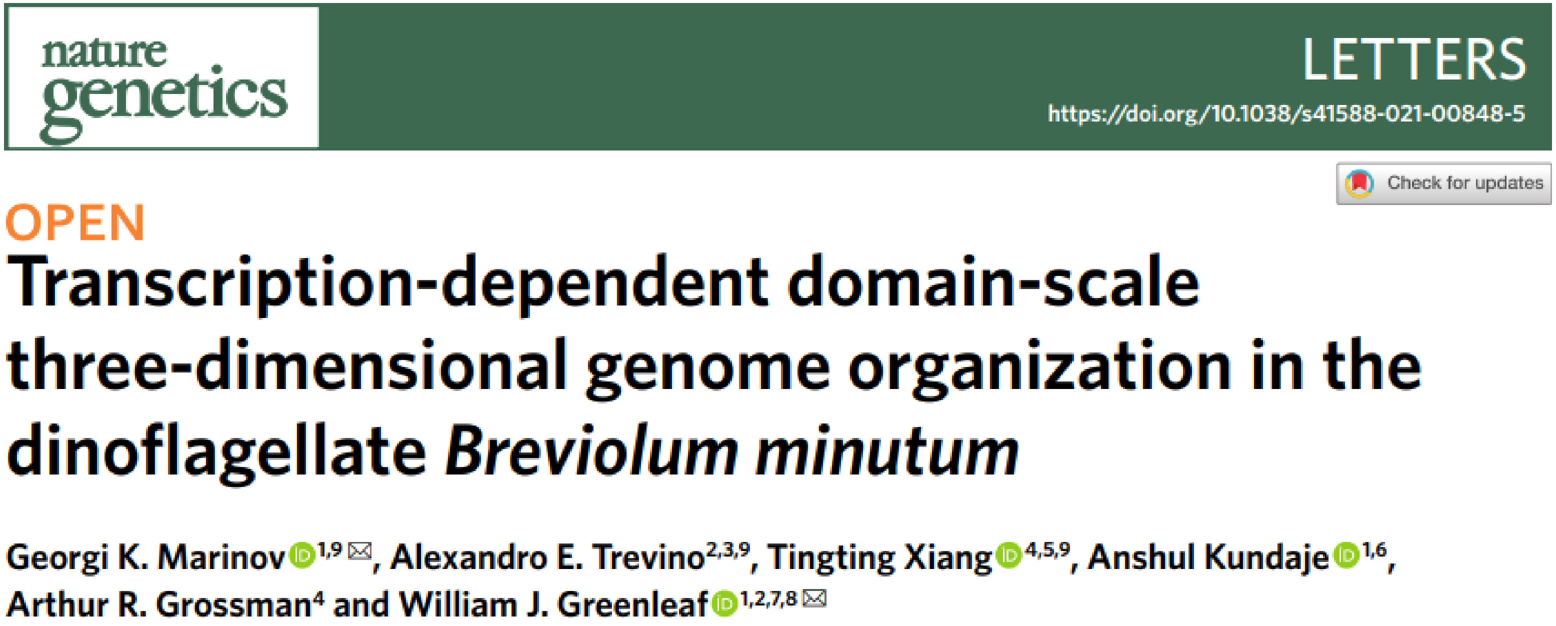


Figure 2. A symbiotic relationship between corals and Symbiodinium



nature genetics

LETTERS

https://doi.org/10.1038/s41588-021-00848-5

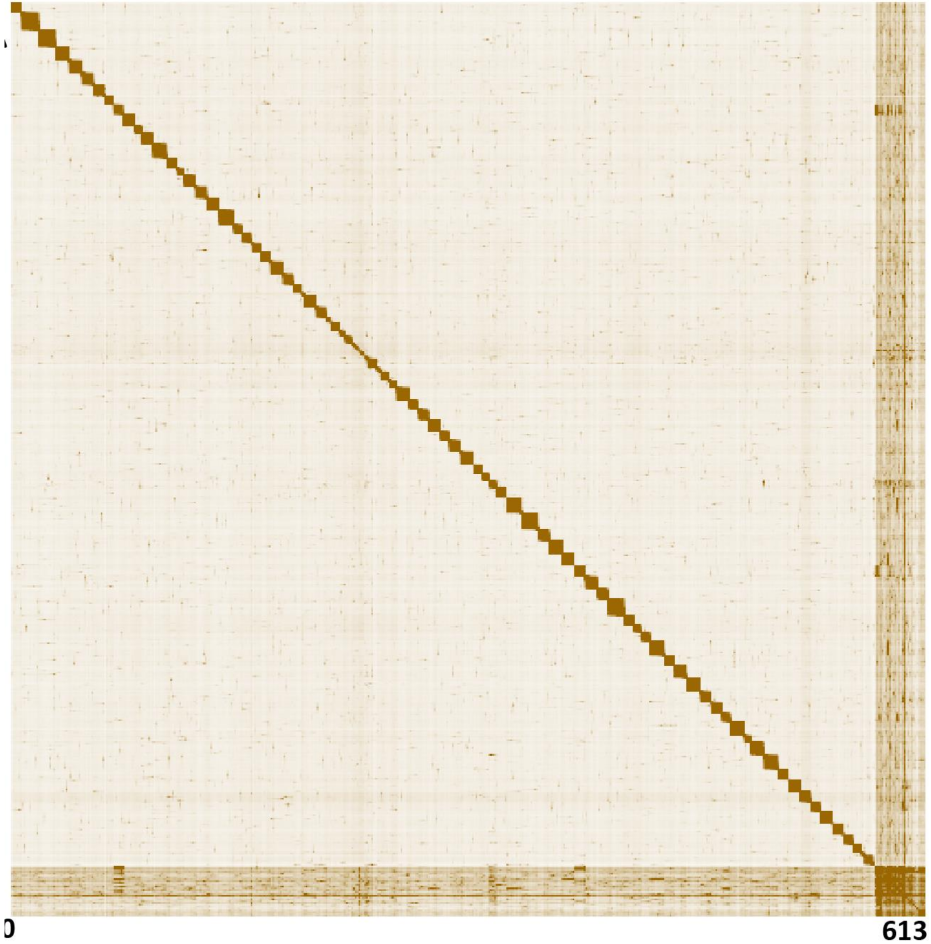
OPEN

Transcription-dependent domain-scale three-dimensional genome organization in the dinoflagellate *Breviolum minutum*

Georgi K. Marinov[1,9] ✉, Alexandro E. Trevino[2,3,9], Tingting Xiang[4,5,9], Anshul Kundaje[1,6],
Arthur R. Grossman[4] and William J. Greenleaf[1,2,7,8] ✉

SYMBIODINIUM/BREVIOLUM MINUTUM HI-C

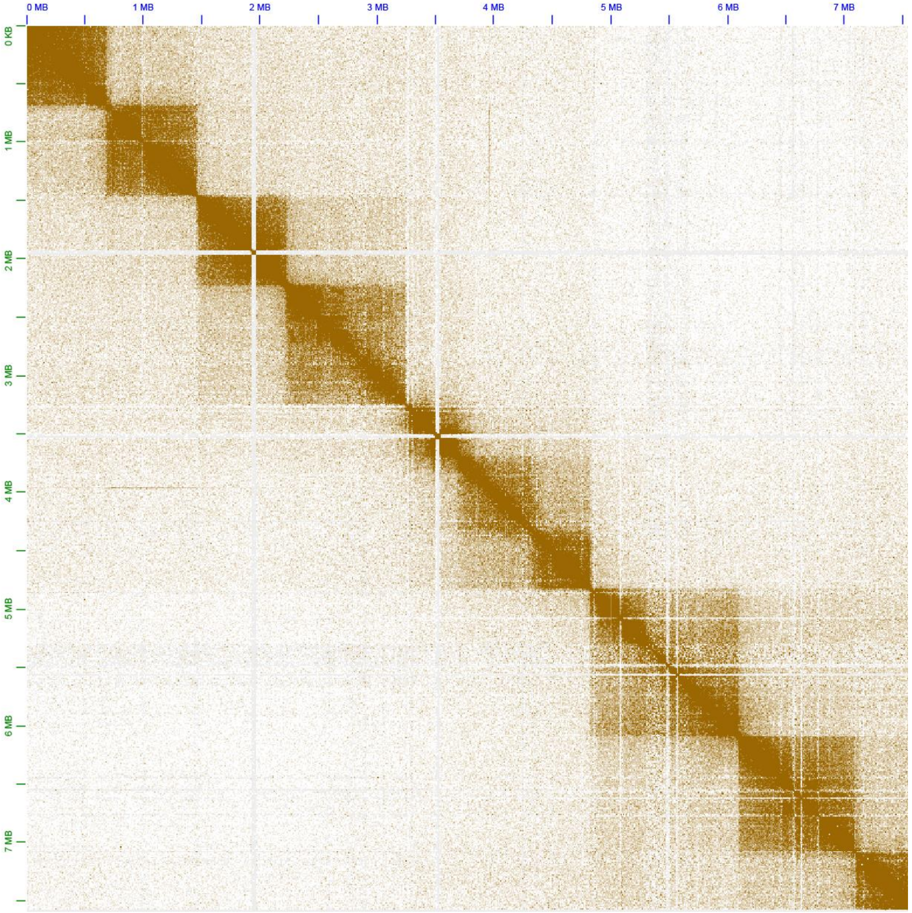
- Extremely poor draft assembly – N50 < 50 kbp
- Carried out deep Hi-C and Hi-C assisted scaffolding
- Hi-C scaffolding identifies ~90 chromosomes



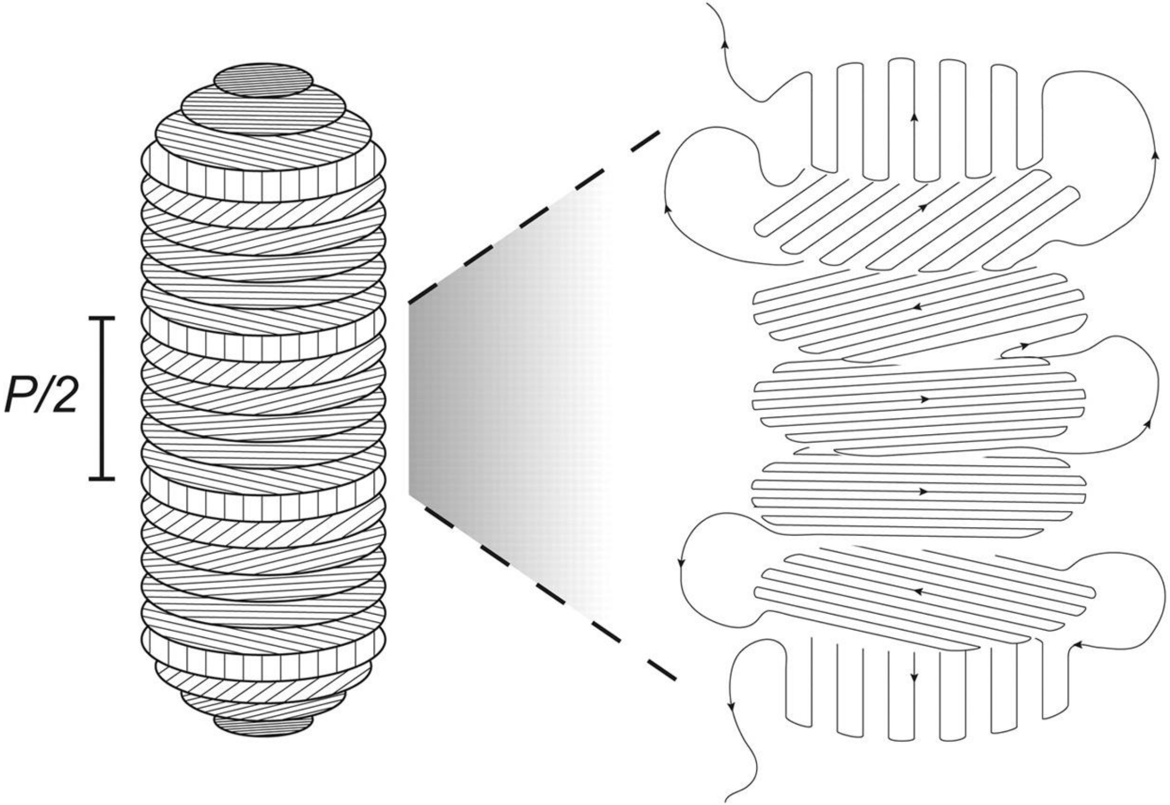
DOMAIN STRUCTURE

- “dinoTADs”

pseudochromosome 10



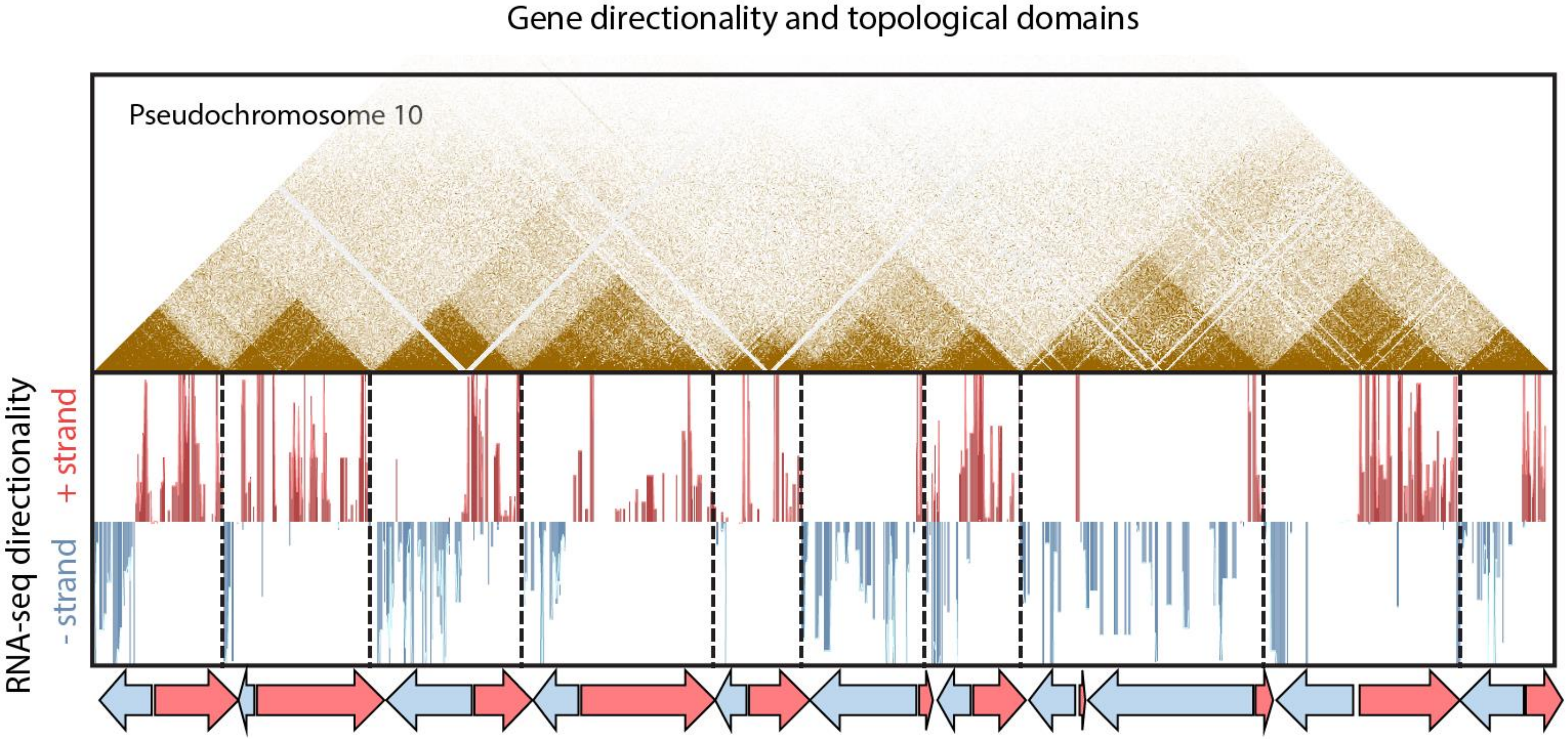
TRADITIONAL MODELS



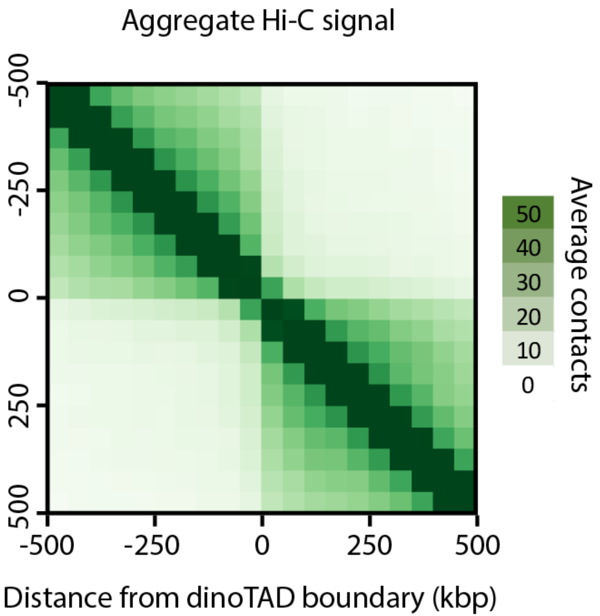
Dinoflagellate chromosome

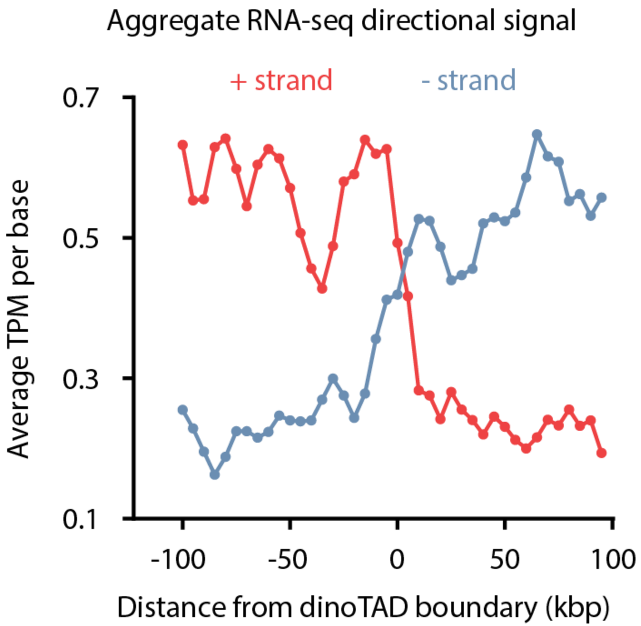
Chow et al. 2010

DOMAIN STRUCTURE AND TRANSCRIPTION

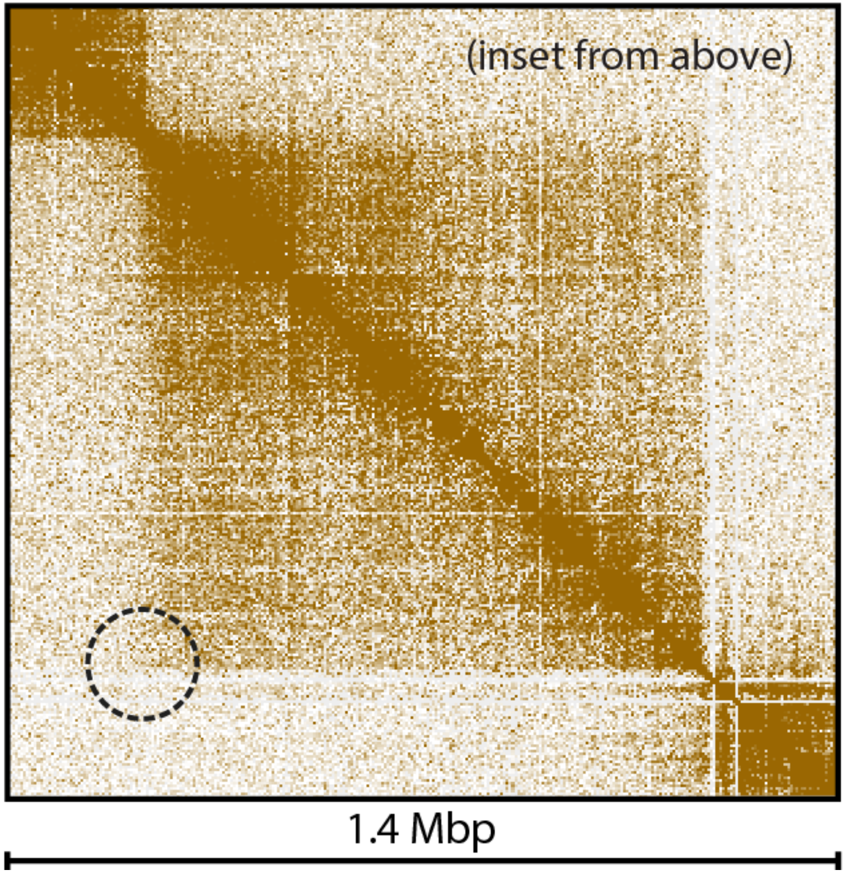


DOMAIN STRUCTURE AND TRANSCRIPTION



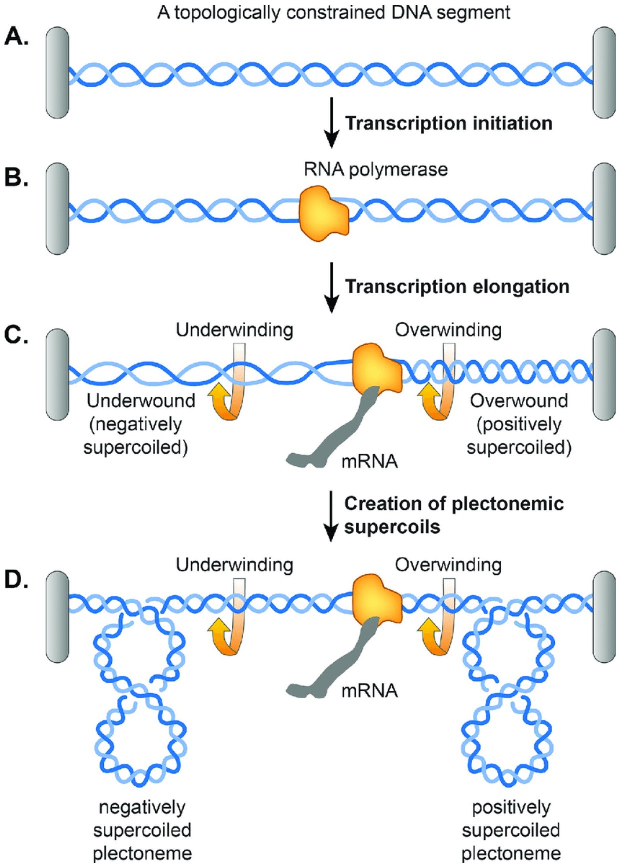


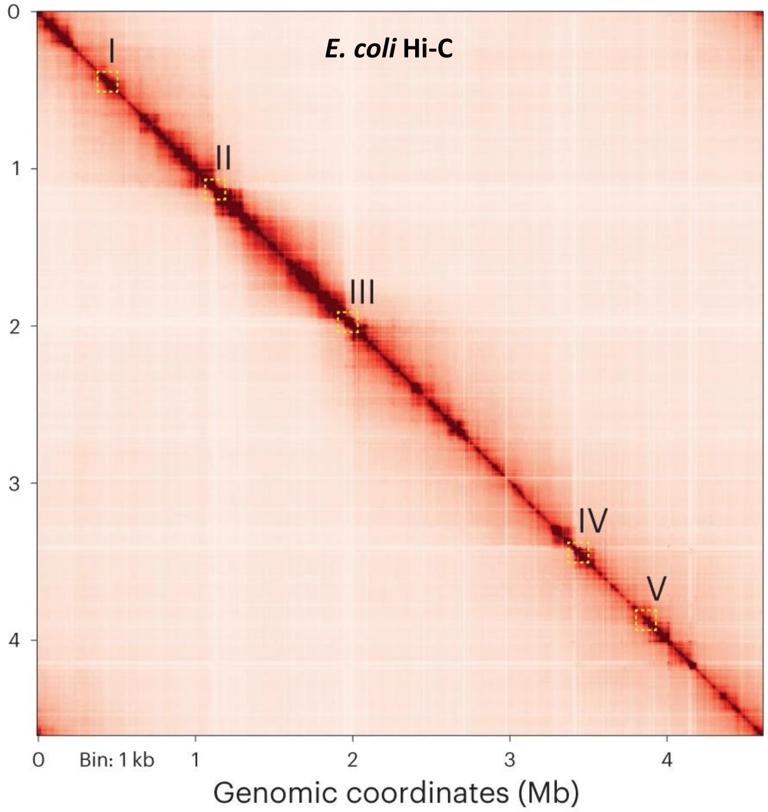
NO CHROMATIN LOOPS OBSERVED


Pseudochromosome 10

WHAT IS DRIVING DOMAIN FORMATION IN SYMBIODINIUM?

- Most likely it is transcription itself as domains coincide with gene arrays

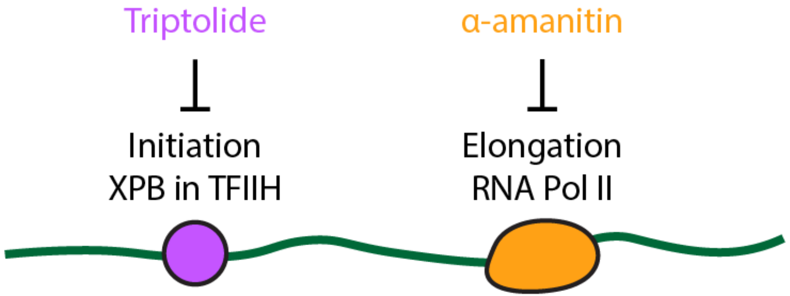




Verma et al. 2019
Bignaud et al. 2024

TRANSCRIPTION INHIBITION EXPERIMENTS

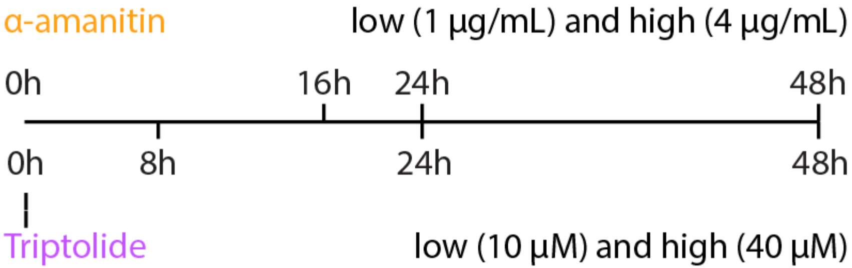
Transcriptional inhibitors



Triptolide ⊥ Initiation XPB in TFIIH

α-amanitin ⊥ Elongation RNA Pol II

Experimental overview



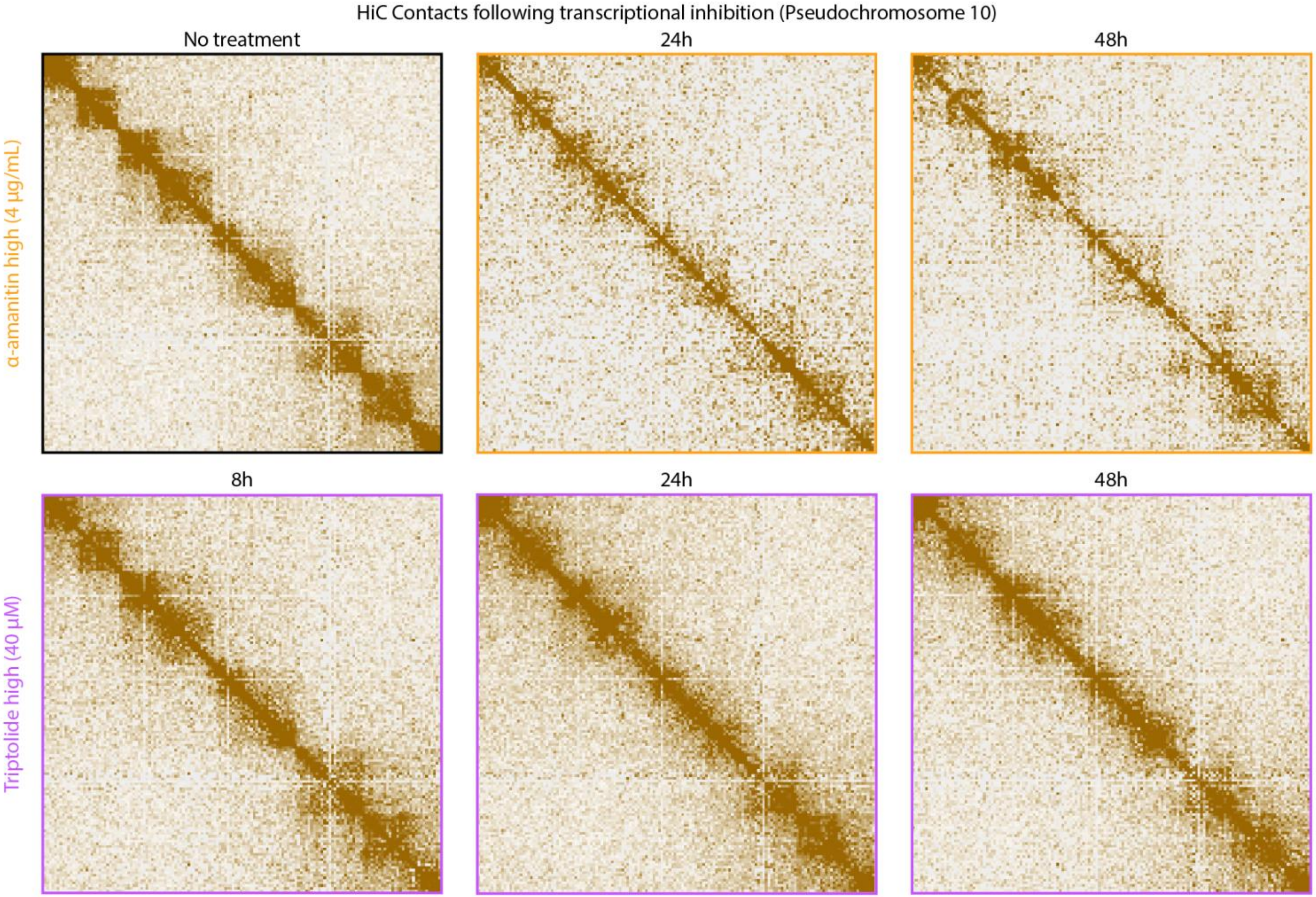
α-amanitin: low (1 μg/mL) and high (4 μg/mL)

0h, 16h, 24h, 48h

Triptolide: low (10 μM) and high (40 μM)

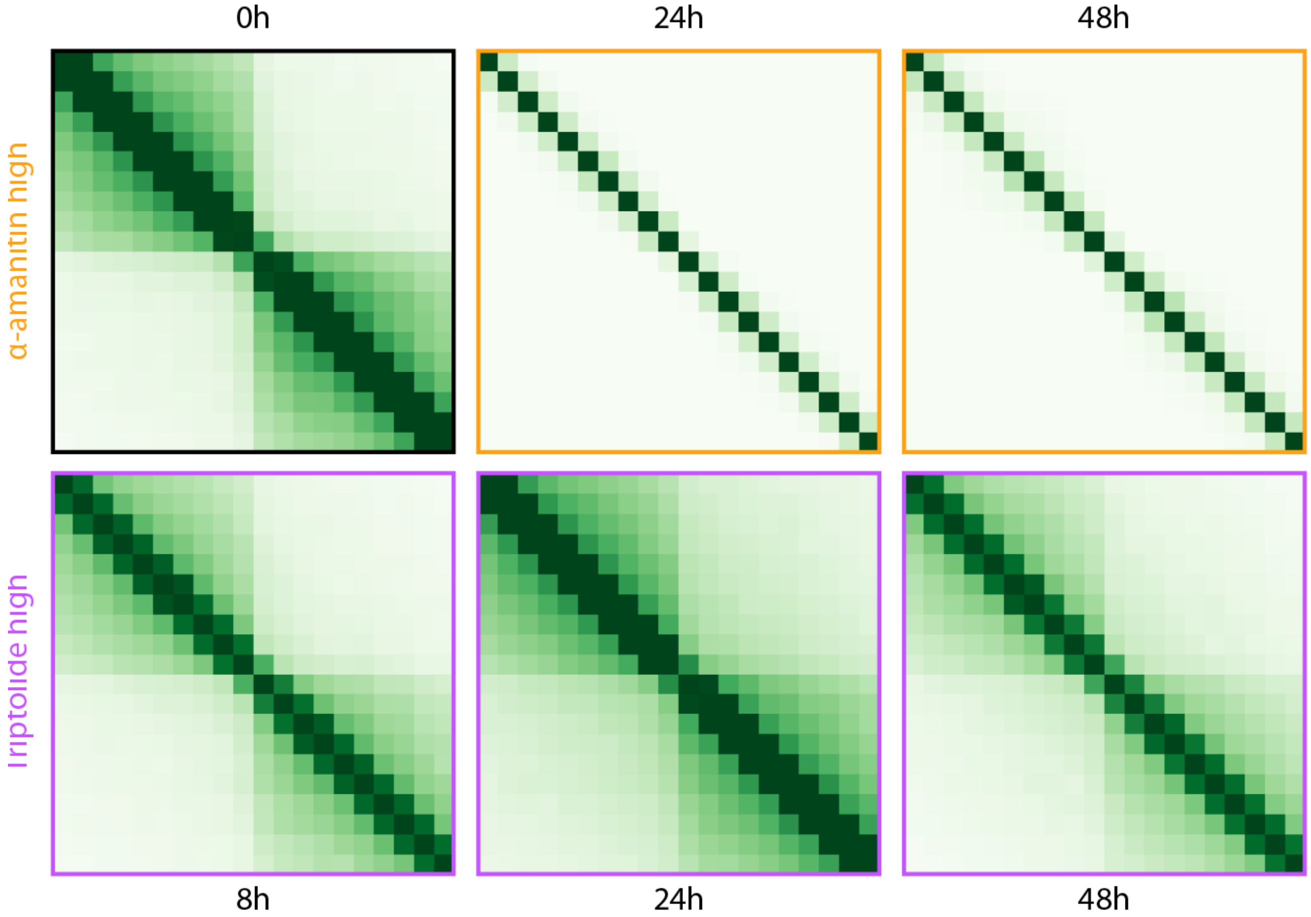
0h, 8h, 24h, 48h

TRANSCRIPTION INHIBITION EXPERIMENTS



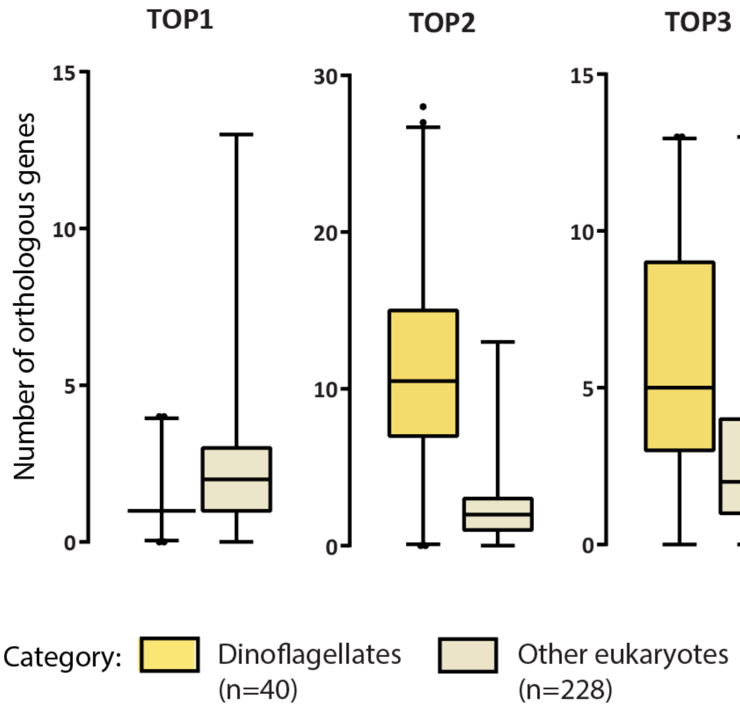
TRANSCRIPTION INHIBITION EXPERIMENTS

Aggregate Hi-C signal across all dinoTADs



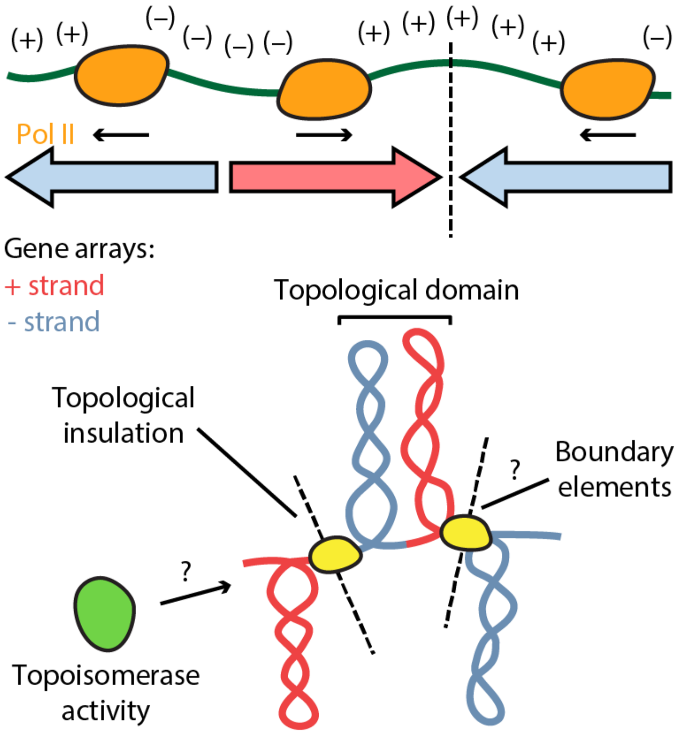
PHYLOGENOMIC EVIDENCE FOR ELEVATED TOPOLOGICAL STRESS

Topoisomerase gene amplification
in dinoflagellates



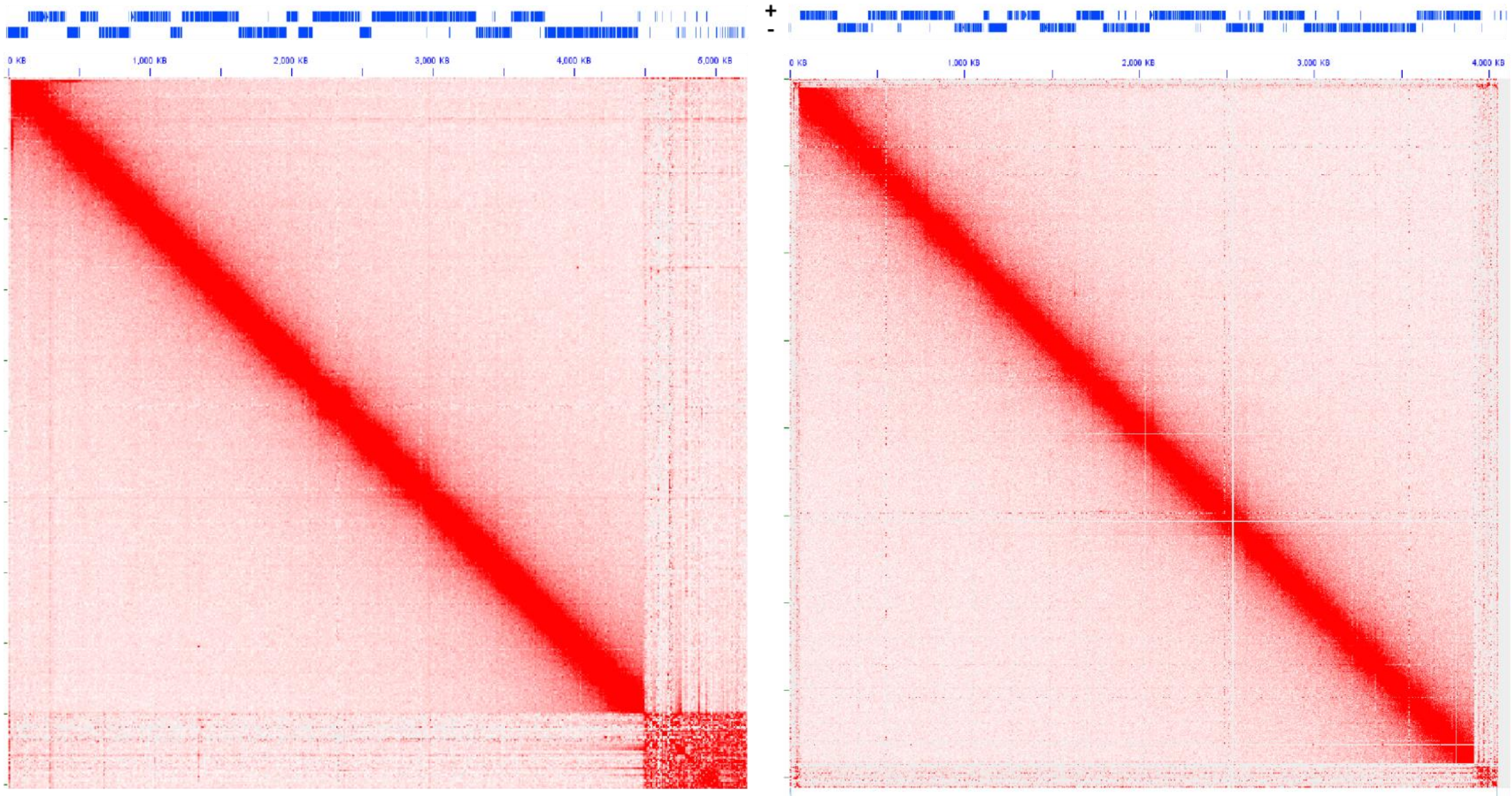
MODEL

Model of transcription-induced plectonemes and domain formation



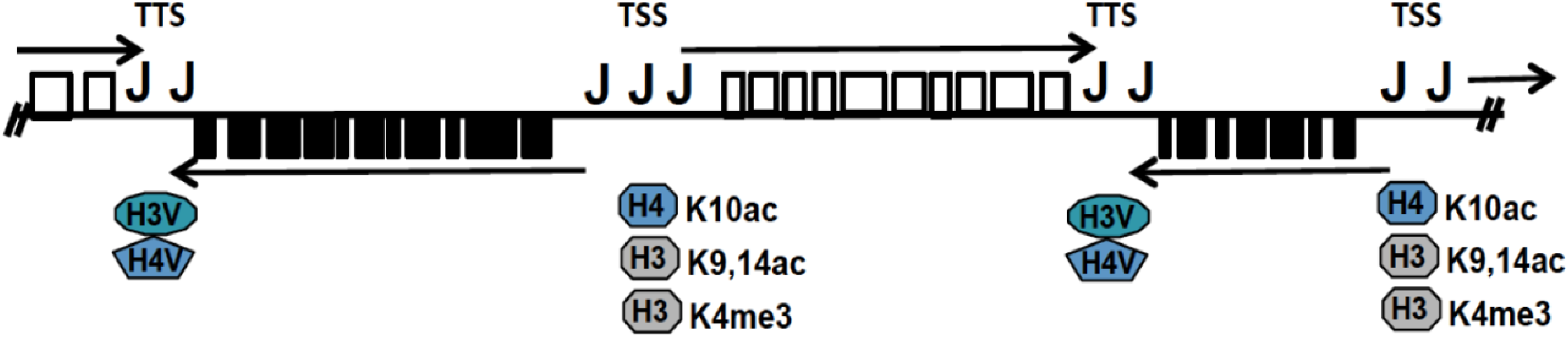
NO TADs IN KINETOPLASTIDS

- Kinetoplastids have a similar gene array organization but have not lost histones



WHAT ABOUT 5-HMU?

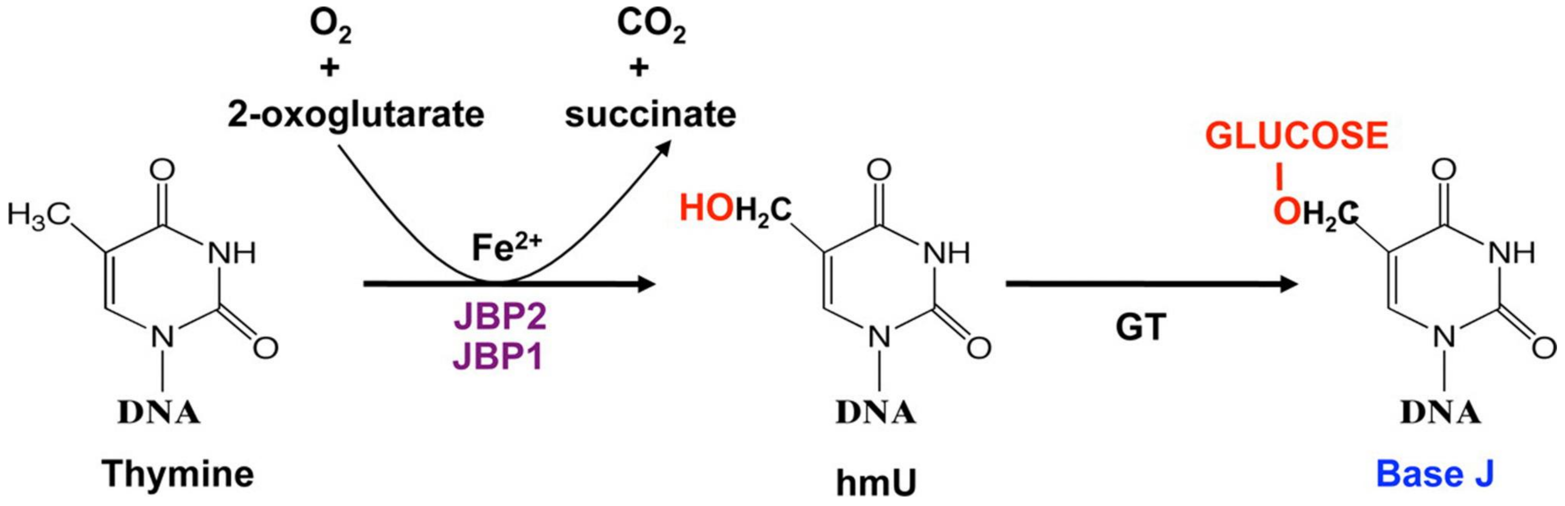
- In kinetoplastids, hmU is an intermediate in the synthesis of Base J
- and Base J is found at boundaries between gene arrays



Reynolds at el. 2016

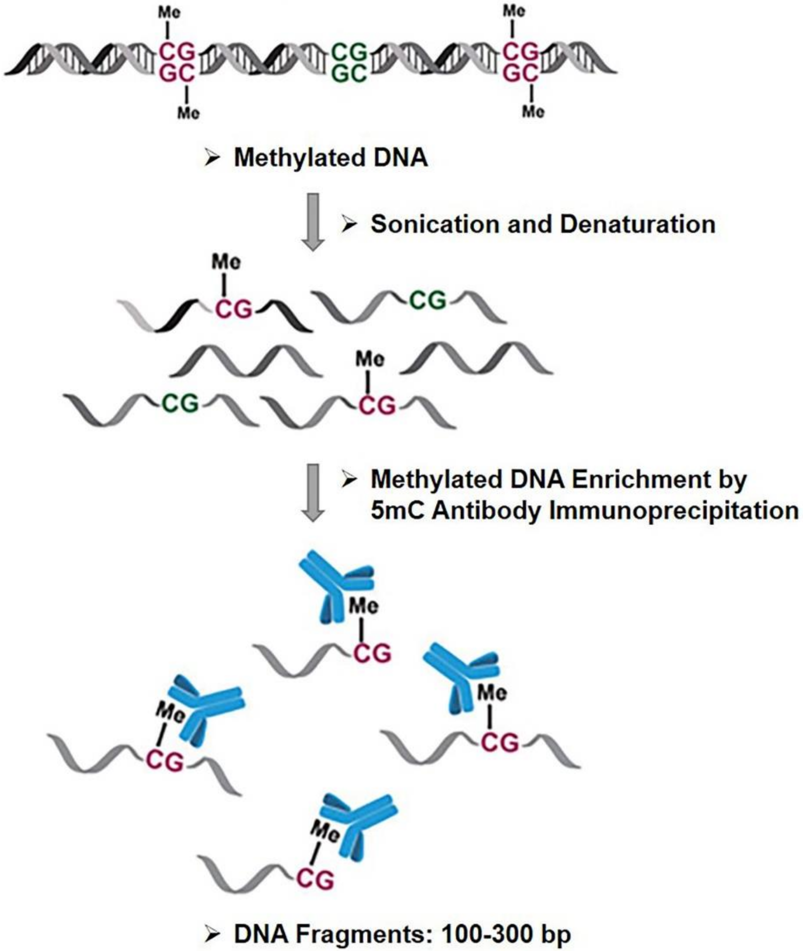
WHAT ABOUT 5-HMU?

- In kinetoplastids, hmU is an intermediate in the synthesis of Base J



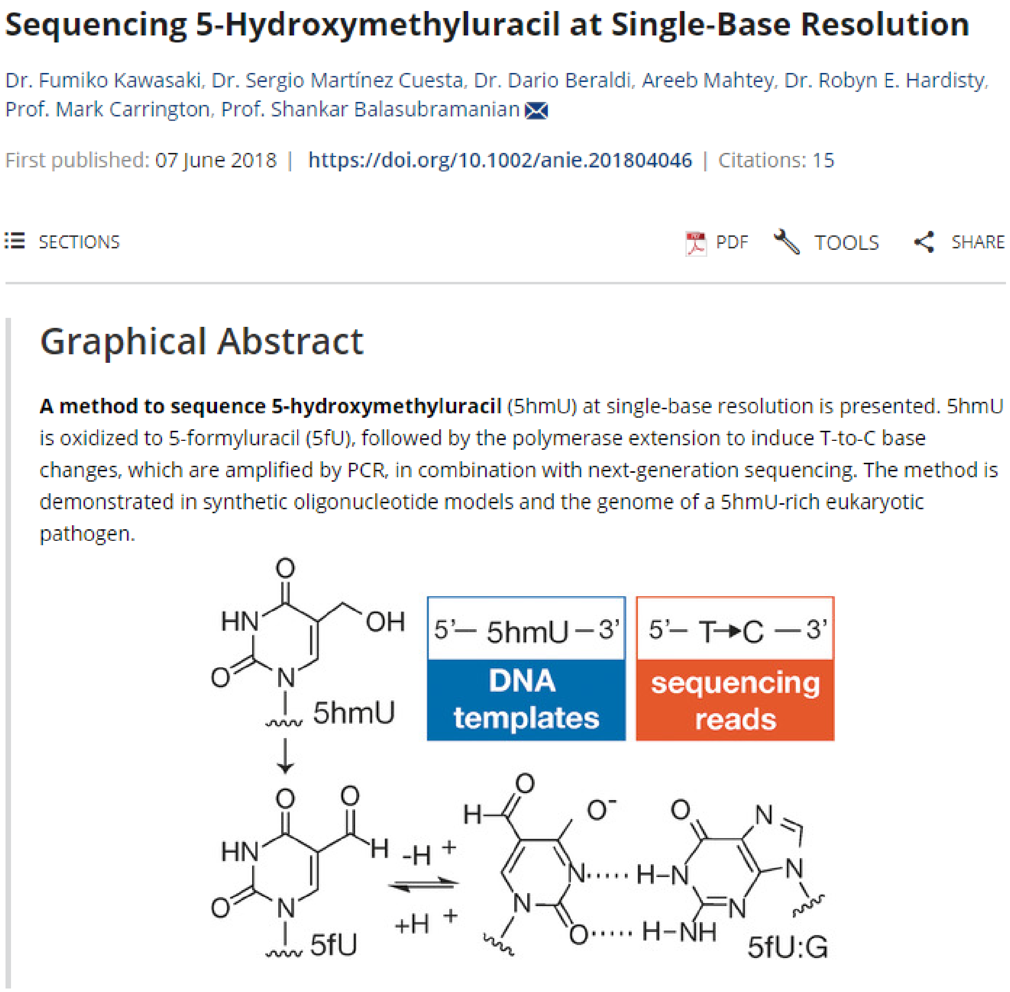
Cliffe at el. 2012

5-HMU MEDIP-SEQ



- ➢ Methylated DNA
- ➢ Sonication and Denaturation
- ➢ Methylated DNA Enrichment by 5mC Antibody Immunoprecipitation
- ➢ DNA Fragments: 100-300 bp

5-HMU CHEMICAL MAPPING



Sequencing 5-Hydroxymethyluracil at Single-Base Resolution

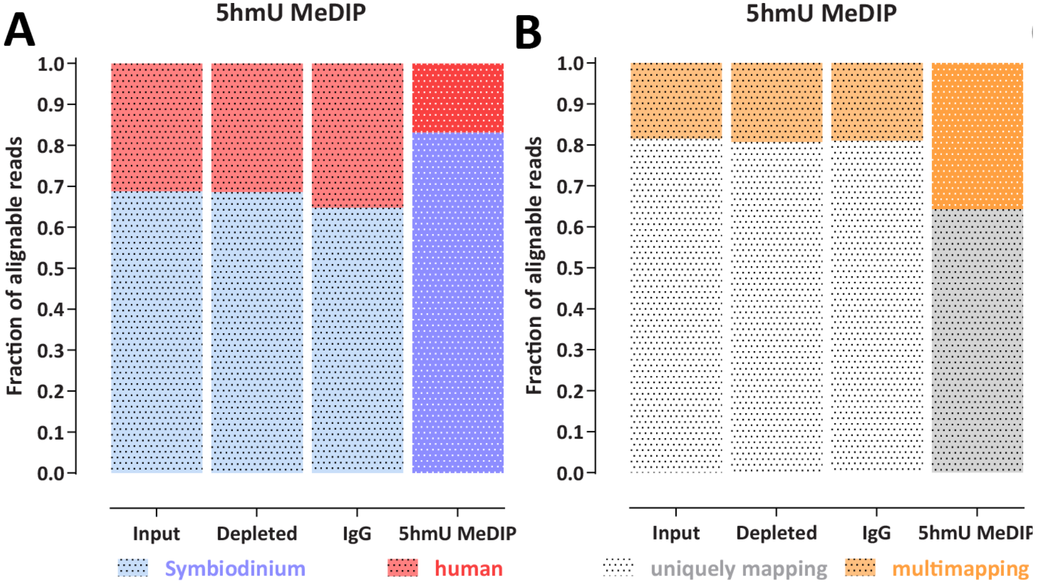
Dr. Fumiko Kawasaki, Dr. Sergio Martínez Cuesta, Dr. Dario Beraldi, Areeb Mahtey, Dr. Robyn E. Hardisty, Prof. Mark Carrington, Prof. Shankar Balasubramanian

First published: 07 June 2018 | https://doi.org/10.1002/anie.201804046 | Citations: 15

Graphical Abstract

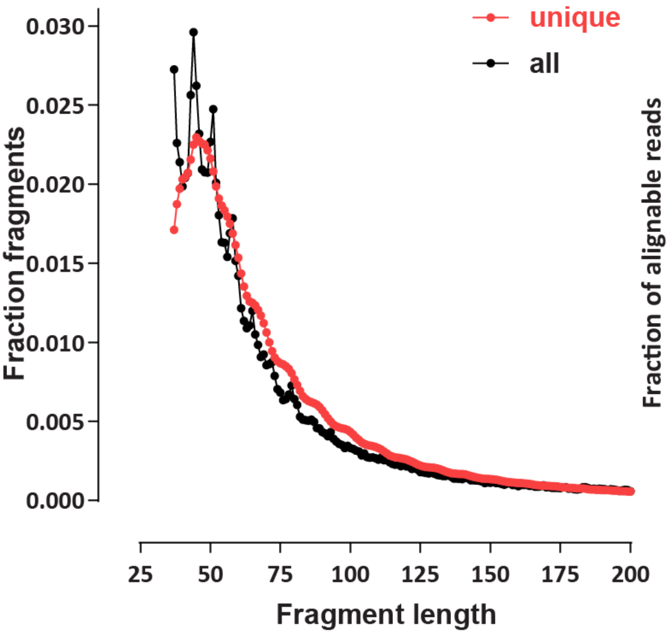
A method to sequence 5-hydroxymethyluracil (5hmU) at single-base resolution is presented. 5hmU is oxidized to 5-formyluracil (5fU), followed by the polymerase extension to induce T-to-C base changes, which are amplified by PCR, in combination with next-generation sequencing. The method is demonstrated in synthetic oligonucleotide models and the genome of a 5hmU-rich eukaryotic pathogen.

5-HMU IS ENRICHED OVER REPEATS

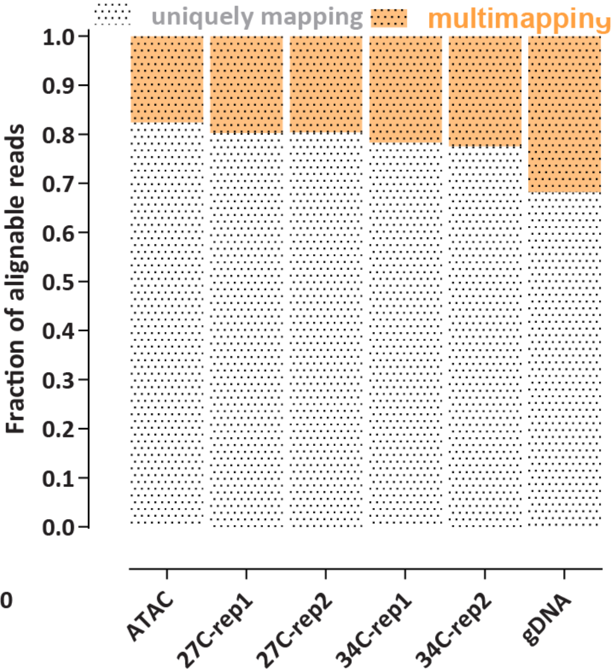


ATAC-SEQ AND GENOME-WIDE CHROMATIN ACCESSIBILITY

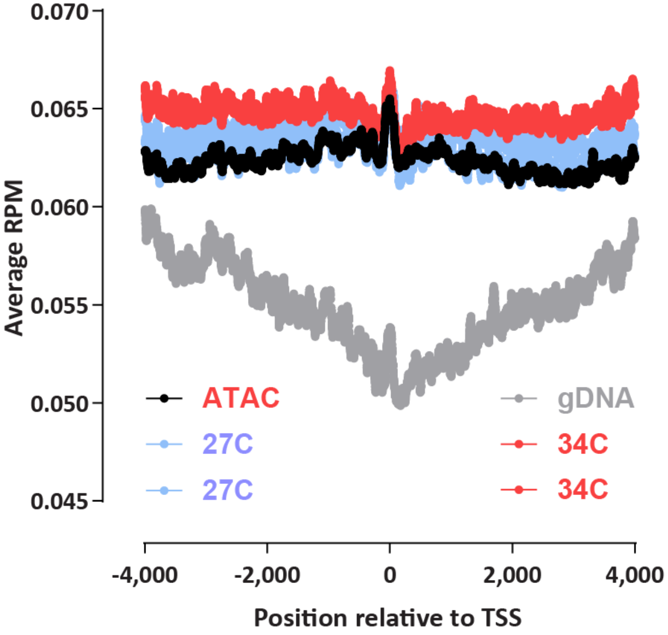
ATAC-seq



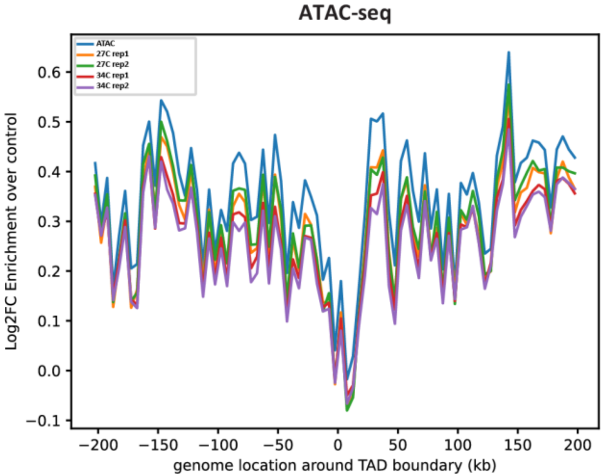
ATAC-seq

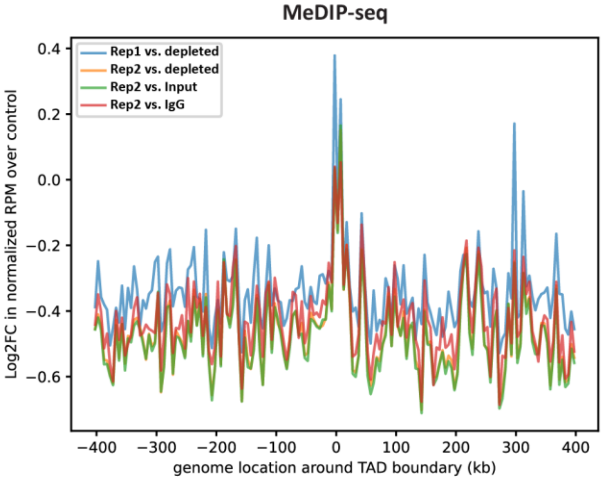


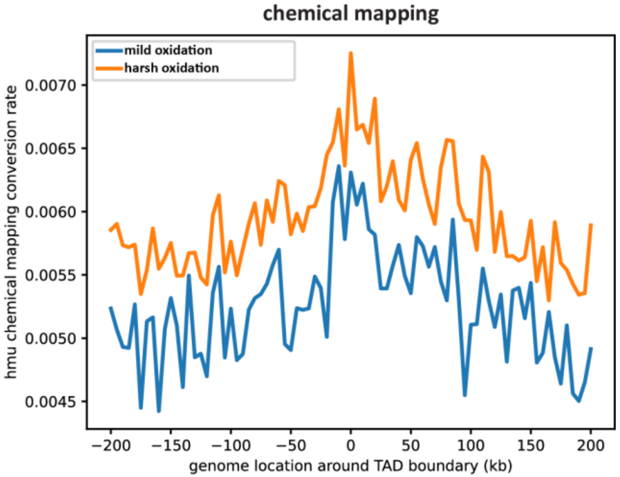
ATAC-seq



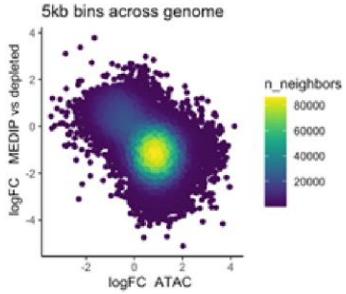
TAD BOUNDARIES AND 5-HMU/ATAC

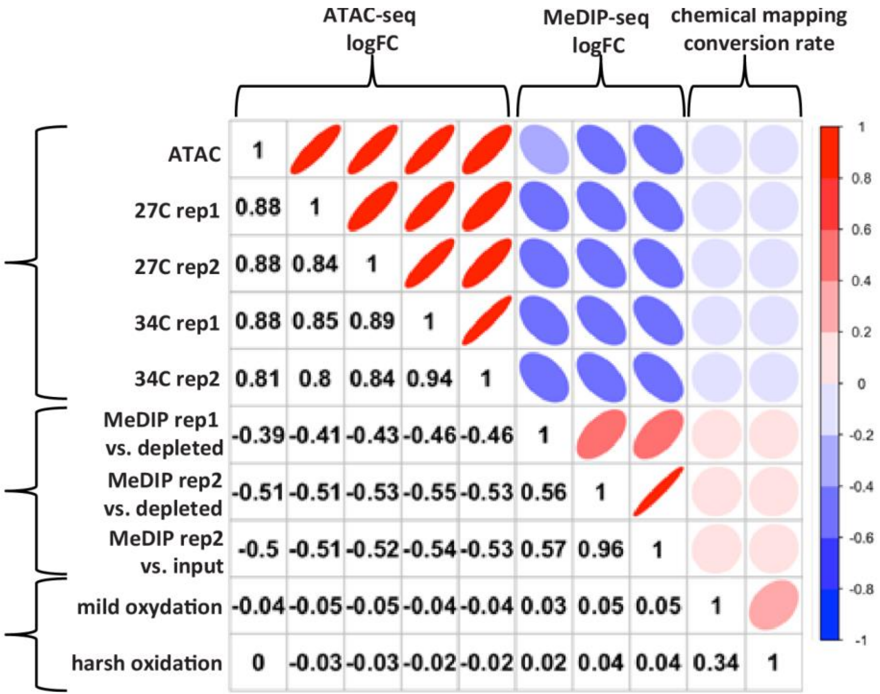

ATAC-seq


MeDIP-seq


chemical mapping

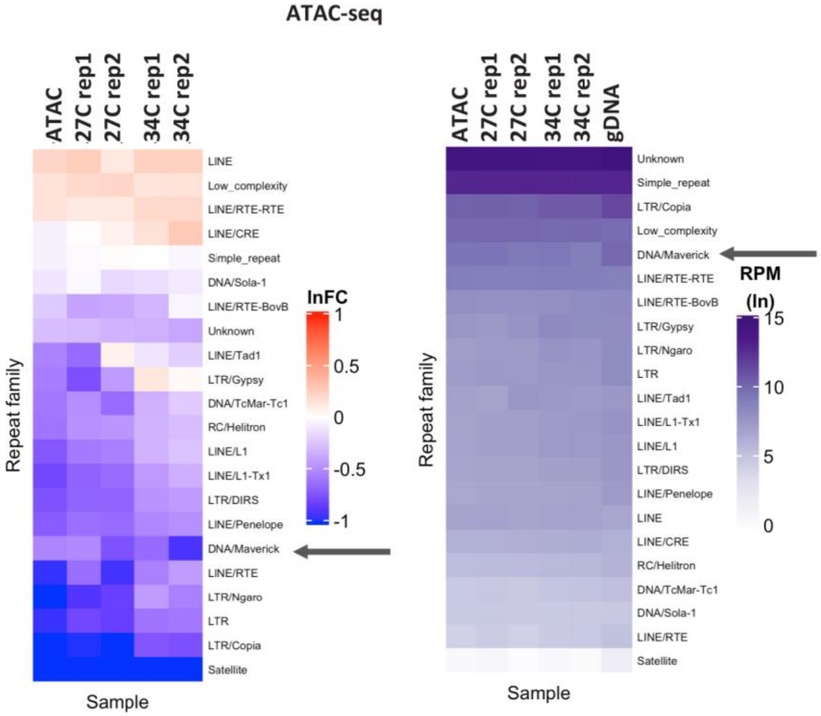
ATAC VS 5-HMU



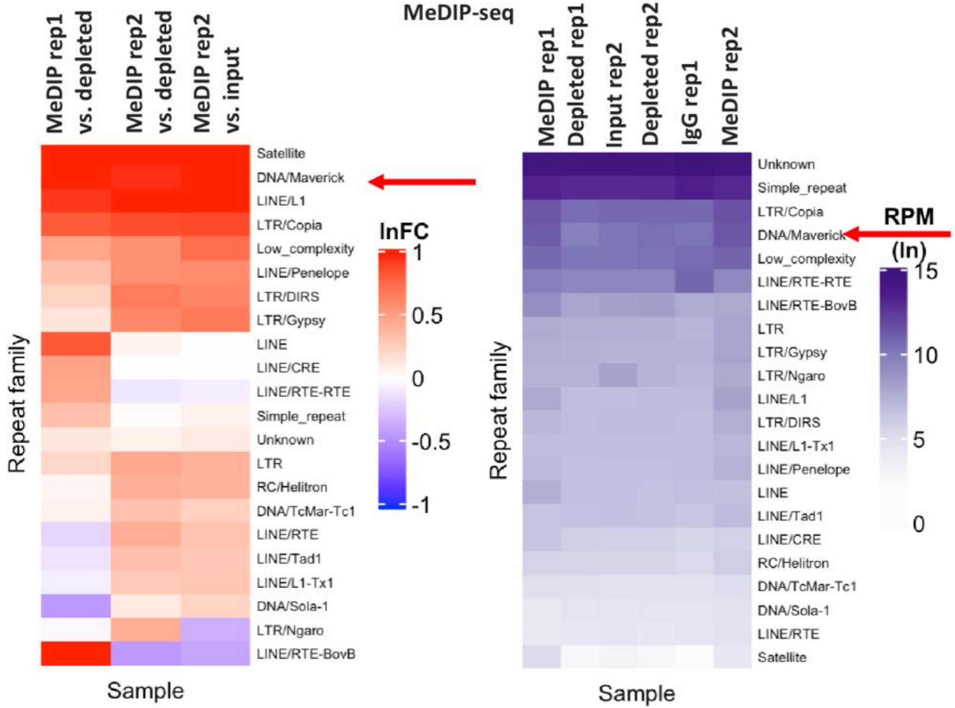


ATAC	1									
27C rep1	0.88	1								
27C rep2	0.88	0.84	1							
34C rep1	0.88	0.85	0.89	1						
34C rep2	0.81	0.8	0.84	0.94	1					
MeDIP rep1 vs. depleted	-0.39	-0.41	-0.43	-0.46	-0.46	1				
MeDIP rep2 vs. depleted	-0.51	-0.51	-0.53	-0.55	-0.53	0.56	1			
MeDIP rep2 vs. input	-0.5	-0.51	-0.52	-0.54	-0.53	0.57	0.96	1		
mild oxydation	-0.04	-0.05	-0.05	-0.04	-0.04	0.03	0.05	0.05	1	
harsh oxidation	0	-0.03	-0.03	-0.02	-0.02	0.02	0.04	0.04	0.34	1

ATAC, 5-HMU AND REPEATS



ATAC-seq



MeDIP-seq

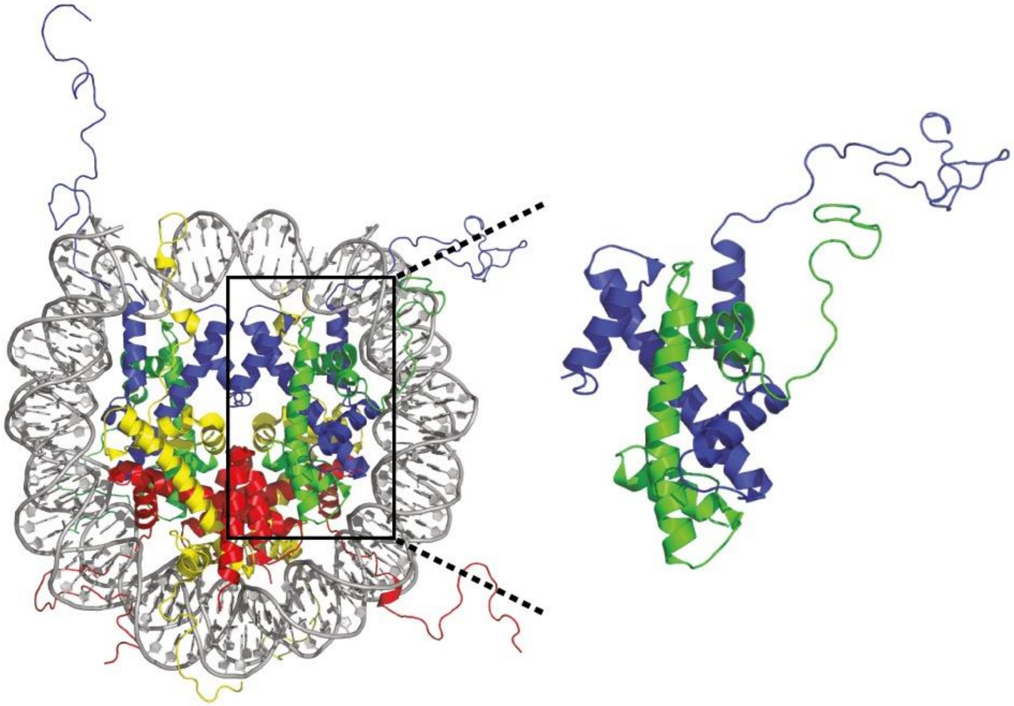
SUMMARY AND CONCLUSIONS

- 3D genome organization in dinoflagellates is novel and unique to eukaryotes
- Large and very strong topological domains are formed
- Driven by transcription-induced supercoiling
- Absence of histones plus the long gene arrays likely key to allowing supercoiling to manifest itself
- The genome appears mostly ``naked’’
- Repeats show decreased accessibility
- 5-hmU is enriched over repeats
- Possible role for 5-hmU and/or repeats around dinoTAD/array boundaries

III. ARCHAEAL CHROMATIN

ARCHAEAL HISTONES

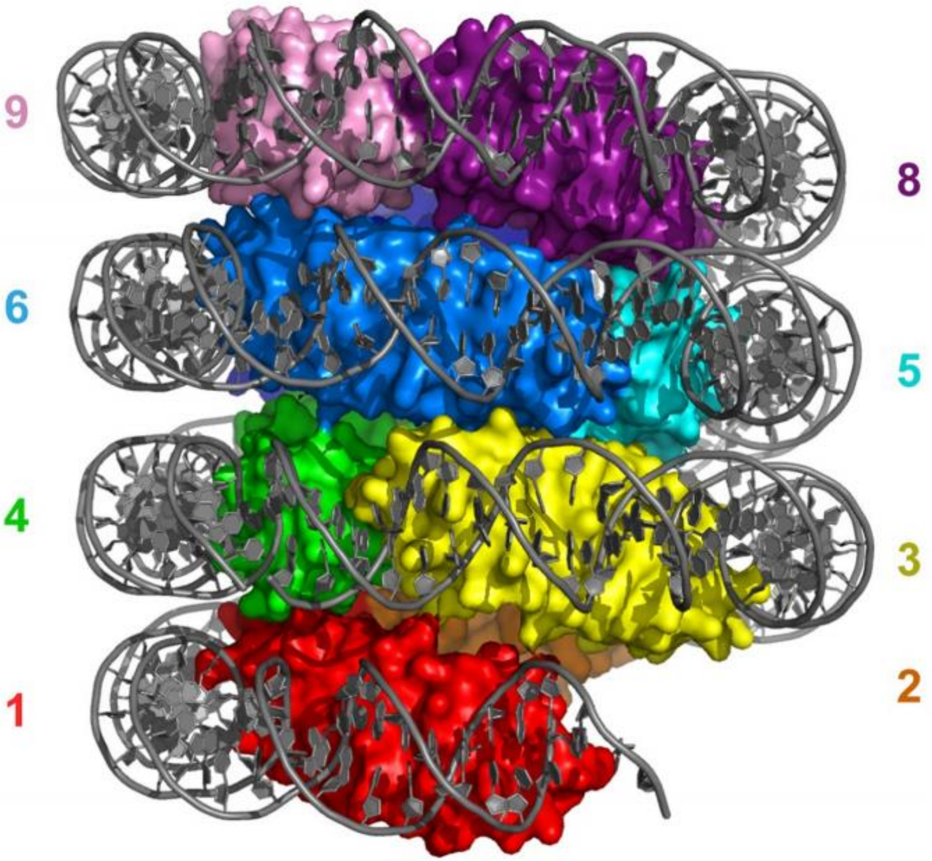
- histone fold with no long tails



eukaryote nucleosome and histones

archaeal histones

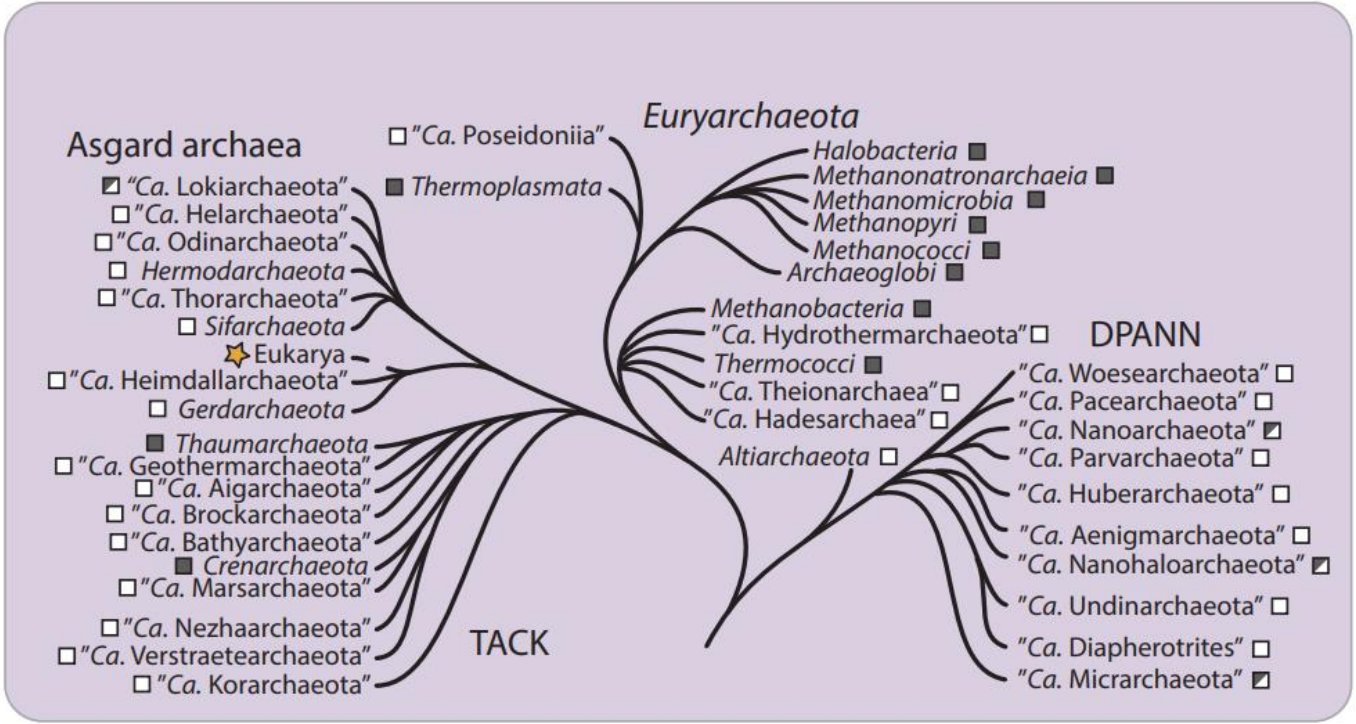
HYPERNUCLEOSOMES



Henneman et al. 2021

Histones missing in:

- Geothermarchaeota
- Parvarchaeota
- some Crenarchaeota
- a few other lineages



Tahon et al. 2021

Marinov *et al. Genome Biology* (2023) 24:253
https://doi.org/10.1186/s13059-023-03095-5

Genome Biology

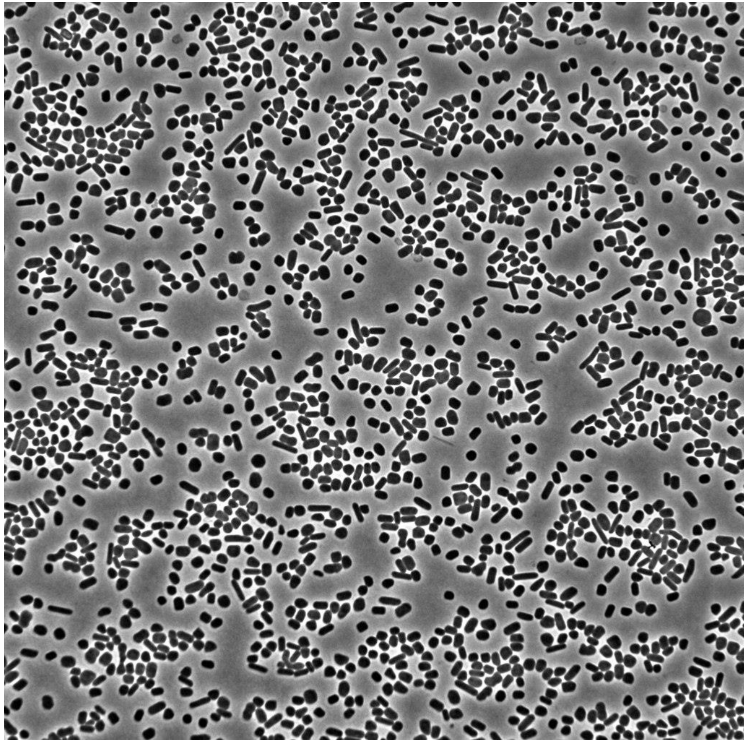
RESEARCH **Open Access**

The chromatin landscape of the euryarchaeon *Haloferax volcanii*

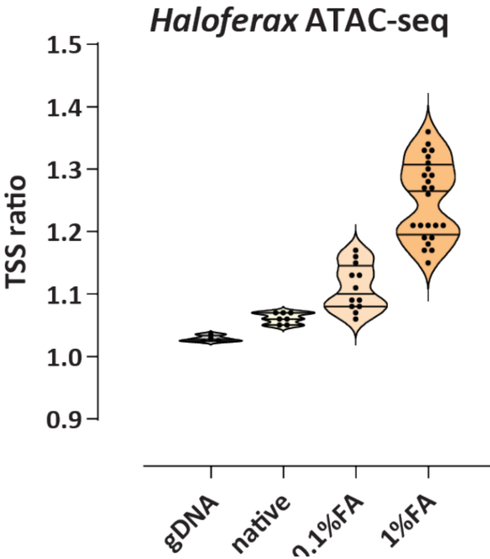
Georgi K. Marinov[1*], S. Tansu Bagdatli[1], Tong Wu[2], Chuan He[2,3,4], Anshul Kundaje[1,5] and William J. Greenleaf[1,6,7,8]

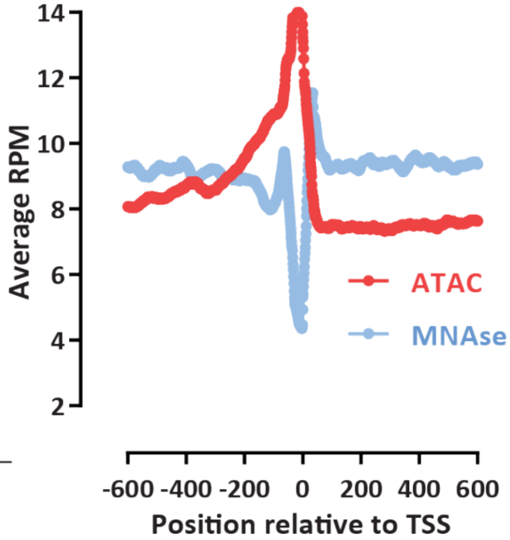
HALOFERAX VOLCANII

- Euryarchaeote
- Possesses histones, but with some doubts about their packaging role
- Genome consists of a main chromosome and five plasmids
- Extremely halophilic – grows in 25% salt

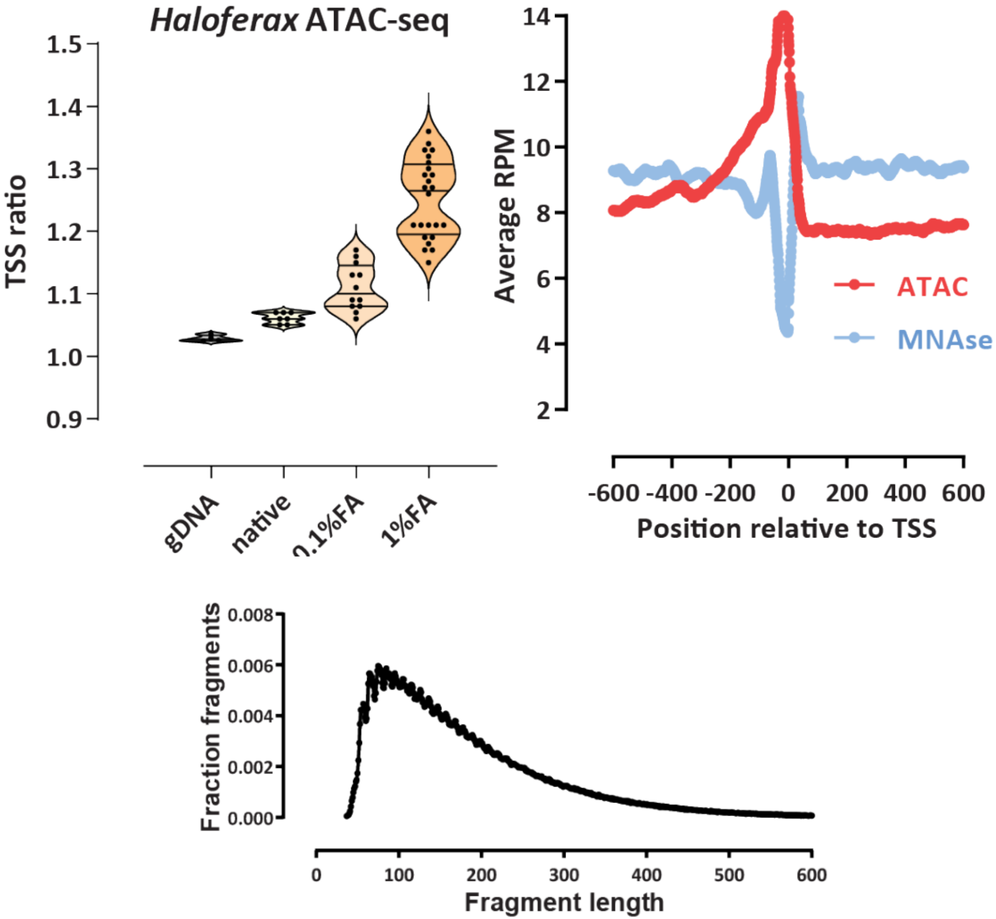


OPTIMIZING ATAC-SEQ FOR ARCHAEA





OPTIMIZING ATAC-SEQ FOR ARCHAEA



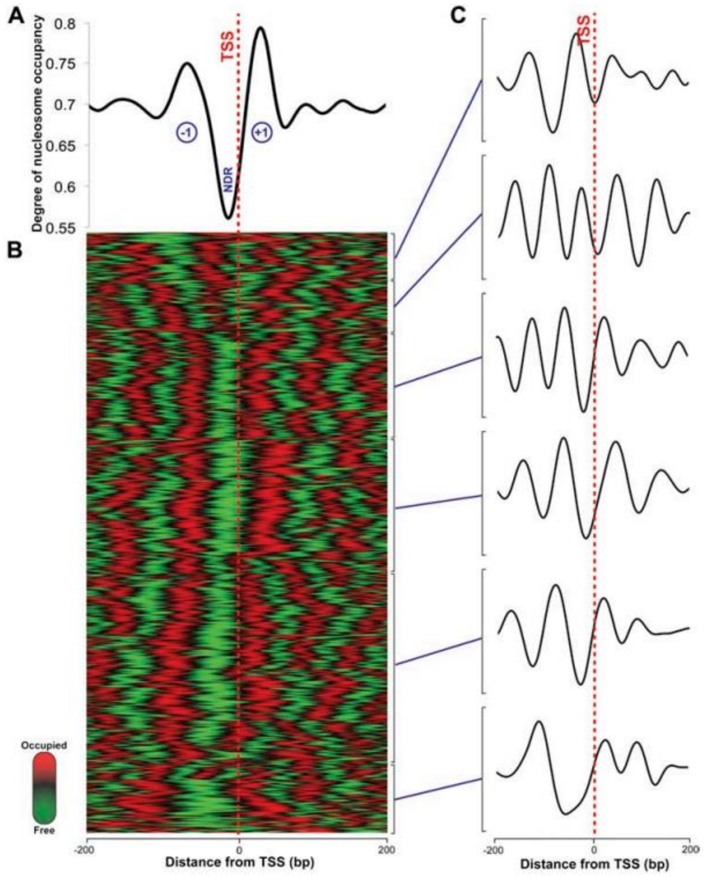
2013 MNASE-SEQ STUDY

eLIFE
elife.elifesciences.org

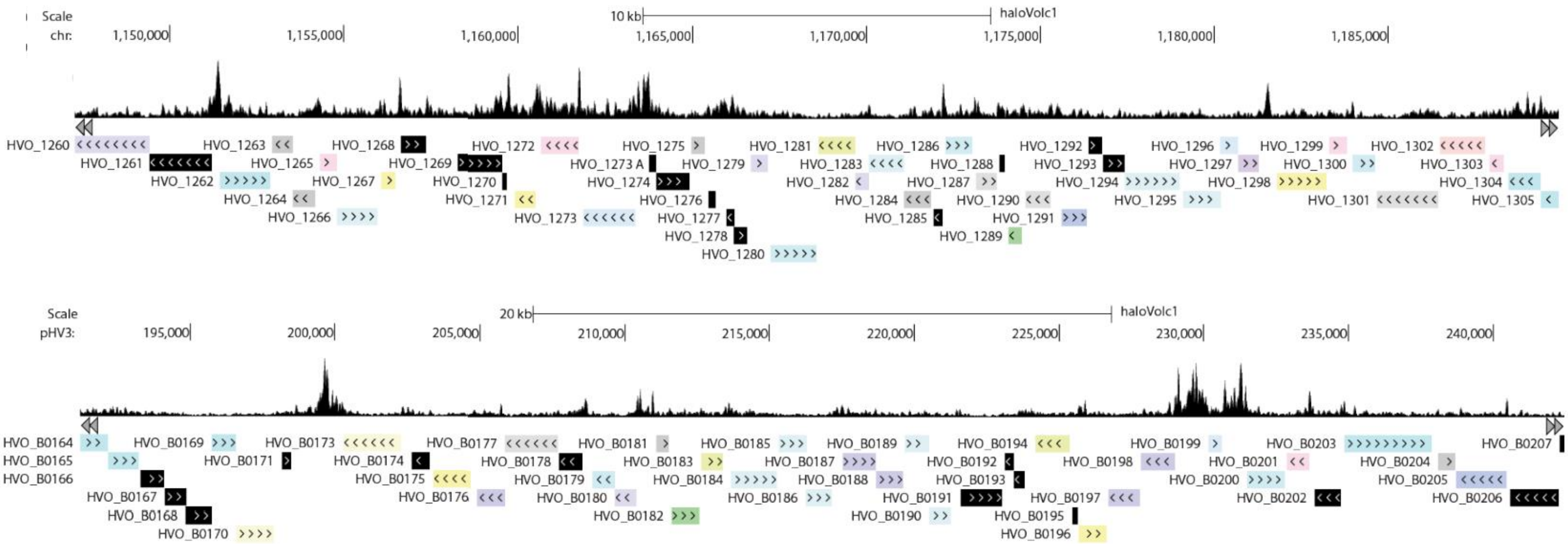
RESEARCH ARTICLE

Chromatin is an ancient innovation conserved between Archaea and Eukarya

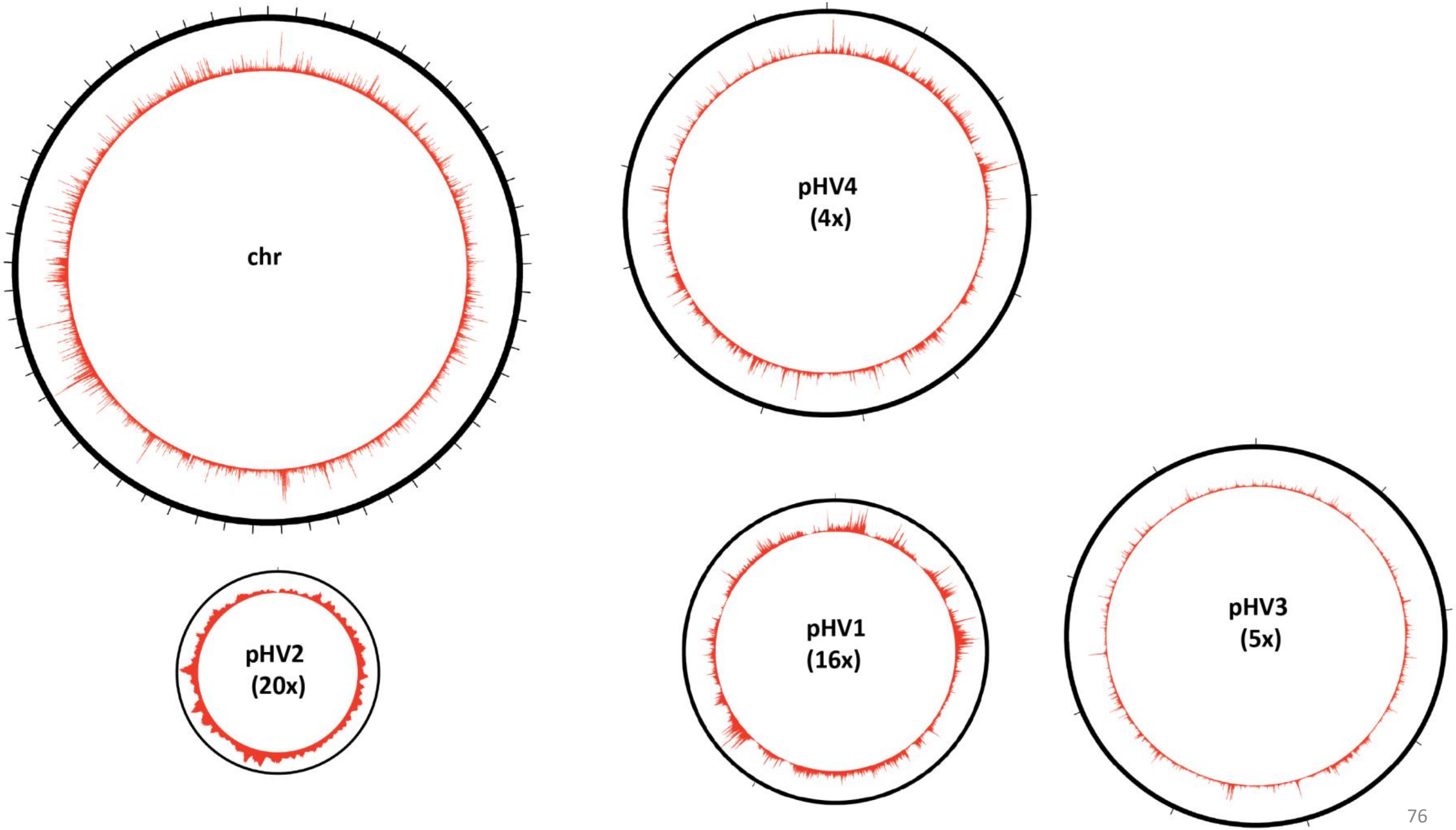
Ron Ammar[1,2], Dax Torti[2], Kyle Tsui[1,3], Marinella Gebbia[1,2], Tanja Durbic[2], Gary D Bader[1,2], Guri Giaever[1,3], Corey Nislow[1,2]*



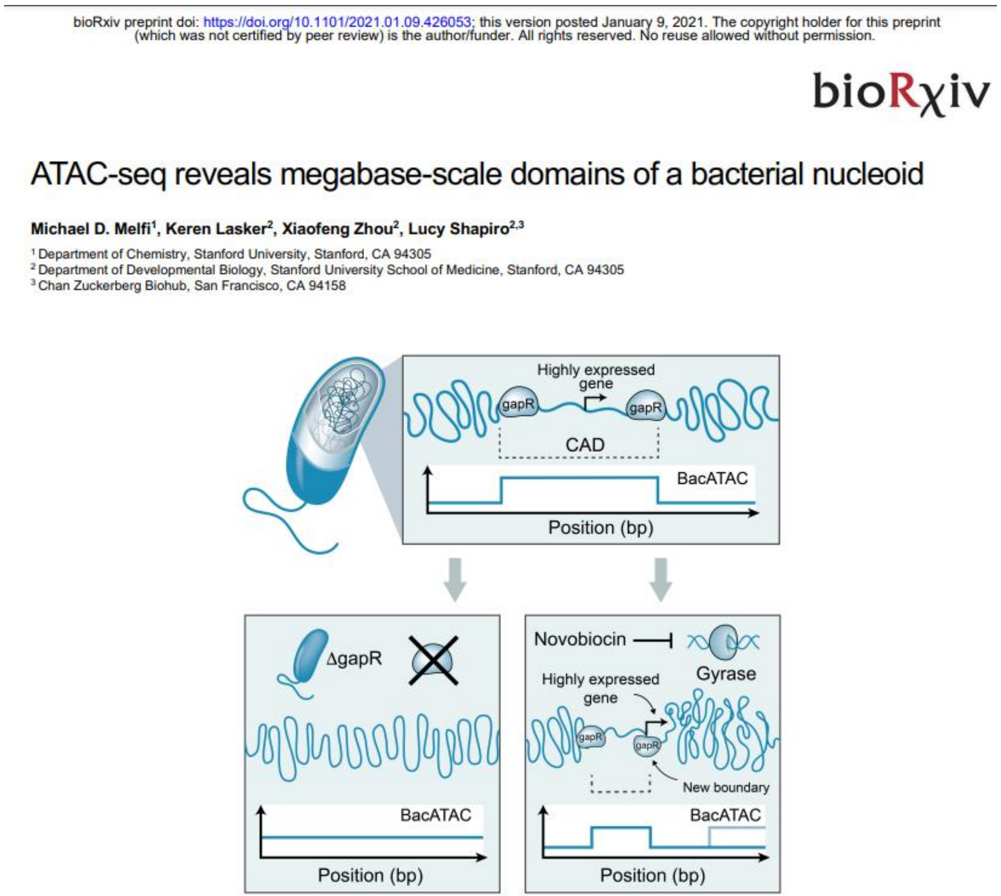
YEAST-LIKE GLOBAL LANDSCAPE



ATAC



NO LARGE DOMAINS AS SEEN IN BACTERIA

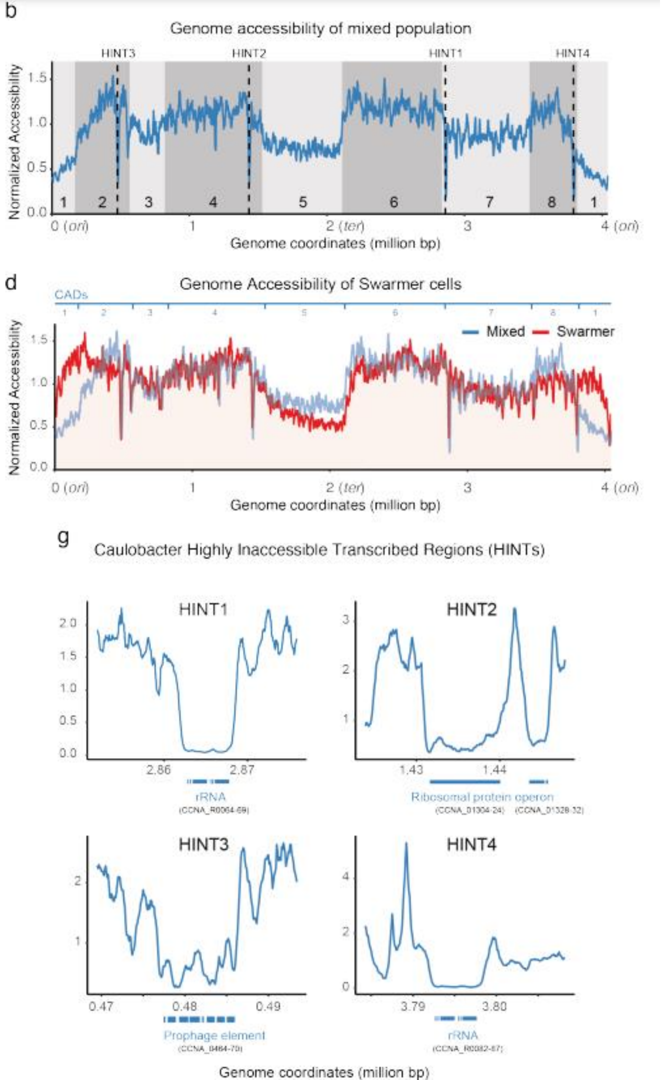


bioRχiv

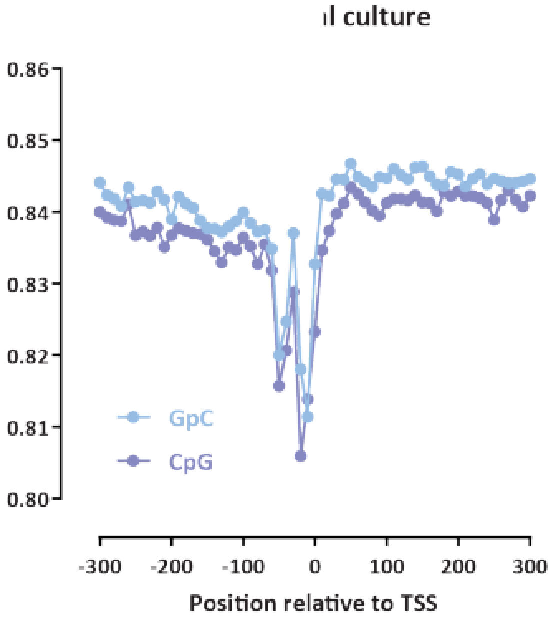
ATAC-seq reveals megabase-scale domains of a bacterial nucleoid

Michael D. Melfi[1], Keren Lasker[2], Xiaofeng Zhou[2], Lucy Shapiro[2,3]

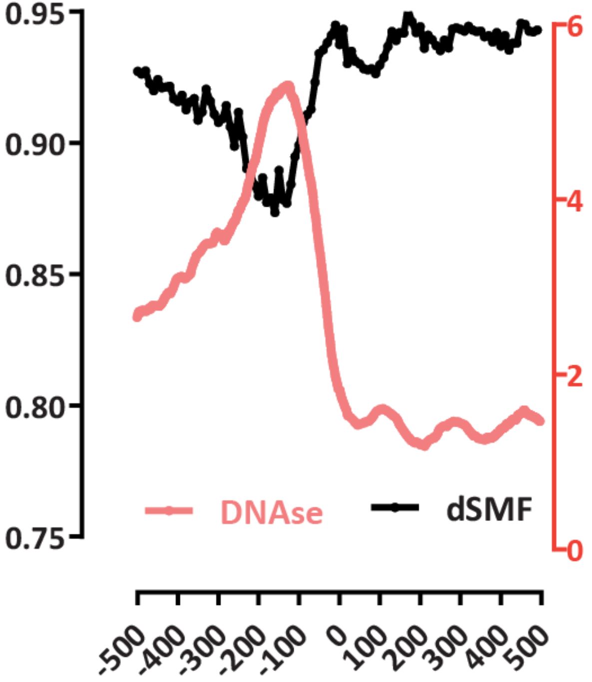
[1] Department of Chemistry, Stanford University, Stanford, CA 94305
[2] Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305
[3] Chan Zuckerberg Biohub, San Francisco, CA 94158



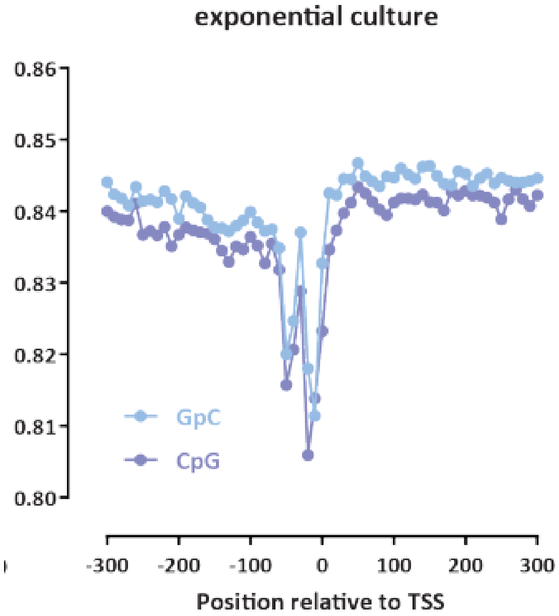
SINGLE MOLECULE MAPPING OF ABSOLUTE OCCUPANCY/ACCESSIBILITY

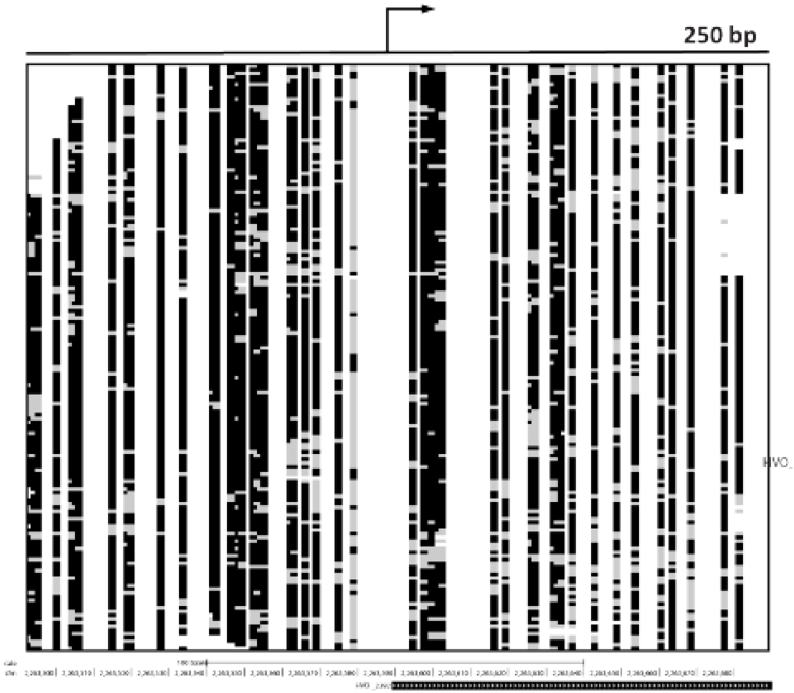


S. cerevisiae for comparison

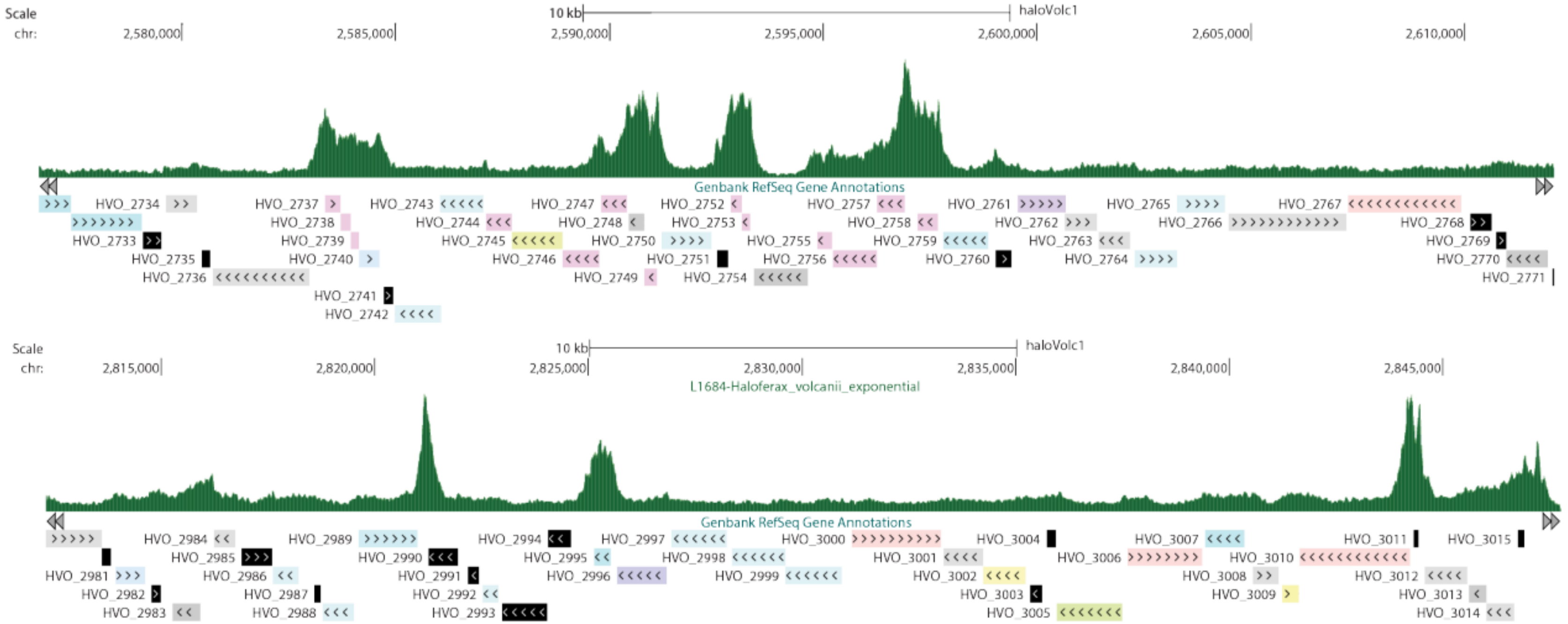


SINGLE MOLECULE MAPPING OF ABSOLUTE OCCUPANCY/ACCESSIBILITY



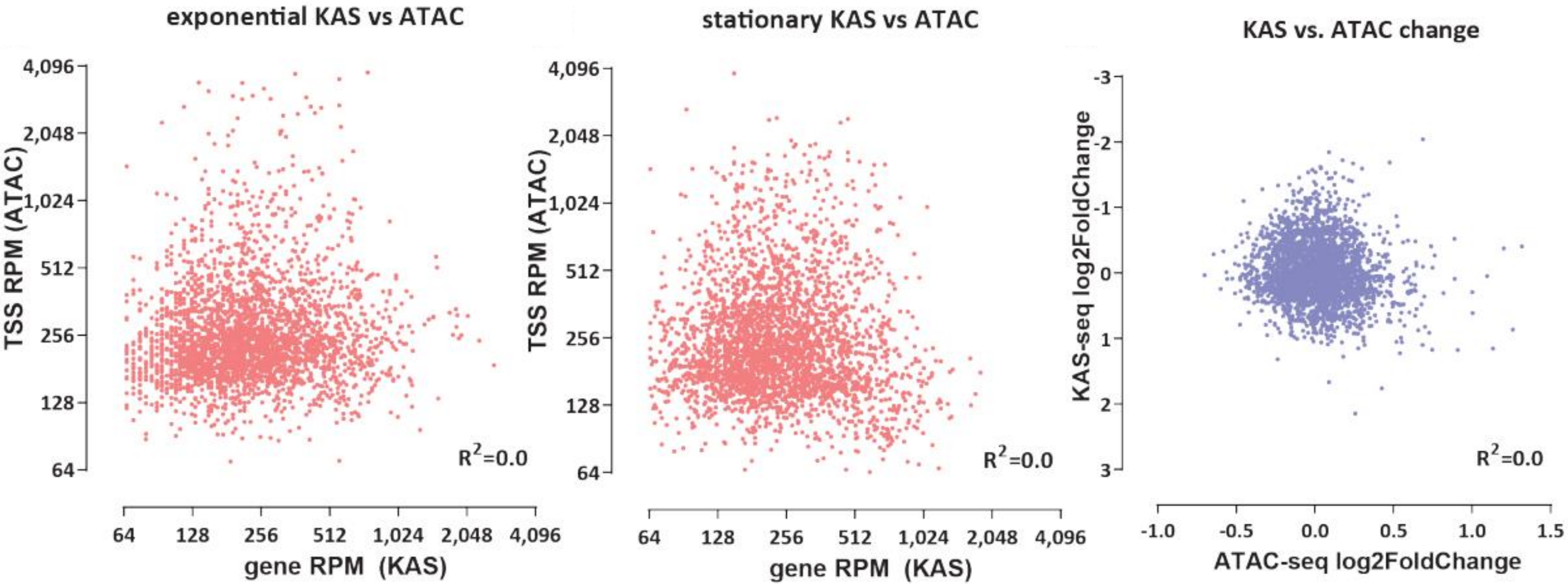


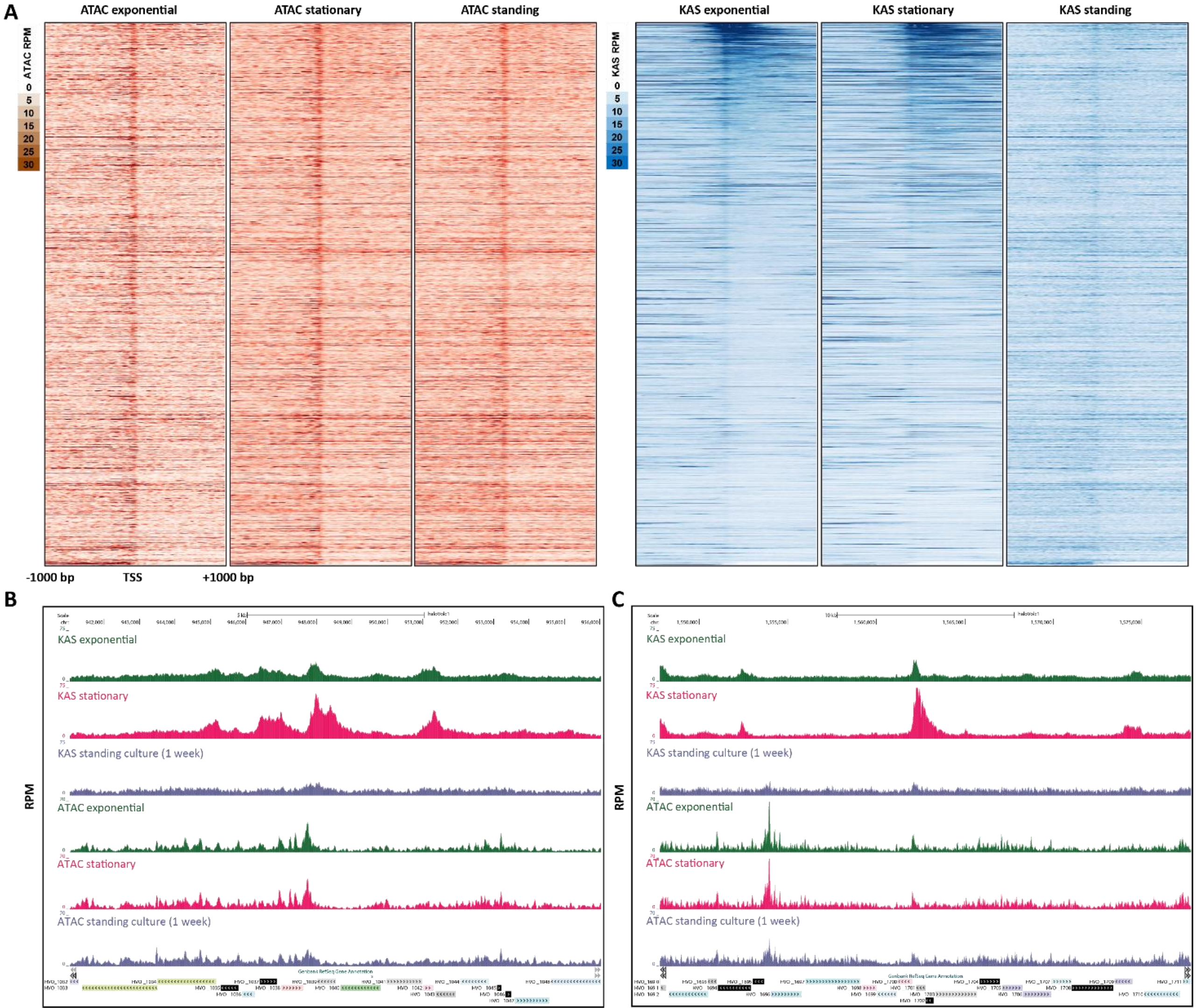
ACTIVE TRANSCRIPTION MAPPING (KAS-SEQ)



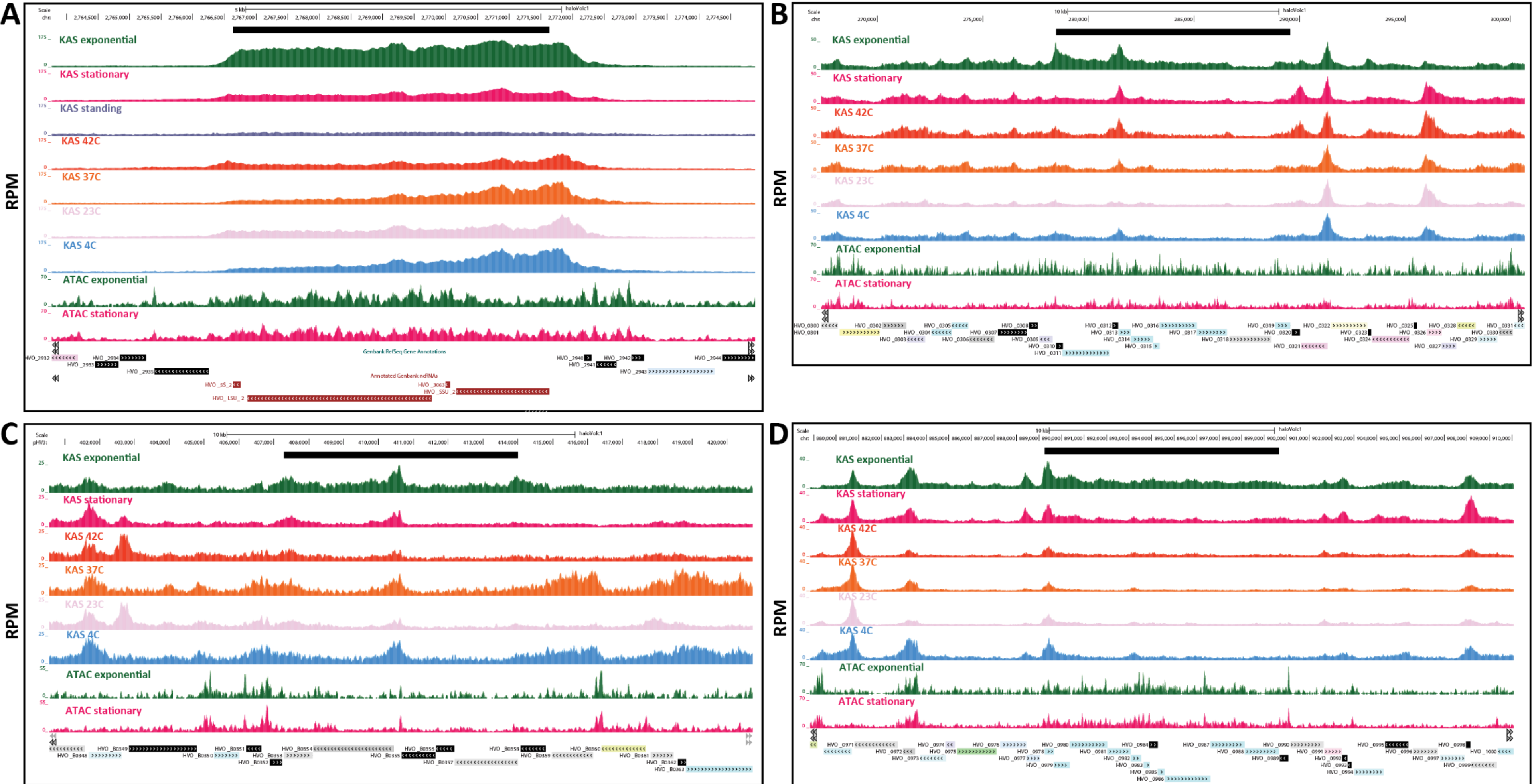
promoter pausing observed for most genes

NO CORRELATION BETWEEN ACCESSIBILITY AND TRANSCRIPTION




A
ATAC exponential
ATAC stationary
ATAC standing
KAS exponential
KAS stationary
KAS standing
ATAC RPM
0
5
10
15
20
25
30
KAS RPM
0
5
10
15
20
25
30
-1000 bp
TSS
+1000 bp
B
RPM
KAS exponential
KAS stationary
KAS standing culture (1 week)
ATAC exponential
ATAC stationary
ATAC standing culture (1 week)
C
RPM
KAS exponential
KAS stationary
KAS standing culture (1 week)
ATAC exponential
ATAC stationary
ATAC standing culture (1 week)

INDEPENDENT REGULATION OF GENES INSIDE OPERONS



SUMMARY

- *Haloferax* exhibits eukaryote-like chromatin organization
  - similar to eukaryotes with compact genomes
  - high levels of absolute protection of DNA
  - not yet clear which protein confers it

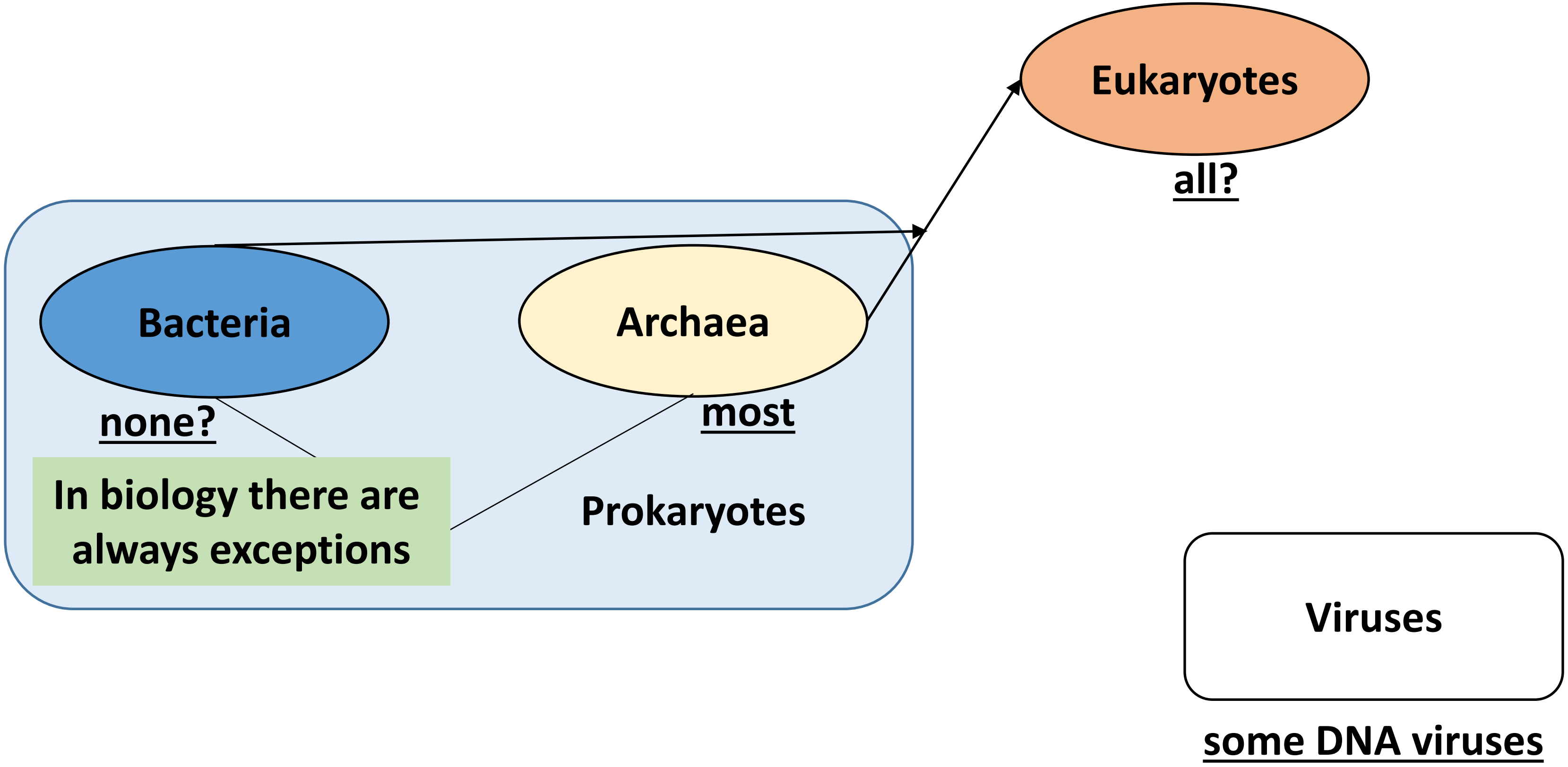
- Promoter pausing is observed
  - suggesting a role for promoter release in expression regulation

- Independent regulation of expression inside operons

- Strikingly, no correlation is observed between transcription and accessibility
  - suggests that promoter opening is not a critical step in gene regulation in *Haloferax*

IV. BACTERIA WITH HISTONES

HISTONES AND CHROMATIN



Eukaryotes

all?

Bacteria

none?

Archaea

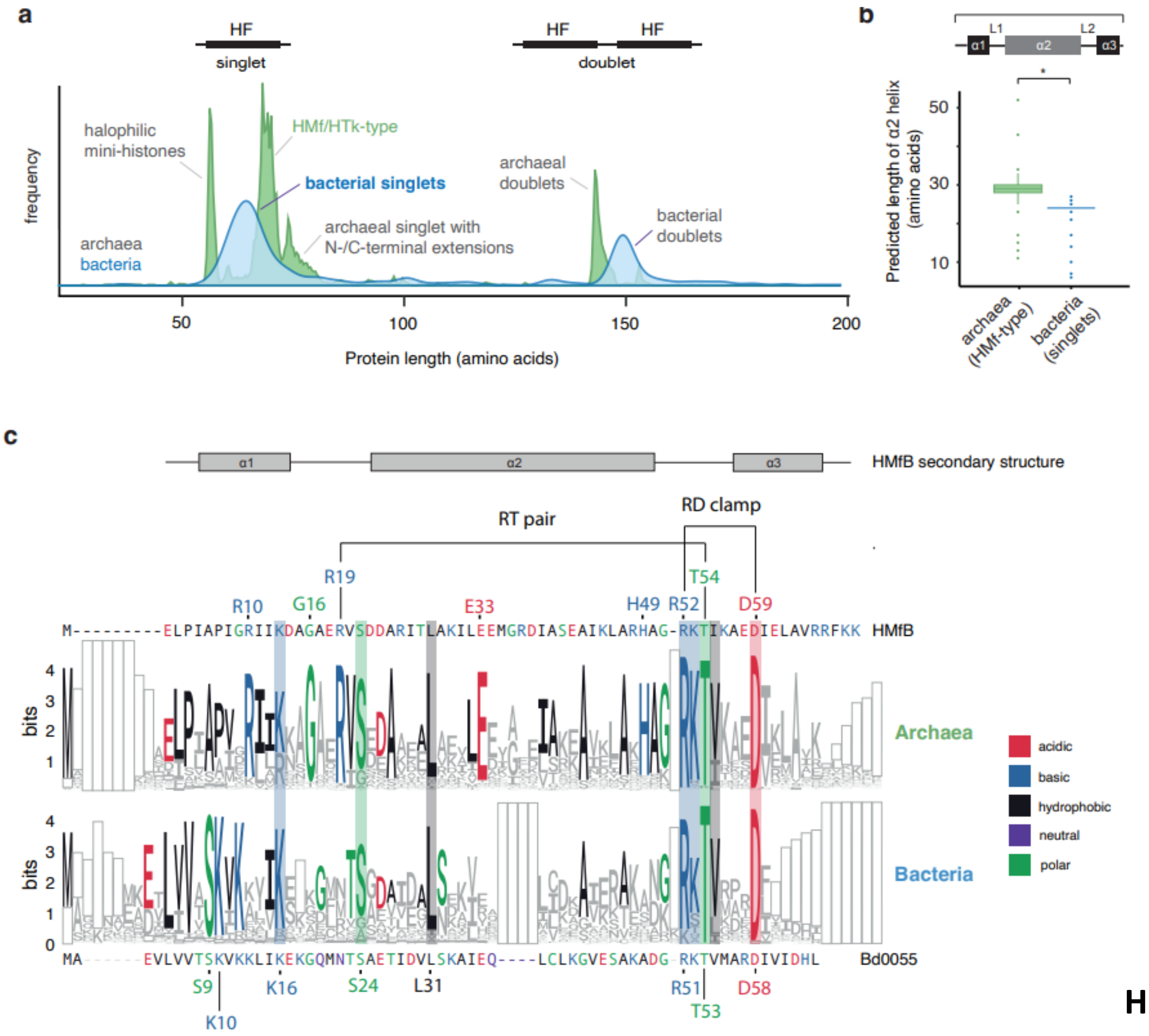
most

In biology there are always exceptions

Prokaryotes

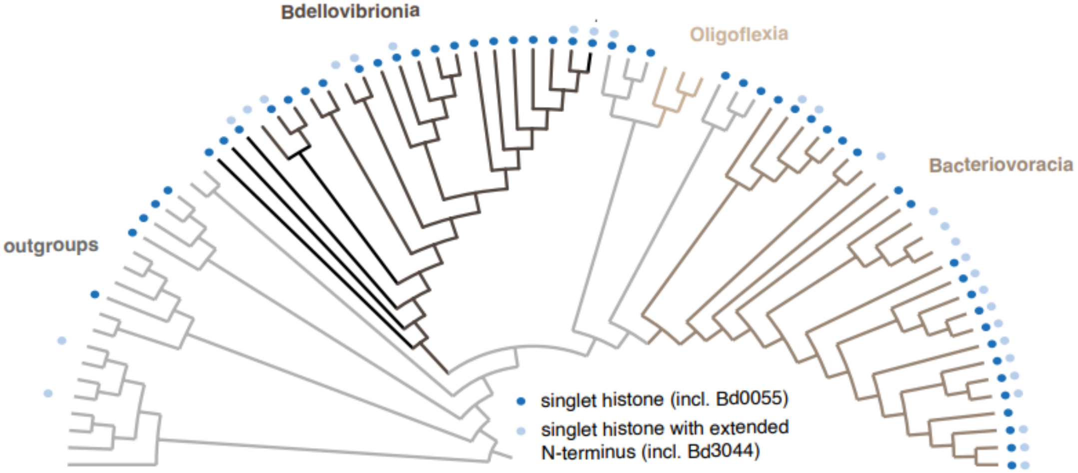
Viruses

some DNA viruses

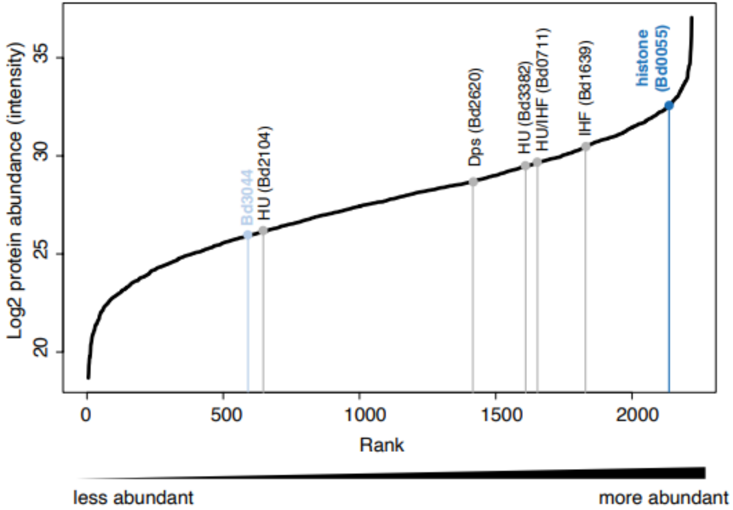


Hocher et al. 2023

a



Cladogram of the Bdellovibrionota and neighbouring clades, GTDB



Hocher et al. 2023

- histones proposed to wrap around DNA

- not DNA around histones as usual!

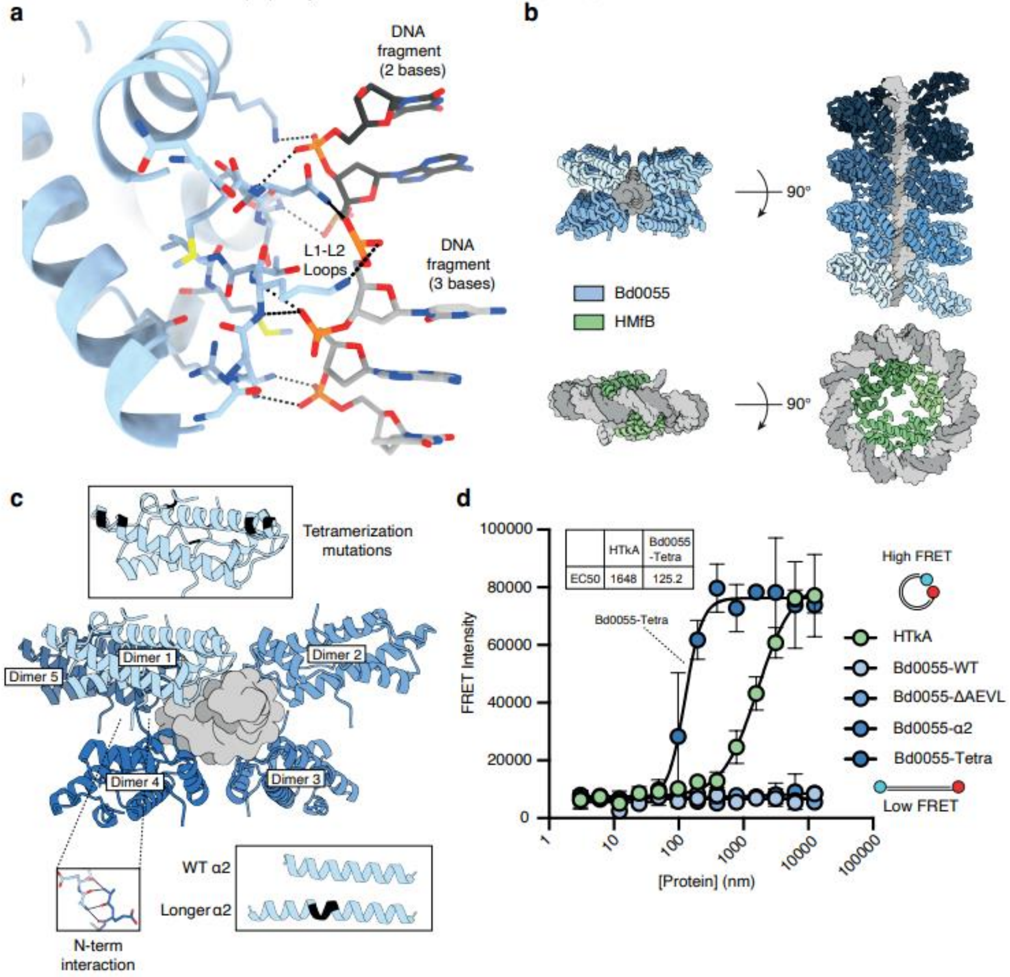
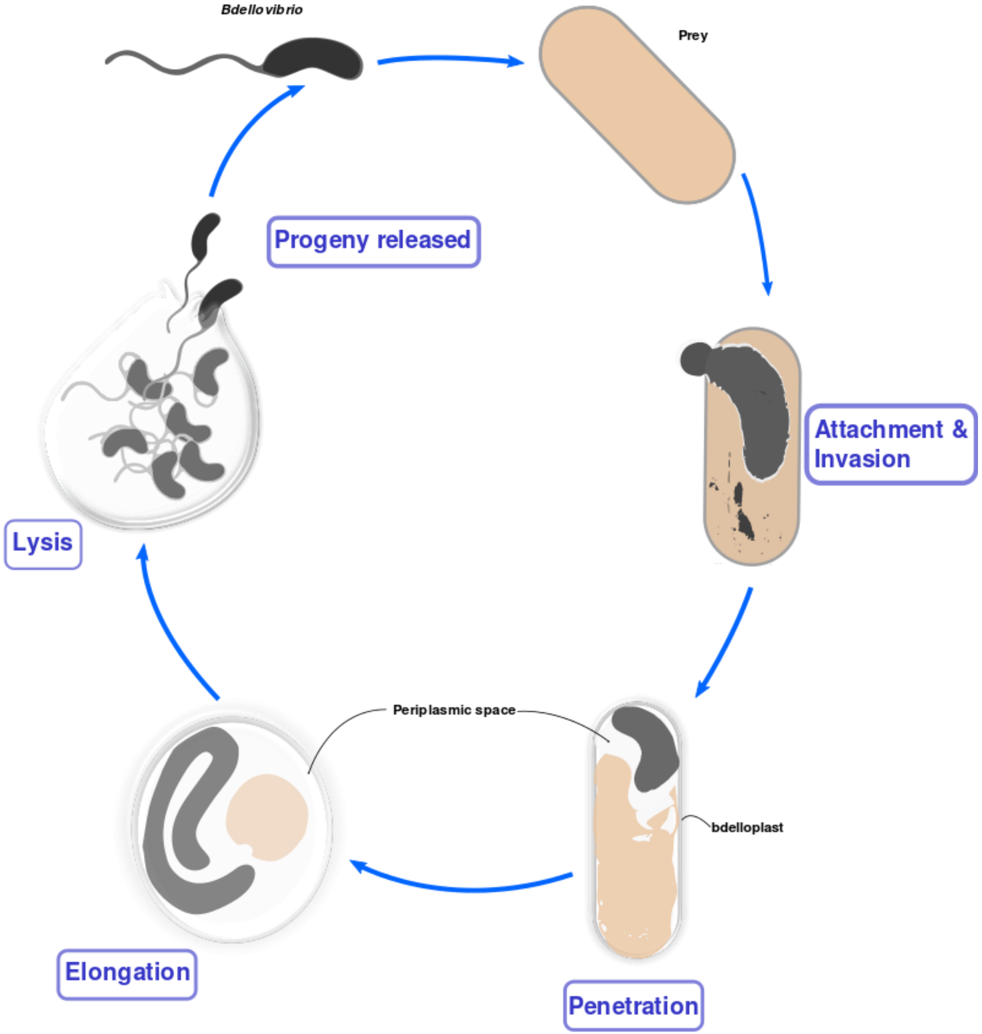


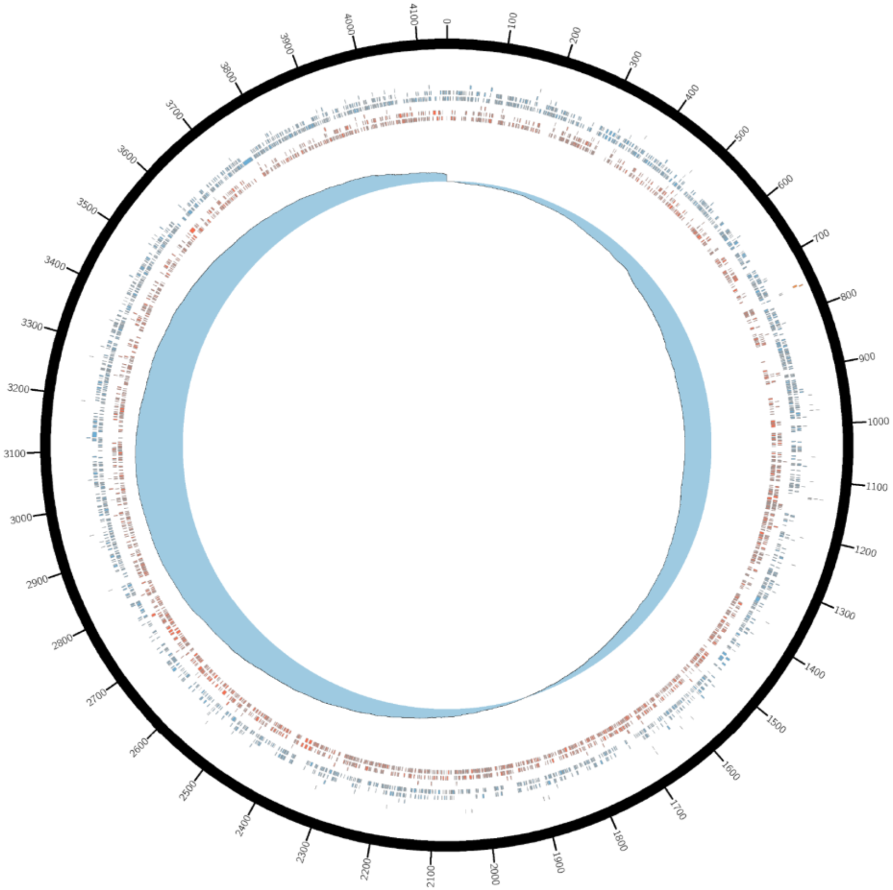
Figure 5. Bd0055 binds DNA end-on and encases straight DNA. (a) Crystal structure of Bd0055 in complex

- *Bdellovibrio* is predatory!
- ATAC-seq does not work reliably because of the presence of the prey
- Fortunately, there are prey-independent strains

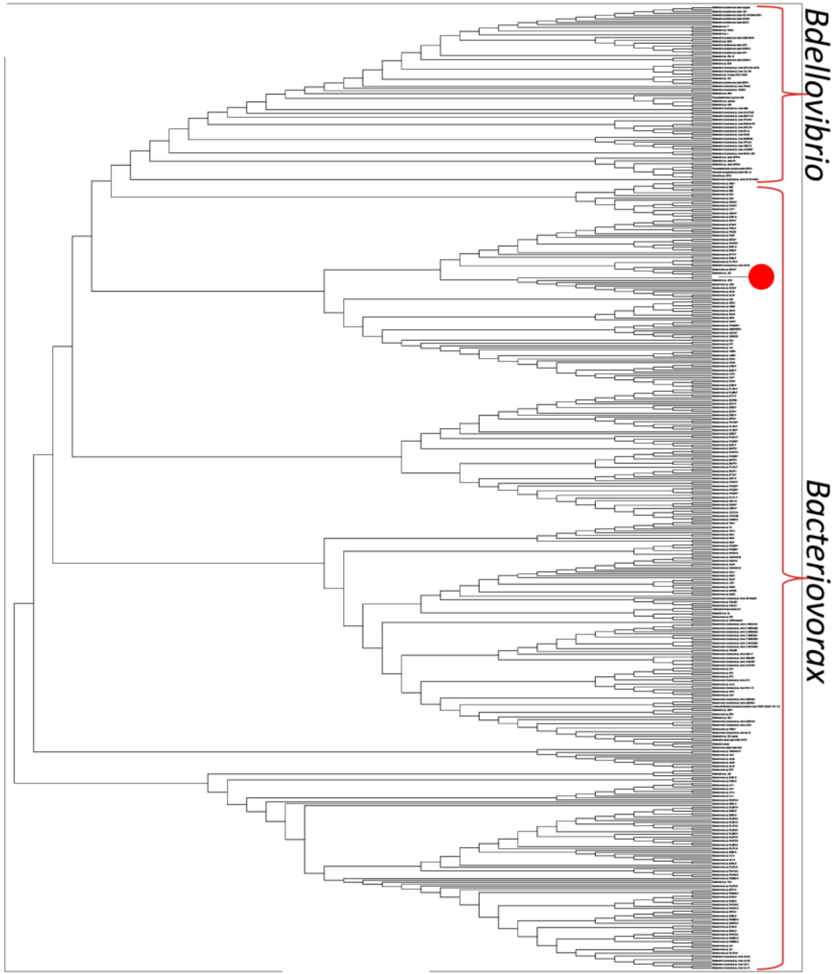


DE NOVO ASSEMBLY ILLUMINA + FLONGLE

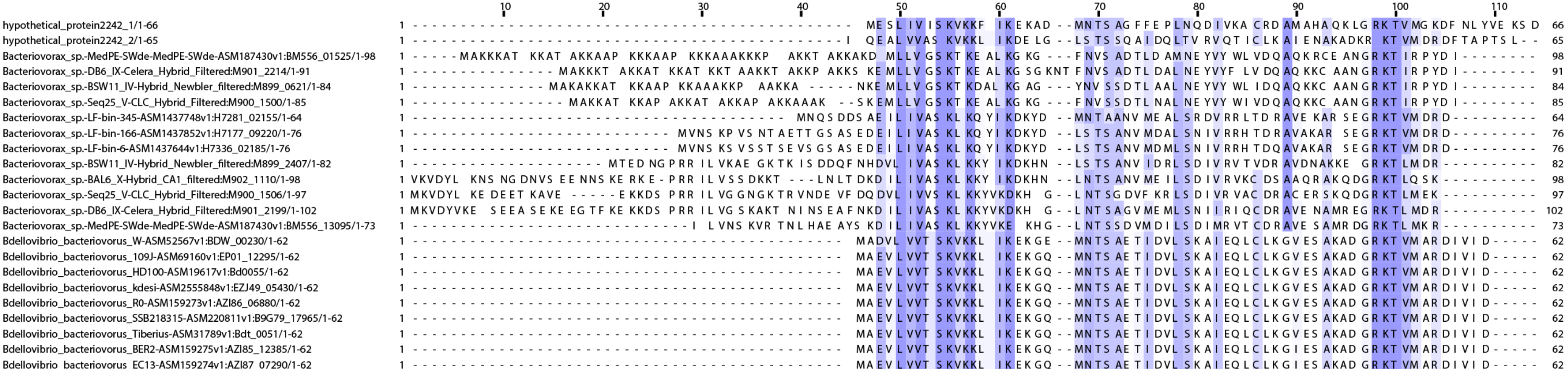
- 667x coverage
- Single contig
- 4,148,738 bp
- 4,127 protein coding genes



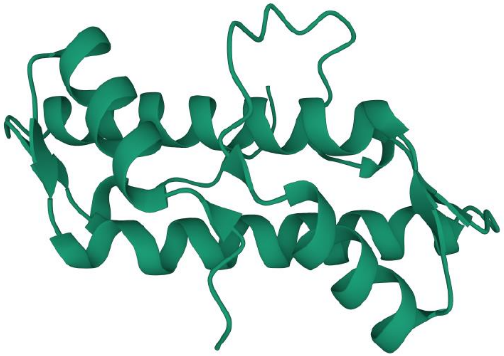
16S RNA PHYLOGENY



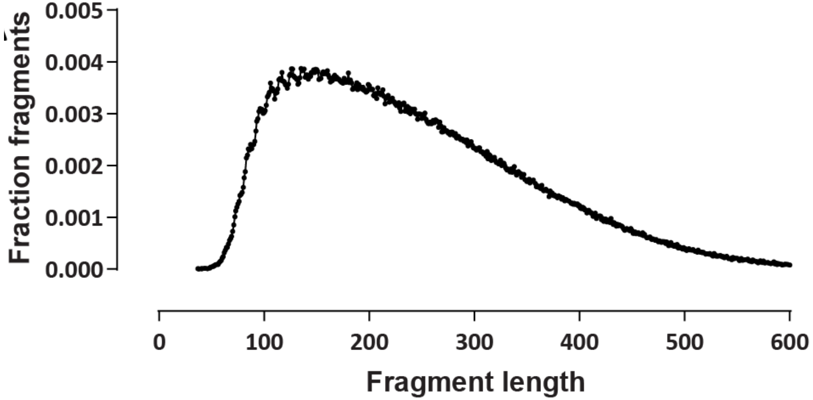
A DOUBLET HISTONE IN BACTERIOVORAX

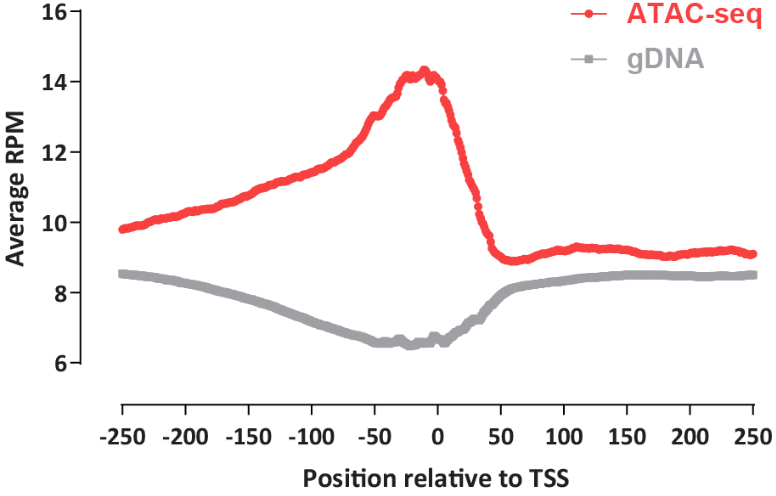


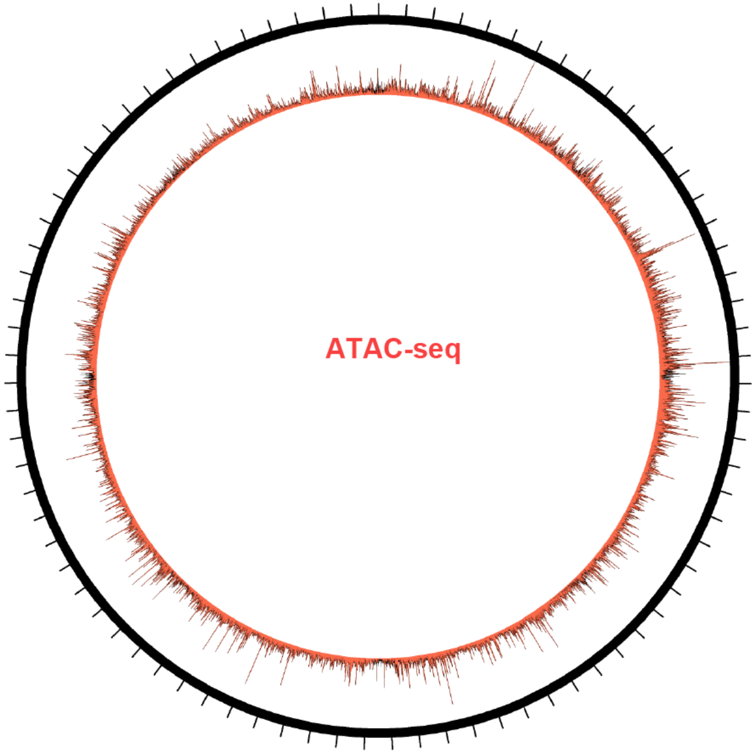
```
10 20 30 40 50 60 70 80 90 100 110
hypothetical_protein2242_1/1-66 1 ----------------------------------------------MESLIVISKVKKFIKEKAD--MNTSAGFFEPLNQDIVKACRDAMAHAQKLGRKTVMGKDFNLYVEKSD 66
hypothetical_protein2242_2/1-65 1 ---------------------------------------------IQEALVVASKVKKLIKDELG--LSTSSQAIDQLTVRVQTICLKAIENAKADKRKTVMDRDFTAPTSL-- 65
Bacteriovorax_sp.-MedPE-SWde-MedPE-SWde-ASM187430v1:BM556_01525/1-98 1 -----MAKKKATKKATAKKAAPKKKAAPKKKAAAKKKPAKKTAKKAKDMLLVGSKTKEALKGKG---FNVSADTLDAMNEYVYWLVDQAQKRCEANGRKTIRPYDI-------- 98
Bacteriovorax_sp.-DB6_IX-Celera_Hybrid_Filtered:M901_2214/1-91 1 ---------------MAKKKTAKKATKKATKKTAAKKTAKKPAKKSKEMLLVGSKTKEALKGSGKNTFNVSADTLDALNEYVYFLVDQAQKKCAANGRKTIRPYDI-------- 91
Bacteriovorax_sp.-BSW11_IV-Hybrid_Newbler_filtered:M899_0621/1-84 1 --------------MAKAKKATKKAAPKKAAAKKPAAKKA-----NKEMLLVGSKTKDALKGAG---YNVSSDTLAALNEYVYWLIDQAQKKCAANGRKTIRPYDI-------- 84
Bacteriovorax_sp.-Seq25_V-CLC_Hybrid_Filtered:M900_1500/1-85 1 ----------------MAKKATKKAPAKKATAKKAPAKKAAAK--SKEMLLVGSKTKEALKGKG---FNVSSDTLNALNEYVYWIVDQAQKKCAANGRKTIRPYDI-------- 85
Bacteriovorax_sp.-LF-bin-345-ASM1437748v1:H7281_02155/1-64 1 ---------------------------------------MNQSDDSAEILIVASKLKQYIKDKYD--MNTAANVMEALSRDVRRLTDRAVEKARSEGRKTVMDRD--------- 64
Bacteriovorax_sp.-LF-bin-166-ASM1437852v1:H7177_09220/1-76 1 ---------------------------MVNSKPVSNTAETTGSASEDEILIVASKLKQYIKDKYD--LSTSANVMDALSNIVRRHTDRAVAKARSEGRKTVMDRD--------- 76
Bacteriovorax_sp.-LF-bin-6-ASM1437644v1:H7336_02185/1-76 1 ---------------------------MVNSKSVSSTSEVSGSASEDEILIVASKLKQYIKDKYD--LSTSANVMDMLSNIVRRHTDQAVAKARSEGRKTVMDRD--------- 76
Bacteriovorax_sp.-BSW11_IV-Hybrid_Newbler_filtered:M899_2407/1-82 1 --------------------MTEDNGPRRILVKAEGKTKISDDQFNHDVLIVASKLKKYIKDKHN--LSTSANVIDRLSDIVRVTVDRAVDNAKKEGRKTLMDR---------- 82
Bacteriovorax_sp.-BAL6_X-Hybrid_CA1_filtered:M902_1110/1-98 1 VKVDYLKNSNGDNVSEENNSKERKE-PRRILVSSDKKT---LNLTDKDILIVASKLKKYIKDKHN--LNTSANVMEILSDIVRVKCDSAAQRAKQDGRKTLQSK---------- 98
Bacteriovorax_sp.-Seq25_V-CLC_Hybrid_Filtered:M900_1506/1-97 1 MKVDYLKEDEETKAVE-----EKKDSPRRILVGGNGKTRVNDEVFDQDVLIVVSKLKKYVKDKHG--LNTSGDVFKRLSDIVRVACDRACERSKQDGRKTLMEK---------- 97
Bacteriovorax_sp.-DB6_IX-Celera_Hybrid_Filtered:M901_2199/1-102 1 MKVDYVKESEEASEKEEGTFKEKKDSPRRILVGSKAKTNINSEAFNKDILIVASKLKKYVKDKHG--LNTSAGVMEMLSNIIRIQCDRAVENAMREGRKTLMDR---------- 102
Bacteriovorax_sp.-MedPE-SWde-MedPE-SWde-ASM187430v1:BM556_13095/1-73 1 -----------------------------ILVNSKVRTNLHAEAYSKDILIVASKLKKYVKEKHG--LNTSSDVMDILSDIMRVTCDRAVESAMRDGRKTLMKR---------- 73
Bdellovibrio_bacteriovorus_W-ASM52567v1:BDW_00230/1-62 1 ---------------------------------------------MADVLVVTSKVKKLIKEKGE--MNTSAETIDVLSKAIEQLCLKGVESAKADGRKTVMARDIVID----- 62
Bdellovibrio_bacteriovorus_109J-ASM69160v1:EP01_12295/1-62 1 ---------------------------------------------MAEVLVVTSKVKKLIKEKGQ--MNTSAETIDVLSKAIEQLCLKGVESAKADGRKTVMARDIVID----- 62
Bdellovibrio_bacteriovorus_HD100-ASM19617v1:Bd0055/1-62 1 ---------------------------------------------MAEVLVVTSKVKKLIKEKGQ--MNTSAETIDVLSKAIEQLCLKGVESAKADGRKTVMARDIVID----- 62
Bdellovibrio_bacteriovorus_kdesi-ASM2555848v1:EZJ49_05430/1-62 1 ---------------------------------------------MAEVLVVTSKVKKLIKEKGQ--MNTSAETIDVLSKAIEQLCLKGVESAKADGRKTVMARDIVID----- 62
Bdellovibrio_bacteriovorus_R0-ASM159273v1:AZI86_06880/1-62 1 ---------------------------------------------MAEVLVVTSKVKKLIKEKGQ--MNTSAETIDVLSKAIEQLCLKGVESAKADGRKTVMARDIVID----- 62
Bdellovibrio_bacteriovorus_SSB218315-ASM220811v1:B9G79_17965/1-62 1 ---------------------------------------------MAEVLVVTSKVKKLIKEKGQ--MNTSAETIDVLSKAIEQLCLKGVESAKADGRKTVMARDIVID----- 62
Bdellovibrio_bacteriovorus_Tiberius-ASM31789v1:Bdt_0051/1-62 1 ---------------------------------------------MAEVLVVTSKVKKLIKEKGQ--MNTSAETIDVLSKAIEQLCLKGVESAKADGRKTVMARDIVID----- 62
Bdellovibrio_bacteriovorus_BER2-ASM159275v1:AZI85_12385/1-62 1 ---------------------------------------------MAEVLVVTSKVKKLIKEKGQ--MNTSAETIDVLSKAIEQLCLKGIESAKADGRKTVMARDIVID----- 62
Bdellovibrio_bacteriovorus_EC13-ASM159274v1:AZI87_07290/1-62 1 ---------------------------------------------MAEVLVVTSKVKKLIKEKGQ--MNTSAETIDVLSKAIEQLCLKGIESAKADGRKTVMARDIVID----- 62
```



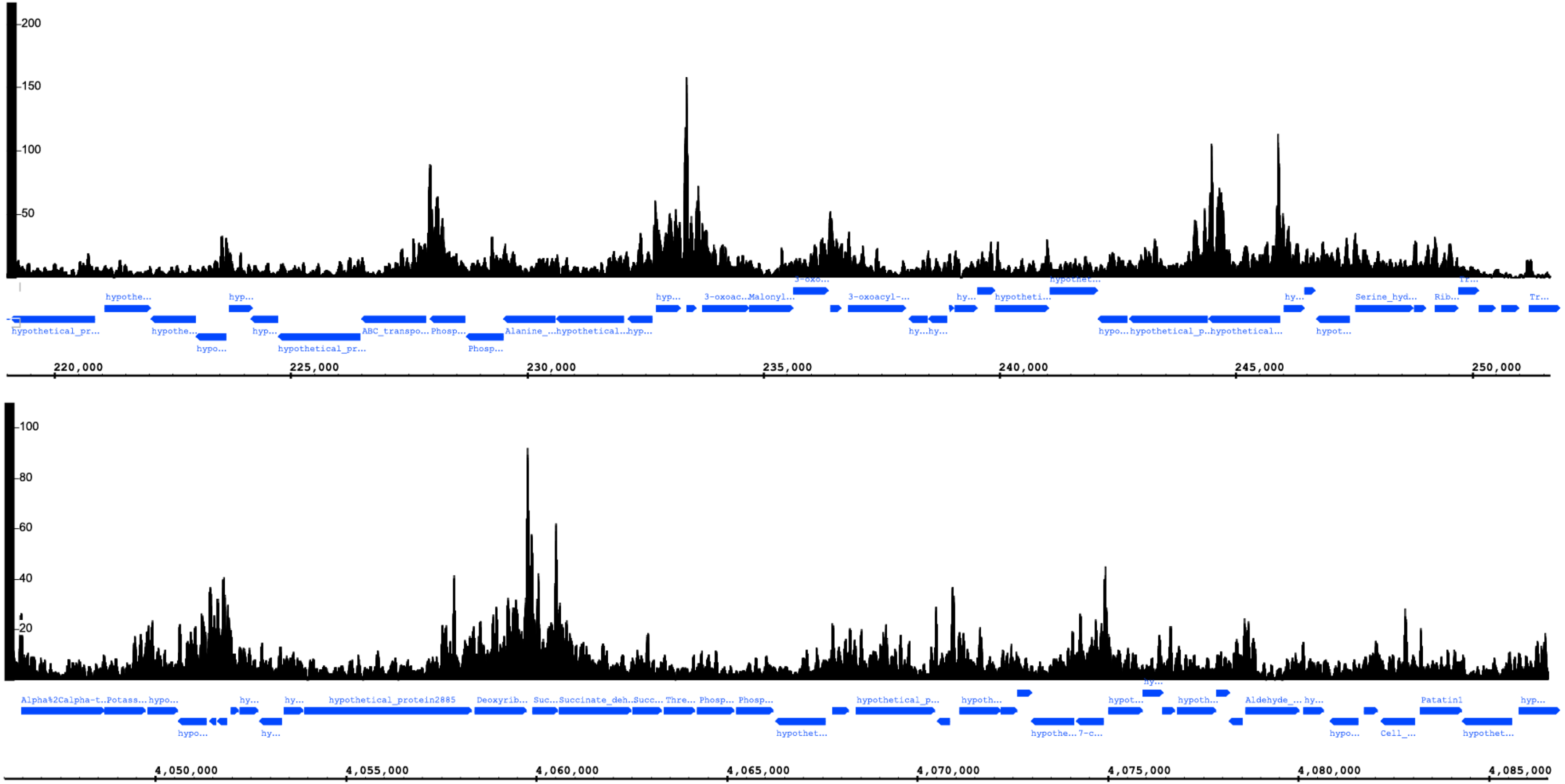
CHROMATIN ACCESSIBILITY (ATAC-SEQ)



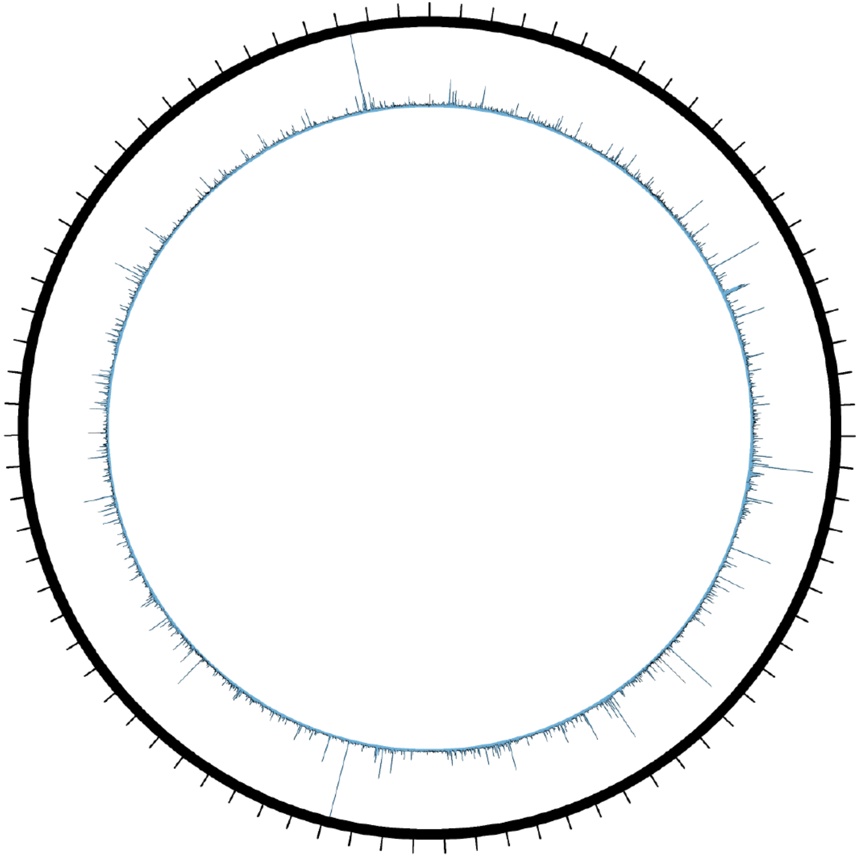




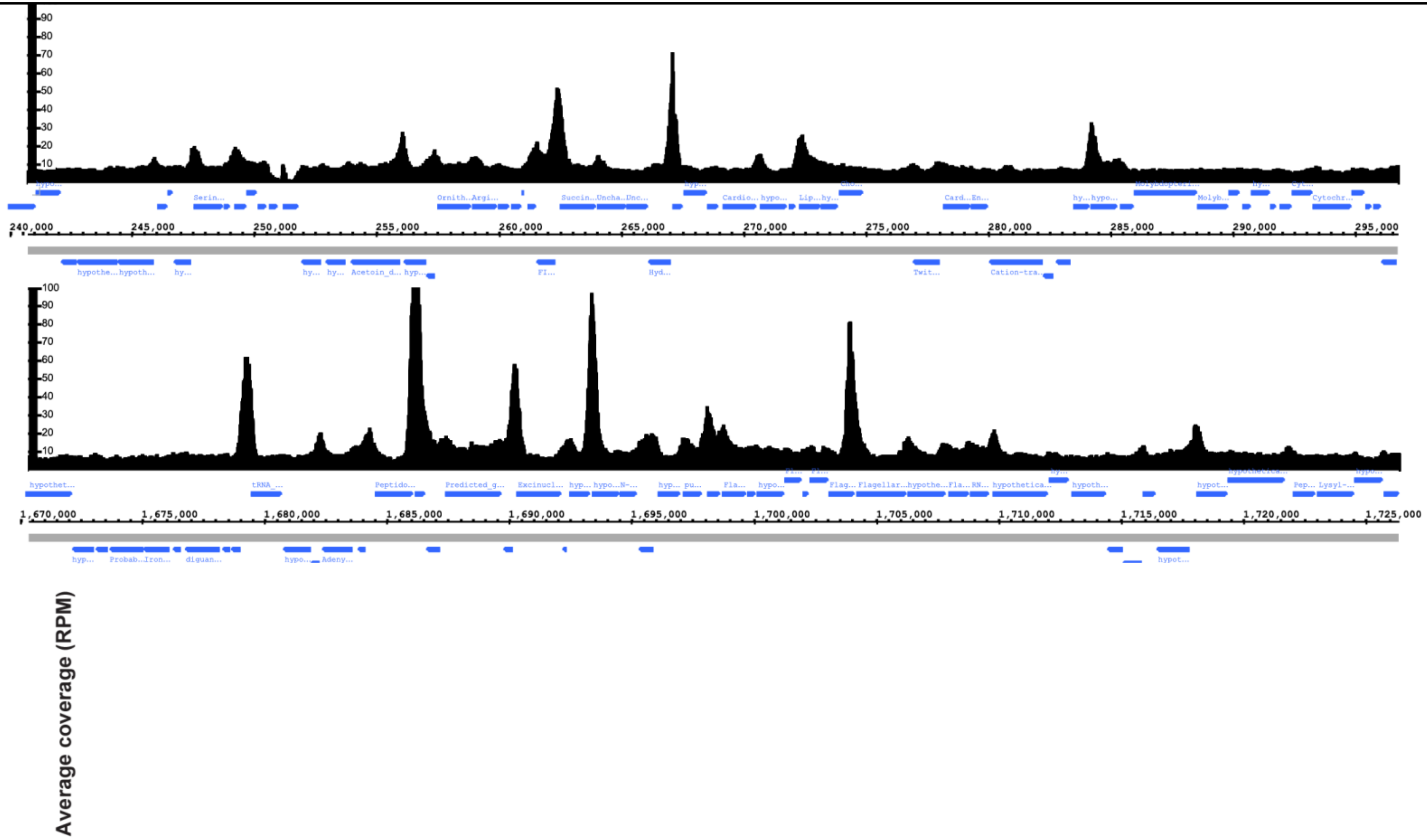
CHROMATIN ACCESSIBILITY (ATAC-SEQ)



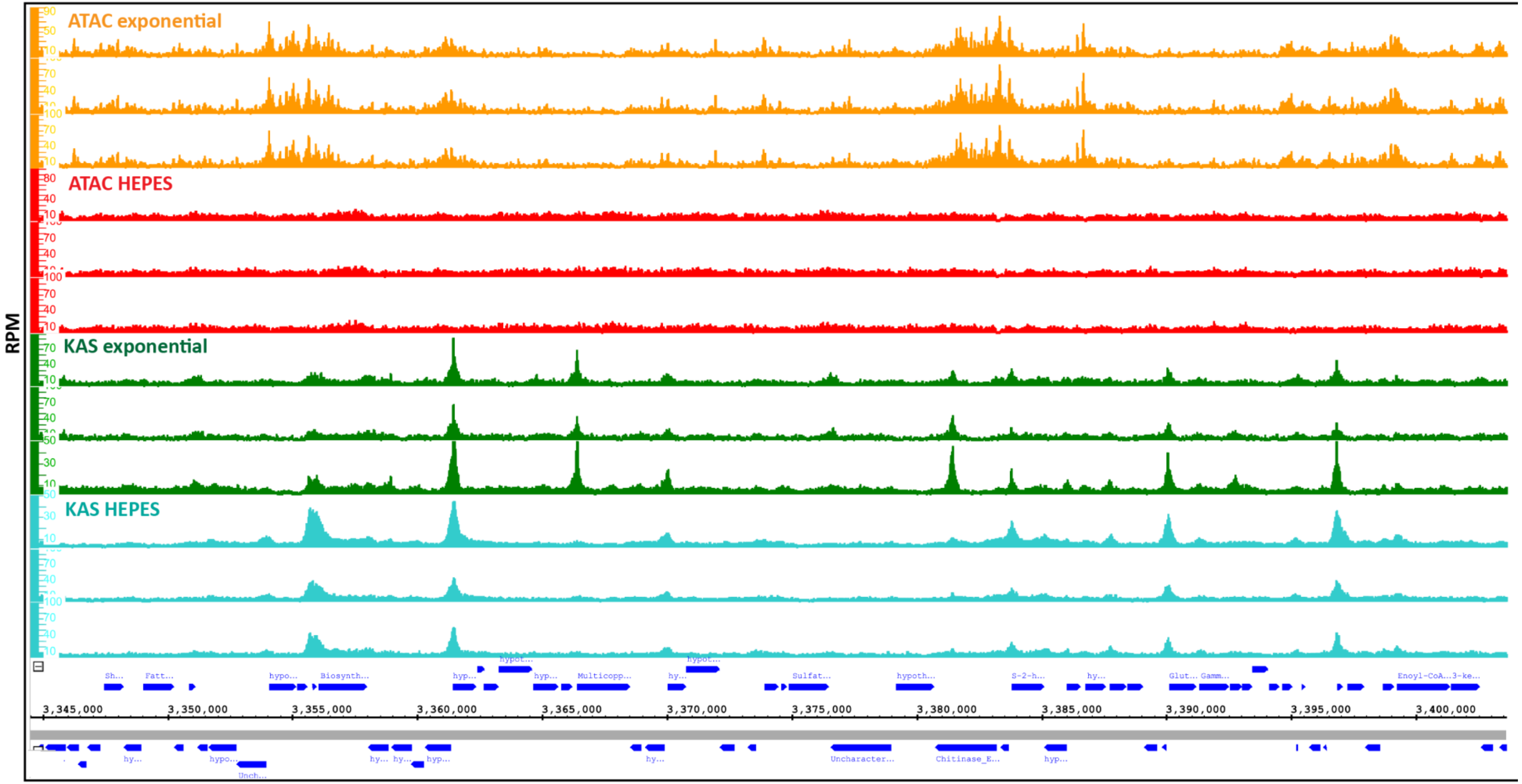
ACTIVE TRANSCRIPTION (KAS-SEQ)



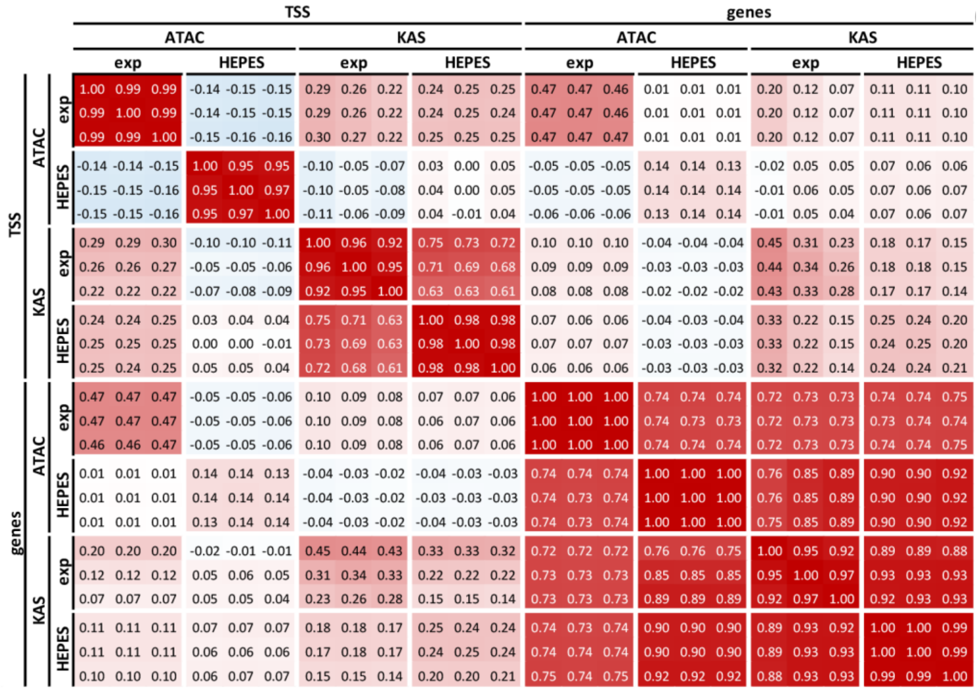
EXTREME PROMOTER PAUSING



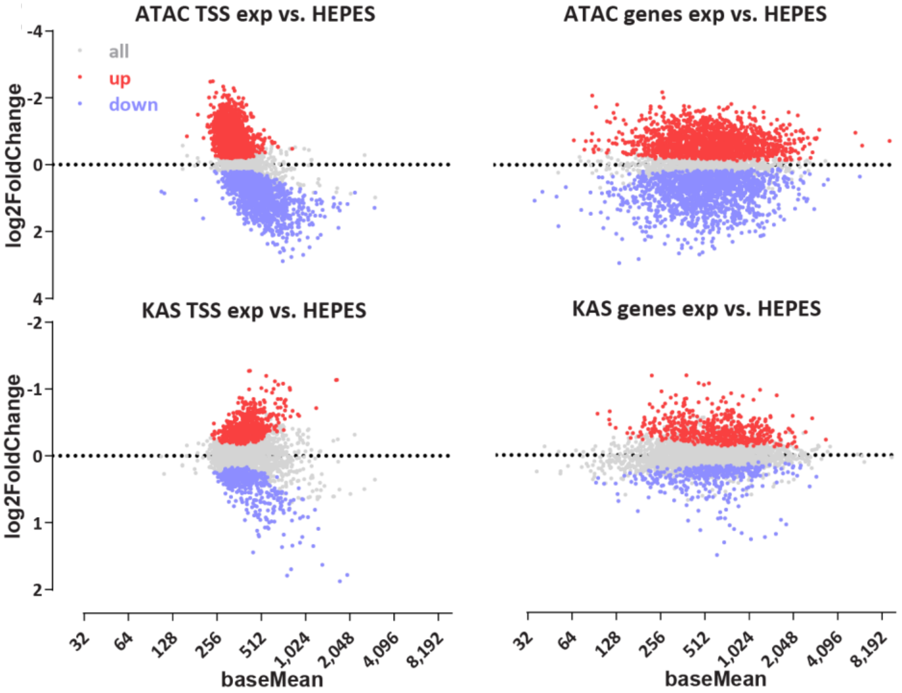
CHROMATIN ACCESSIBILITY AND TRANSCRIPTIONAL ACTIVITY



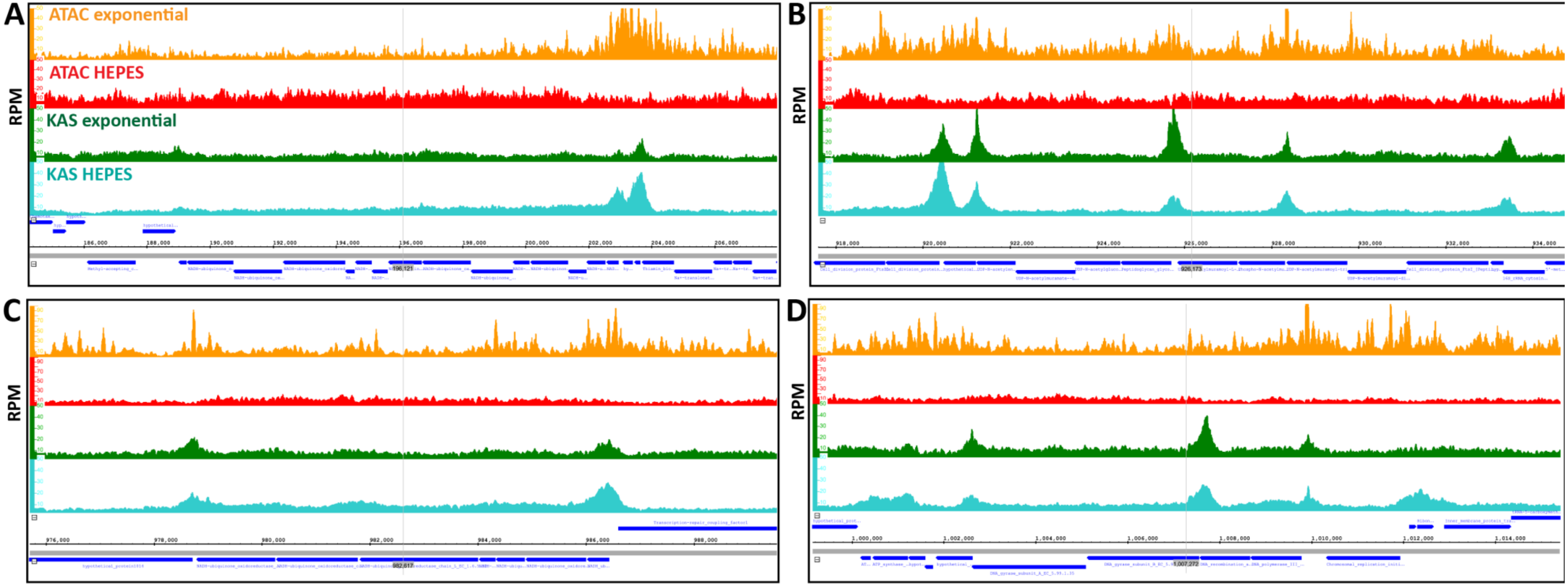
CHROMATIN ACCESSIBILITY AND TRANSCRIPTIONAL ACTIVITY



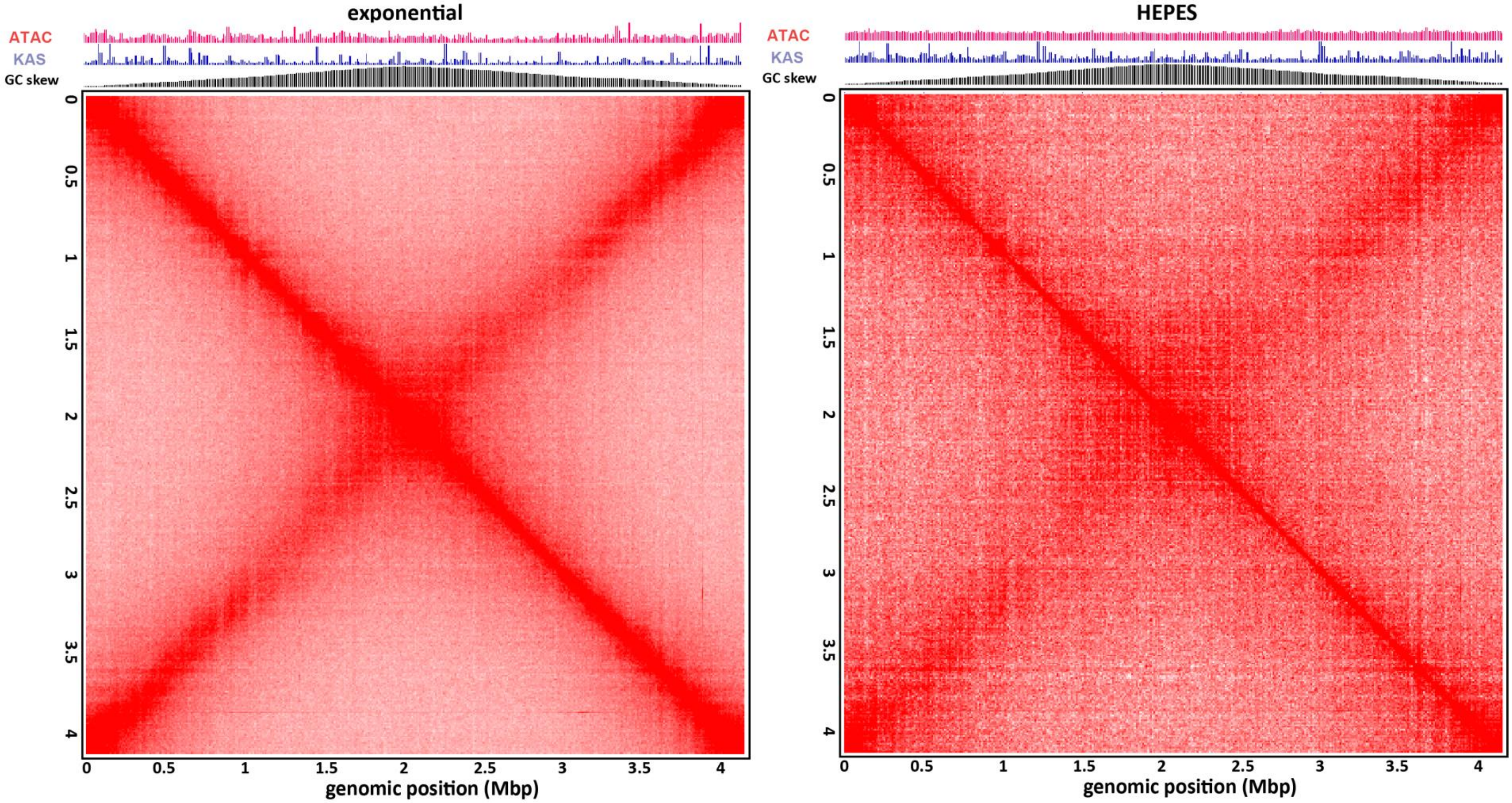
			TSS												genes											
			ATAC						KAS						ATAC						KAS					
			exp			HEPES			exp			HEPES			exp			HEPES			exp			HEPES		
TSS	ATAC	exp	1.00	0.99	0.99	-0.14	-0.15	-0.15	0.29	0.26	0.22	0.24	0.25	0.25	0.47	0.47	0.46	0.01	0.01	0.01	0.20	0.12	0.07	0.11	0.11	0.10
			0.99	1.00	0.99	-0.14	-0.15	-0.15	0.29	0.26	0.22	0.24	0.25	0.24	0.47	0.47	0.46	0.01	0.01	0.01	0.20	0.12	0.07	0.11	0.11	0.10
			0.99	0.99	1.00	-0.15	-0.16	-0.16	0.30	0.27	0.22	0.25	0.25	0.25	0.47	0.47	0.47	0.01	0.01	0.01	0.20	0.12	0.07	0.11	0.11	0.10
		HEPES	-0.14	-0.14	-0.15	1.00	0.95	0.95	-0.10	-0.05	-0.07	0.03	0.00	0.05	-0.05	-0.05	-0.05	0.14	0.14	0.13	-0.02	0.05	0.05	0.07	0.06	0.06
			-0.15	-0.15	-0.16	0.95	1.00	0.97	-0.10	-0.05	-0.08	0.04	0.00	0.05	-0.05	-0.05	-0.05	0.14	0.14	0.14	-0.01	0.06	0.05	0.07	0.06	0.07
			-0.15	-0.15	-0.16	0.95	0.97	1.00	-0.11	-0.06	-0.09	0.04	-0.01	0.04	-0.06	-0.06	-0.06	0.13	0.14	0.14	-0.01	0.05	0.04	0.07	0.06	0.07
	KAS	exp	0.29	0.29	0.30	-0.10	-0.10	-0.11	1.00	0.96	0.92	0.75	0.73	0.72	0.10	0.10	0.10	-0.04	-0.04	-0.04	0.45	0.31	0.23	0.18	0.17	0.15
			0.26	0.26	0.27	-0.05	-0.05	-0.06	0.96	1.00	0.95	0.71	0.69	0.68	0.09	0.09	0.09	-0.03	-0.03	-0.03	0.44	0.34	0.26	0.18	0.18	0.15
			0.22	0.22	0.22	-0.07	-0.08	-0.09	0.92	0.95	1.00	0.63	0.63	0.61	0.08	0.08	0.08	-0.02	-0.02	-0.02	0.43	0.33	0.28	0.17	0.17	0.14
		HEPES	0.24	0.24	0.25	0.03	0.04	0.04	0.75	0.71	0.63	1.00	0.98	0.98	0.07	0.06	0.06	-0.04	-0.03	-0.04	0.33	0.22	0.15	0.25	0.24	0.20
			0.25	0.25	0.25	0.00	0.00	-0.01	0.73	0.69	0.63	0.98	1.00	0.98	0.07	0.07	0.07	-0.03	-0.03	-0.03	0.33	0.22	0.15	0.24	0.25	0.20
			0.25	0.24	0.25	0.05	0.05	0.04	0.72	0.68	0.61	0.98	0.98	1.00	0.06	0.06	0.06	-0.03	-0.03	-0.03	0.32	0.22	0.14	0.24	0.24	0.21
genes	ATAC	exp	0.47	0.47	0.47	-0.05	-0.05	-0.06	0.10	0.09	0.08	0.07	0.07	0.06	1.00	1.00	1.00	0.74	0.74	0.74	0.72	0.73	0.73	0.74	0.74	0.75
			0.47	0.47	0.47	-0.05	-0.05	-0.06	0.10	0.09	0.08	0.06	0.07	0.06	1.00	1.00	1.00	0.74	0.73	0.73	0.72	0.73	0.73	0.73	0.74	0.74
			0.46	0.46	0.47	-0.05	-0.05	-0.06	0.10	0.09	0.08	0.06	0.07	0.06	1.00	1.00	1.00	0.74	0.74	0.74	0.72	0.73	0.73	0.74	0.74	0.75
		HEPES	0.01	0.01	0.01	0.14	0.14	0.13	-0.04	-0.03	-0.02	-0.04	-0.03	-0.03	0.74	0.74	0.74	1.00	1.00	1.00	0.76	0.85	0.89	0.90	0.90	0.92
			0.01	0.01	0.01	0.14	0.14	0.14	-0.04	-0.03	-0.02	-0.03	-0.03	-0.03	0.74	0.73	0.74	1.00	1.00	1.00	0.76	0.85	0.89	0.90	0.90	0.92
			0.01	0.01	0.01	0.13	0.14	0.14	-0.04	-0.03	-0.02	-0.04	-0.03	-0.03	0.74	0.73	0.74	1.00	1.00	1.00	0.75	0.85	0.89	0.90	0.90	0.92
	KAS	exp	0.20	0.20	0.20	-0.02	-0.01	-0.01	0.45	0.44	0.43	0.33	0.33	0.32	0.72	0.72	0.72	0.76	0.76	0.75	1.00	0.95	0.92	0.89	0.89	0.88
			0.12	0.12	0.12	0.05	0.06	0.05	0.31	0.34	0.33	0.22	0.22	0.22	0.73	0.73	0.73	0.85	0.85	0.85	0.95	1.00	0.97	0.93	0.93	0.93
			0.07	0.07	0.07	0.05	0.05	0.04	0.23	0.26	0.28	0.15	0.15	0.14	0.73	0.73	0.73	0.89	0.89	0.89	0.92	0.97	1.00	0.92	0.93	0.93
		HEPES	0.11	0.11	0.11	0.07	0.07	0.07	0.18	0.18	0.17	0.25	0.24	0.24	0.74	0.73	0.74	0.90	0.90	0.90	0.89	0.93	0.92	1.00	1.00	0.99
			0.11	0.11	0.11	0.06	0.06	0.06	0.17	0.18	0.17	0.24	0.25	0.24	0.74	0.74	0.74	0.90	0.90	0.90	0.89	0.93	0.93	1.00	1.00	0.99
			0.10	0.10	0.10	0.06	0.07	0.07	0.15	0.15	0.14	0.20	0.20	0.21	0.75	0.74	0.75	0.92	0.92	0.92	0.88	0.93	0.93	0.99	0.99	1.00



OPERONS


A
B
C
D
ATAC exponential
ATAC HEPES
KAS exponential
KAS HEPES
RPM
186,000
188,000
190,000
192,000
194,000
196,000
198,000
200,000
202,000
204,000
206,000
918,000
920,000
922,000
924,000
926,000
928,000
930,000
932,000
934,000
976,000
978,000
980,000
982,000
984,000
986,000
988,000
1,000,000
1,002,000
1,004,000
1,006,000
1,008,000
1,010,000
1,012,000
1,014,000
196,121
926,173
982,617
1,007,272
Transcription-repair_coupling_factor1

3D GENOME ORGANIZATION



SUMMARY

- Contrary to e.g. *Caulobacter*, Bdellovbrionota exhibit eukaryote-like chromatin organization
  - similar to *Haloferax*
  - similar to eukaryotes with compact genomes
  - this is despite the proposed inverted physical relationship between DNA and histones

- Extremely strong promoter pausing is observed
  - suggesting major role for promoter release in expression regulation

- 3D organization similar to that of other bacteria

BASIC CHROMATIN PROPERTIES ACROSS THE TREE OF LIFE

	clade	histones	histone-based chromatin	genome size	gene length		intergenic space		ATAC TSS score	promoter pausing
					mean	median	mean	median		
Bacteriovorax sp.	**Bacteria**	prokaryotic	✓	4,148,738	939	785	94	65	1.35-1.4	✓
Caulobacter crescentus		×	×	4,042,929	944	800	128	97	low	?
Haloferax volcanii	**Archaea**	prokaryotic	?	4,012,900	833	740	167	108	1.35-1.4	✓
Sulfolobus islandicus		×	×	2,522,992	823	722	194	77	1.1	?
Saccharomyces cerevisiae	**Eukaryotes**	H2A, H2B, H3, H4	✓	12,157,105	1,300	1,019	498	347	1.4-2	×
Drosophila melanogaster		H2A, H2B, H3, H4	✓	143,726,002	5,753	1,709	3,597	644	1.6-2.3	✓
Homo sapiens		H2A, H2B, H3, H4	✓	3,099,750,718	67,046	26,780	113,388	20,430	10-25	✓

V. ADVANCING THE ASSAY TOOLKIT

nature methods

ARTICLES

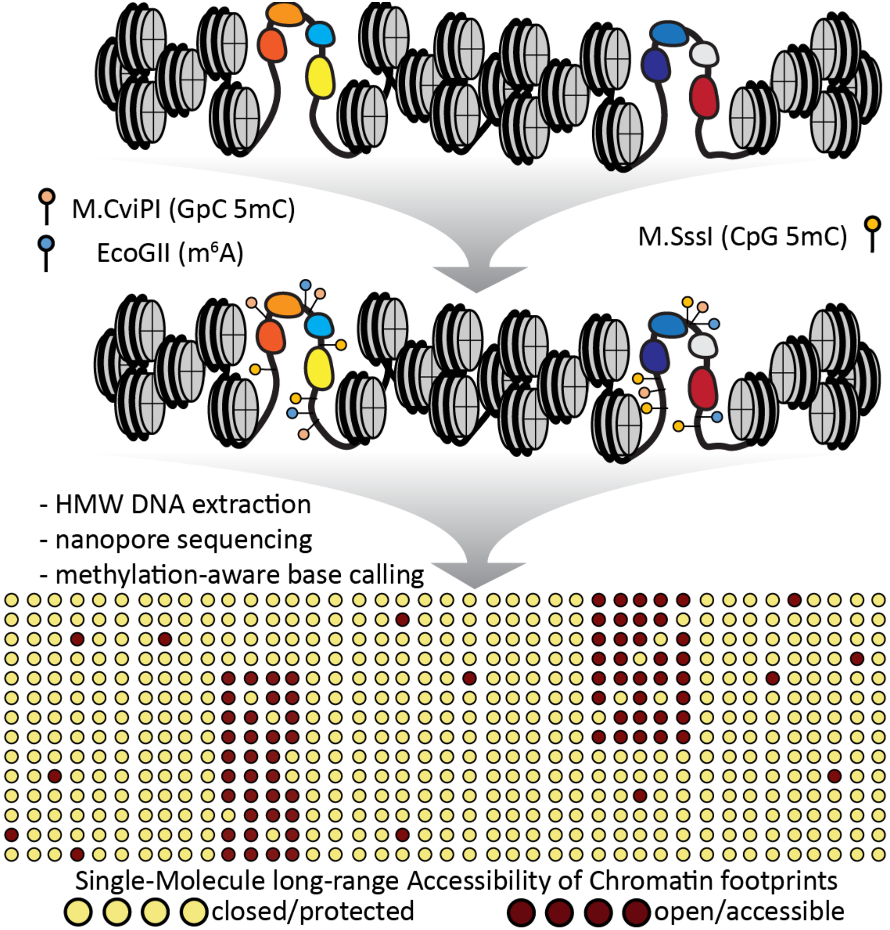
https://doi.org/10.1038/s41592-019-0730-2

Long-range single-molecule mapping of chromatin accessibility in eukaryotes

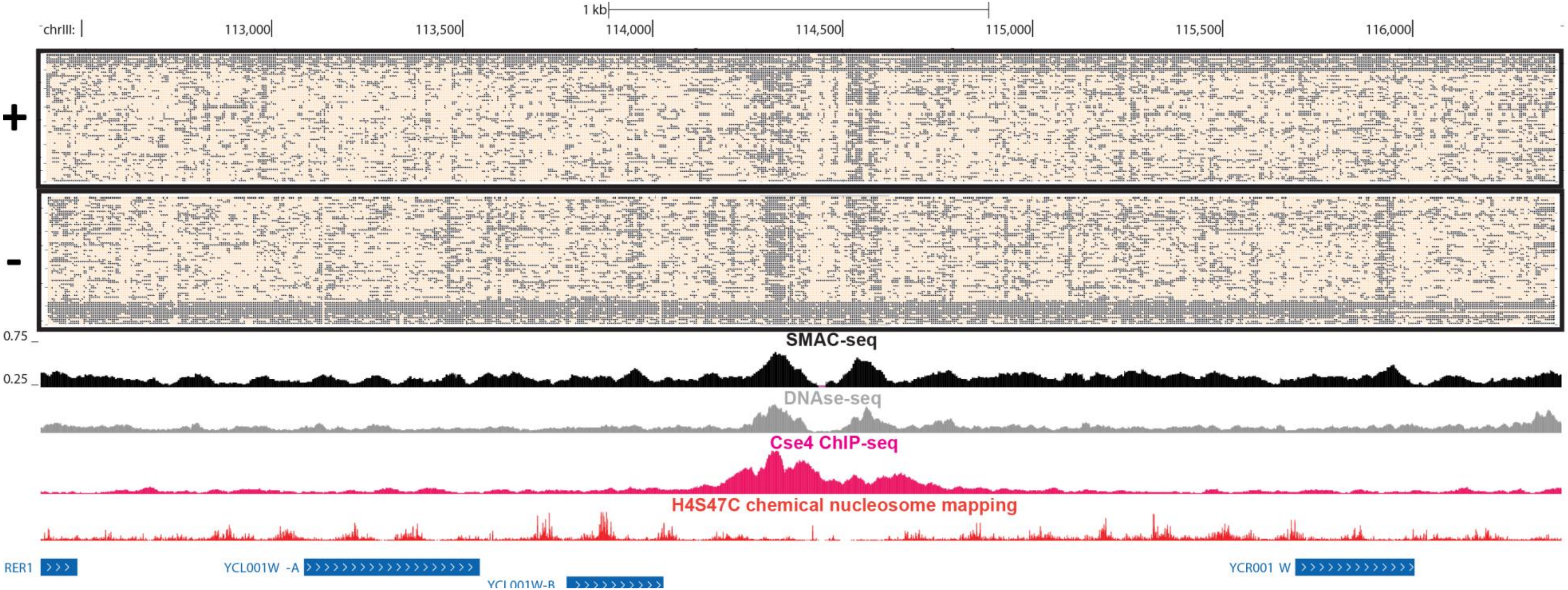
Zohar Shipony[1,6], Georgi K. Marinov[1,6], Matthew P. Swaffer[2], Nicholas A. Sinnott-Armstrong[1], Jan M. Skotheim[2], Anshul Kundaje[1,3] and William J. Greenleaf[1,4,5]*

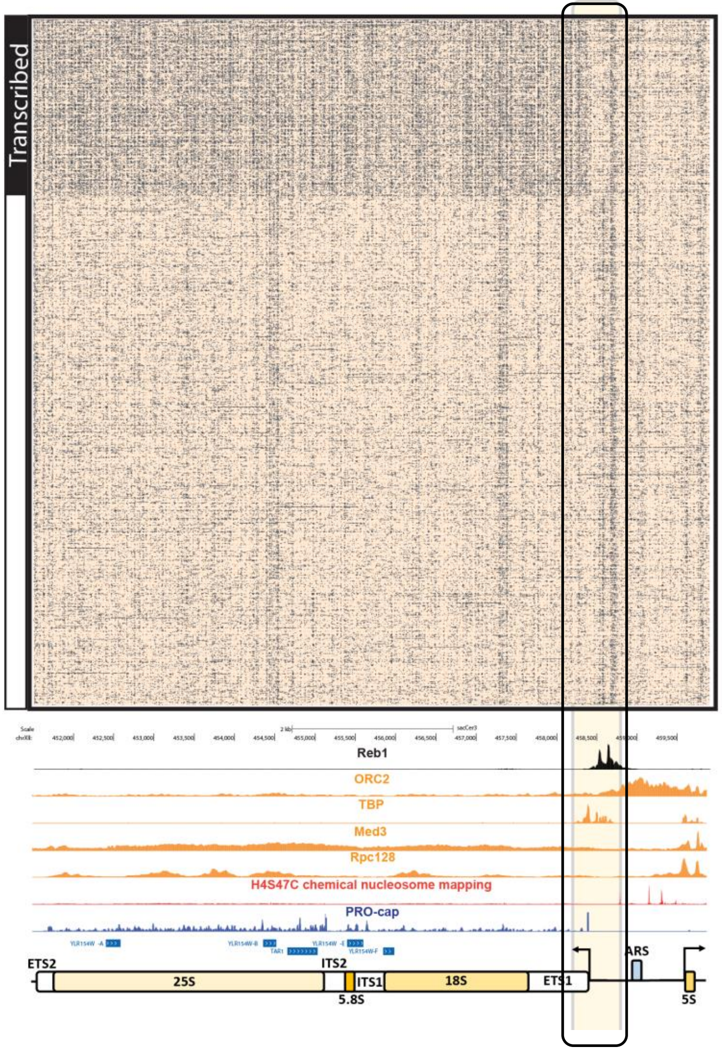
SINGLE-MOLECULE LONG READ ACCESSIBLE CHROMATIN MAPPING

SMAC-SEQ

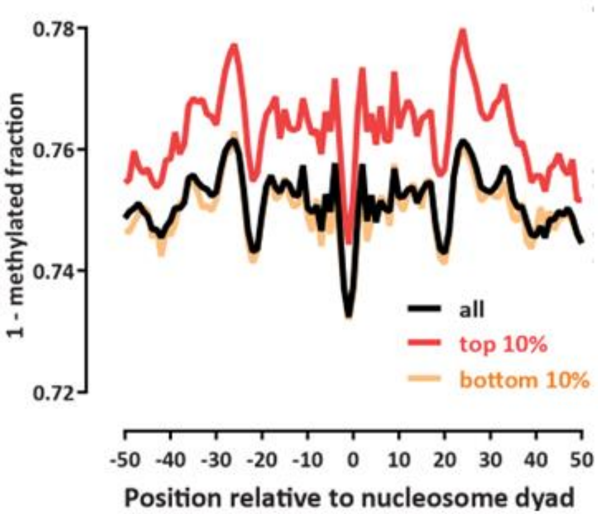


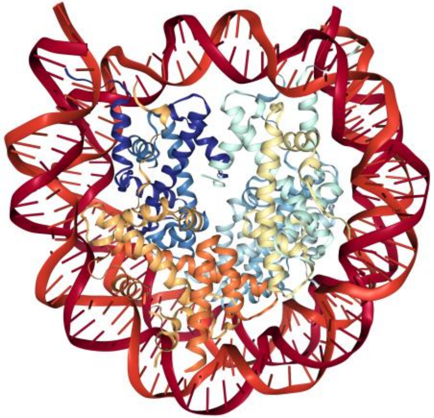
SINGLE-MOLECULE POPULATION-SCALE VIEW OF CHROMATIN HAPLOTYPES



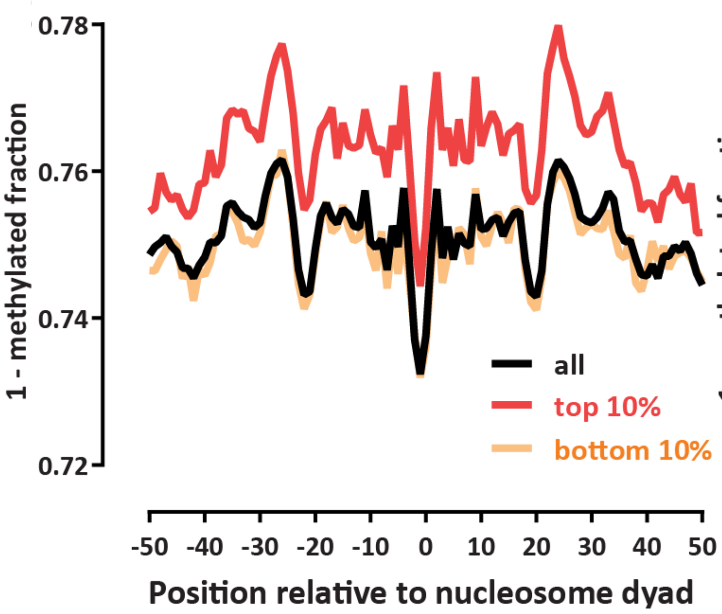

Transcribed
Scale
chrXII:
2 kb
sacCer3
452,000
452,500
453,000
453,500
454,000
454,500
455,000
455,500
456,000
456,500
457,000
457,500
458,000
458,500
459,000
459,500
Reb1
ORC2
TBP
Med3
Rpc128
H4S47C chemical nucleosome mapping
PRO-cap
YLR154W -A
YLR154W-B
TAR1
YLR154W -E
YLR154W-F
ETS2
25S
ITS2
5.8S
ITS1
18S
ETS1
ARS
5S

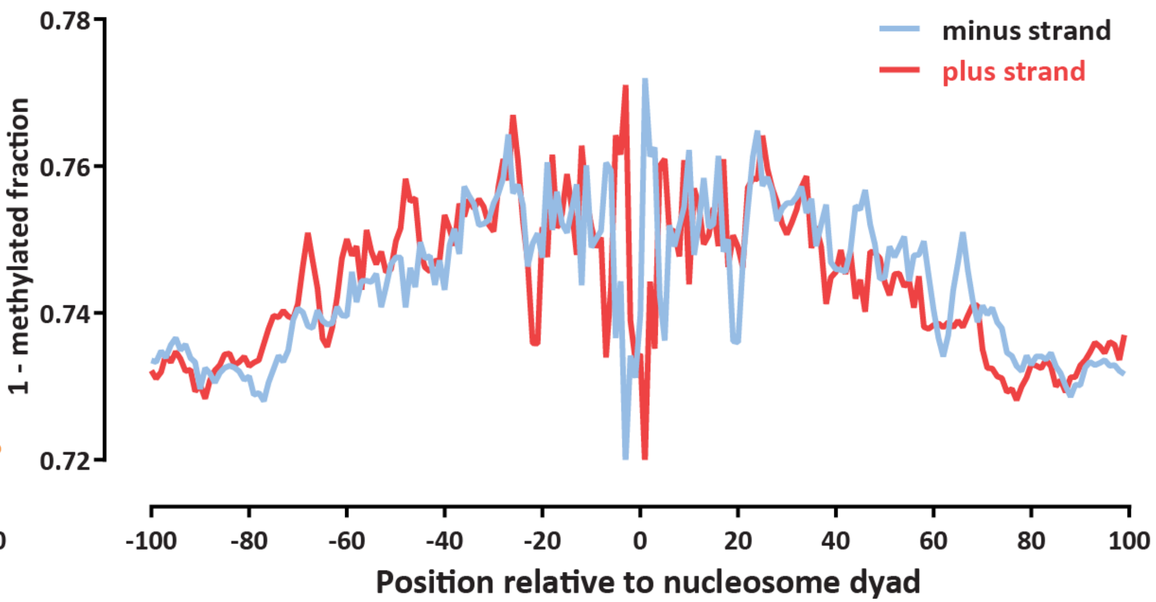
SMAC-SEQ REVEALS FINE-SCALE STRAND-SPECIFIC OCCUPANCY FEATURES





SMAC-SEQ REVEALS FINE-SCALE STRAND-SPECIFIC OCCUPANCY FEATURES





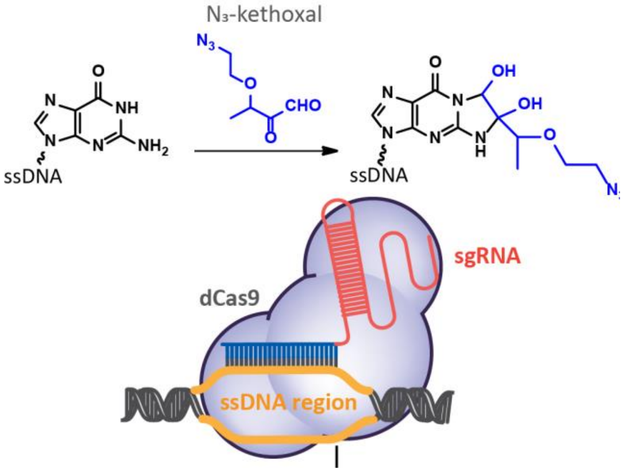
Marinov *et al. Genome Biology* (2023) 24:85
https://doi.org/10.1186/s13059-023-02930-z

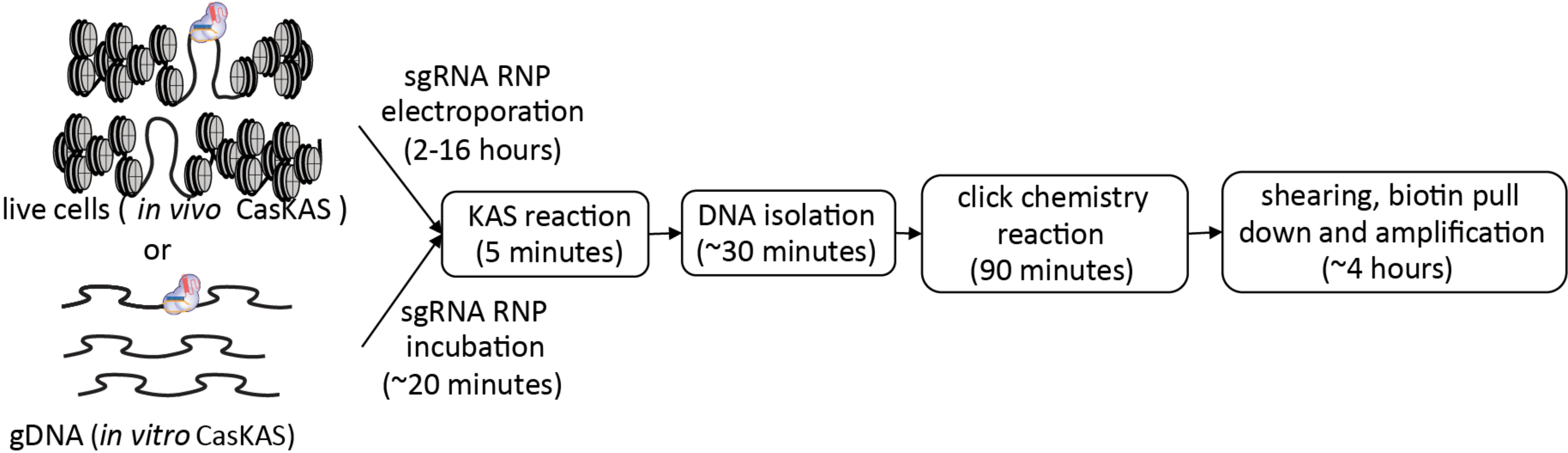
Genome Biology

METHOD **Open Access**

CasKAS: direct profiling of genome-wide dCas9 and Cas9 specificity using ssDNA mapping

Georgi K. Marinov[1*†], Samuel H. Kim[2†], S. Tansu Bagdatli[1], Soon Il Higashino[1], Alexandro E. Trevino[3,4], Josh Tycko[1], Tong Wu[5], Lacramioara Bintu[4], Michael C. Bassik[1,6], Chuan He[5,7,8], Anshul Kundaje[1,9] and William J. Greenleaf[1,10,11,12*]


N₃-kethoxal
N3
O
CHO
O
N
NH
N
N
NH2
ssDNA
O
OH
OH
N
N
N
N
N
H
O
ssDNA
sgRNA
dCas9
ssDNA region



live cells (*in vivo* CasKAS)

or

gDNA (*in vitro* CasKAS)

sgRNA RNP electroporation (2-16 hours)

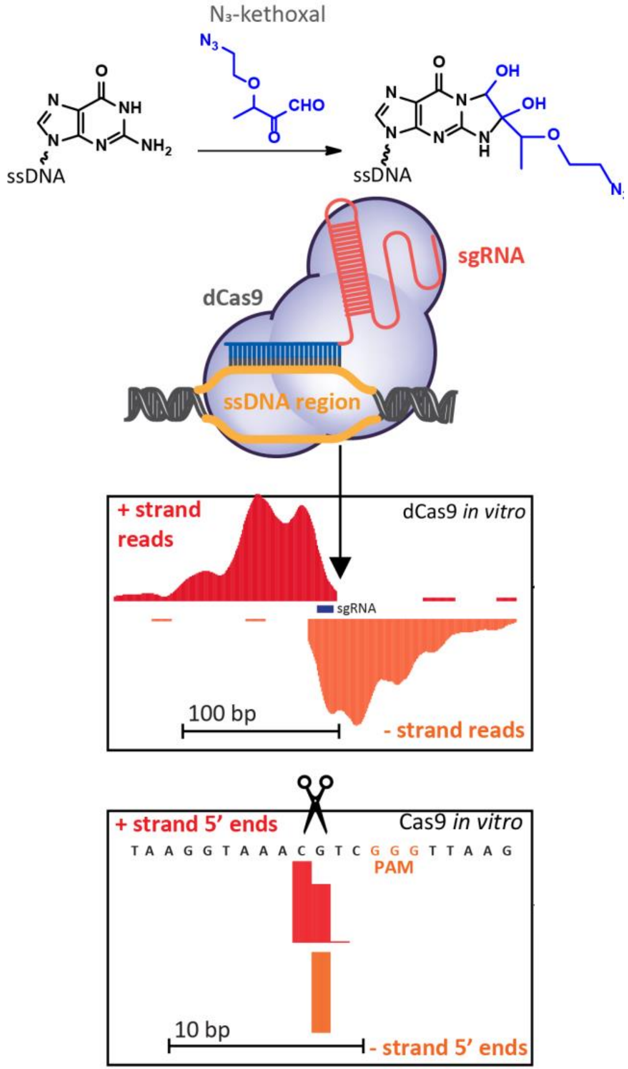
sgRNA RNP incubation (~20 minutes)

KAS reaction (5 minutes)

DNA isolation (~30 minutes)

click chemistry reaction (90 minutes)

shearing, biotin pull down and amplification (~4 hours)

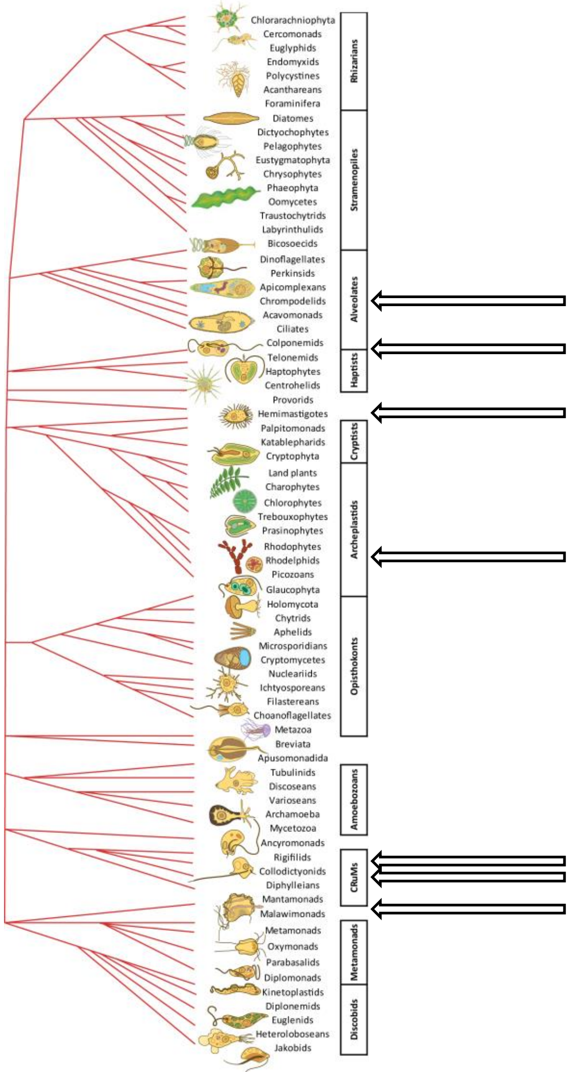

N₃-kethoxal
ssDNA
ssDNA
sgRNA
dCas9
ssDNA region
+ strand reads
dCas9 *in vitro*
sgRNA
100 bp
- strand reads
+ strand 5' ends
Cas9 *in vitro*
T A A G G T A A A C G T C G G G T T A A G
PAM
10 bp
- strand 5' ends

VI. FUTURE DIRECTIONS

CHALLENGES

- Most of the eukaryotic diversity is uncultured/unculturable

- Much of it is not well known



major eukaryotic clades discovered or defined in the last couple decades

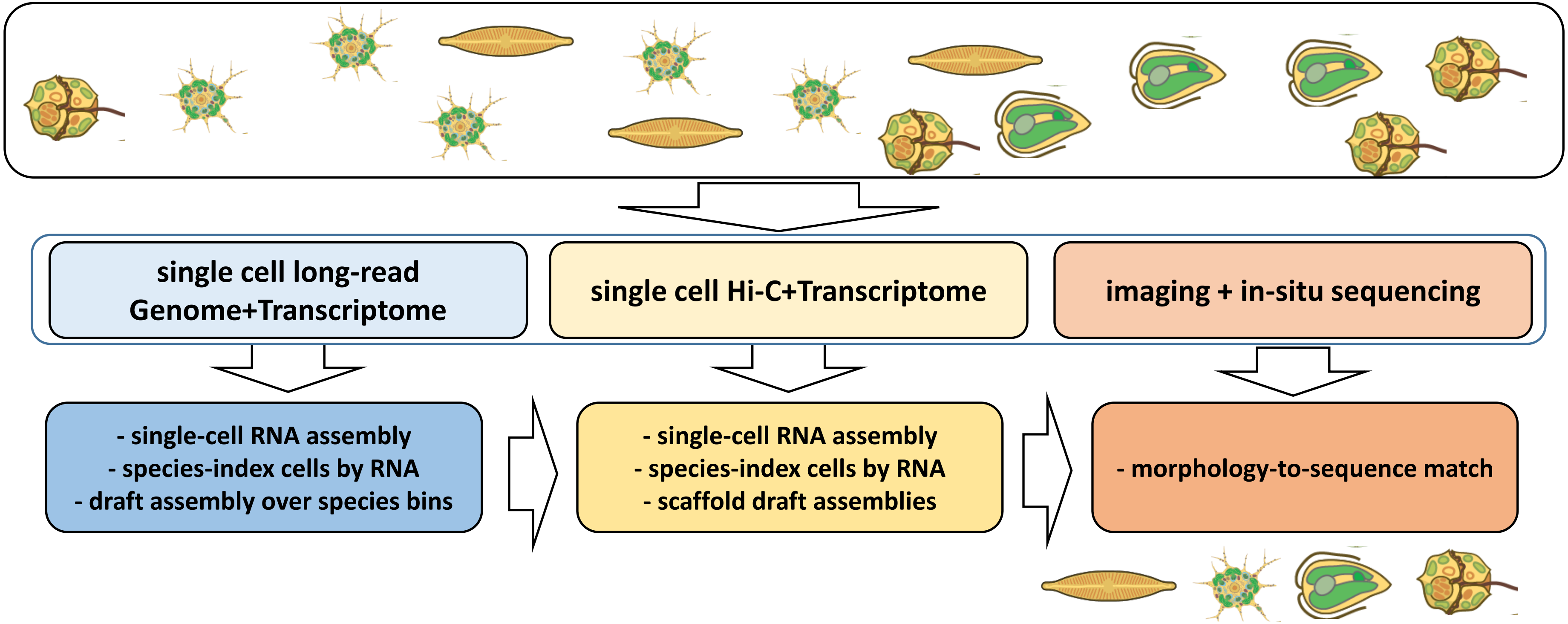
CHALLENGES

- Most of the eukaryotic diversity is uncultured/unculturable

- Much of it is not well known

- Single-cell genome sequencing for assembly does not work well currently, and it requires cell-by-cell isolation

ENVIRONMENTAL SAMPLE SPECIES-RESOLVED ASSEMBLY



single cell long-read Genome+Transcriptome

single cell Hi-C+Transcriptome

imaging + in-situ sequencing

- single-cell RNA assembly
- species-index cells by RNA
- draft assembly over species bins

- single-cell RNA assembly
- species-index cells by RNA
- scaffold draft assemblies

- morphology-to-sequence match

Throughput: <u>millions of cells at a time</u>

TREE-OF-LIFE REGULATORY ATLAS

- Goals:
  - Annotated ENCODE-like regulatory atlases for all major eukaryotic lineages

  - Identifying cell lineages, their trajectories and specifiers in all multicellular clades

  - Long-term: mapping gene regulation at the single-cell level in the environment

TECHNICAL CHALLENGES

- ATAC-seq does not work if there are bacteria around
- Tn5 is a magnet for transposase and all reads end up bacterial because of no nucleosomes in bacteria
- Environmental samples are impossible/very hard to profile
- So are macroscopic species that have tightly associating epibionts and other symbionts
- So are predatory protozoans, unless they are subjected to prolonged starvation (not desirable)

ENHANCED SMF/SMAC-SEQ

- dense coverage, e.g. all As, Cs, Ts, or Gs
- direct base conversion
- can be amplified but can also be read directly
- can do both short and long reads
- can do both short and long reads in a multiome format (with other modalities too)
- allows for environmental and contaminated single-cell multiome profiling
- dense coverage allows for precise nucleosome and TF footprinting, thus regulatory atlas building
- long read formats and dense absolute occupancy maps allow charting exotic genomic landscapes like organellar genomes in protists

EXPANDING THE MOLECULAR TOOLKIT

- Single-molecule readouts of active transcription

- Single-molecule readouts of base-pair protein-DNA association

- RNA single-molecule footprinting

- Single-molecule multiomics (DNA and RNA)

ACKNOWLEDGEMENTS

William Greenleaf
Anshul Kundaje

Zohar Shipony
Alex E. Trevino
Samuel H. Kim
Soon-Il Higashino

Arthur R. Grossman
John R. Pringle
Tingting Xiang

Chuan He
Tong Wu

Erez-Lieberman Aiden
Olga Dudchenko

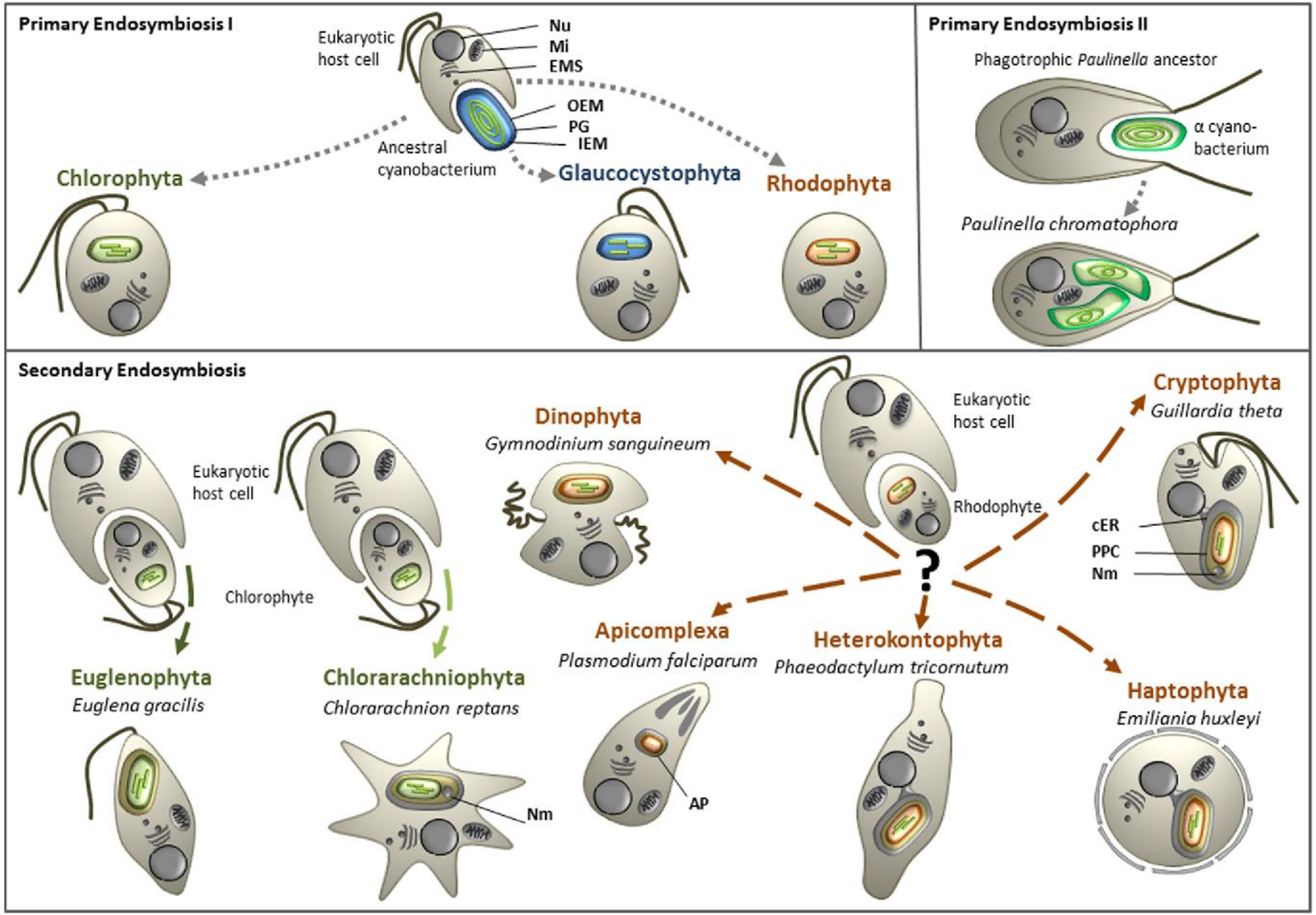
ENCODE

Center For Personal Dynamic Regulomes

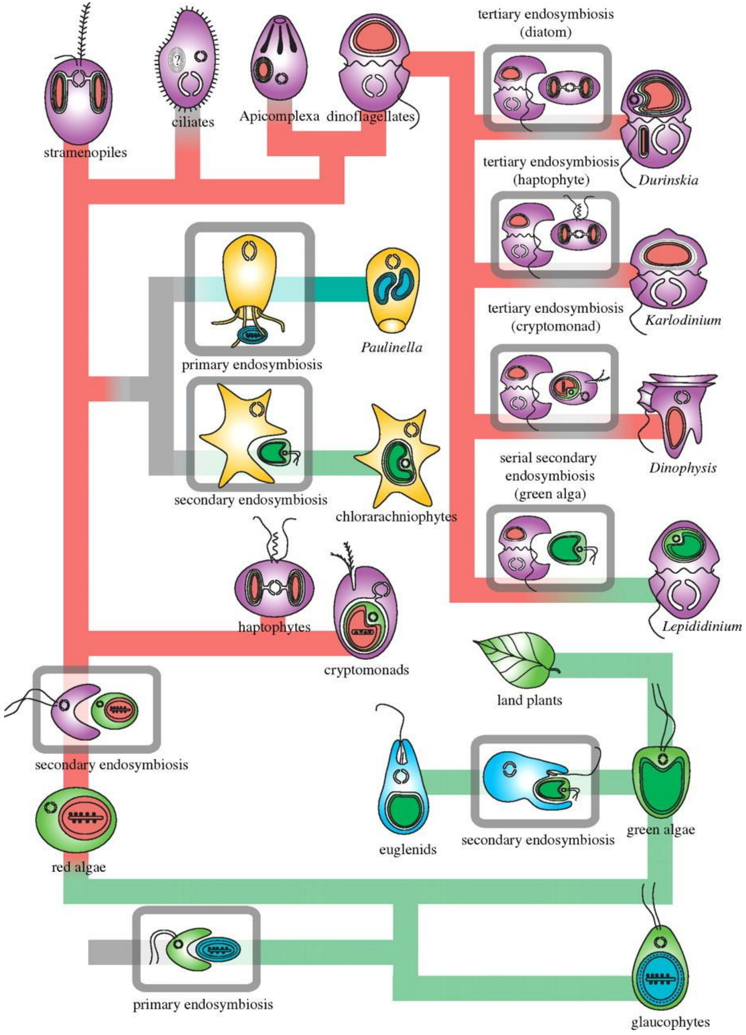
CHAN ZUCKERBERG BIOHUB

NUCLEOMORPHS

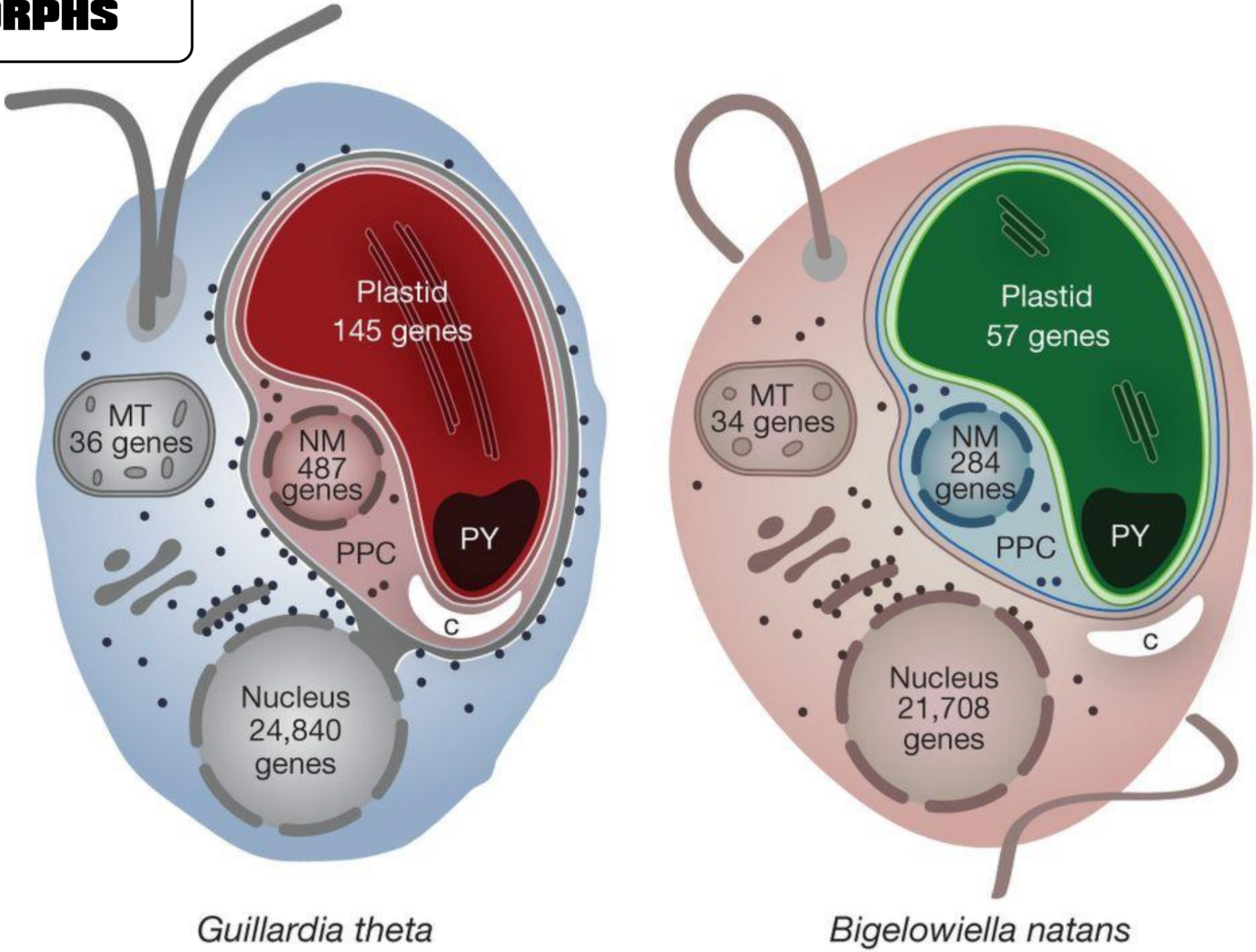
SECONDARY ENDOSYMBIOSIS



TERTIARY ENDOSYMBIOSIS

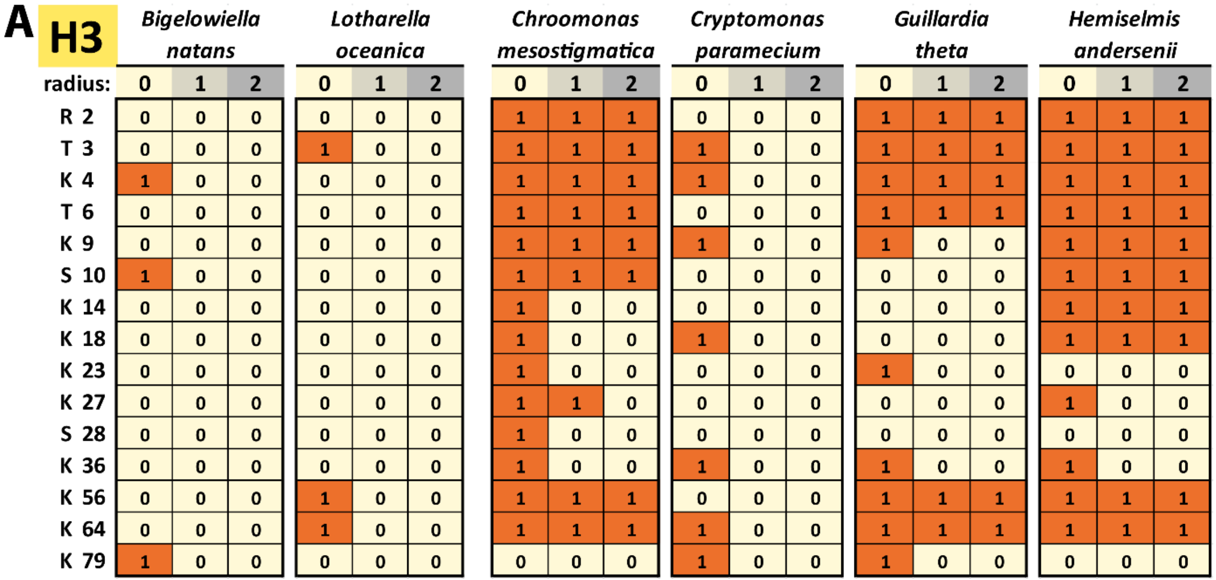


NUCLEOMORPHS

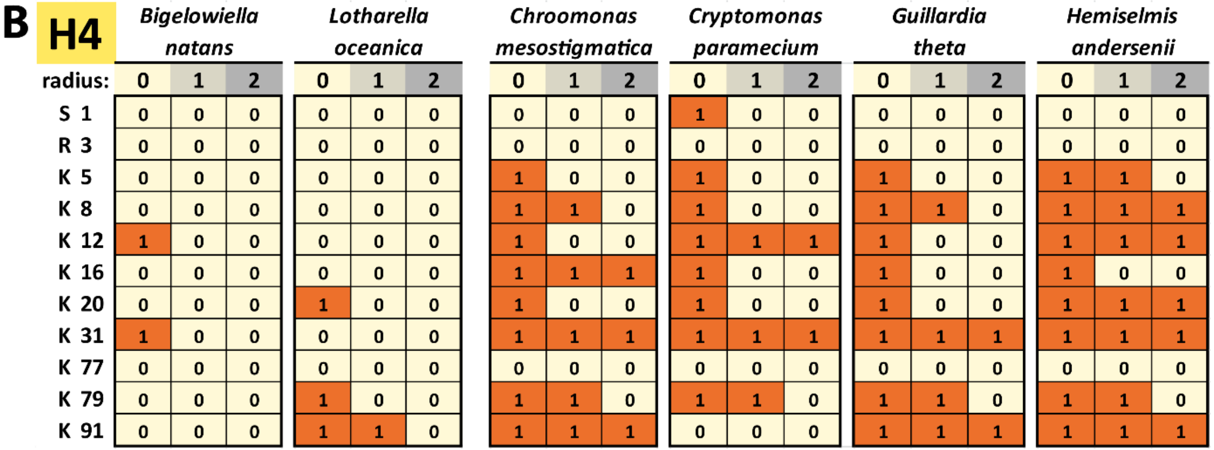


SPECIAL PROPERTIES:

- Most reduced known eukaryote genomes
- Overlapping genes, extremely little intergenic space
- Surprising convergence between the two groups – 3 chromosomes with subtelomeric rDNA arrays
- Divergent histone code and RNA Pol2 CTD

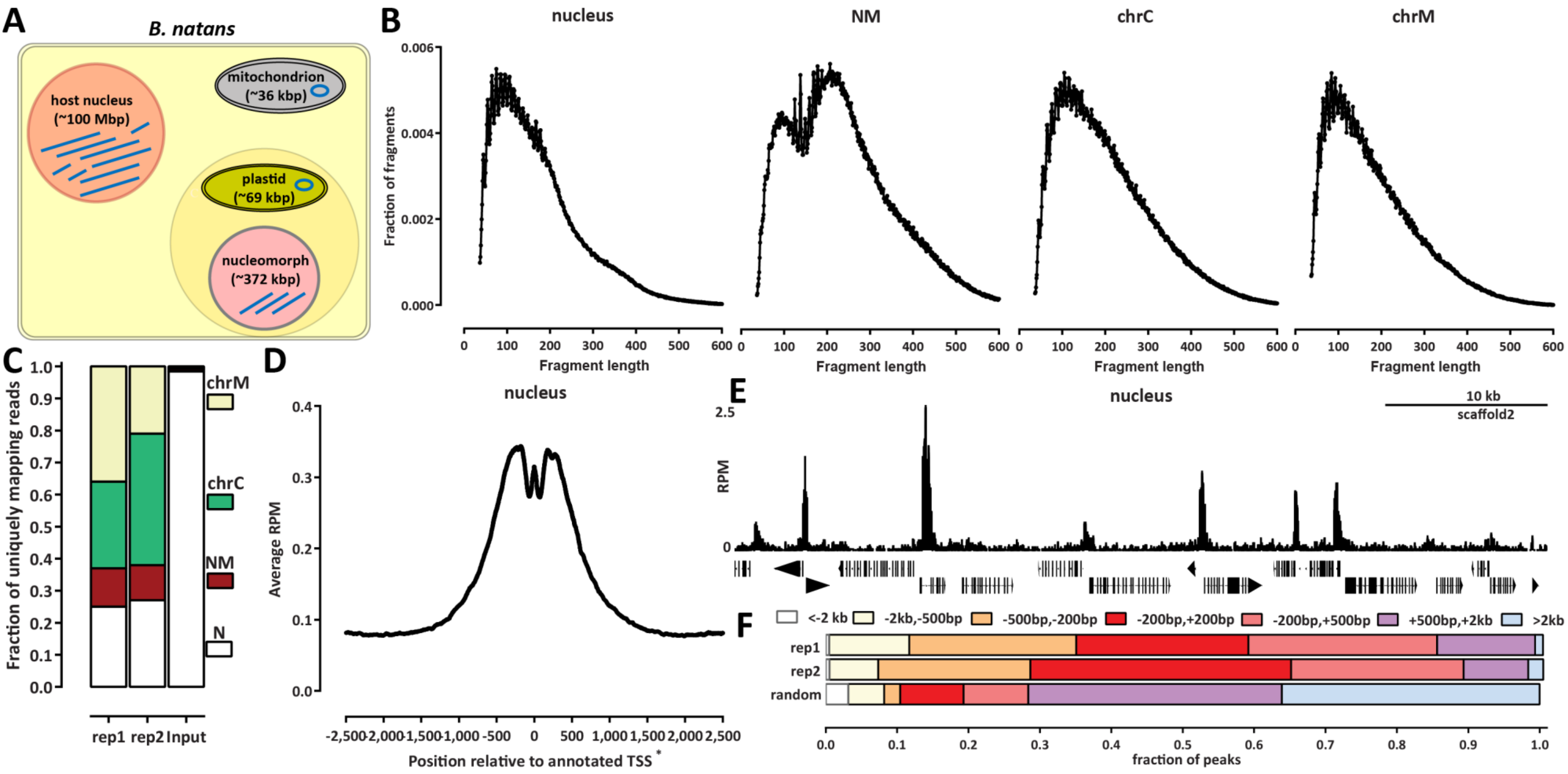


	Bigelowiella natans			*Lotharella oceanica*			*Chroomonas mesostigmatica*			*Cryptomonas paramecium*			*Guillardia theta*			*Hemiselmis andersenii*		
radius:	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2
R 2	0	0	0	0	0	0	1	1	1	0	0	0	1	1	1	1	1	1
T 3	0	0	0	1	0	0	1	1	1	1	0	0	1	1	1	1	1	1
K 4	1	0	0	0	0	0	1	1	1	1	0	0	1	1	1	1	1	1
T 6	0	0	0	0	0	0	1	1	1	0	0	0	1	1	1	1	1	1
K 9	0	0	0	0	0	0	1	1	1	1	0	0	1	0	0	1	1	1
S 10	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	1	1	1
K 14	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	1
K 18	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	1	1
K 23	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
K 27	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0
S 28	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
K 36	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0
K 56	0	0	0	1	0	0	1	1	1	0	0	0	1	1	1	1	1	1
K 64	0	0	0	1	0	0	1	1	1	1	0	0	1	1	1	1	1	1
K 79	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0



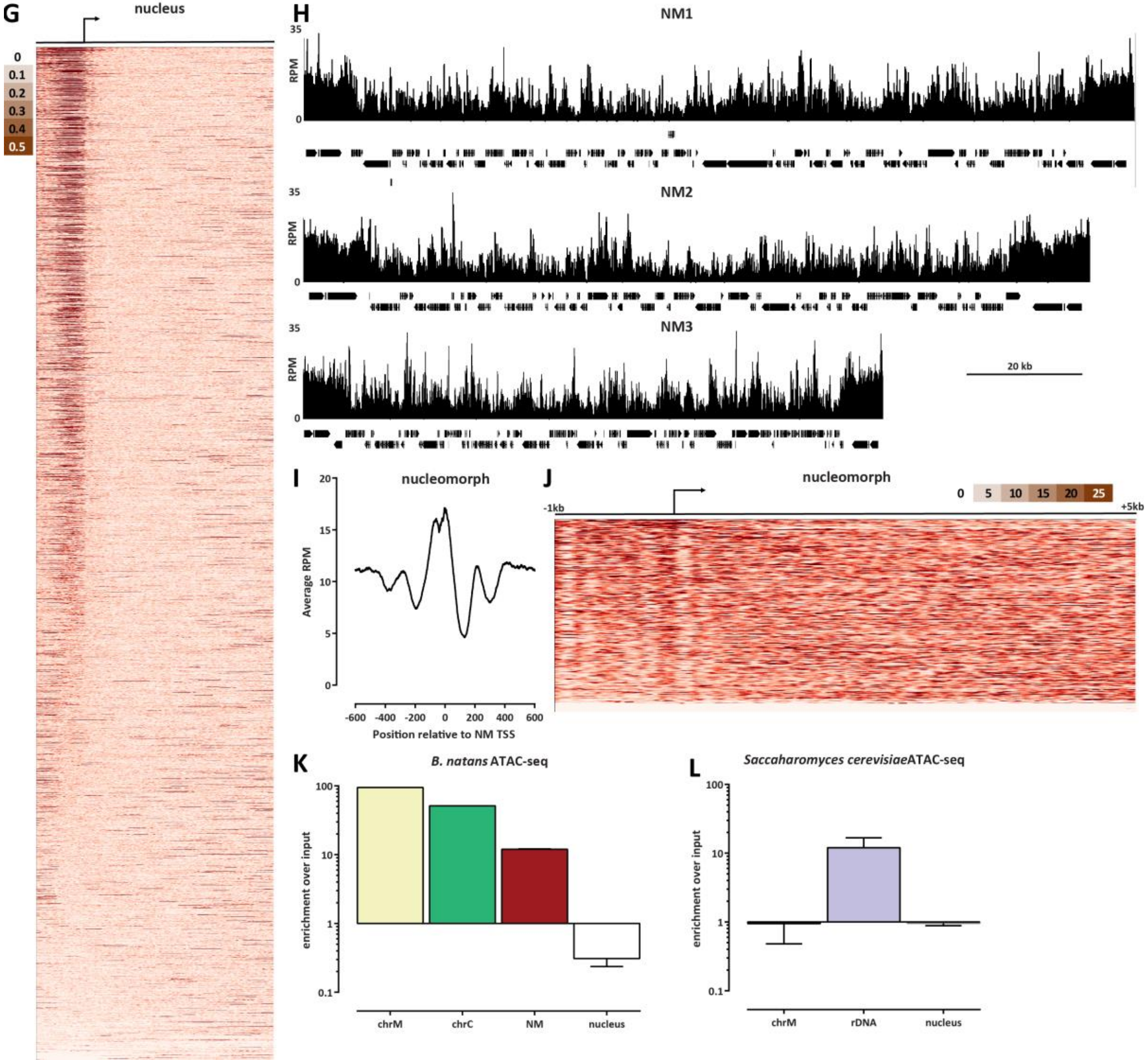
	Bigelowiella natans			*Lotharella oceanica*			*Chroomonas mesostigmatica*			*Cryptomonas paramecium*			*Guillardia theta*			*Hemiselmis andersenii*		
radius:	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2
S 1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
R 3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K 5	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	1	1	0
K 8	0	0	0	0	0	0	1	1	0	1	0	0	1	1	0	1	1	1
K 12	1	0	0	0	0	0	1	0	0	1	1	1	1	0	0	1	1	1
K 16	0	0	0	0	0	0	1	1	1	1	0	0	1	0	0	1	0	0
K 20	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	1	1
K 31	1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
K 77	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K 79	0	0	0	1	0	0	1	1	0	1	1	0	1	1	0	1	1	0
K 91	0	0	0	1	1	0	1	1	1	0	0	0	1	1	1	1	1	1

ATAC-SEQ

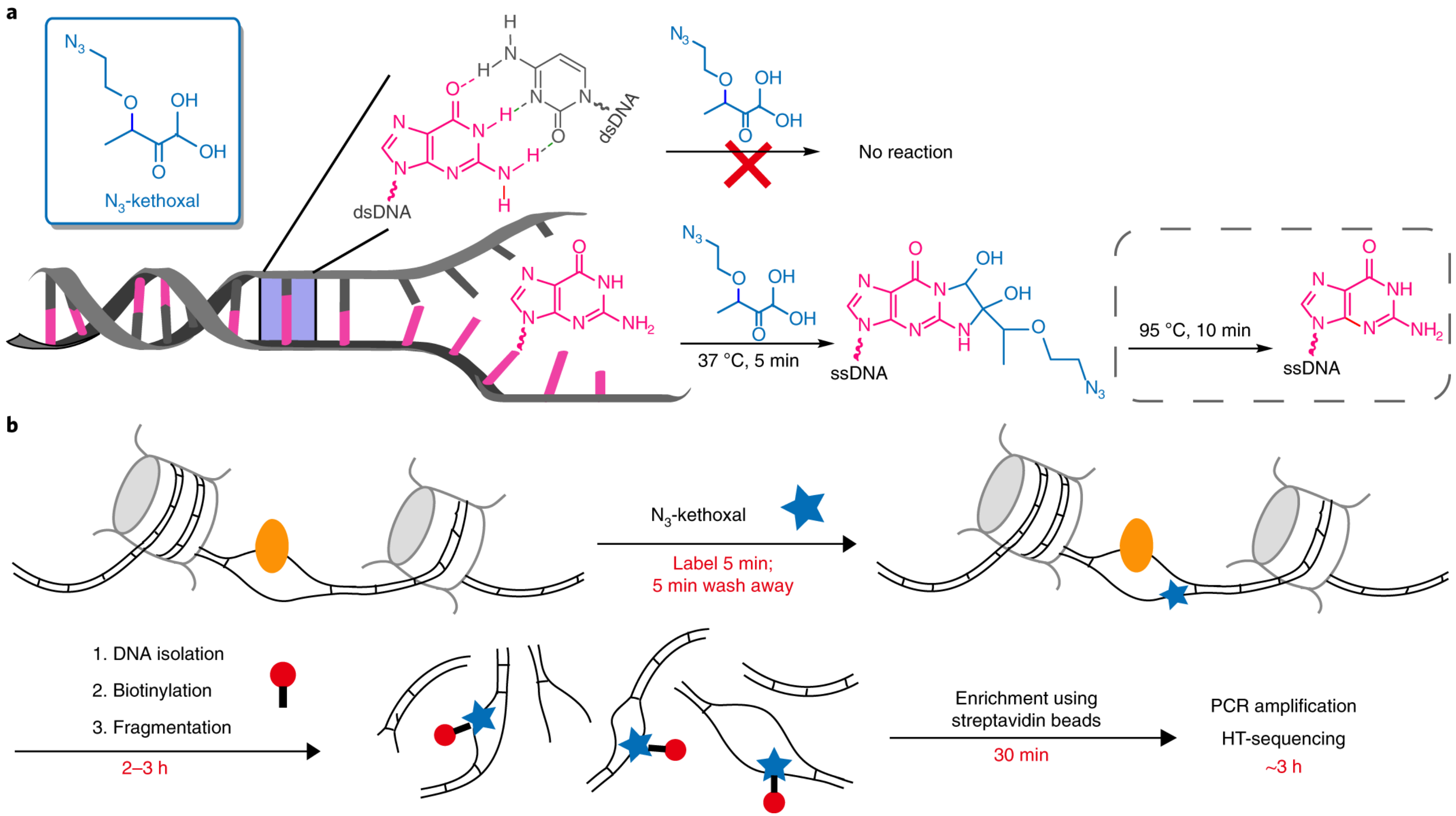


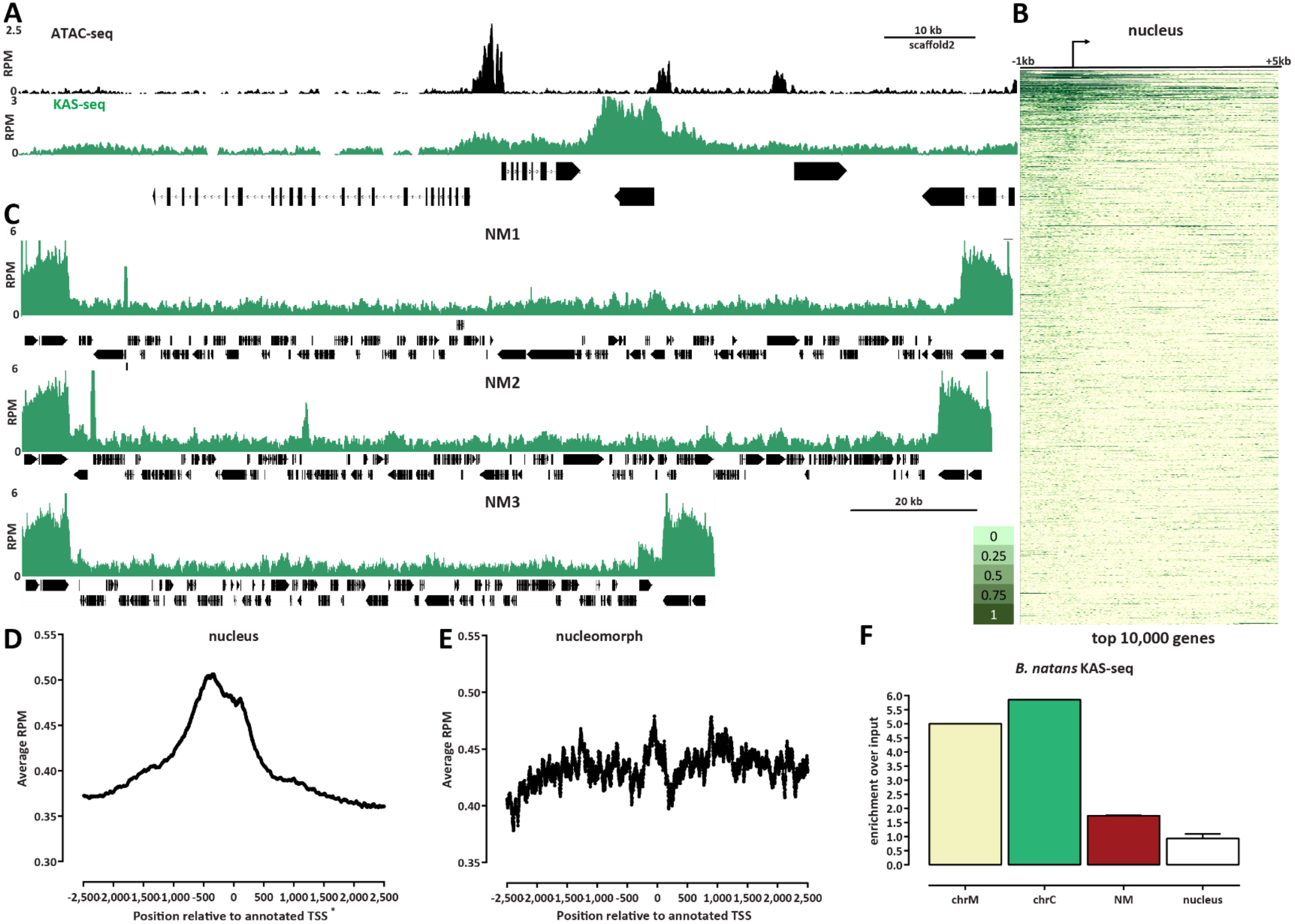
A
B. natans
host nucleus
(~100 Mbp)
mitochondrion
(~36 kbp)
plastid
(~69 kbp)
nucleomorph
(~372 kbp)
B
nucleus
NM
chrC
chrM
Fraction of fragments
0.006
0.004
0.002
0.000
0
100
200
300
400
500
600
Fragment length
C
Fraction of uniquely mapping reads
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
rep1
rep2
Input
chrM
chrC
NM
N
D
nucleus
Average RPM
0.4
0.3
0.2
0.1
0.0
-2,500
-2,000
-1,500
-1,000
-500
0
500
1,000
1,500
2,000
2,500
Position relative to annotated TSS *
E
nucleus
10 kb
scaffold2
2.5
RPM
0
F
<-2 kb
-2kb,-500bp
-500bp,-200bp
-200bp,+200bp
-200bp,+500bp
+500bp,+2kb
>2kb
rep1
rep2
random
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
fraction of peaks

ATAC

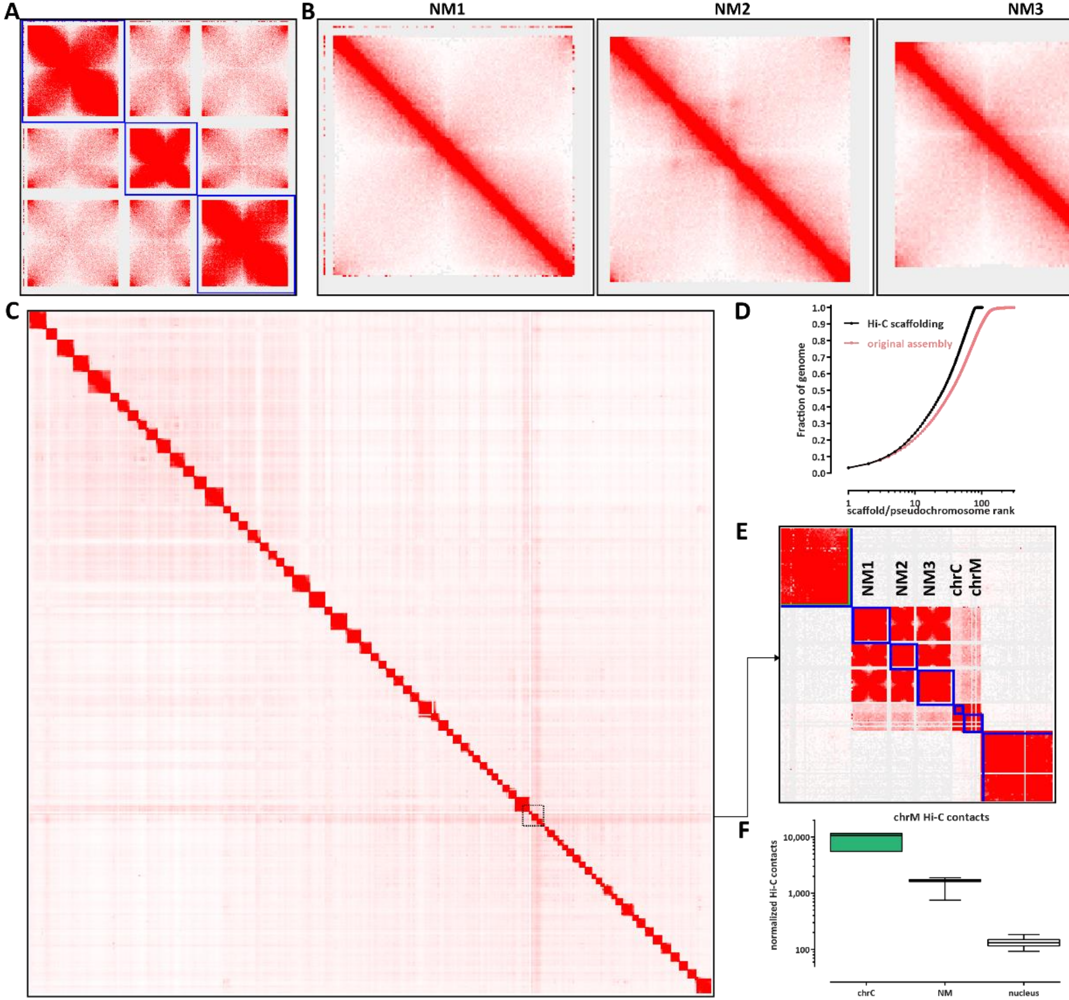

G
nucleus
0
0.1
0.2
0.3
0.4
0.5
H
NM1
35
RPM
0
NM2
35
RPM
0
NM3
35
RPM
0
20 kb
I
nucleomorph
20
15
10
5
0
Average RPM
-600
-400
-200
0
200
400
600
Position relative to NM TSS
J
nucleomorph
0
5
10
15
20
25
-1kb
+5kb
K
B. natans ATAC-seq
100
10
1
0.1
enrichment over input
chrM
chrC
NM
nucleus
L
Saccaharomyces cerevisiaeATAC-seq
100
10
1
0.1
enrichment over input
chrM
rDNA
nucleus

KAS-SEQ

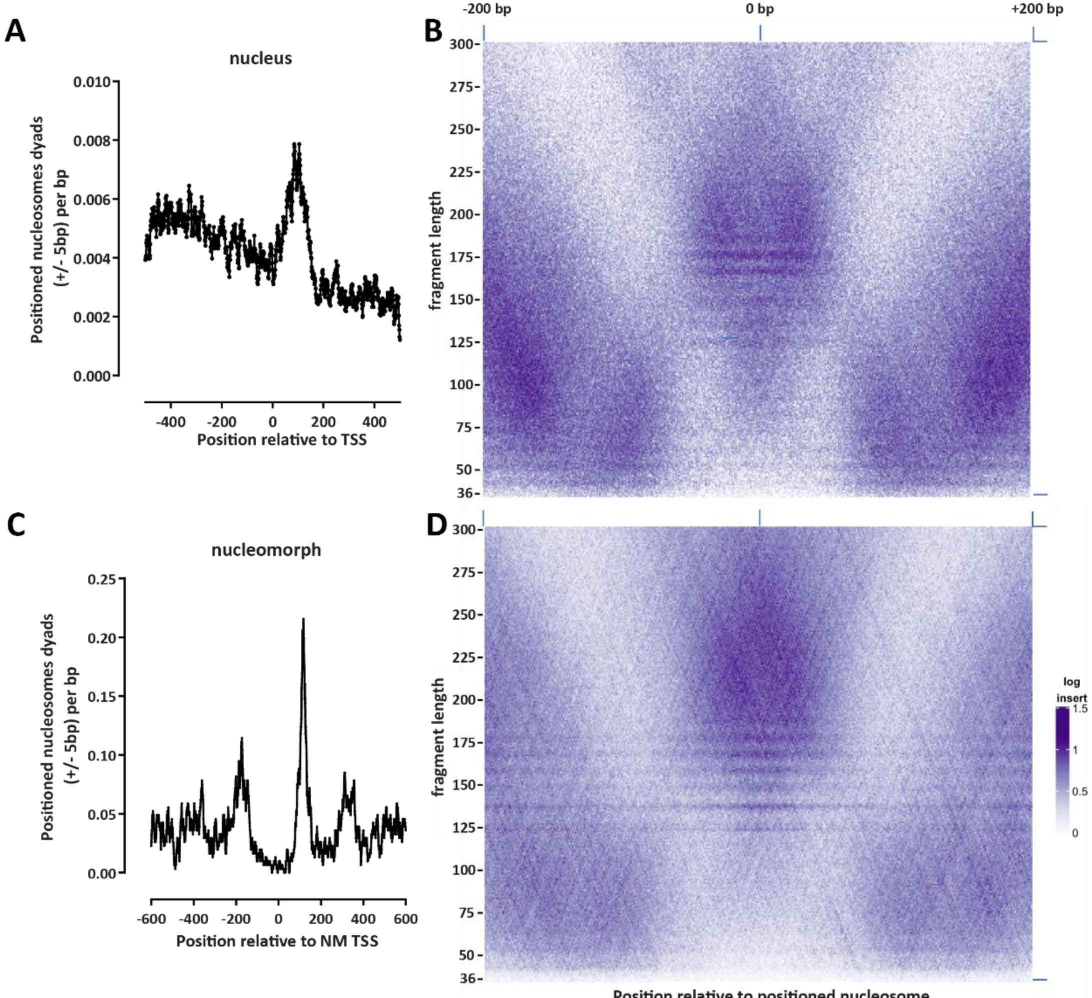

a
N_3-kethoxal
dsDNA
dsDNA
No reaction
37 °C, 5 min
ssDNA
95 °C, 10 min
ssDNA
b
N_3-kethoxal
Label 5 min;
5 min wash away
1. DNA isolation
2. Biotinylation
3. Fragmentation
2–3 h
Enrichment using
streptavidin beads
30 min
PCR amplification
HT-sequencing
~3 h


A
2.5
ATAC-seq
RPM
0
10 kb
scaffold2
3
KAS-seq
RPM
0
B
nucleus
-1kb
+5kb
C
NM1
6
RPM
0
NM2
6
RPM
0
NM3
6
RPM
0
20 kb
0
0.25
0.5
0.75
1
top 10,000 genes
D
nucleus
0.55
0.50
0.45
0.40
0.35
0.30
Average RPM
-2,500 -2,000 -1,500 -1,000 -500 0 500 1,000 1,500 2,000 2,500
Position relative to annotated TSS
E
nucleomorph
0.55
0.50
0.45
0.40
0.35
Average RPM
-2,500 -2,000 -1,500 -1,000 -500 0 500 1,000 1,500 2,000 2,500
Position relative to annotated TSS
F
B. natans KAS-seq
enrichment over input
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
chrM
chrC
NM
nucleus

HI-C


A
B
NM1
NM2
NM3
C
D
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Fraction of genome
Hi-C scaffolding
original assembly
1
10
100
scaffold/pseudochromosome rank
E
NM1
NM2
NM3
chrC
chrM
F
chrM Hi-C contacts
normalized Hi-C contacts
10,000
1,000
100
chrC
NM
nucleus

POSITIONED NUCLEOSOMES

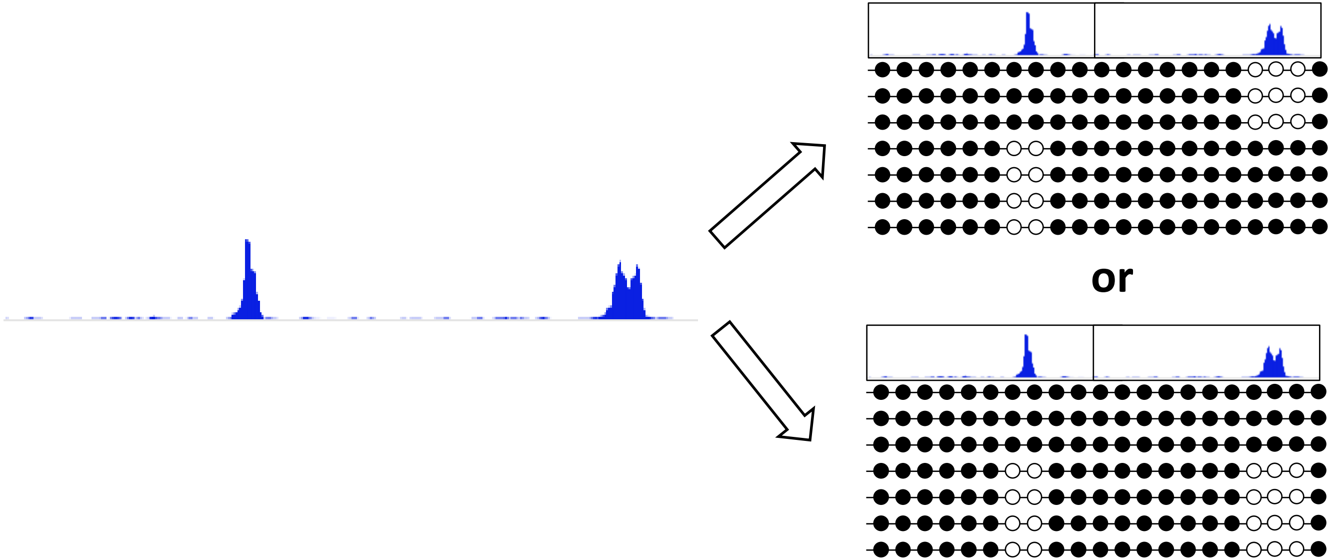


V. ADVANCING THE ASSAY TOOLKIT

SINGLE-MOLECULE CHROMATIN STATE MAPPING

- Is the chromatin status of distant regulatory regions in the same chromatin fiber correlated?

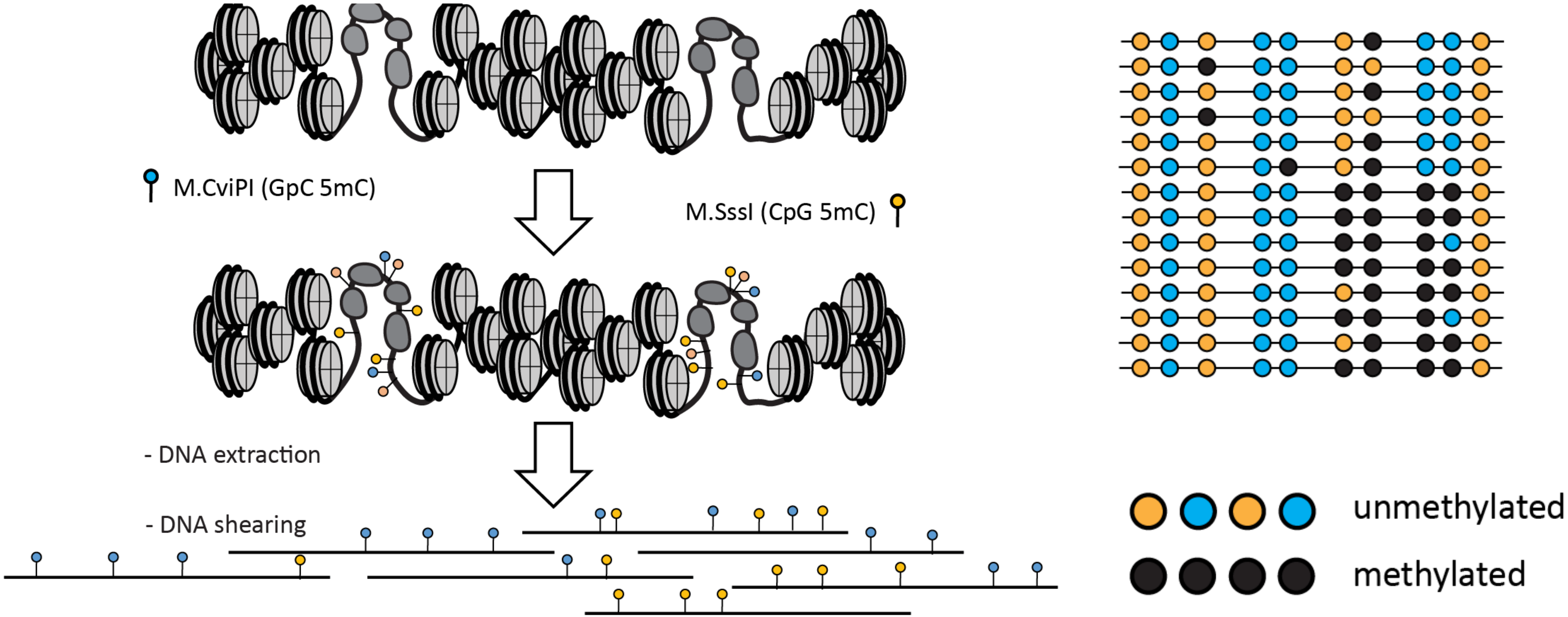
- What is the distribution of open chromatin states (chromatin haplotypes) within the population?



- scATAC does not answer these questions because of sparseness, noise and diploidy

LIMITED SINGLE MOLECULE MAPPING USING SHORT READS

NOME-SEQ/dSMF



M.CviPI (GpC 5mC)

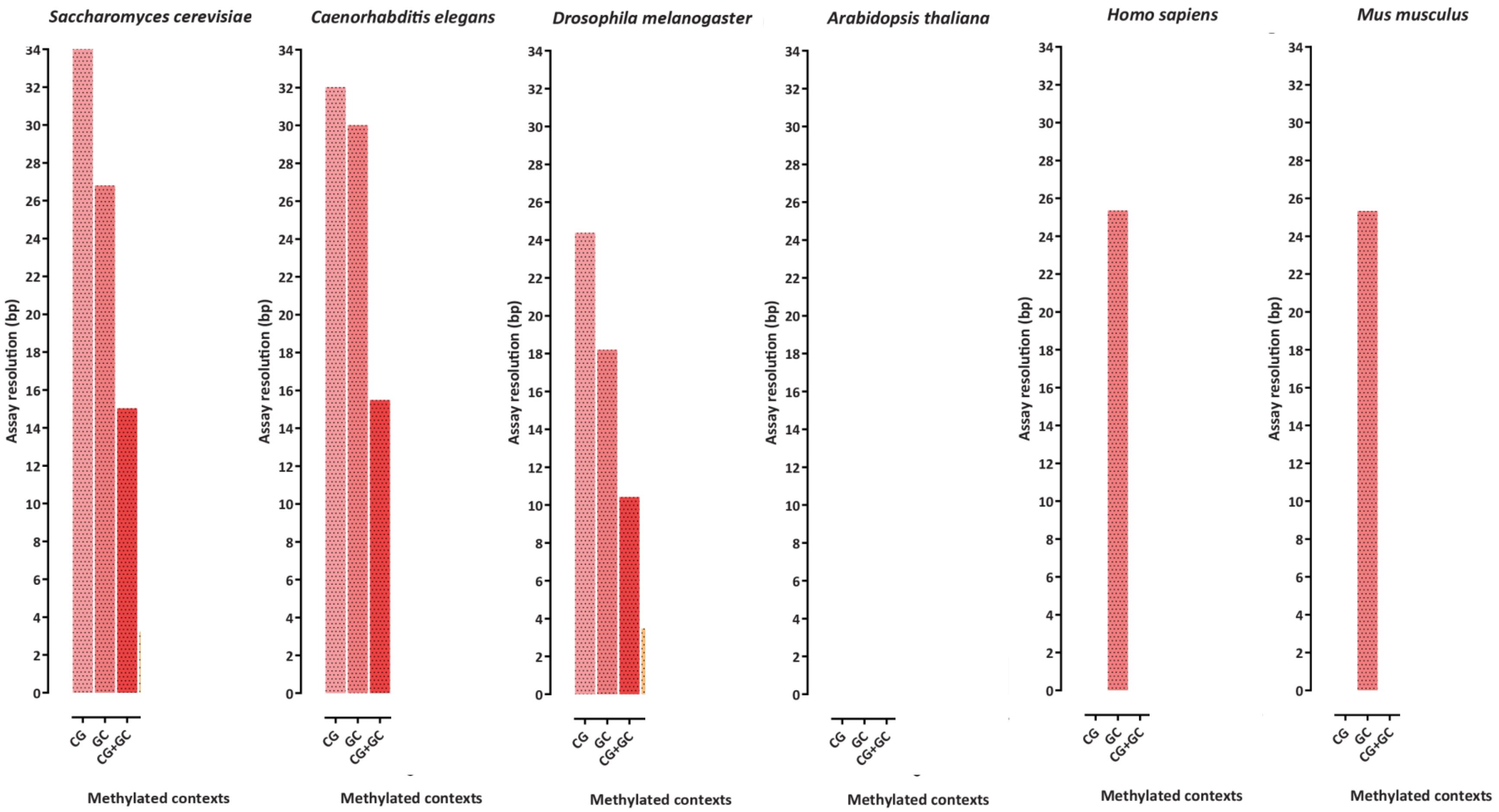
M.SssI (CpG 5mC)

- DNA extraction
- DNA shearing

unmethylated

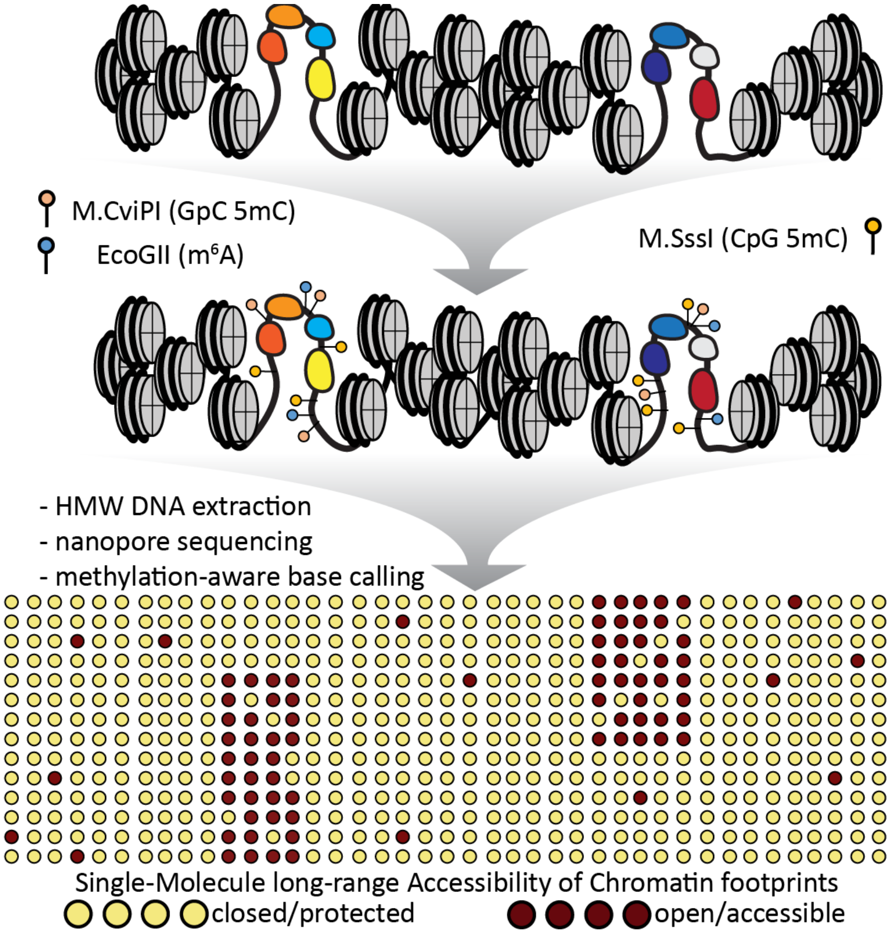
methylated

LIMITED RESOLUTION OF CG/GC METHYLTRANSFERASE-BASED ASSAYS



SINGLE-MOLECULE LONG READ ACCESSIBLE CHROMATIN MAPPING

SMAC-SEQ



M.CviPI (GpC 5mC)

EcoGII (m^{6}A)

M.SssI (CpG 5mC)

- HMW DNA extraction
- nanopore sequencing
- methylation-aware base calling

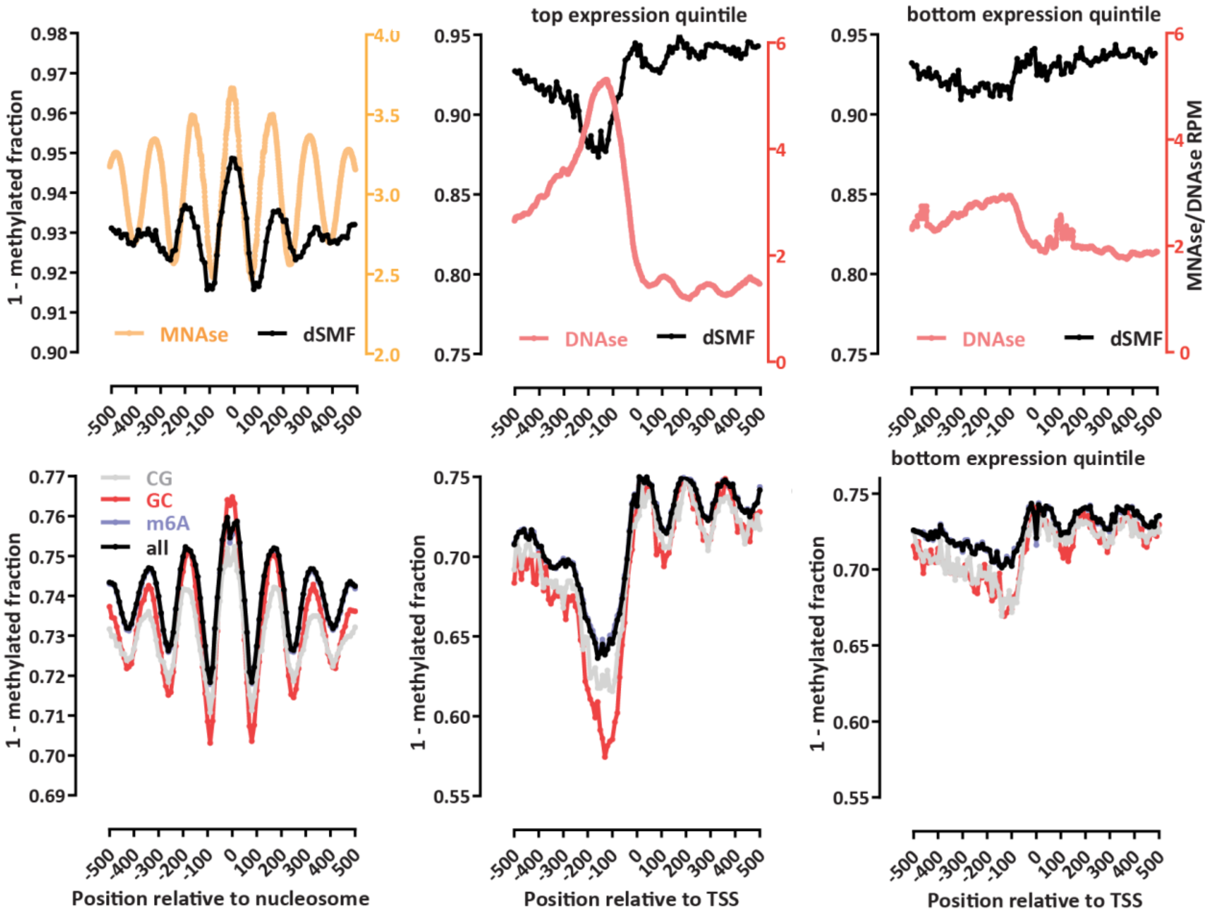
Single-Molecule long-range Accessibility of Chromatin footprints

○○○○ closed/protected ●●●● open/accessible

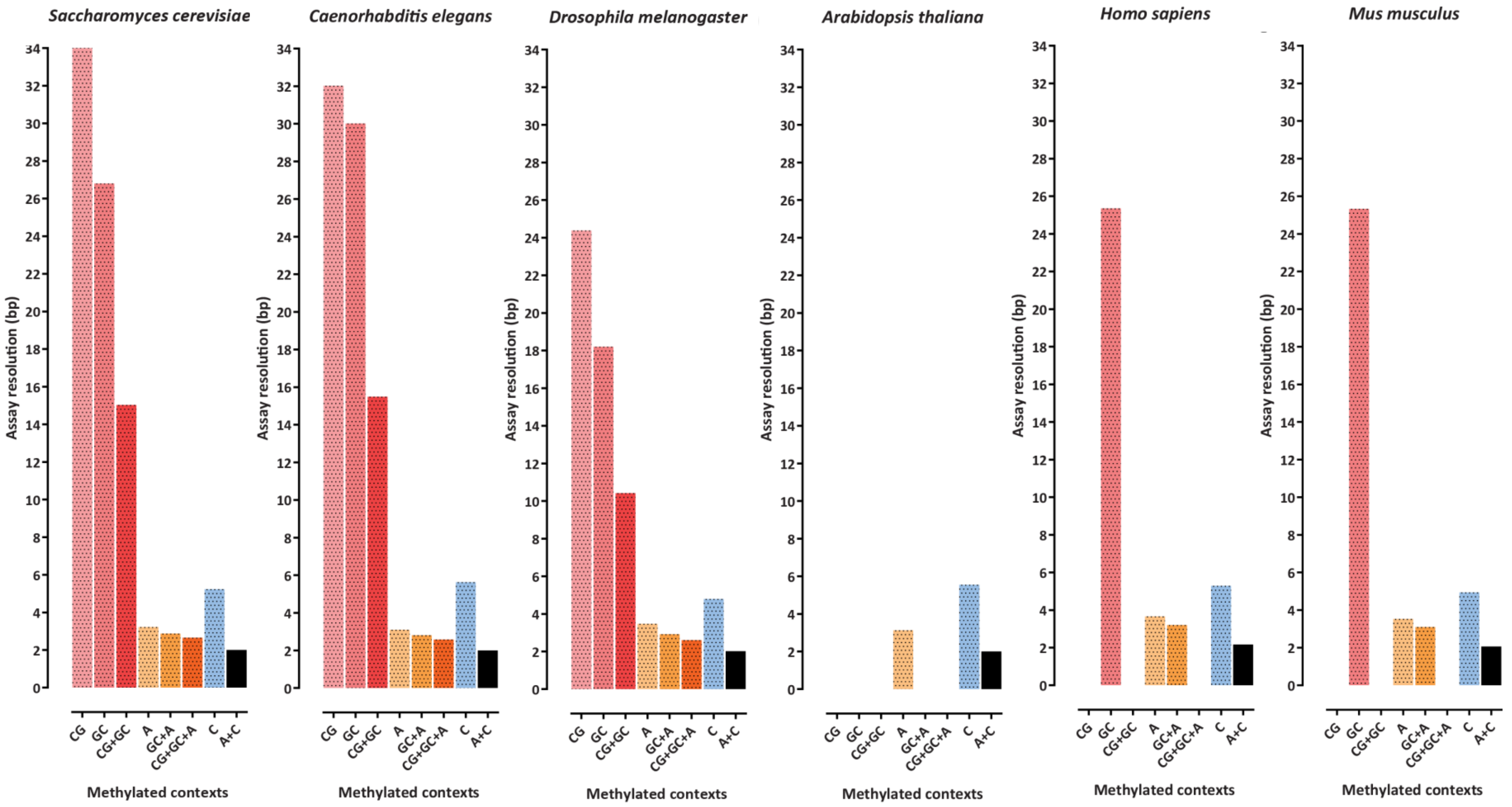
SMAC-SEQ CAPTURES THE BUDDING YEAST CHROMATIN LANDSCAPE

dSMF

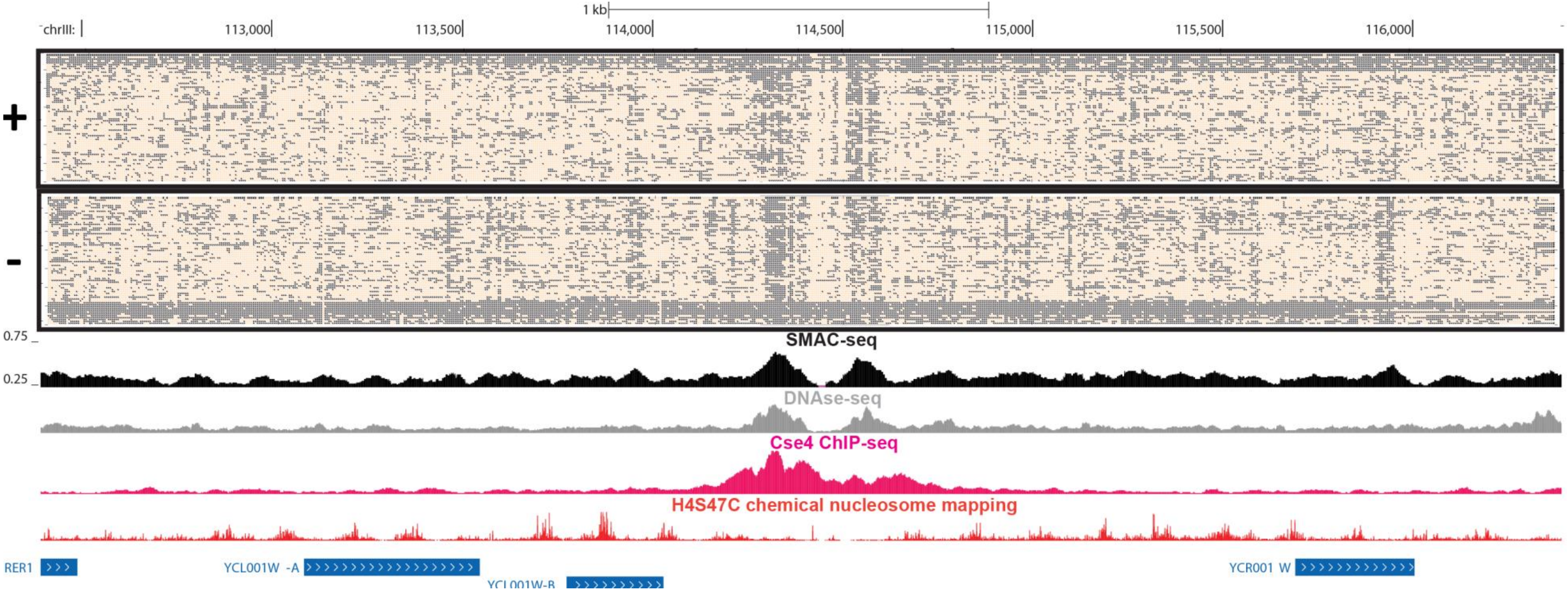
SMAC



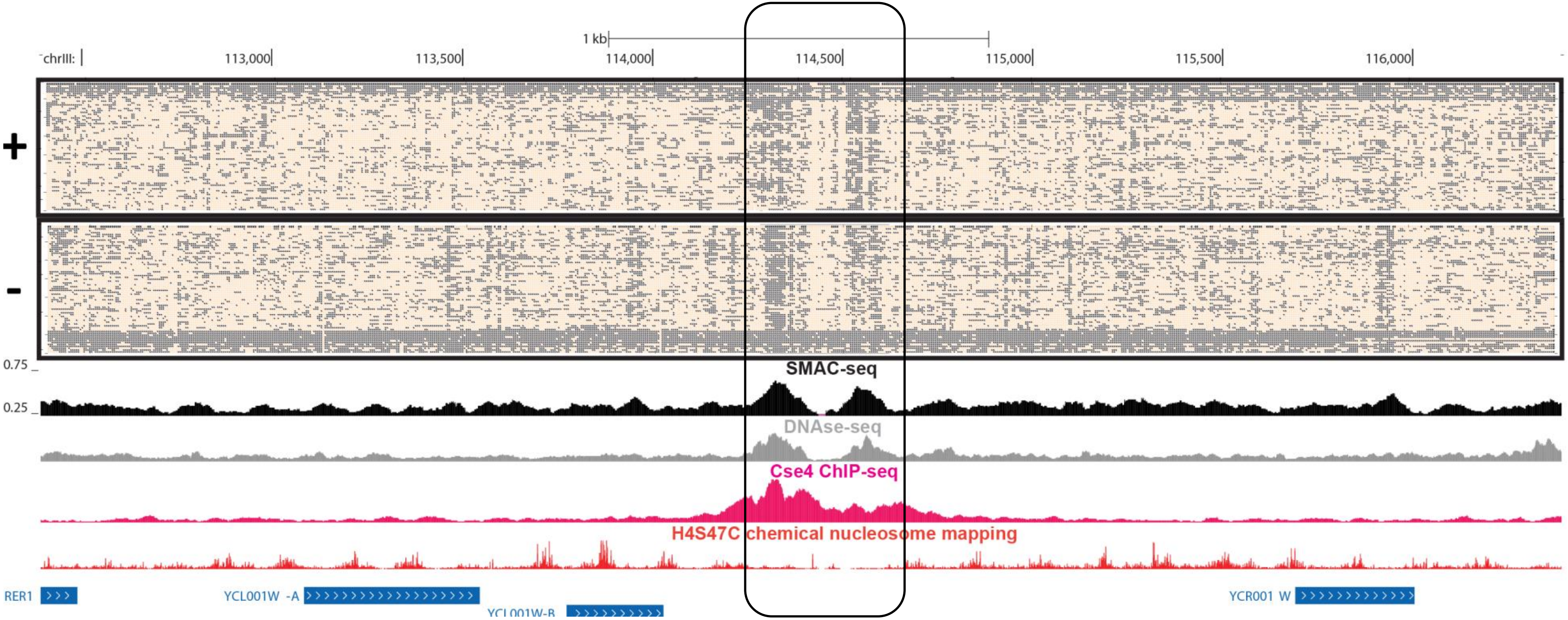
INCREASED ASSAY RESOLUTION WITH m6A



SINGLE-MOLECULE POPULATION-SCALE VIEW OF CHROMATIN HAPLOTYPES

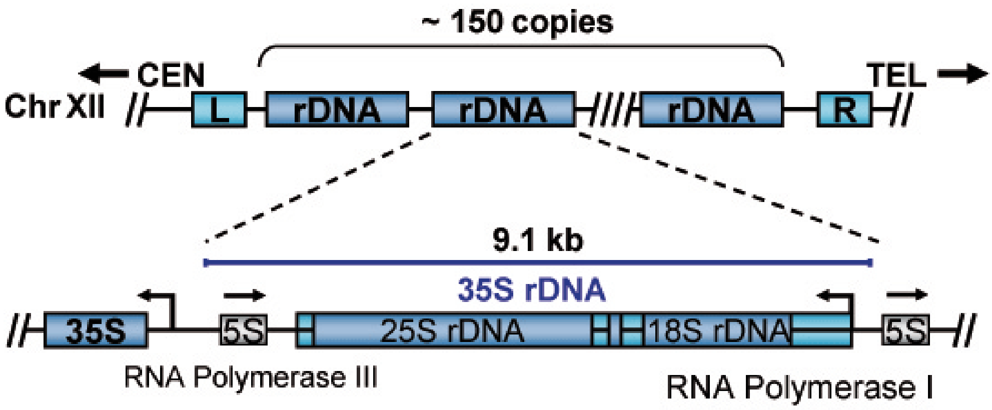


SINGLE-MOLECULE POPULATION-SCALE VIEW OF CHROMATIN HAPLOTYPES

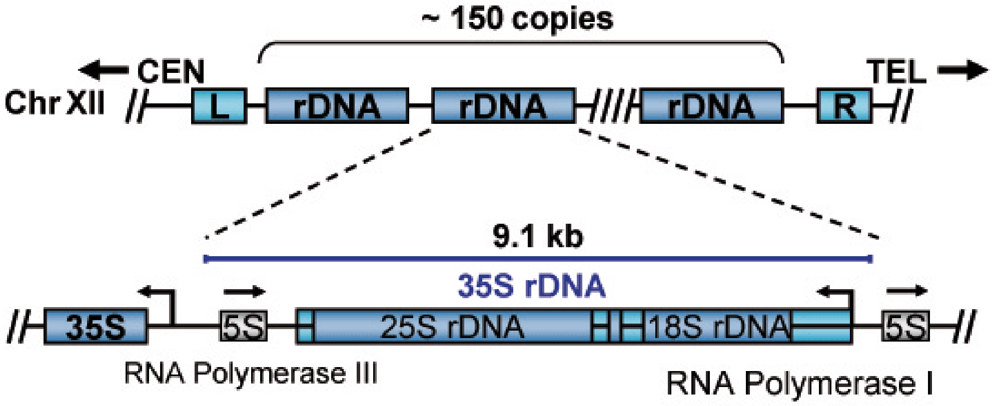


centromere

RIBOSOMAL DNA CASE STUDY

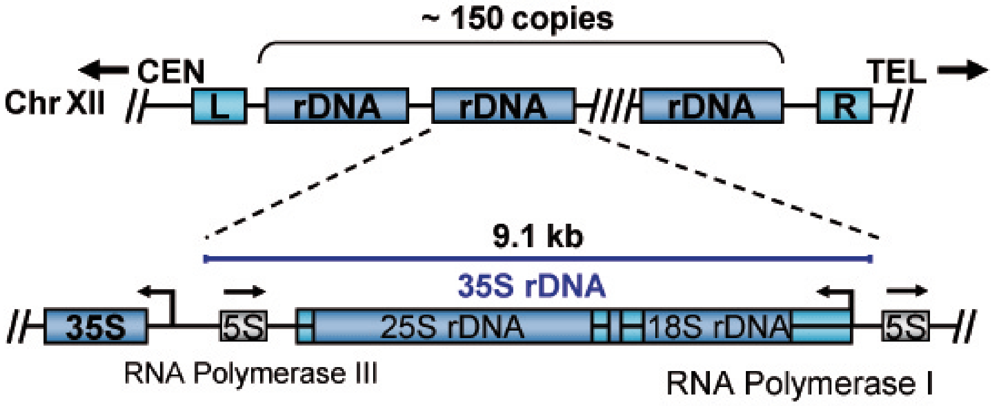


RIBOSOMAL DNA CASE STUDY



- Not uniquely mappable

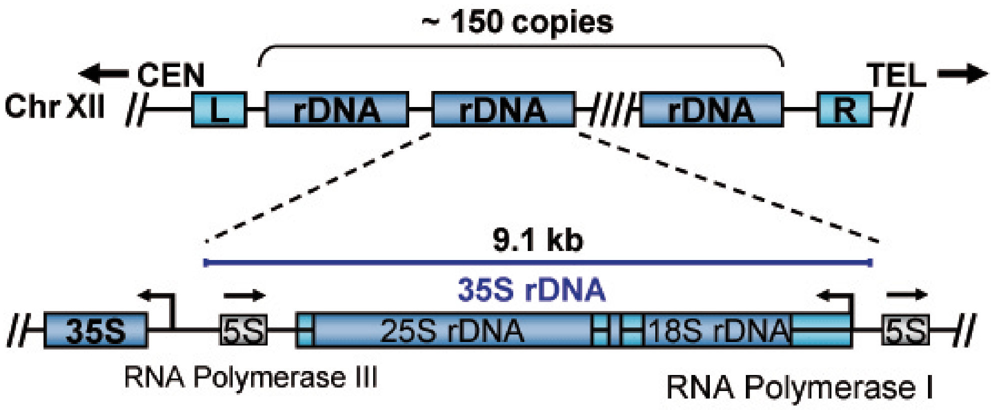
RIBOSOMAL DNA CASE STUDY



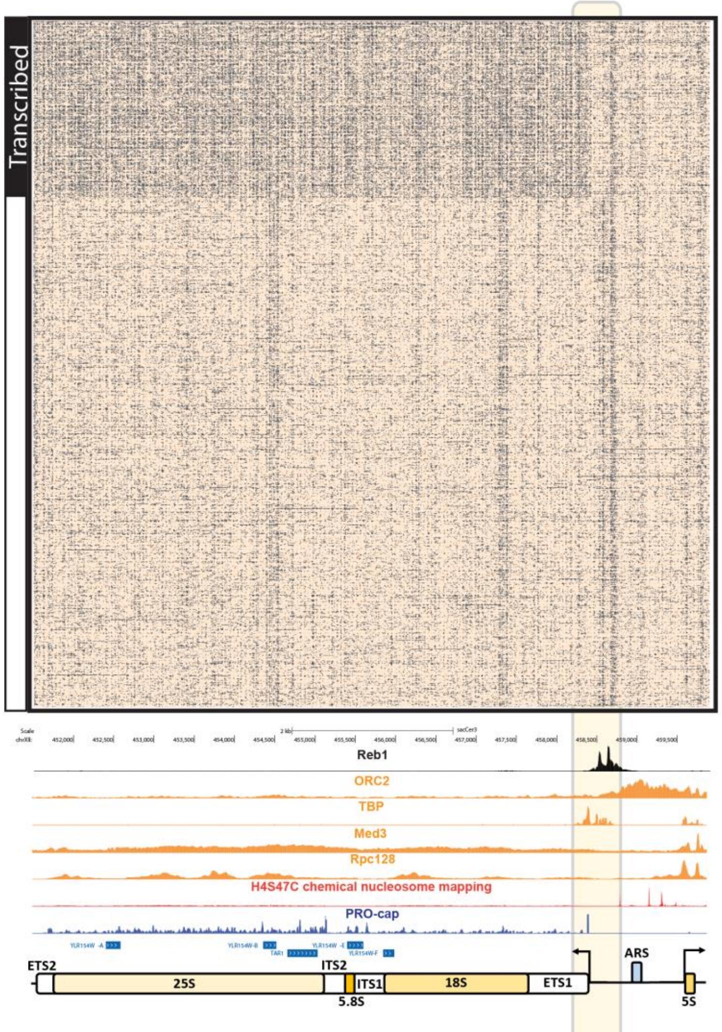
- Not uniquely mappable

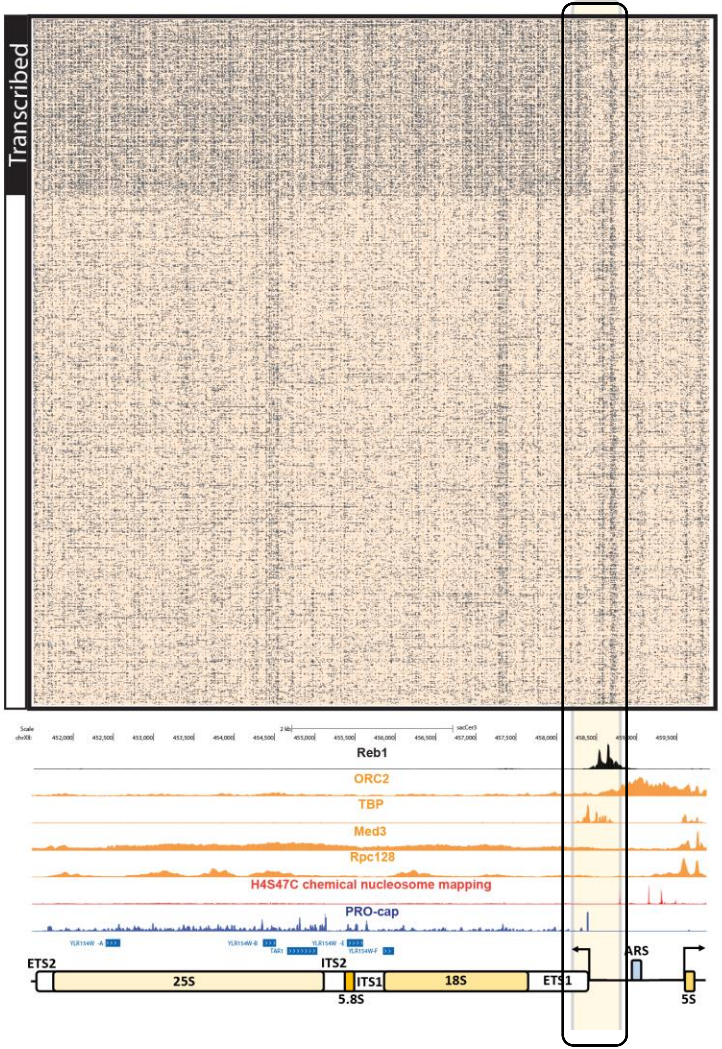
- Present in arrays of 150-200 copies in the genome (but the whole arrays are not included in the genome assembly)

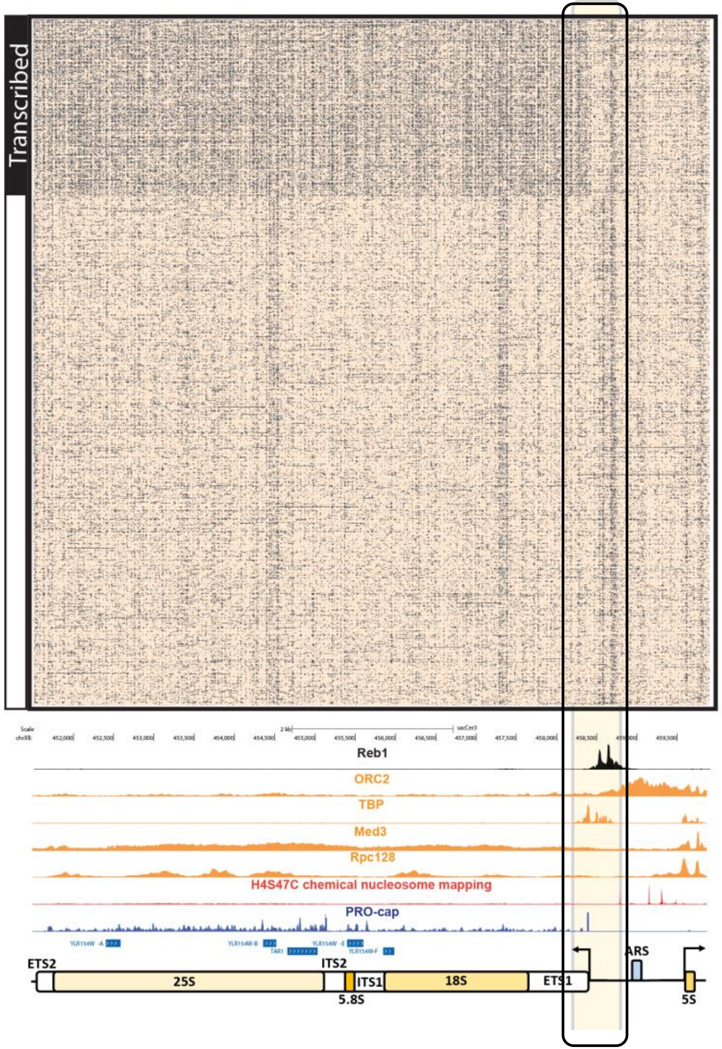
RIBOSOMAL DNA CASE STUDY



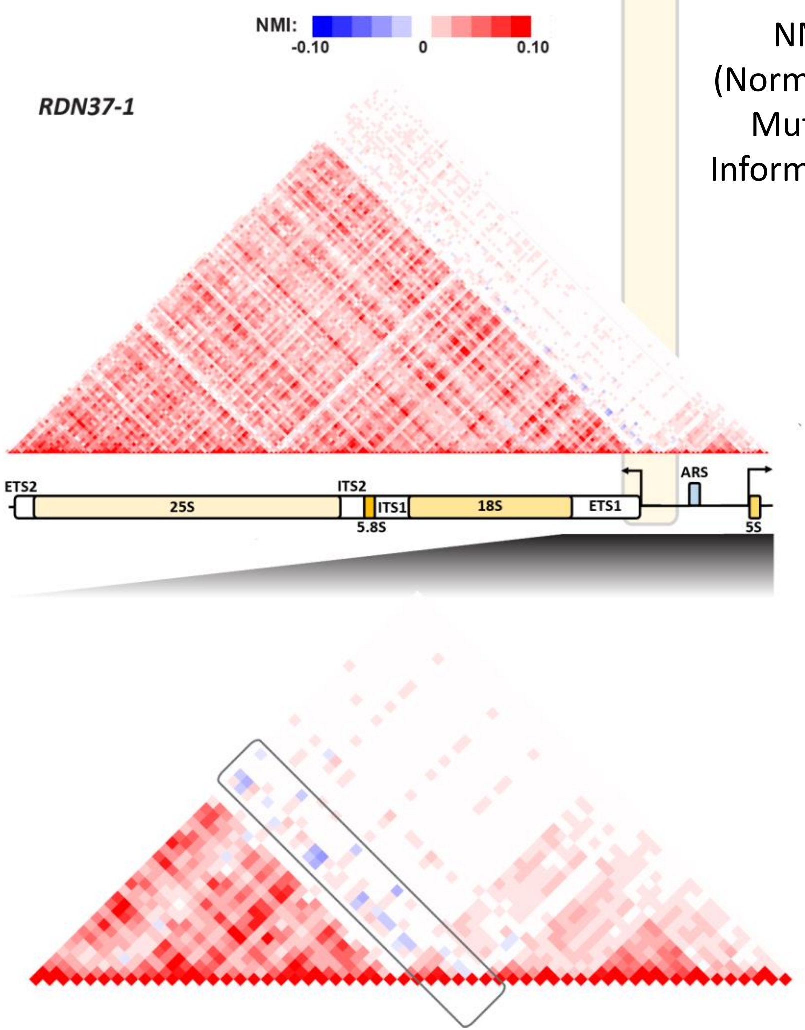
- Not uniquely mappable
- Present in arrays of 150-200 copies in the genome (but the whole arrays are not included in the genome assembly)
- Extremely highly transcribed and thought to be almost devoid of chromatin when active


Transcribed
Scale
chrXII:
452,000
452,500
453,000
453,500
454,000
454,500
2 kb
sacCer3
455,000
455,500
456,000
456,500
457,000
457,500
458,000
458,500
459,000
459,500
Reb1
ORC2
TBP
Med3
Rpc128
H4S47C chemical nucleosome mapping
PRO-cap
YLR154W -A
YLR154W-B
TAR1
YLR154W -E
YLR154W-F
ETS2
25S
ITS2
5.8S
ITS1
18S
ETS1
ARS
5S

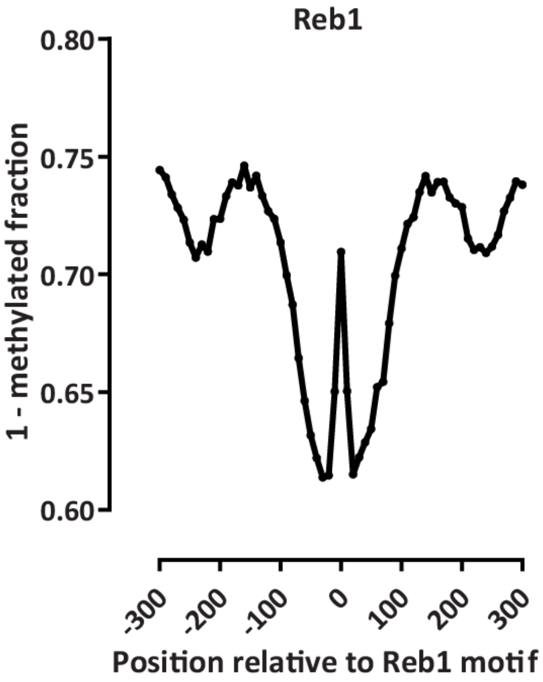

Transcribed
Scale
chrXII:
452,000
452,500
453,000
453,500
454,000
454,500
2 kb
455,000
455,500
456,000
456,500
sacCer3
457,000
457,500
458,000
458,500
459,000
459,500
Reb1
ORC2
TBP
Med3
Rpc128
H4S47C chemical nucleosome mapping
PRO-cap
YLR154W -A
YLR154W-B
TAR1
YLR154W -E
YLR154W-F
ARS
ETS2
ITS2
25S
ITS1
18S
ETS1
5.8S
5S

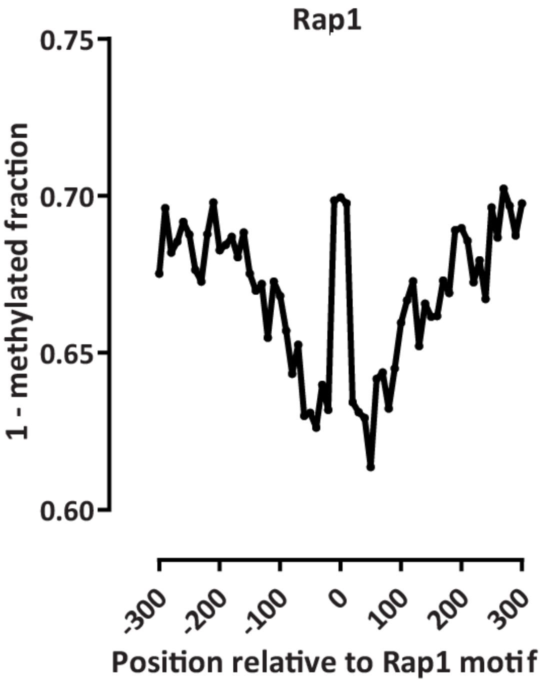


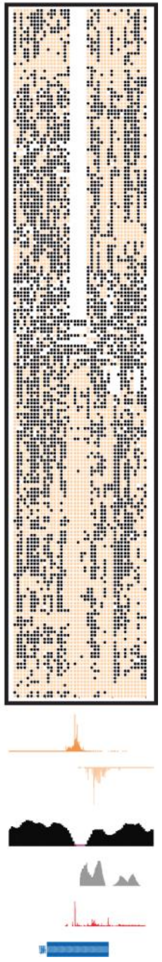
NMI
(Normalized
Mutual
Information)



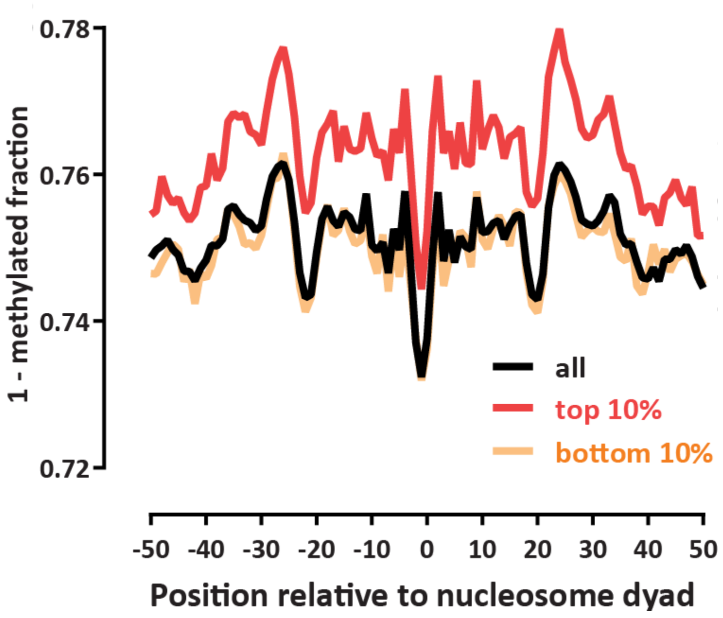
TRANSCRIPTION FACTOR FOOTPRINTING



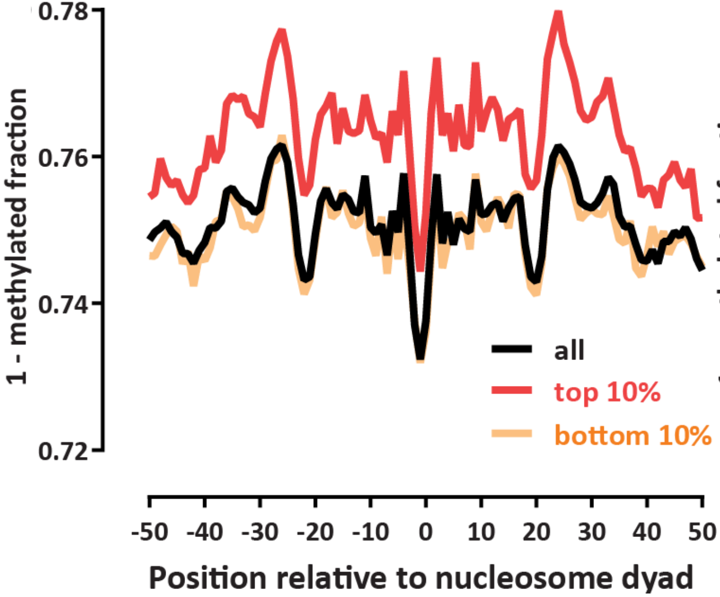


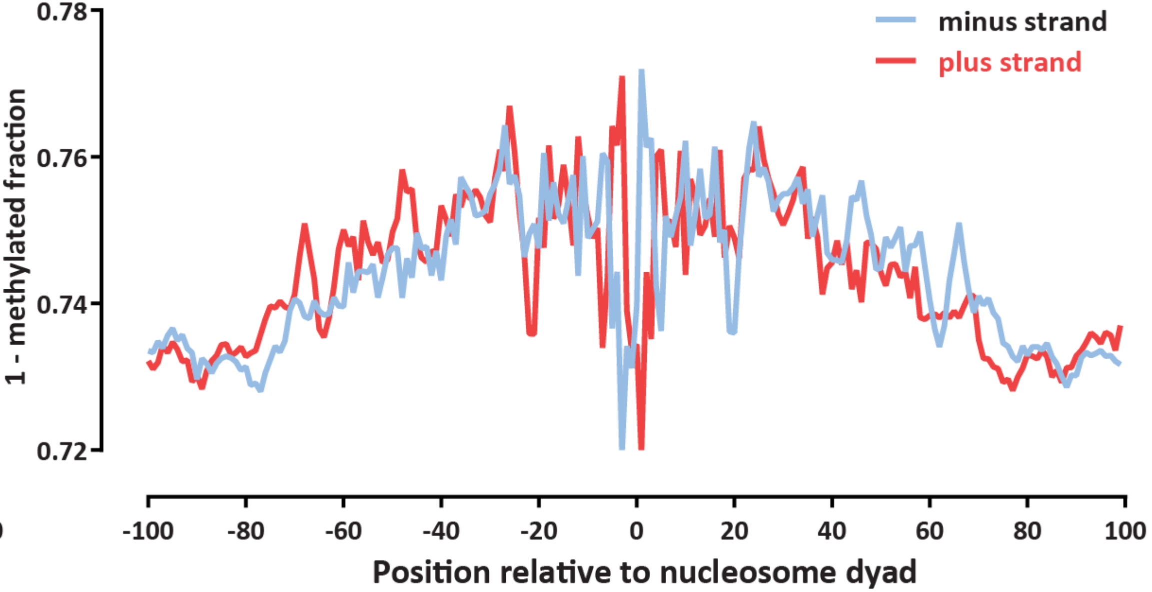


SMAC-SEQ REVEALS FINE-SCALE STRAND-SPECIFIC OCCUPANCY FEATURES

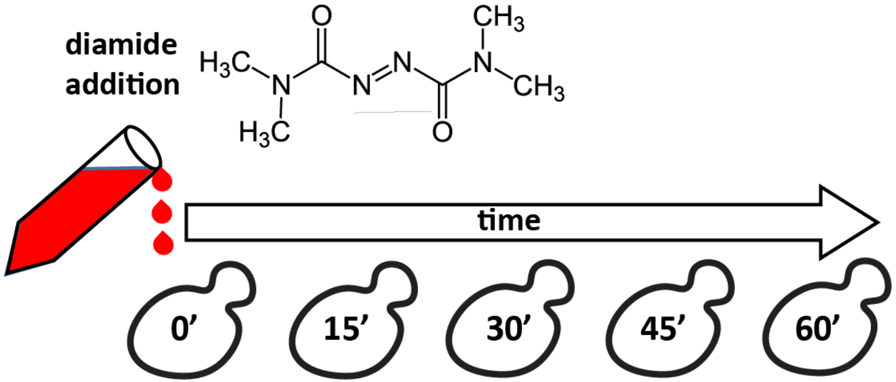


SMAC-SEQ REVEALS FINE-SCALE STRAND-SPECIFIC OCCUPANCY FEATURES

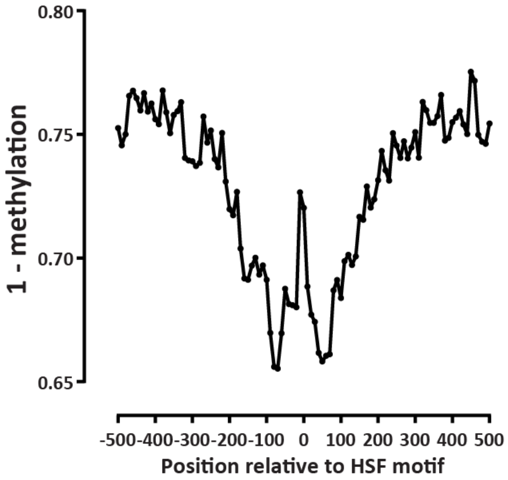


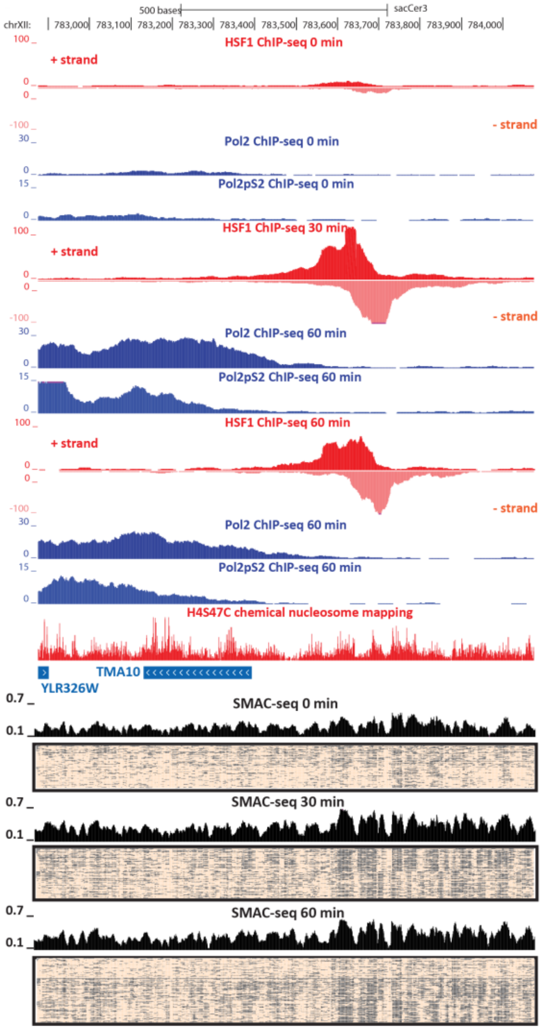


LONG-RANGE CHROMATIN DYNAMICS
ASSOCITED WITH TRANSCRIPTONAL REGULATION



	0'	15'	30'	45'	60'
SMAC-seq	✓		✓		✓
ATAC-seq	✓	✓	✓	✓	✓
RNA-seq	✓	✓	✓	✓	✓
Pol2 ChIP-seq	✓		✓		✓
Pol2pS2 ChIP-seq	✓		✓		✓
HSF1 ChIP-seq	✓		✓		✓




500 bases
sacCer3
chrXII:
783,000
783,100
783,200
783,300
783,400
783,500
783,600
783,700
783,800
783,900
784,000
HSF1 ChIP-seq 0 min
+ strand
- strand
100
0
-100
Pol2 ChIP-seq 0 min
30
Pol2pS2 ChIP-seq 0 min
15
HSF1 ChIP-seq 30 min
Pol2 ChIP-seq 60 min
Pol2pS2 ChIP-seq 60 min
HSF1 ChIP-seq 60 min
Pol2 ChIP-seq 60 min
Pol2pS2 ChIP-seq 60 min
H4S47C chemical nucleosome mapping
TMA10
YLR326W
SMAC-seq 0 min
0.7
0.1
SMAC-seq 30 min
SMAC-seq 60 min