GREENLEAF LAB GROUP MEETING

GEORGI K. MARINOV

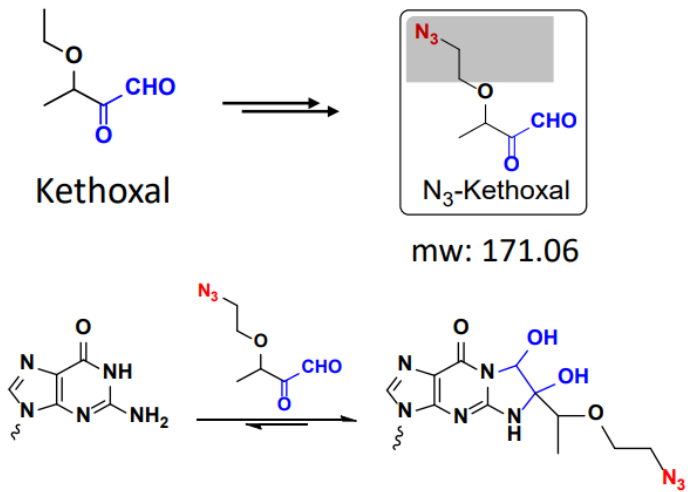
November 27th 2019

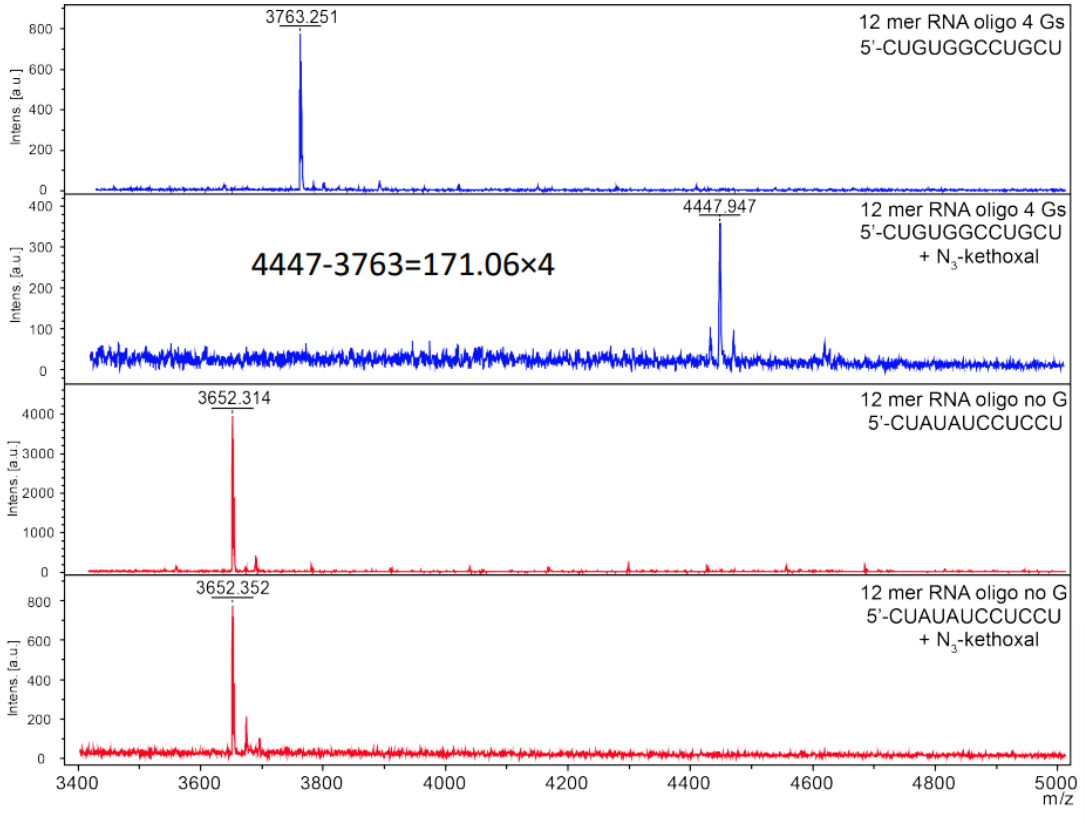
1. KAS-SEQ

Initial version of method developed by Tong Wu & Ruitu Lyv in Chuan He's lab

1.1. KAS-SEQ SLIDES FROM THE HE LAB

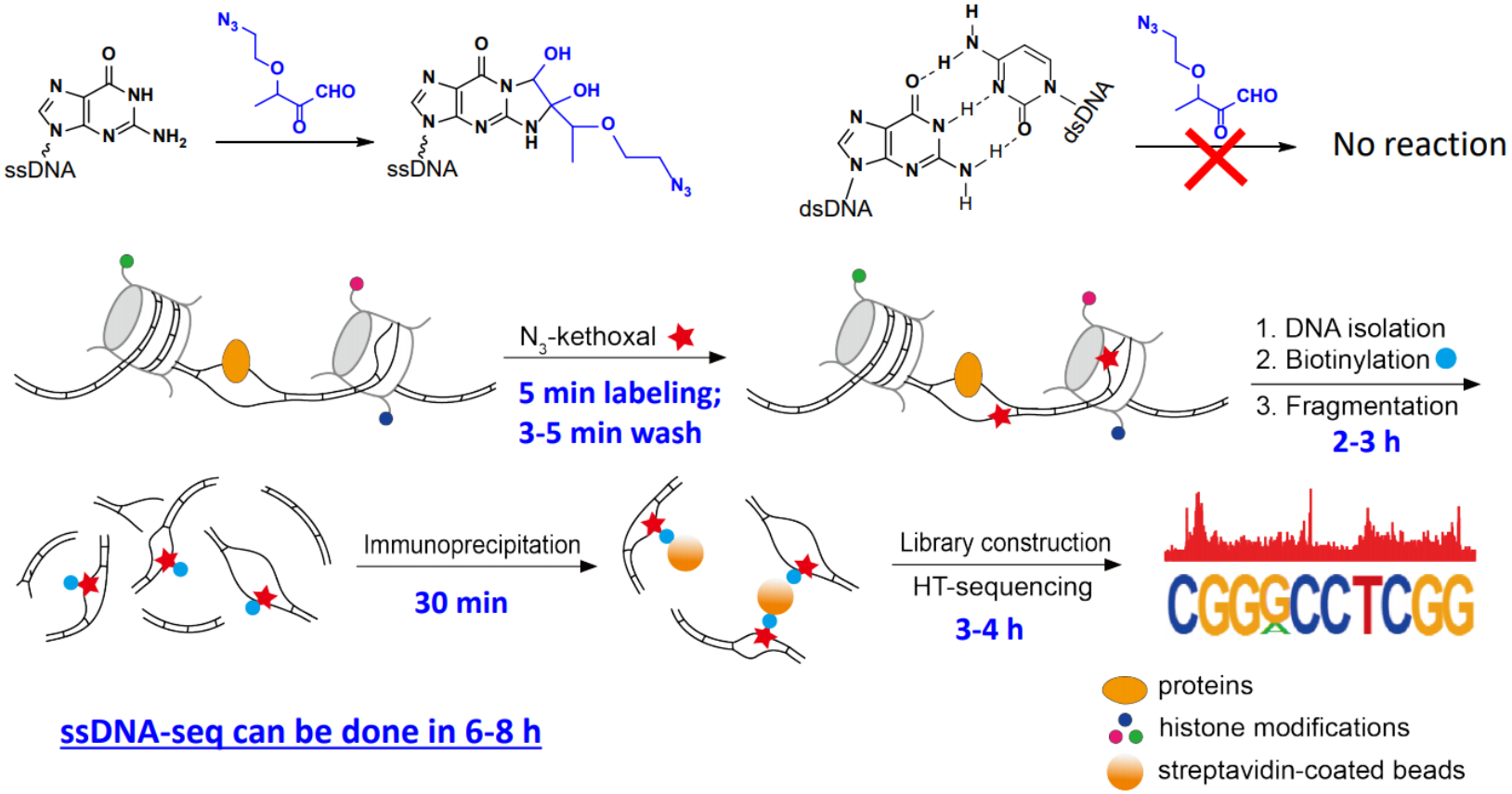
N_3-kethoxal labels guanine efficiently at the Waston-Crick interface





Wu & Lyv

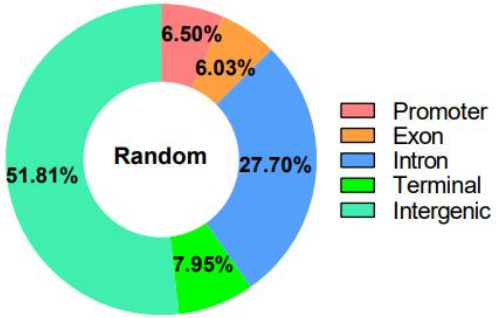
The scheme of ssDNA-seq

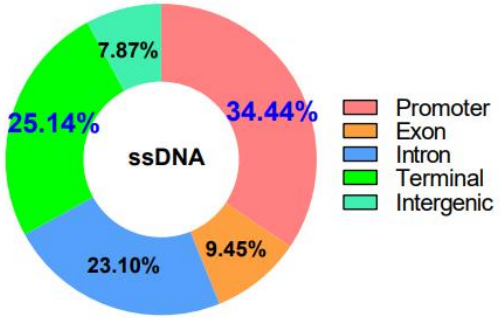


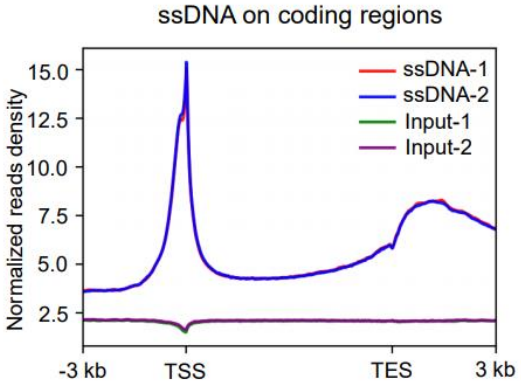
ssDNA-seq can be done in 6-8 h

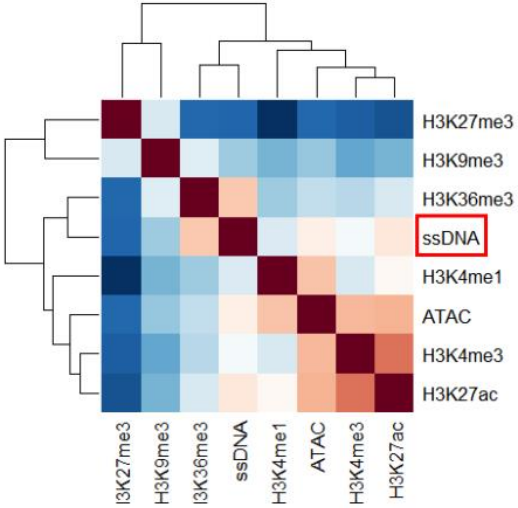
Wu & Lyv

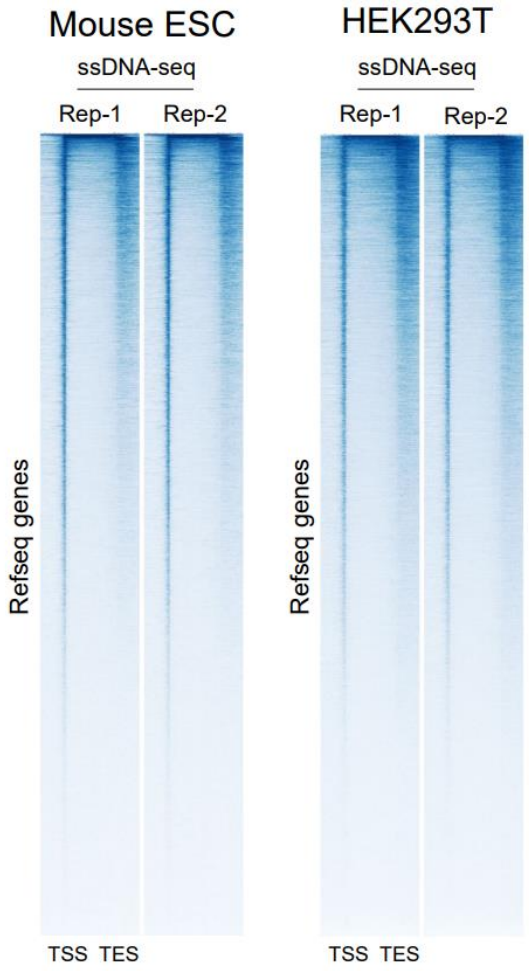
An overview of ssDNA-seq profile





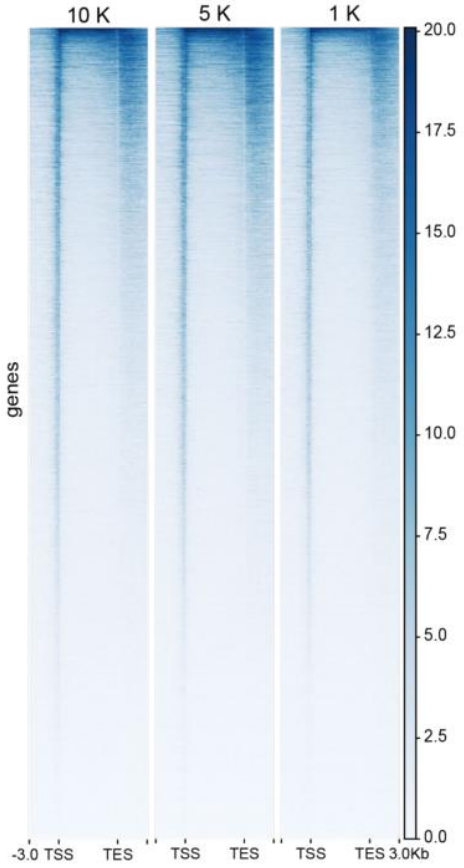


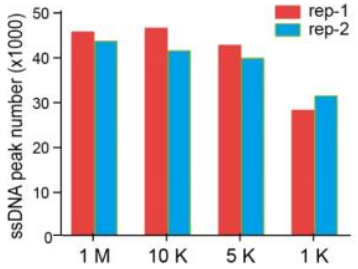


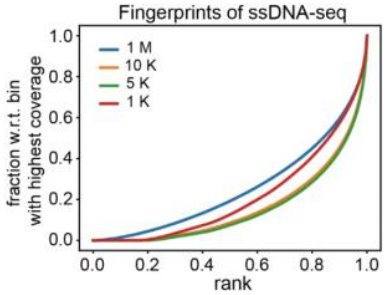


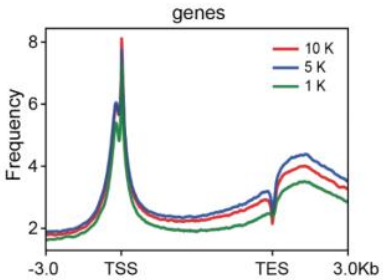
Wu & Lyv

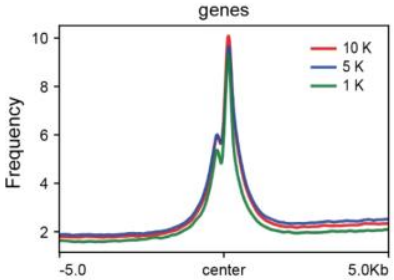
ssDNA-seq works by using 1,000 cells

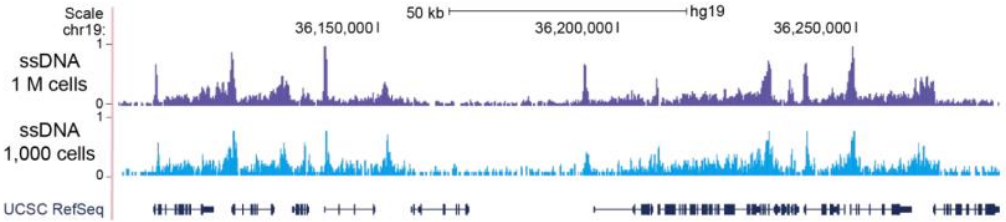






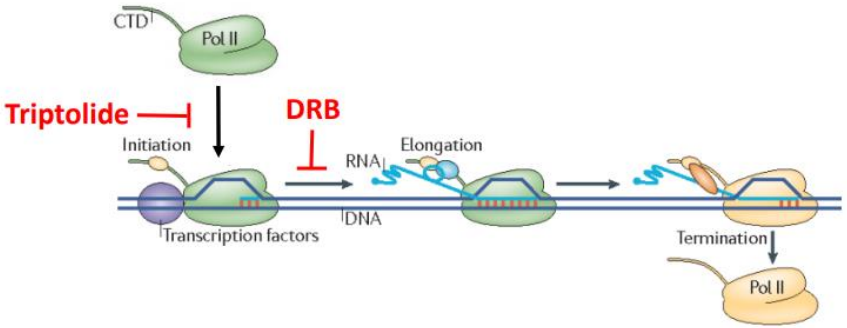


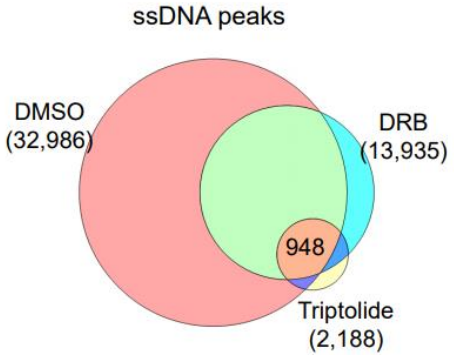


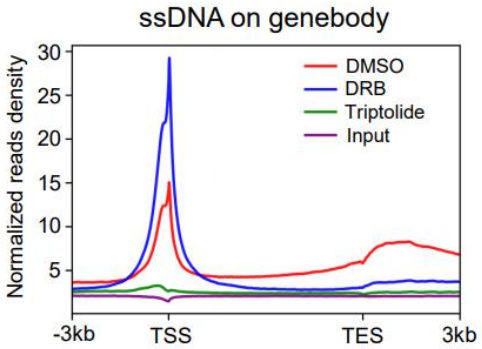


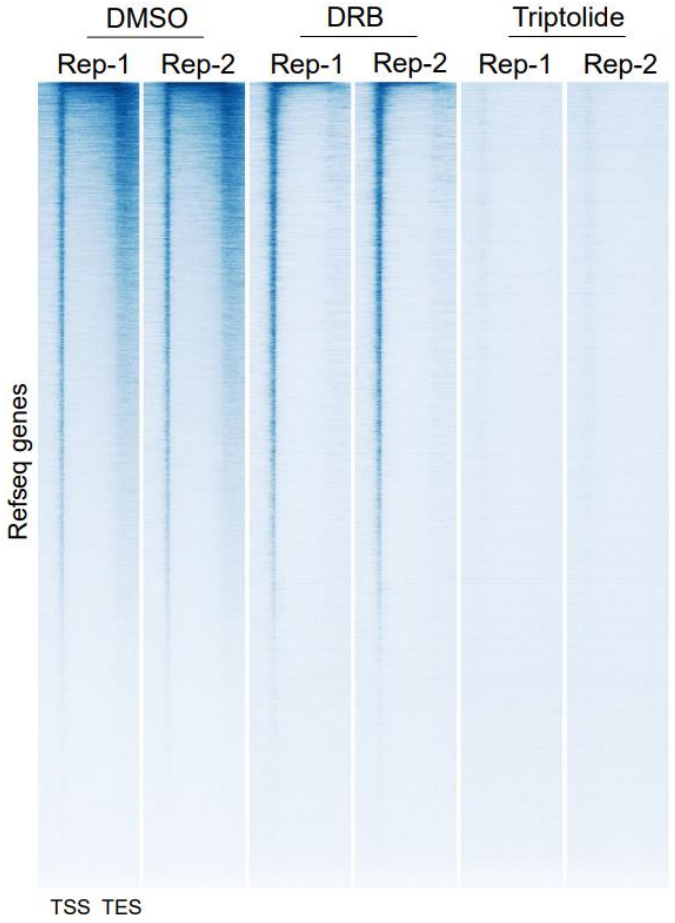
Wu & Lyv

ssDNA signals reveals Pol II dynamics





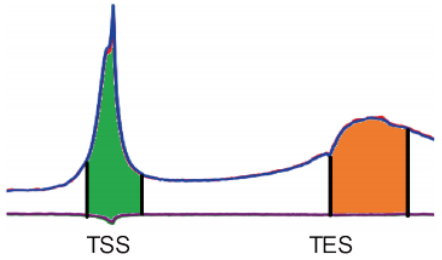


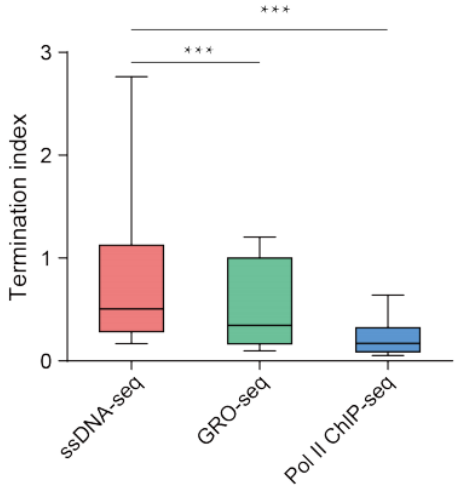


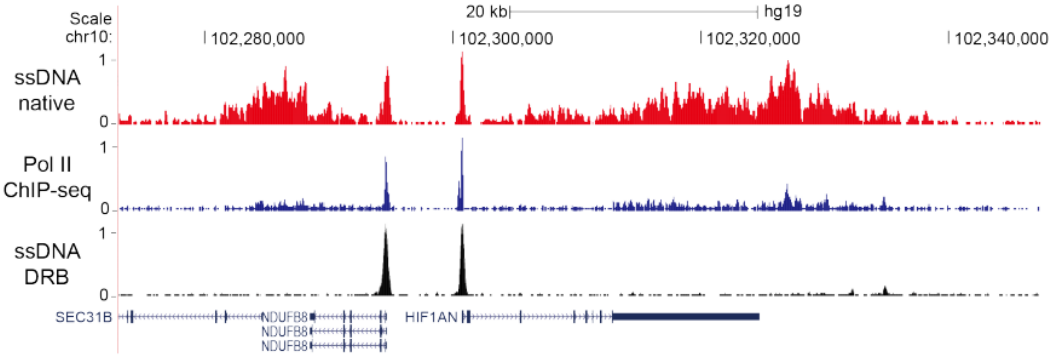
Wu & Lyv

ssDNA-seq reveals strong Pol II pausing during termination

$$\text{Termination index} = \frac{\text{Termination (TES to + 2 kb)}}{\text{TSS (-200 bp to +400 bp)}}$$

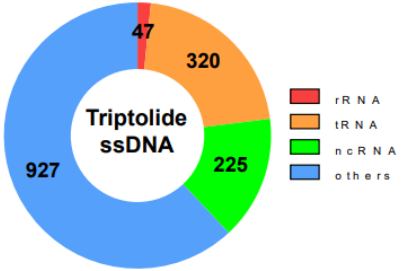




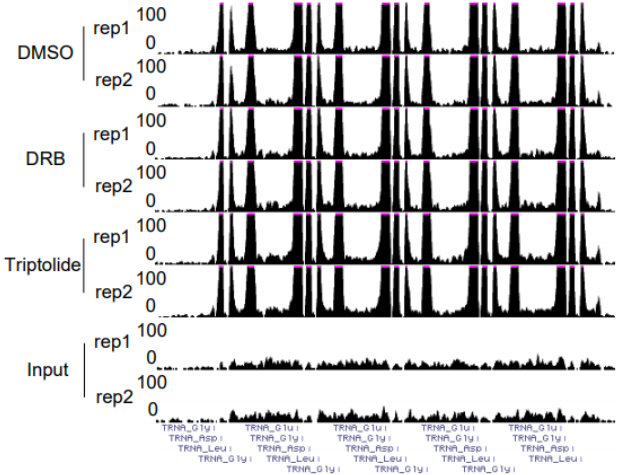


Wu & Lyv

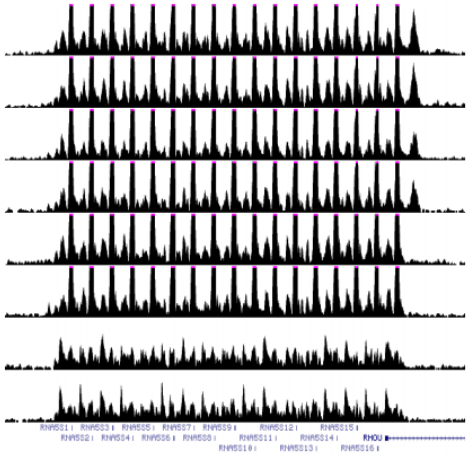
Other ssDNA hotspots



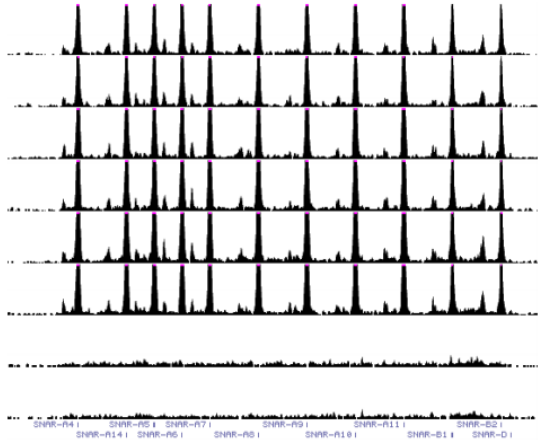
Transfer RNA (411/606)



Ribosome RNA (47/47)

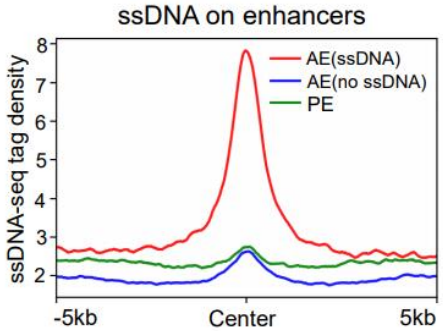


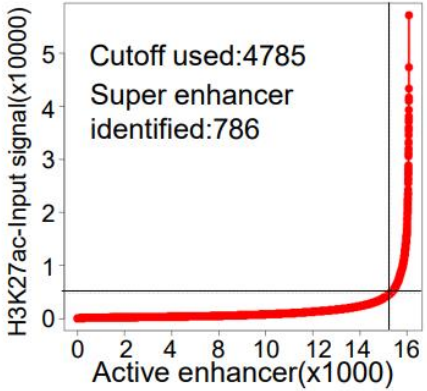
small NF90-associated RNAs

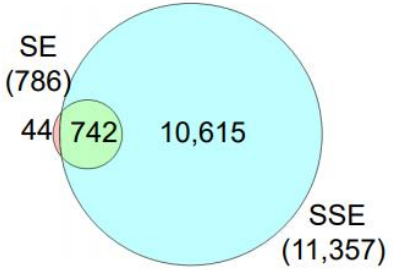


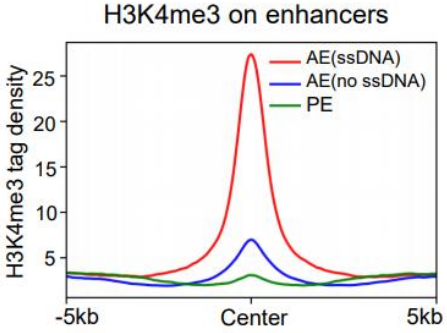
Wu & Lyv

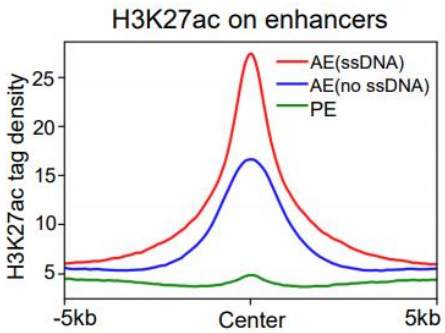
A group of enhancers (single-stranded enhancers, SSE) process ssDNA signals

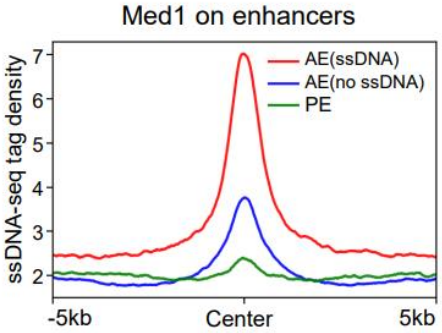






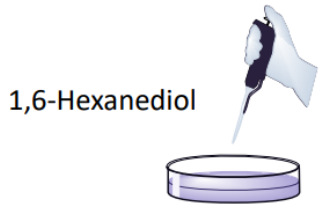


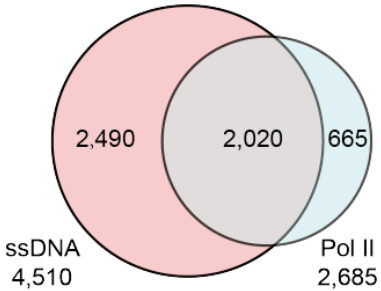


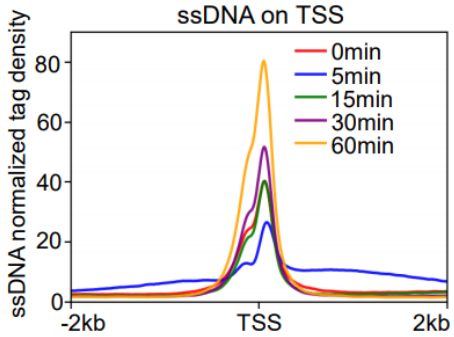


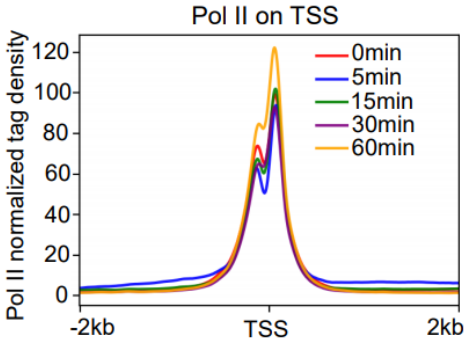
Wu & Lyv

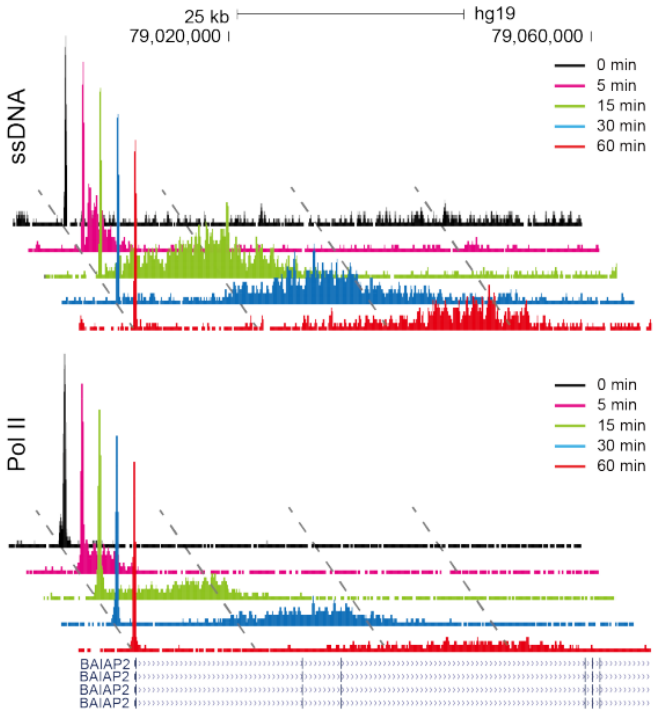
ssDNA-seq is more sensitive than Pol II ChIP-seq in the detection of dynamics











Wu & Lyv

1.2. KAS-SEQ IN OUR HANDS

Georgi, Alex, Zohar

PROTOCOL OVERVIEW

1. Dissolve ketoxal at 37C in media or PBS
2. Dump ketoxal on cells, incubate for 10 minutes
3. Wash away ketoxal
4. Isolate DNA
5. Click chemistry on modified DNA, attach biotin (1.5 hours)
6. Fragment DNA (Covaris)
7. Biotin pull down
8. PCR amplification

PROTOCOL MODIFICATIONS

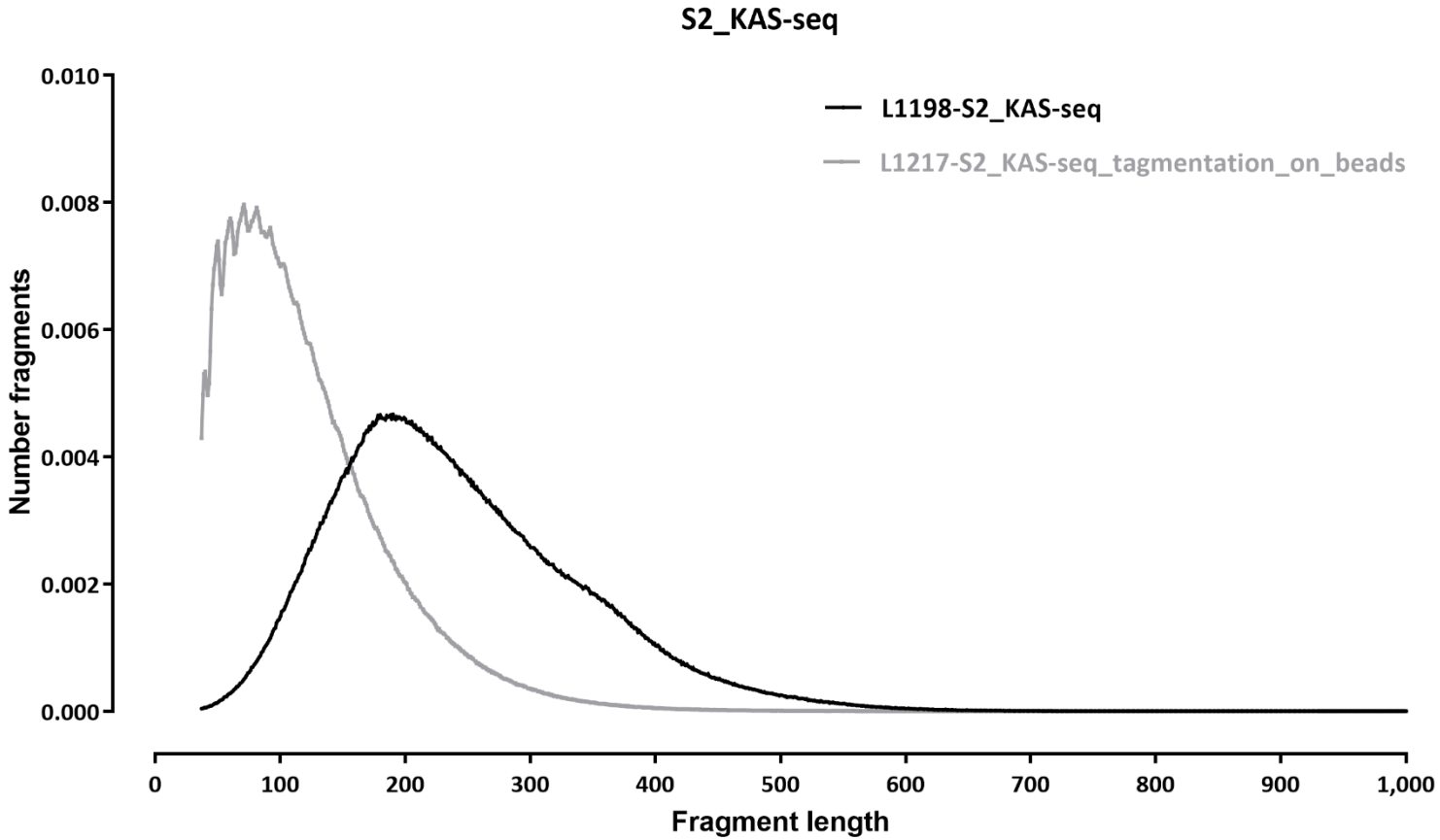
- I am doing library prep on the streptavidin beads – end repair, ligation and PCR

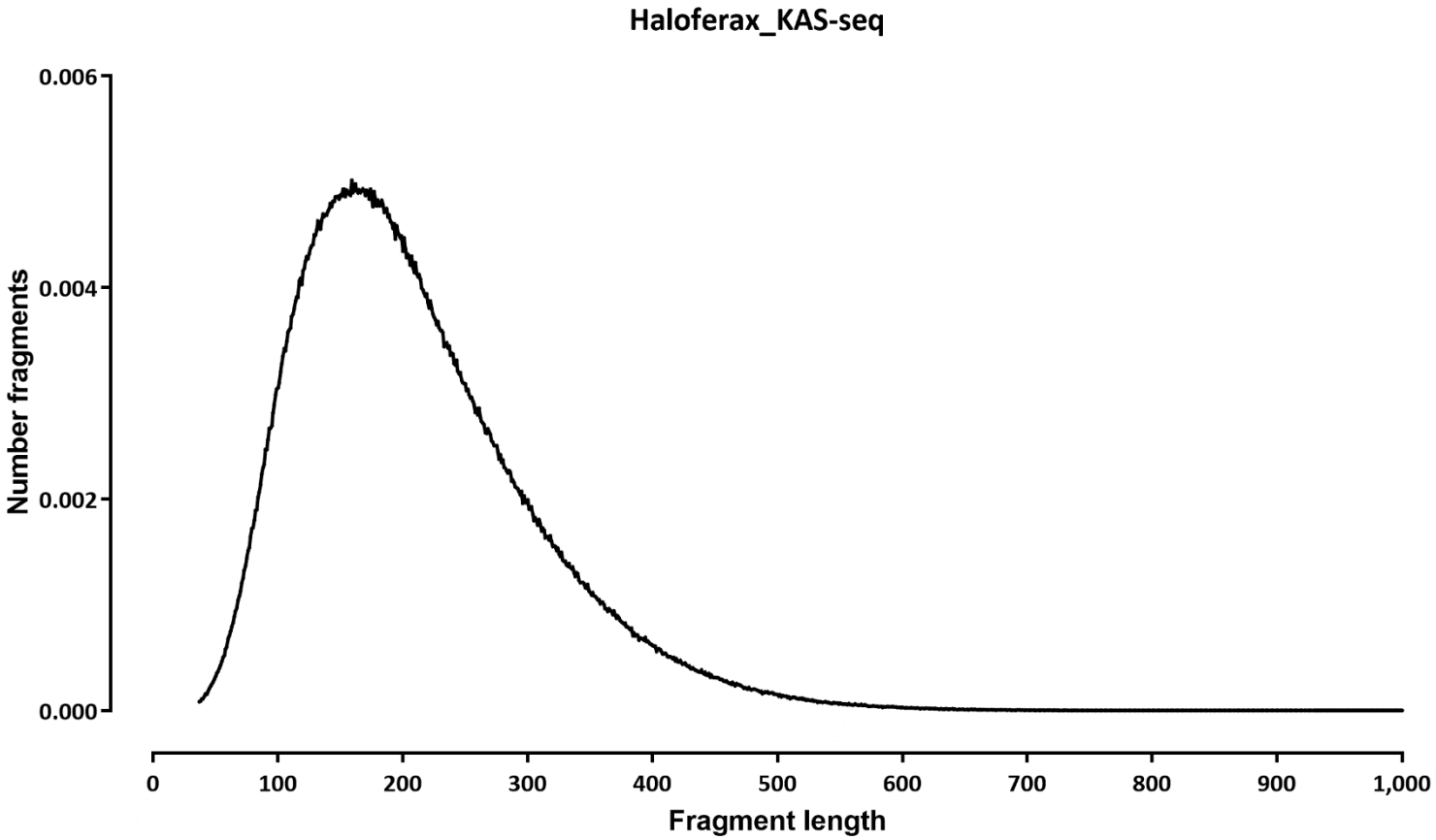
- Alex tried tagmentation on the streptavidin beads

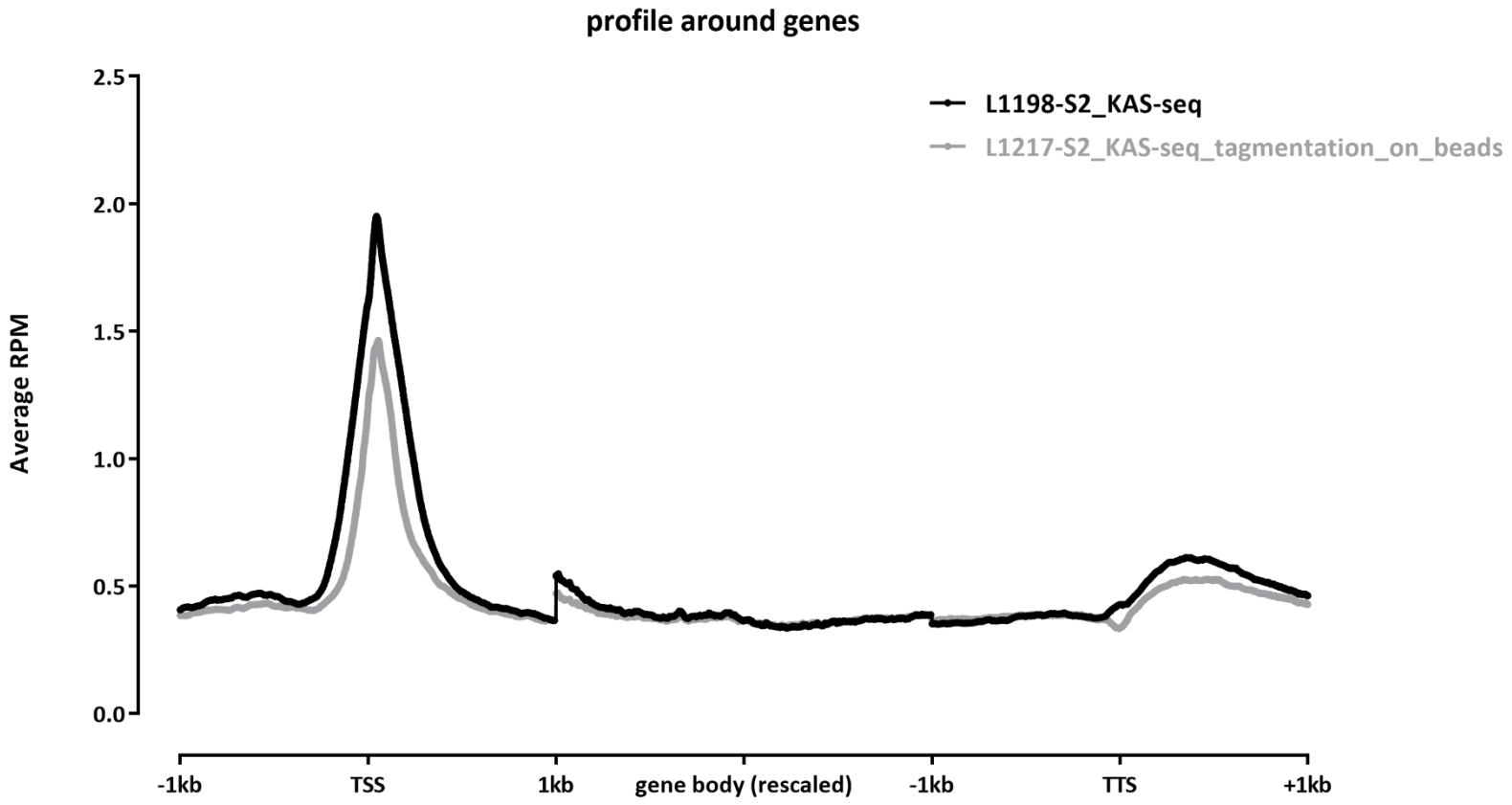
DATASET STATS

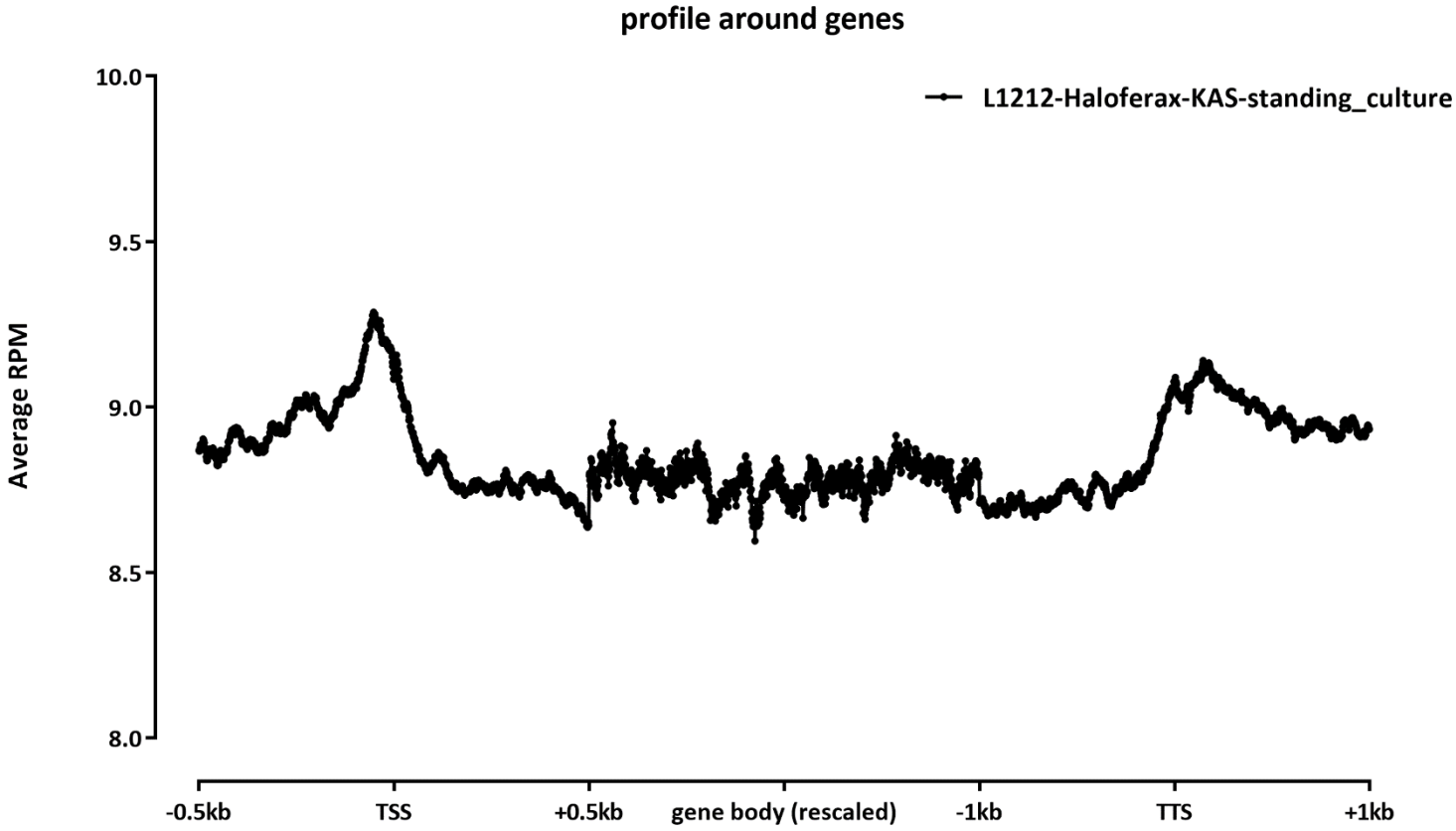
Species	#	Complexity	MACS2 NumPeaks	MACS2 RPM	MACS2 noBL NumPeaks	MACS2 noBL RPM	Read Length	Unique	Unque dedup	Raw fragments
Drosophila melanogaster dm6	L1198-S2_KAS-seq	0.61	6,213	312,578			2x36	25,271,558	17,536,742	23,071,415
Drosophila melanogaster dm6	L1217-S2_KAS-seq_tagmentation_on_beads	0.65	8,735	223,411			2x36	39,376,934	30,745,498	30,428,387
Drosophila melanogaster dm3	L1198-S2_KAS-seq	0.62	6,266	303,522			2x36	24,754,264	17,190,696	23,071,415
Drosophila melanogaster dm3	L1217-S2_KAS-seq_tagmentation_on_beads	0.66	8,704	213,479			2x36	38,486,900	30,333,572	30,428,387
Haloferax volcanii	L1212-Haloferax-KAS-standing_culture	0.67	597	148,246			2x36	9,815,872	8,407,646	5,869,879

FRAGMENT LENGTH



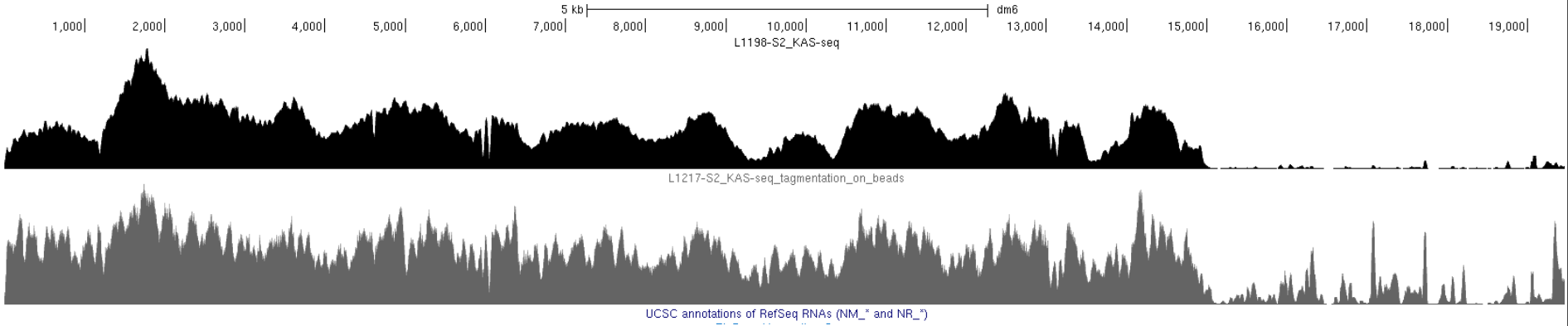

Haloferax_KAS-seq
0.006
0.004
0.002
0.000
Number fragments
0
100
200
300
400
500
600
700
800
900
1,000
Fragment length


profile around genes
L1198-S2_KAS-seq
L1217-S2_KAS-seq_tagmentation_on_beads
Average RPM
2.5
2.0
1.5
1.0
0.5
0.0
-1kb
TSS
1kb
gene body (rescaled)
-1kb
TTS
+1kb

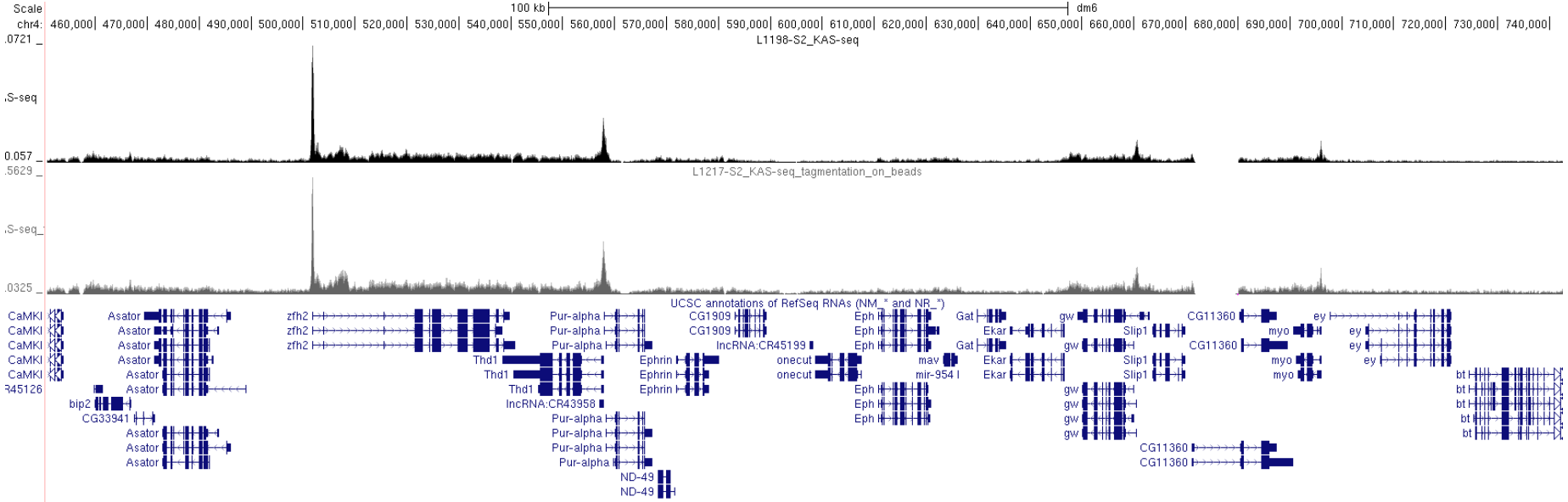

profile around genes
L1212-Haloferax-KAS-standing_culture
Average RPM
10.0
9.5
9.0
8.5
8.0
-0.5kb
TSS
+0.5kb
gene body (rescaled)
-1kb
TTS
+1kb

DROSOPHILA

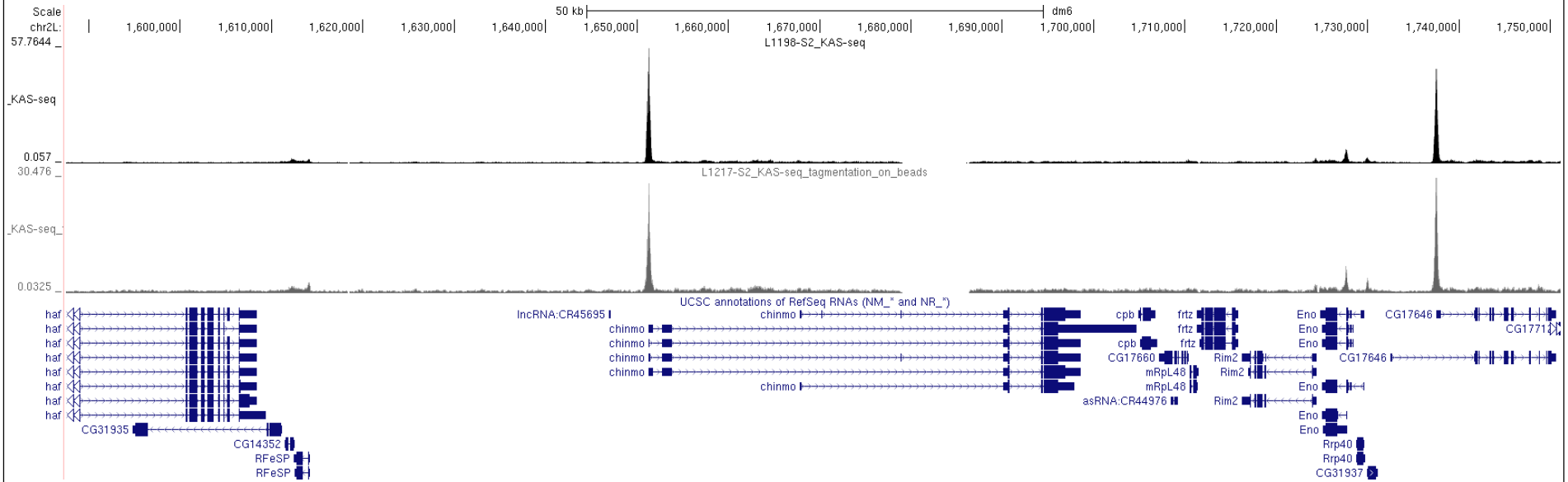
chrM:



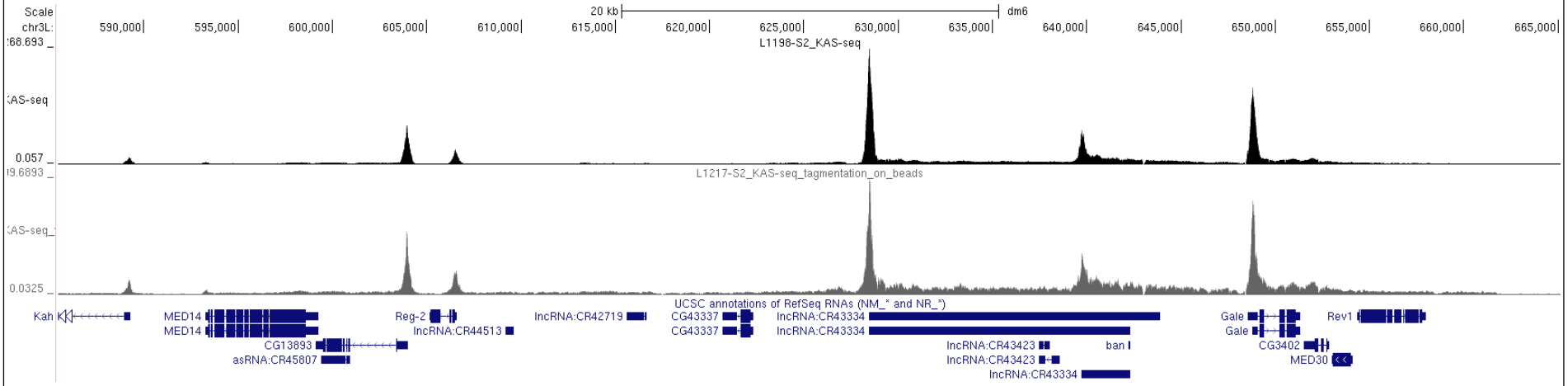
DROSOPHILA


Scale
100 kb
dm6
chr4: 460,000 470,000 480,000 490,000 500,000 510,000 520,000 530,000 540,000 550,000 560,000 570,000 580,000 590,000 600,000 610,000 620,000 630,000 640,000 650,000 660,000 670,000 680,000 690,000 700,000 710,000 720,000 730,000 740,000
L1198-S2_KAS-seq
L1217-S2_KAS-seq_tagmentation_on_beads
UCSC annotations of RefSeq RNAs (NM_* and NR_*)
CaMKI
Asator
zfh2
Pur-alpha
CG1909
Eph
Gat
gw
CG11360
ey
Thd1
Ephrin
lncRNA:CR45199
onecut
Ekar
Slip1
myo
bip2
CG33941
lncRNA:CR43958
mav
mir-954
bt
ND-49

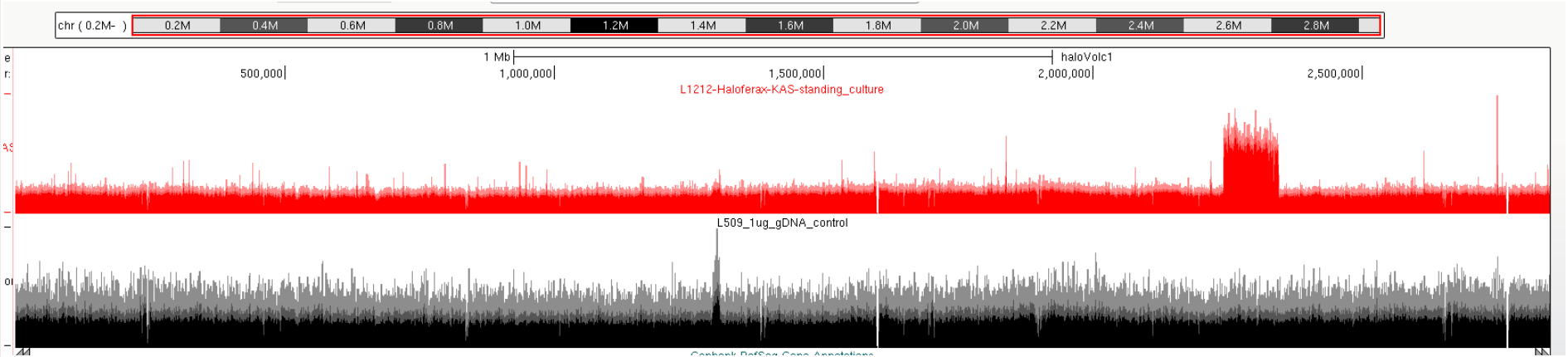
DROSOPHILA


Scale
chr2L:
50 kb
dm6
1,600,000
1,610,000
1,620,000
1,630,000
1,640,000
1,650,000
1,660,000
1,670,000
1,680,000
1,690,000
1,700,000
1,710,000
1,720,000
1,730,000
1,740,000
1,750,000
57.7644
L1198-S2_KAS-seq
_KAS-seq
0.057
30.476
L1217-S2_KAS-seq_tagmentation_on_beads
KAS-seq
0.0325
UCSC annotations of RefSeq RNAs (NM_* and NR_*)
haf
haf
haf
haf
haf
haf
haf
haf
CG31935
CG14352
RFeSP
RFeSP
lncRNA:CR45695
chinmo
chinmo
chinmo
chinmo
chinmo
chinmo
cpb
cpb
CG17660
mRpL48
mRpL48
asRNA:CR44976
frtz
frtz
frtz
Rim2
Rim2
Rim2
Eno
Eno
Eno
Eno
Eno
Rrp40
Rrp40
CG31937
CG17646
CG17646
CG17714

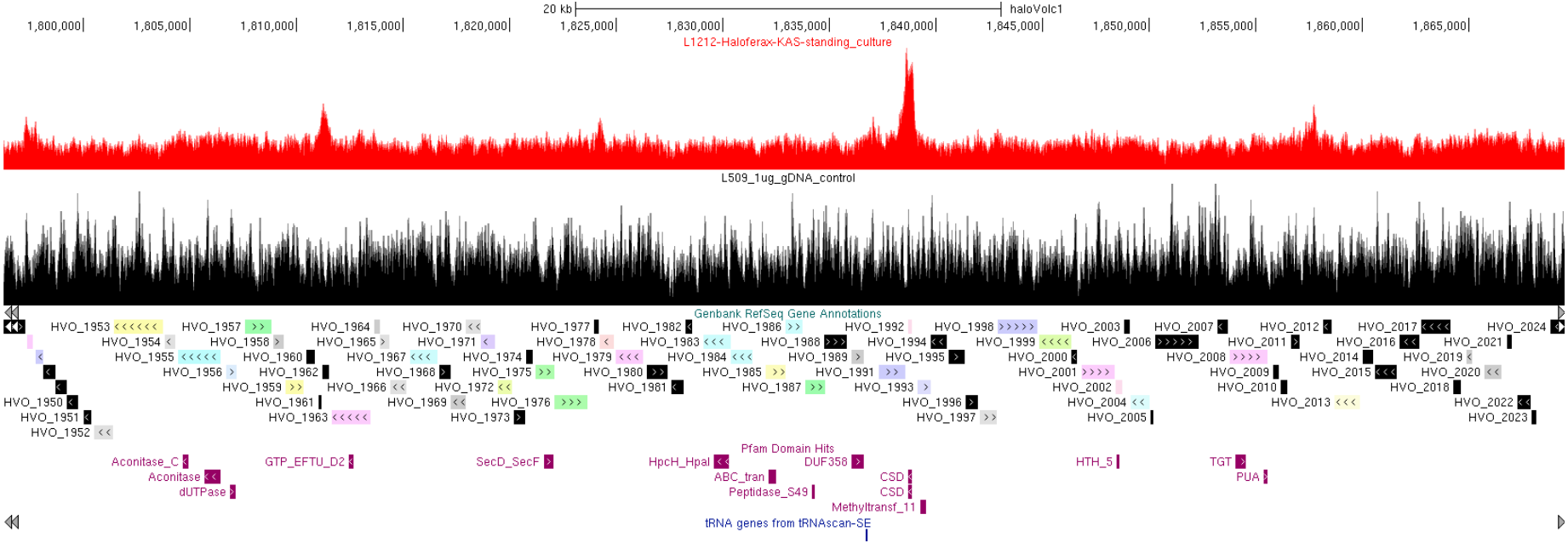
DROSOPHILA


Scale
chr3L:
20 kb
dm6
590,000
595,000
600,000
605,000
610,000
615,000
620,000
625,000
630,000
635,000
640,000
645,000
650,000
655,000
660,000
665,000
L1198-S2_KAS-seq
0.057
L1217-S2_KAS-seq_tagmentation_on_beads
0.0325
UCSC annotations of RefSeq RNAs (NM_* and NR_*)
Kah K
MED14
MED14
CG13893
asRCR45807
Reg-2
lncRNA:CR44513
lncRNA:CR42719
CG43337
CG43337
lncRNA:CR43334
lncRNA:CR43334
lncRNA:CR43423
lncRNA:CR43423
lncRNA:CR43334
ban
Gale
Gale
Rev1
CG3402
MED30

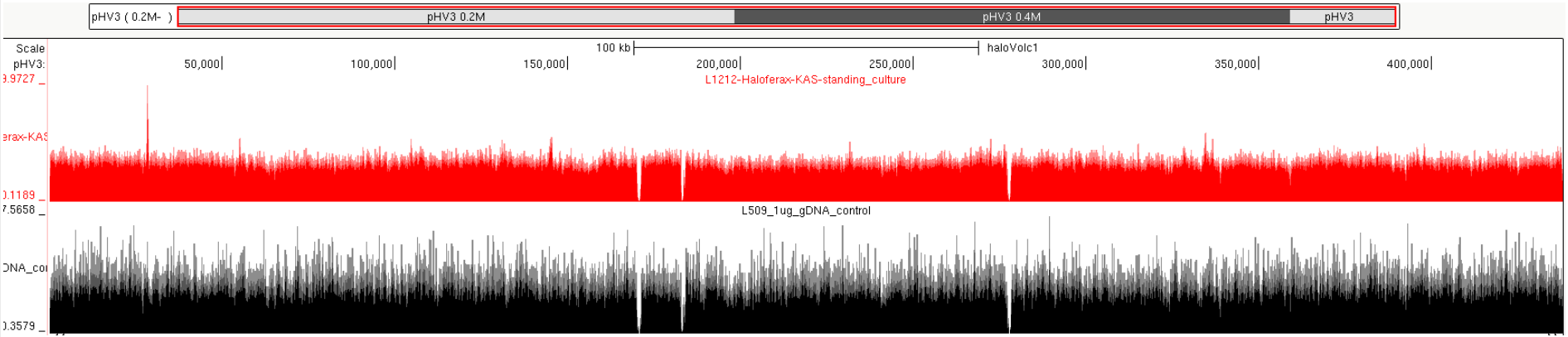
HALOFERAX


chr (0.2M-)
0.2M
0.4M
0.6M
0.8M
1.0M
1.2M
1.4M
1.6M
1.8M
2.0M
2.2M
2.4M
2.6M
2.8M
1 Mb
haloVolc1
500,000
1,000,000
1,500,000
2,000,000
2,500,000
L1212-Haloferax-KAS-standing_culture
L509_1ug_gDNA_control

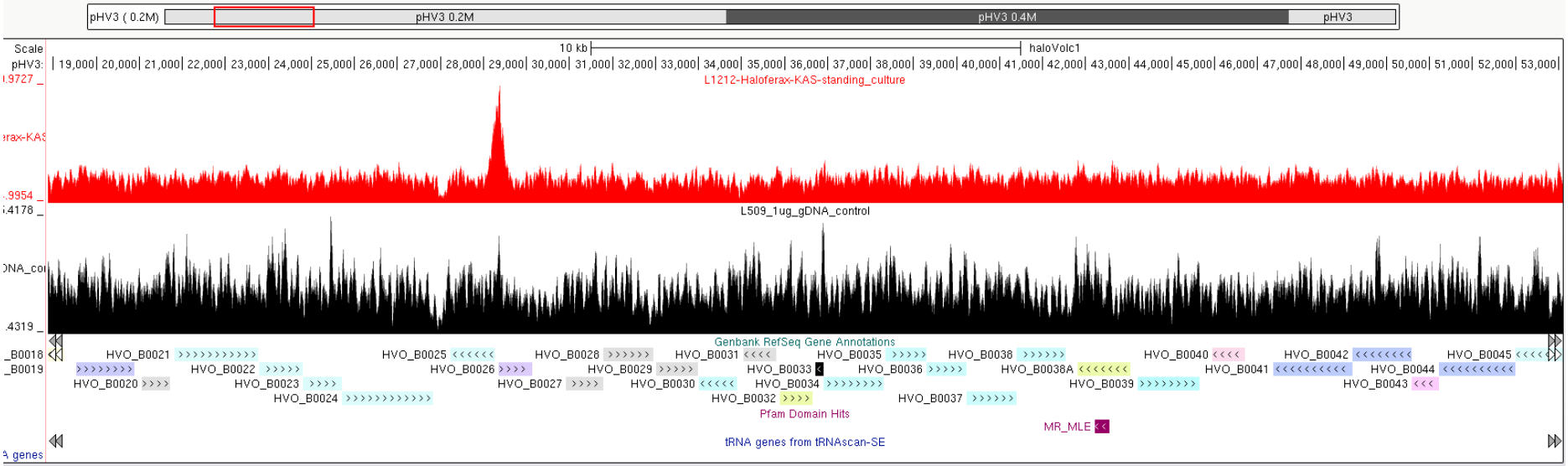
HALOFERAX


20 kb
haloVolc1
1,800,000
1,805,000
1,810,000
1,815,000
1,820,000
1,825,000
1,830,000
1,835,000
1,840,000
1,845,000
1,850,000
1,855,000
1,860,000
1,865,000
L1212-Haloferax-KAS-standing_culture
L509_1ug_gDNA_control
Genbank RefSeq Gene Annotations
HVO_1953
HVO_1954
HVO_1955
HVO_1956
HVO_1957
HVO_1958
HVO_1959
HVO_1960
HVO_1961
HVO_1962
HVO_1963
HVO_1964
HVO_1965
HVO_1966
HVO_1967
HVO_1968
HVO_1969
HVO_1970
HVO_1971
HVO_1972
HVO_1973
HVO_1974
HVO_1975
HVO_1976
HVO_1977
HVO_1978
HVO_1979
HVO_1980
HVO_1981
HVO_1982
HVO_1983
HVO_1984
HVO_1985
HVO_1986
HVO_1987
HVO_1988
HVO_1989
HVO_1991
HVO_1992
HVO_1993
HVO_1994
HVO_1995
HVO_1996
HVO_1997
HVO_1998
HVO_1999
HVO_2000
HVO_2001
HVO_2002
HVO_2003
HVO_2004
HVO_2005
HVO_2006
HVO_2007
HVO_2008
HVO_2009
HVO_2010
HVO_2011
HVO_2012
HVO_2013
HVO_2014
HVO_2015
HVO_2016
HVO_2017
HVO_2018
HVO_2019
HVO_2020
HVO_2021
HVO_2022
HVO_2023
HVO_2024
Pfam Domain Hits
Aconitase_C
Aconitase
dUTPase
GTP_EFTU_D2
SecD_SecF
HpcH_HpaI
ABC_tran
Peptidase_S49
DUF358
CSD
CSD
Methyltransf_11
HTH_5
TGT
PUA
tRNA genes from tRNAscan-SE

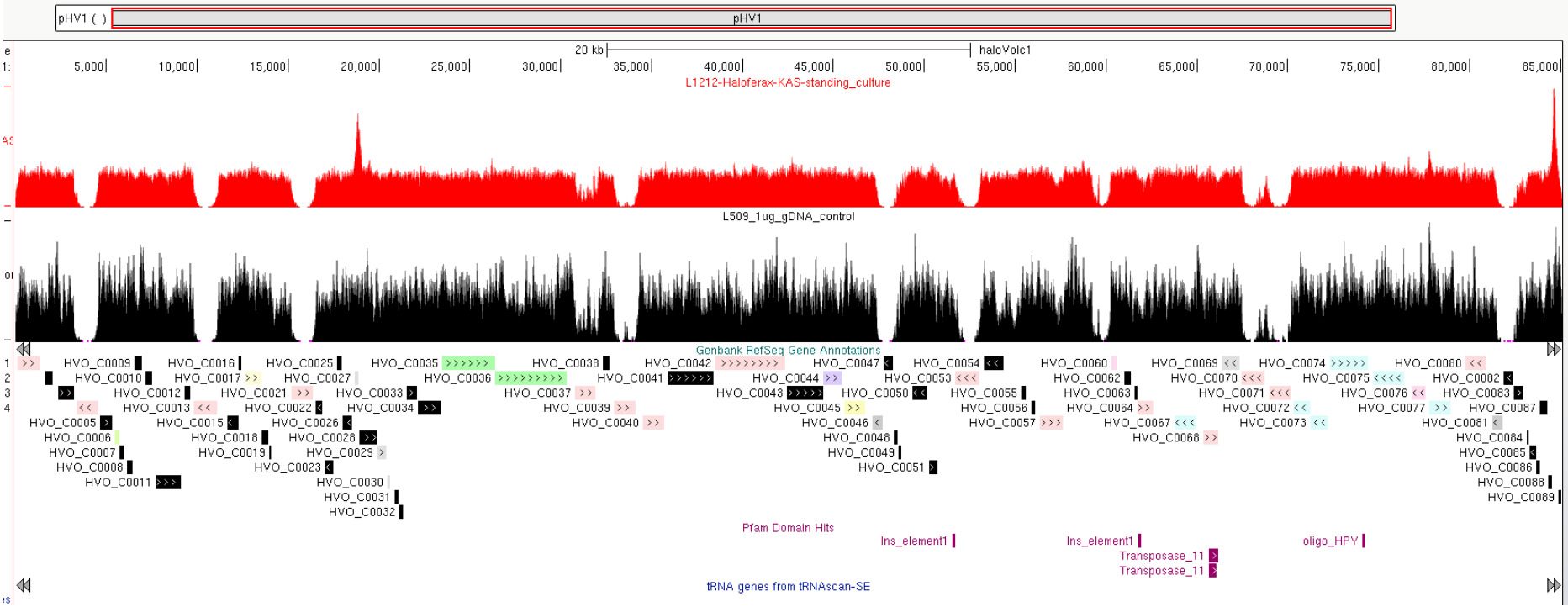
HALOFERAX


pHV3 (0.2M-)
pHV3 0.2M
pHV3 0.4M
pHV3
Scale
100 kb
haloVolc1
50,000
100,000
150,000
200,000
250,000
300,000
350,000
400,000
L1212-Haloferax-KAS-standing_culture
L509_1ug_gDNA_control

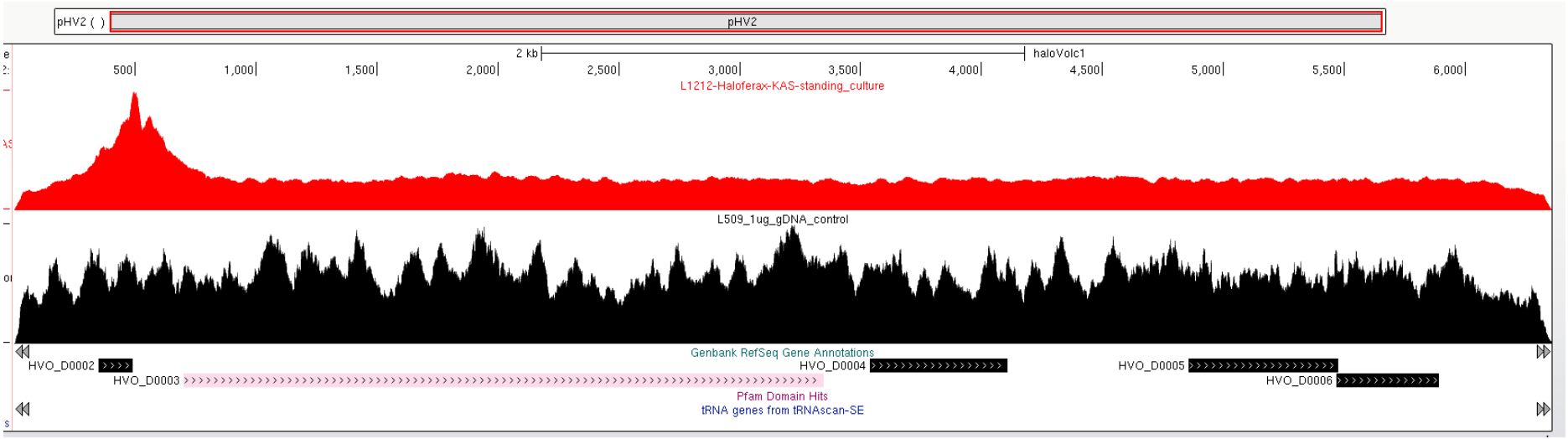
HALOFERAX


pHV3 (0.2M)
pHV3 0.2M
pHV3 0.4M
pHV3
Scale
10 kb
haloVolc1
pHV3:
19,000 20,000 21,000 22,000 23,000 24,000 25,000 26,000 27,000 28,000 29,000 30,000 31,000 32,000 33,000 34,000 35,000 36,000 37,000 38,000 39,000 40,000 41,000 42,000 43,000 44,000 45,000 46,000 47,000 48,000 49,000 50,000 51,000 52,000 53,000
L1212-Haloferax-KAS-standing_culture
L509_1ug_gDNA_control
Genbank RefSeq Gene Annotations
HVO_B0018
HVO_B0019
HVO_B0020
HVO_B0021
HVO_B0022
HVO_B0023
HVO_B0024
HVO_B0025
HVO_B0026
HVO_B0027
HVO_B0028
HVO_B0029
HVO_B0030
HVO_B0031
HVO_B0032
HVO_B0033
HVO_B0034
HVO_B0035
HVO_B0036
HVO_B0037
HVO_B0038
HVO_B0038A
HVO_B0039
HVO_B0040
HVO_B0041
HVO_B0042
HVO_B0043
HVO_B0044
HVO_B0045
Pfam Domain Hits
MR_MLE
tRNA genes from tRNAscan-SE

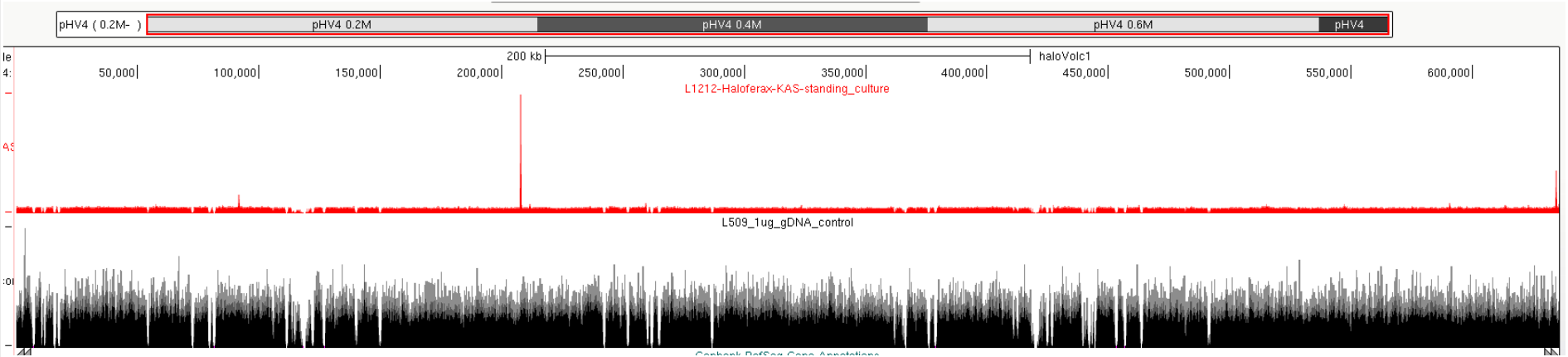
HALOFERAX


pHV1 ( )
pHV1
20 kb
haloVolc1
5,000
10,000
15,000
20,000
25,000
30,000
35,000
40,000
45,000
50,000
55,000
60,000
65,000
70,000
75,000
80,000
85,000
L1212-Haloferax-KAS-standing_culture
L509_1ug_gDNA_control
Genbank RefSeq Gene Annotations
HVO_C0009
HVO_C0016
HVO_C0025
HVO_C0035
HVO_C0038
HVO_C0042
HVO_C0047
HVO_C0054
HVO_C0060
HVO_C0069
HVO_C0074
HVO_C0080
HVO_C0010
HVO_C0017
HVO_C0027
HVO_C0036
HVO_C0041
HVO_C0044
HVO_C0053
HVO_C0062
HVO_C0070
HVO_C0075
HVO_C0082
HVO_C0012
HVO_C0021
HVO_C0033
HVO_C0037
HVO_C0043
HVO_C0050
HVO_C0055
HVO_C0063
HVO_C0071
HVO_C0076
HVO_C0083
HVO_C0013
HVO_C0022
HVO_C0034
HVO_C0039
HVO_C0045
HVO_C0056
HVO_C0064
HVO_C0072
HVO_C0077
HVO_C0087
HVO_C0005
HVO_C0015
HVO_C0026
HVO_C0040
HVO_C0046
HVO_C0057
HVO_C0067
HVO_C0073
HVO_C0081
HVO_C0006
HVO_C0018
HVO_C0028
HVO_C0048
HVO_C0068
HVO_C0084
HVO_C0007
HVO_C0019
HVO_C0029
HVO_C0049
HVO_C0085
HVO_C0008
HVO_C0023
HVO_C0051
HVO_C0086
HVO_C0011
HVO_C0030
HVO_C0088
HVO_C0031
HVO_C0089
HVO_C0032
Pfam Domain Hits
Ins_element1
Ins_element1
oligo_HPY
Transposase_11
Transposase_11
tRNA genes from tRNAscan-SE

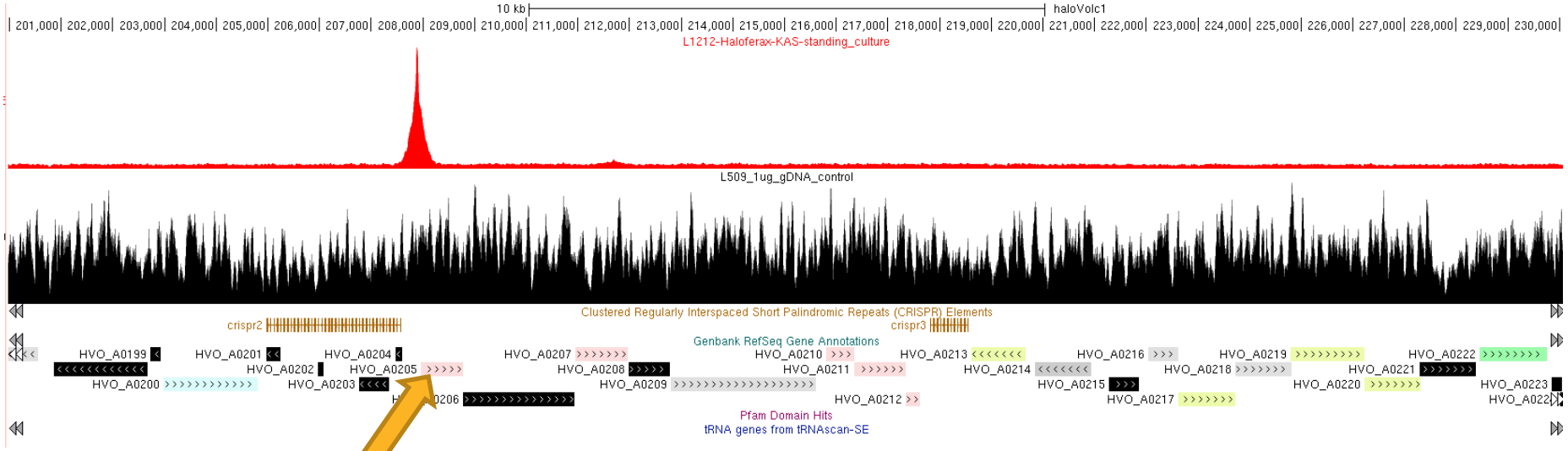
HALOFERAX


pHV2 ( )
pHV2
2 kb
haloVolc1
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
4,500
5,000
5,500
6,000
L1212-Haloferax-KAS-standing_culture
L509_1ug_gDNA_control
Genbank RefSeq Gene Annotations
HVO_D0002
HVO_D0003
HVO_D0004
HVO_D0005
HVO_D0006
Pfam Domain Hits
tRNA genes from tRNAscan-SE

HALOFERAX


pHV4 (0.2M-)
pHV4 0.2M
pHV4 0.4M
pHV4 0.6M
pHV4
200 kb
haloVolc1
50,000
100,000
150,000
200,000
250,000
300,000
350,000
400,000
450,000
500,000
550,000
600,000
L1212-Haloferax-KAS-standing_culture
L509_1ug_gDNA_control

HALOFERAX



This is Cas6

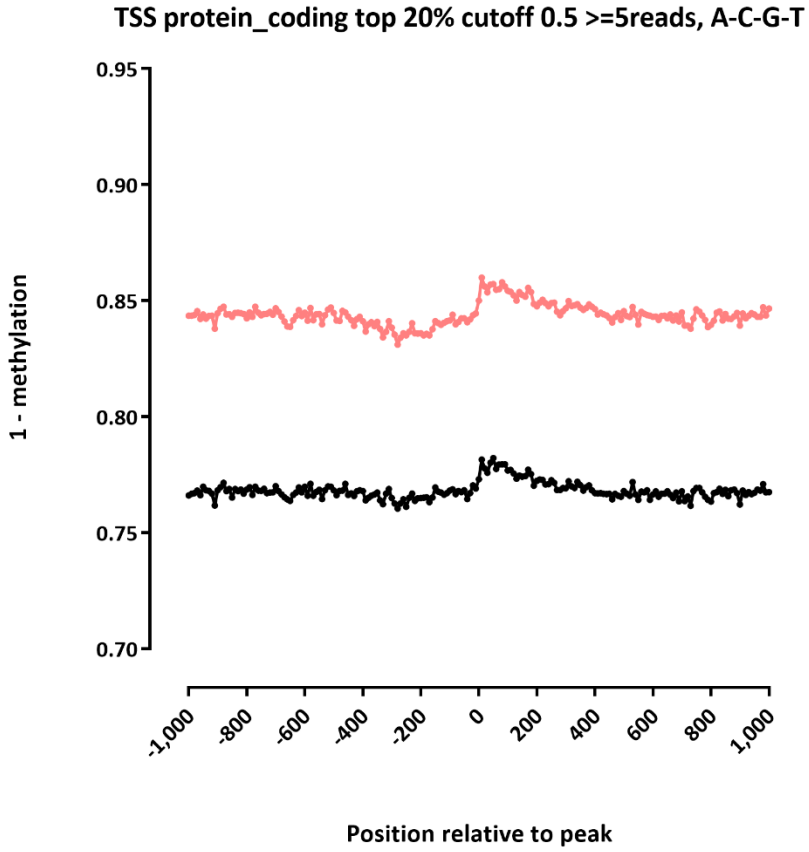
1.3. nanoKAS ATTEMPT

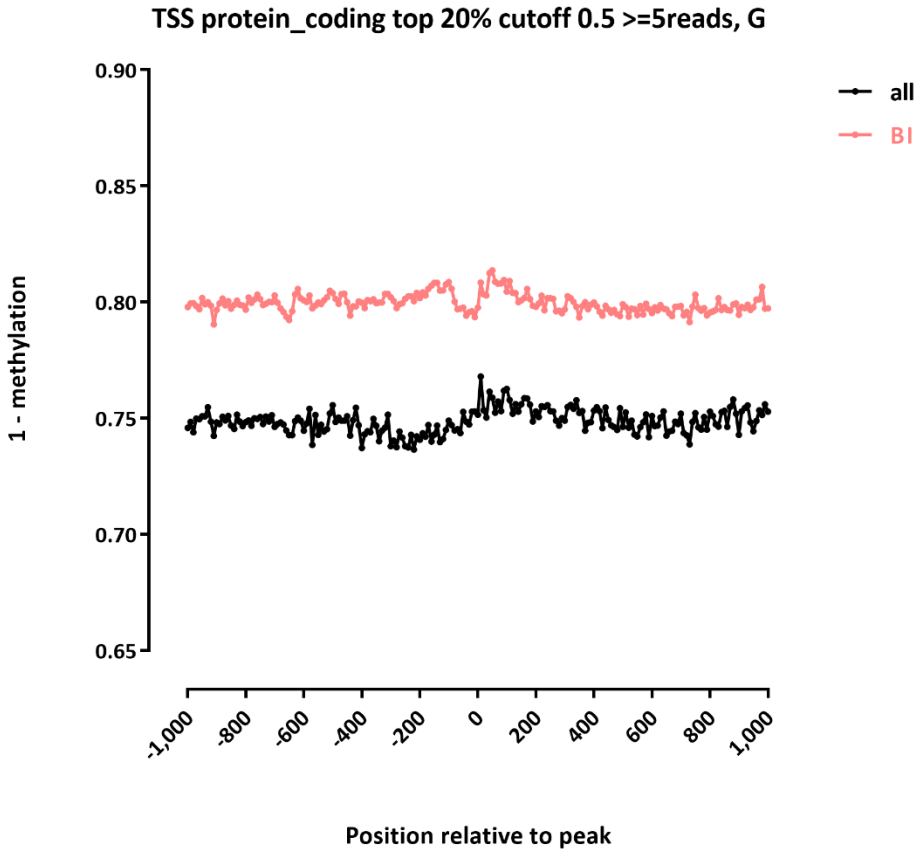
POSSIBILITIES AND ADVANTAGES

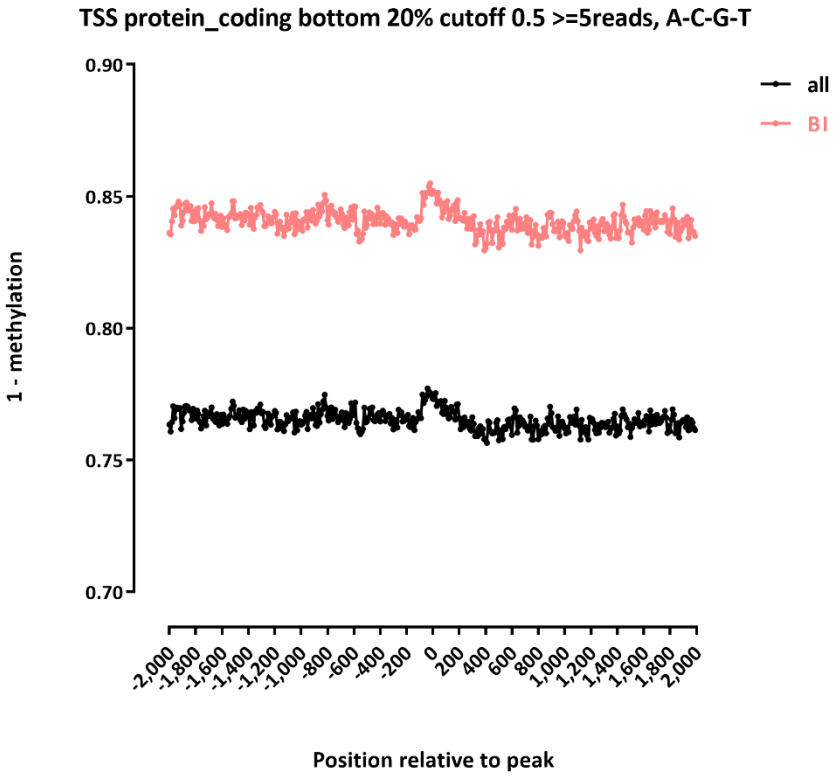
- Ketoxal is a bulky modification
- Should be detectable very well using nanopore sequencing
- Would provide basepair readout of ssDNA
- Would provide linked long-range single-molecule readout of ssDNA

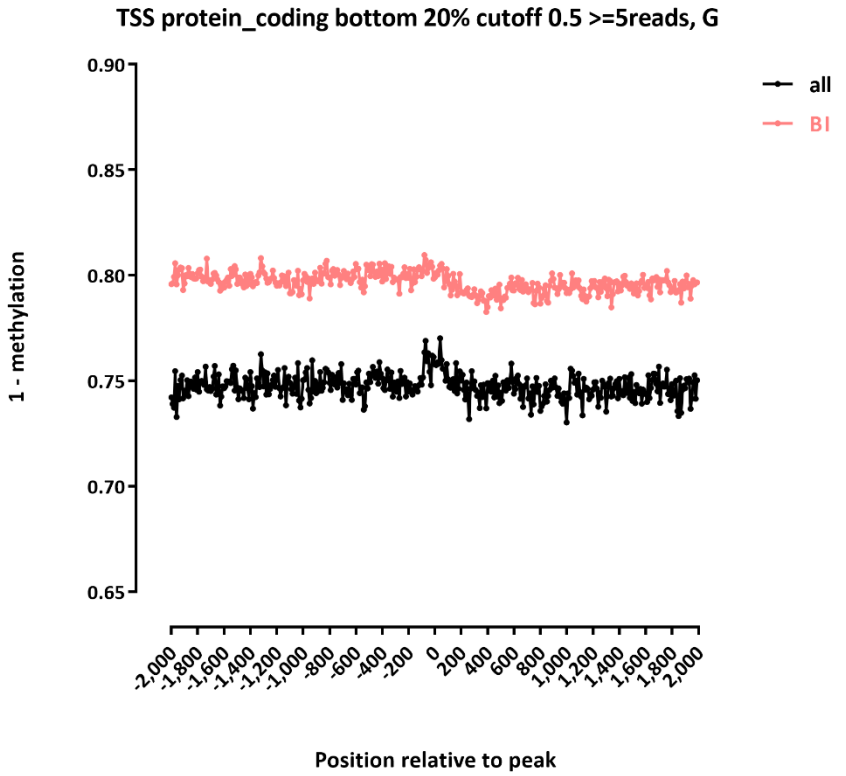
FIRST 800K READS:

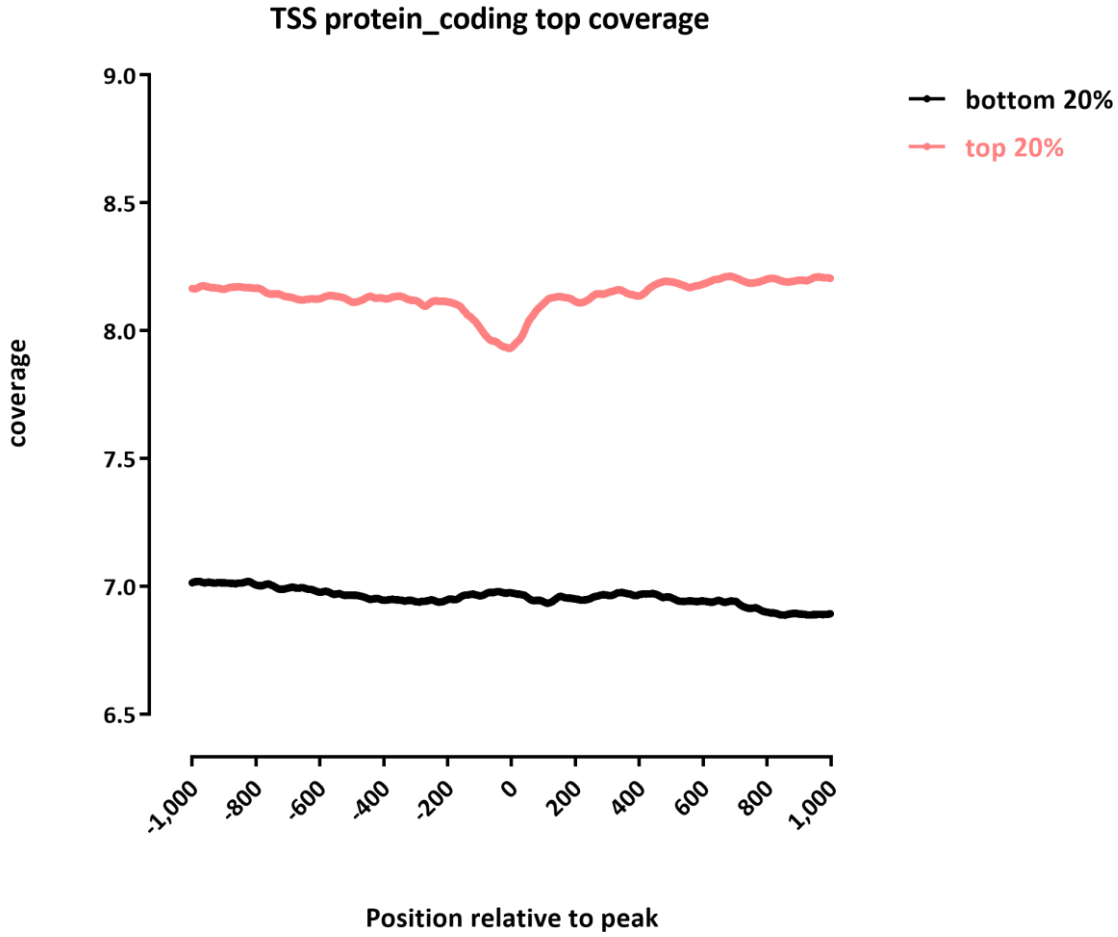
Species	Assay	#	Mapping	Base Contexts	Reads	Total Bases	Mean Read Length	Median Read Length
dm3	**ssDNA**	2019-11-06-S2-ssDNA-800k	**Tombo 1.5 denovo**	**generic-A-C-T-G**	533,199	975,589,084	1,830	806
dm3	**ssDNA**	2019-11-06-S2-ssDNA-800k	**Tombo 1.5 denovo**	**generic-G**	532,664	969,620,818	1,820	797


TSS protein_coding top 20% cutoff 0.5 >=5reads, A-C-G-T
1 - methylation
0.95
0.90
0.85
0.80
0.75
0.70
-1,000
-800
-600
-400
-200
0
200
400
600
800
1,000
Position relative to peak

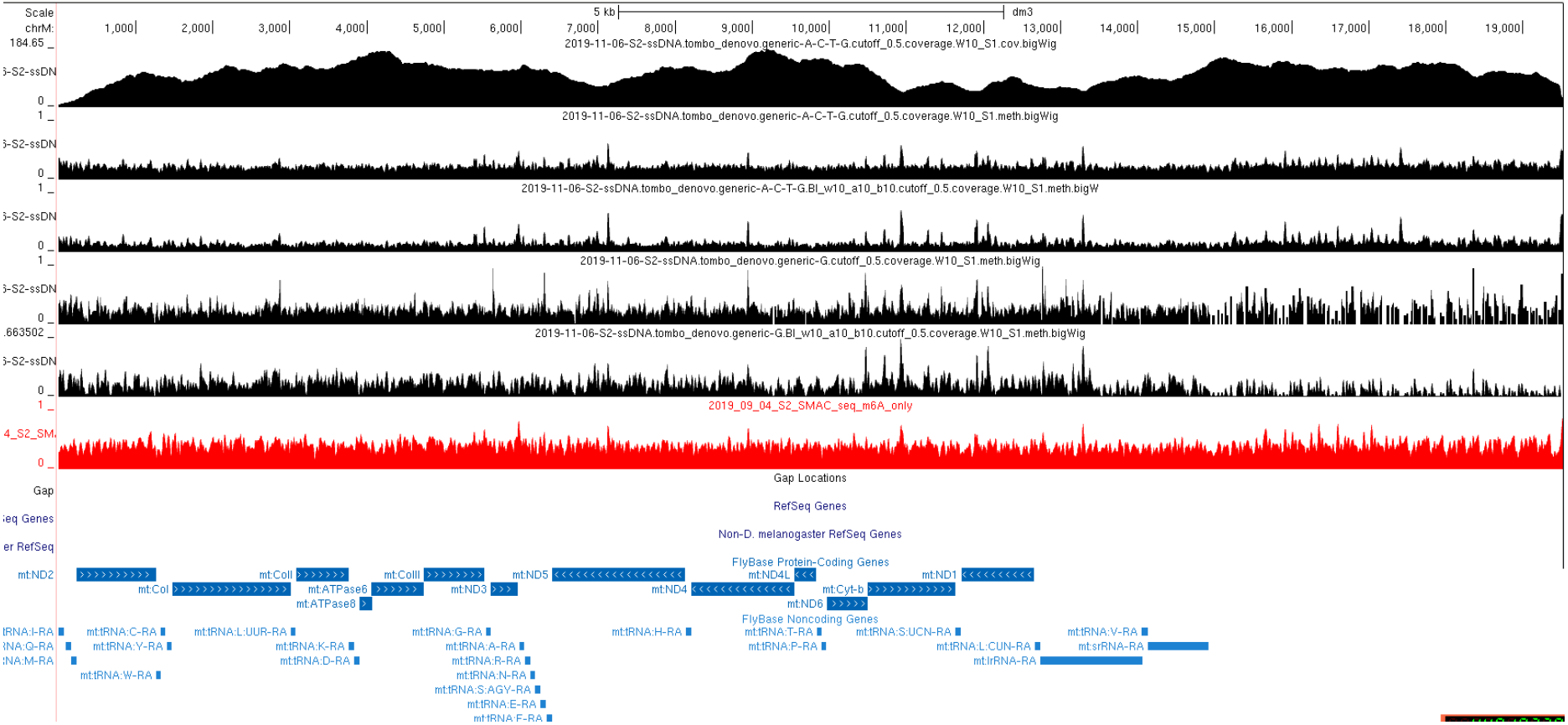

TSS protein_coding top 20% cutoff 0.5 >=5reads, G
1 - methylation
0.90
0.85
0.80
0.75
0.70
0.65
all
BI
-1,000
-800
-600
-400
-200
0
200
400
600
800
1,000
Position relative to peak


TSS protein_coding bottom 20% cutoff 0.5 >=5reads, A-C-G-T
all
BI
1 - methylation
0.90
0.85
0.80
0.75
0.70
-2,000
-1,800
-1,600
-1,400
-1,200
-1,000
-800
-600
-400
-200
0
200
400
600
800
1,000
1,200
1,400
1,600
1,800
2,000
Position relative to peak

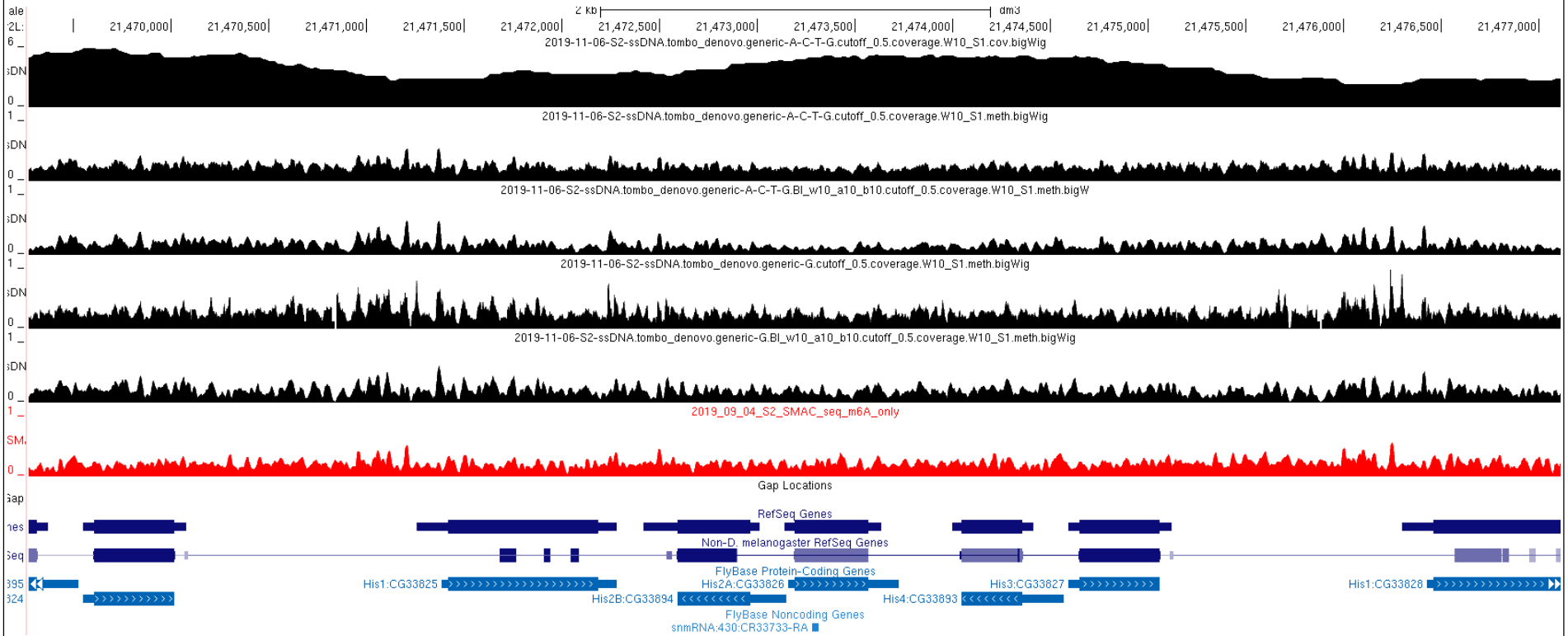

TSS protein_coding bottom 20% cutoff 0.5 >=5reads, G
all
BI
1 - methylation
0.90
0.85
0.80
0.75
0.70
0.65
-2,000
-1,800
-1,600
-1,400
-1,200
-1,000
-800
-600
-400
-200
0
200
400
600
800
1,000
1,200
1,400
1,600
1,800
2,000
Position relative to peak


TSS protein_coding top coverage
9.0
8.5
8.0
7.5
7.0
6.5
coverage
bottom 20%
top 20%
-1,000
-800
-600
-400
-200
0
200
400
600
800
1,000
Position relative to peak

chrM:


Scale
chrM:
5 kb
dm3
1,000
2,000
3,000
4,000
5,000
6,000
7,000
8,000
9,000
10,000
11,000
12,000
13,000
14,000
15,000
16,000
17,000
18,000
19,000
2019-11-06-S2-ssDNA.tombo_denovo.generic-A-C-T-G.cutoff_0.5.coverage.W10_S1.cov.bigWig
2019-11-06-S2-ssDNA.tombo_denovo.generic-A-C-T-G.cutoff_0.5.coverage.W10_S1.meth.bigWig
2019-11-06-S2-ssDNA.tombo_denovo.generic-A-C-T-G.BI_w10_a10_b10.cutoff_0.5.coverage.W10_S1.meth.bigW
2019-11-06-S2-ssDNA.tombo_denovo.generic-G.cutoff_0.5.coverage.W10_S1.meth.bigWig
2019-11-06-S2-ssDNA.tombo_denovo.generic-G.BI_w10_a10_b10.cutoff_0.5.coverage.W10_S1.meth.bigWig
2019_09_04_S2_SMAC_seq_m6A_only
Gap
Gap Locations
RefSeq Genes
Non-D. melanogaster RefSeq Genes
FlyBase Protein-Coding Genes
mt:ND2
mt:COI
mt:COII
mt:ATPase8
mt:ATPase6
mt:COIII
mt:ND3
mt:ND5
mt:ND4
mt:ND4L
mt:ND6
mt:ND1
mt:Cyt-b
FlyBase Noncoding Genes
mt:tRNA:I-RA
mt:tRNA:C-RA
mt:tRNA:Y-RA
mt:tRNA:W-RA
mt:tRNA:L:UUR-RA
mt:tRNA:K-RA
mt:tRNA:D-RA
mt:tRNA:G-RA
mt:tRNA:A-RA
mt:tRNA:R-RA
mt:tRNA:N-RA
mt:tRNA:S:AGY-RA
mt:tRNA:E-RA
mt:tRNA:F-RA
mt:tRNA:H-RA
mt:tRNA:T-RA
mt:tRNA:P-RA
mt:tRNA:S:UCN-RA
mt:tRNA:L:CUN-RA
mt:tRNA:V-RA
mt:srRNA-RA
mt:lrRNA-RA

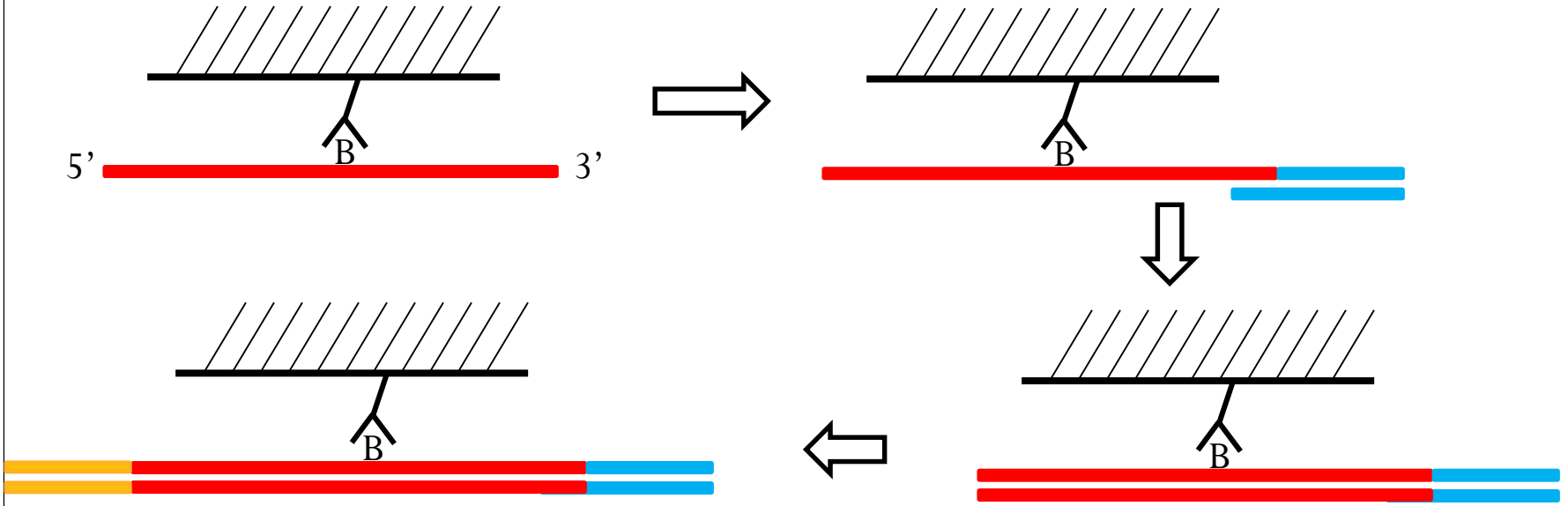
Histone genes:



1.4. TO DO LIST

SS-KAS-SEQ

- It would potentially be very useful to know which strand is actually single-stranded
- This might distinguish between R-loops/transcriptional bubbles and DNA that is single-stranded for other reasons (if any)
- There is a straightforward strategy to do that, by pulling down denatured ssDNA and ligating adaptors on beads:



scKAS-SEQ

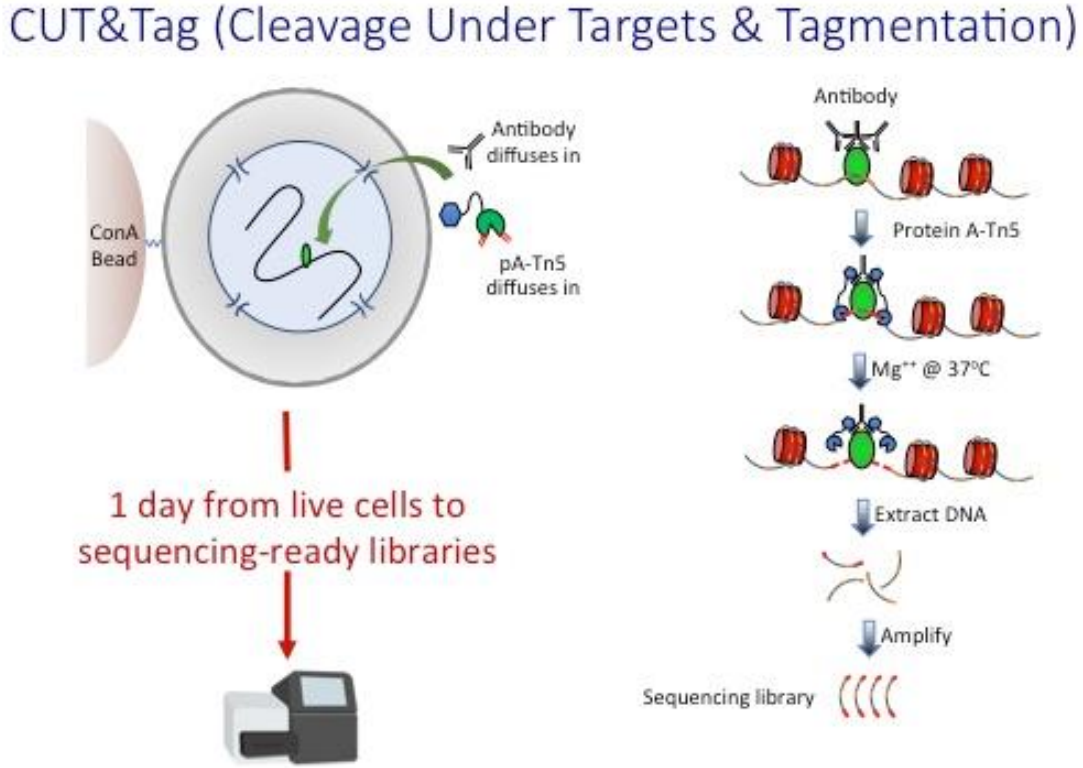
- The main objective behind the collaboration is the development of a single-cell version of KAS-seq

- Key problem – how do we do enrichment while also barcoding cells?

- Idea #1: we do cell barcoding on the 10X, then we do click chemistry and biotin pulldown
  - would require crosslinking and denaturation prior to transposition
  - losses would probably be substantial

- Idea #2: adapt CUT&Tag to KAS-seq, then to scKAS-seq and potentially to sciKAS-seq

CUT&TAG



PMID: 31036827

GENERAL STRATEGY AND VARIATIONS TO TEST:

1. Start with a KAS-seq reaction in bulk

2. Stop the reaction by washing away the ketoxal, then we crosslink in order to "freeze" things

3. Denature cells with SDS the way we do it for Hi-C

4. Do click reaction *in situ*

5. CUT&Tag variations:

    5.1 Use an anti-biotin antibody to bring the pA-Tn5 to ssDNA sites
    5.2 Incubate with streptavidin, then use an anti-streptavidin antibody to bring pA-Tn5

6. For first scKAS-seq experiment, distribute cells into ICELL8 wells for PCR

7. Longer-term, do pool split using ligation and/or indexed pA-Tn5
    - need indexed pA-Tn5 and/or phosphorylated pA-Tn5 for that)

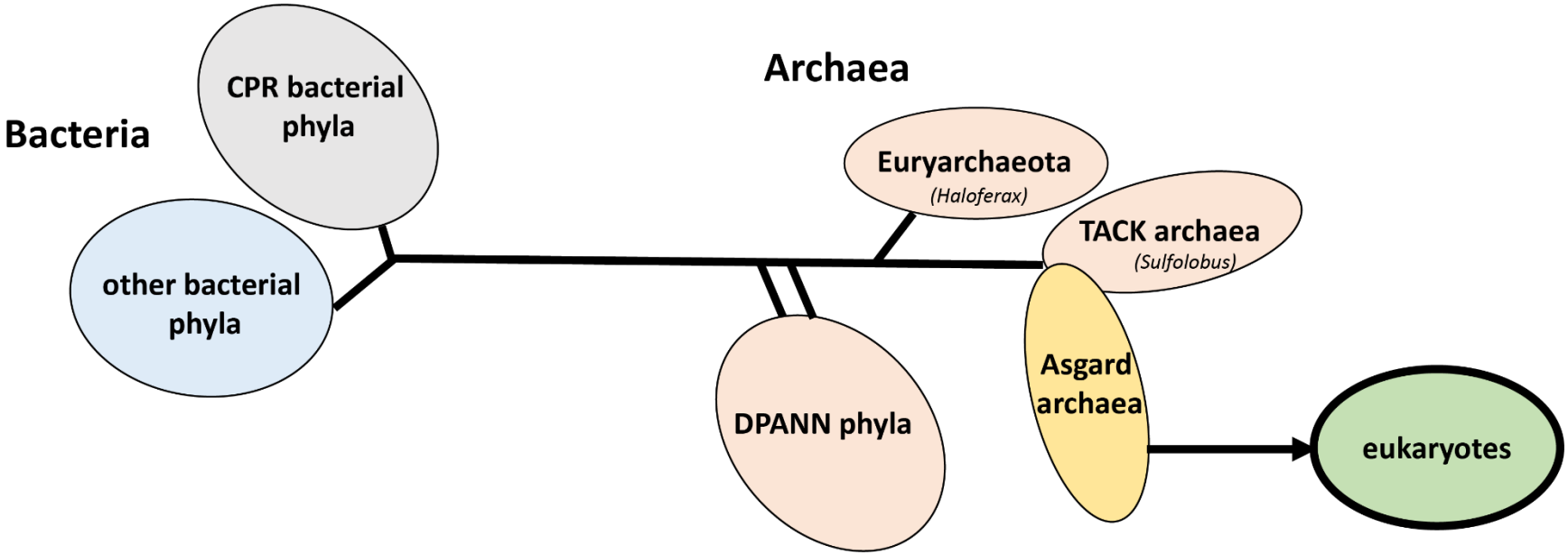
CRISPR OFF-TARGETS PROFILING?

1. Incubate gDNA with a dCas9 RNP
2. Treat with ketoxal
3. Isolate DNA
4. Click biotin onto DNA
5. Make libraries, sequence

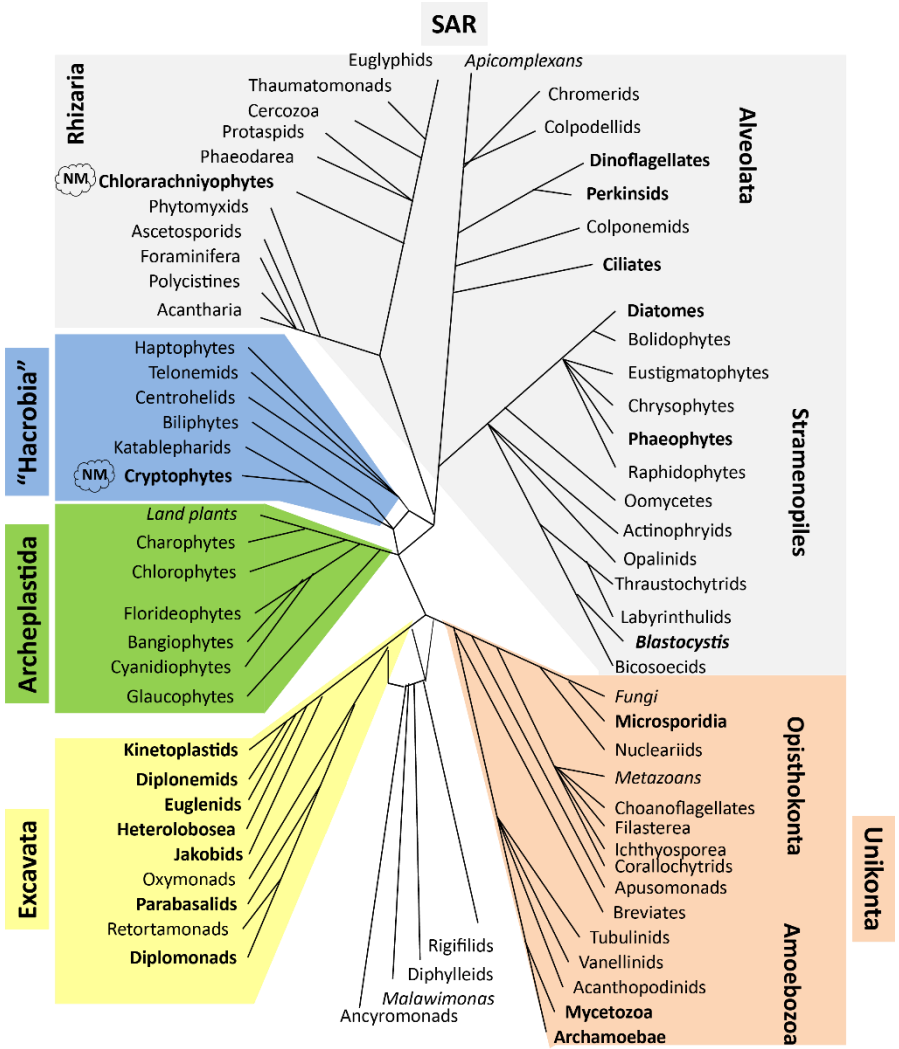
2. THE PHYSICAL GENOME ACRROSS EVOLUTION

2.1. OVERVIEW

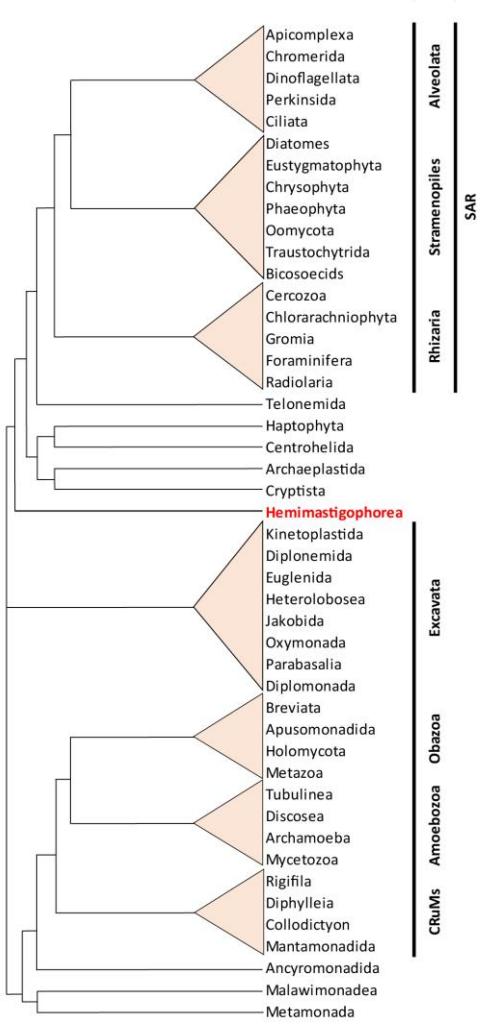
LIFE ON EARTH:



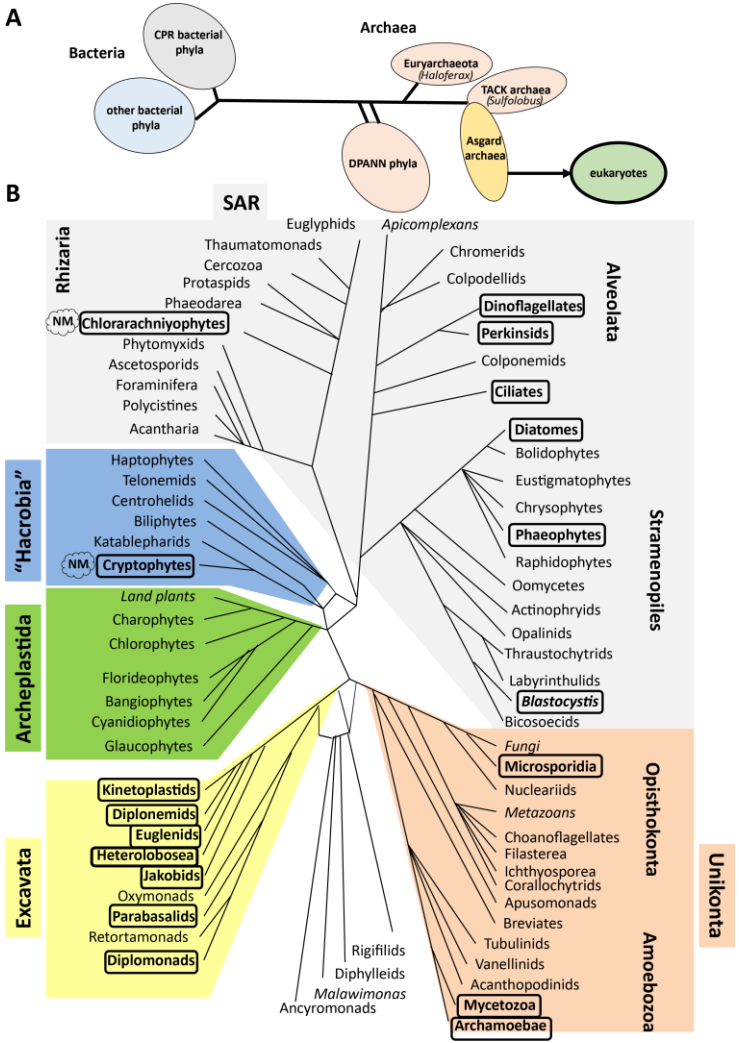
EUKARYOTE TREE



EUKARYOTE TREE, UPDATED VERSION



TARGETS

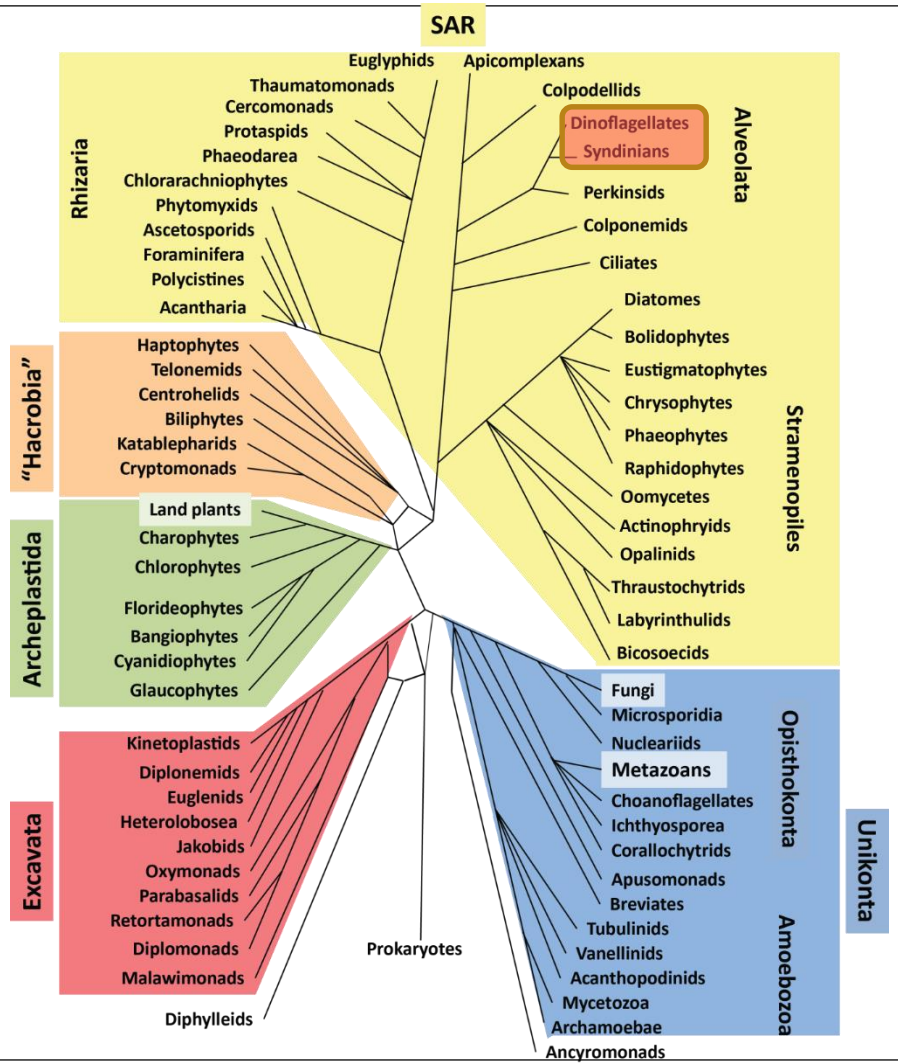

A
Bacteria
CPR bacterial phyla
other bacterial phyla
Archaea
Euryarchaeota (Haloferax)
TACK archaea (Sulfolobus)
Asgard archaea
DPANN phyla
eukaryotes
B
SAR
Rhizaria
Euglyphids
Thaumatomonads
Cercozoa
Protaspids
Phaeodarea
NM
Chlorarachniyophytes
Phytomyxids
Ascetosporids
Foraminifera
Polycistines
Acantharia
Apicomplexans
Alveolata
Chromerids
Colpodellids
Dinoflagellates
Perkinsids
Colponemids
Ciliates
Stramenopiles
Diatomes
Bolidophytes
Eustigmatophytes
Chrysophytes
Phaeophytes
Raphidophytes
Oomycetes
Actinophryids
Opalinids
Thraustochytrids
Labyrinthulids
Blastocystis
Bicosoecids
"Hacrobia"
Haptophytes
Telonemids
Centrohelids
Biliphytes
Katablepharids
NM
Cryptophytes
Archeplastida
Land plants
Charophytes
Chlorophytes
Florideophytes
Bangiophytes
Cyanidiophytes
Glaucophytes
Unikonta
Opisthokonta
Fungi
Microsporidia
Nucleariids
Metazoans
Choanoflagellates
Filasterea
Ichthyosporea
Corallochytrids
Apusomonads
Breviates
Amoebozoa
Tubulinids
Vanellinids
Acanthopodinids
Mycetozoa
Archamoebae
Rigifilids
Diphylleids
Malawimonas
Ancyromonads
Excavata
Kinetoplastids
Diplonemids
Euglenids
Heterolobosea
Jakobids
Oxymonads
Parabasalids
Retortamonads
Diplomonads

WHY ARE THESE THINGS INTERESTING?

GENERAL QUESTIONS:

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

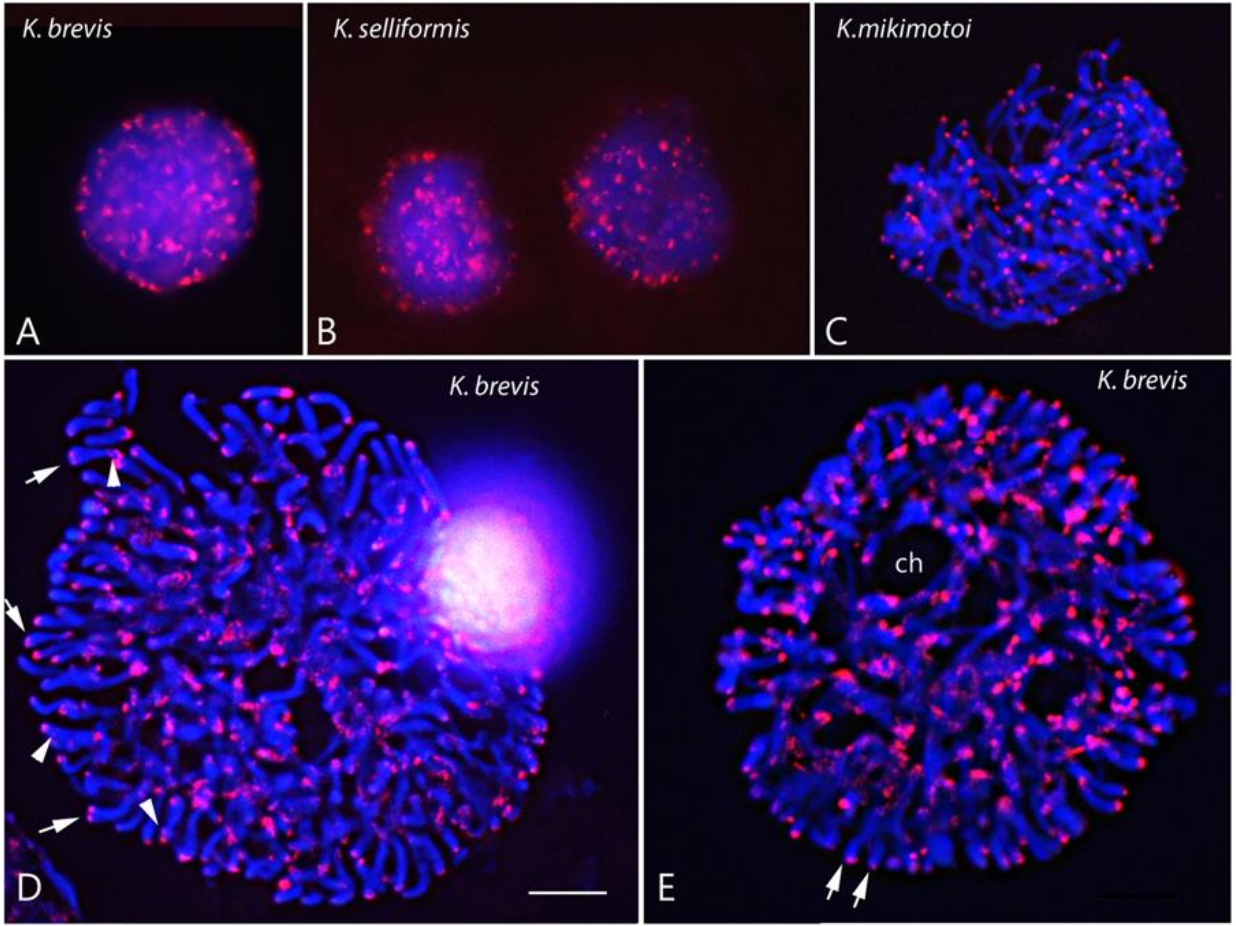
DINOFLAGELLATES



DINOFLAGELLATES AND THEIR SPECIAL FEATURES

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

DINOFLAGELLATE CHROMOSOMES



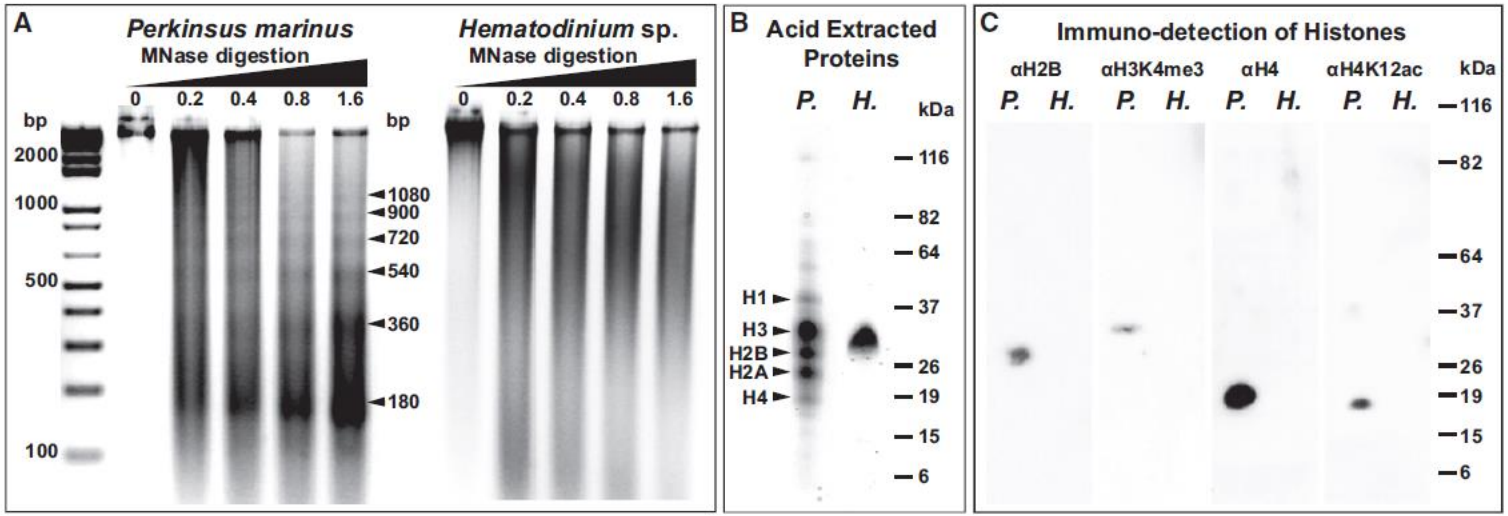
Cuadrado et al. 2019

- DVNPs (Dinoflagellate Viral Nucleoproteins, unrelated to histones) and HLPs (histone-like proteins) are thought to be the main packaging components

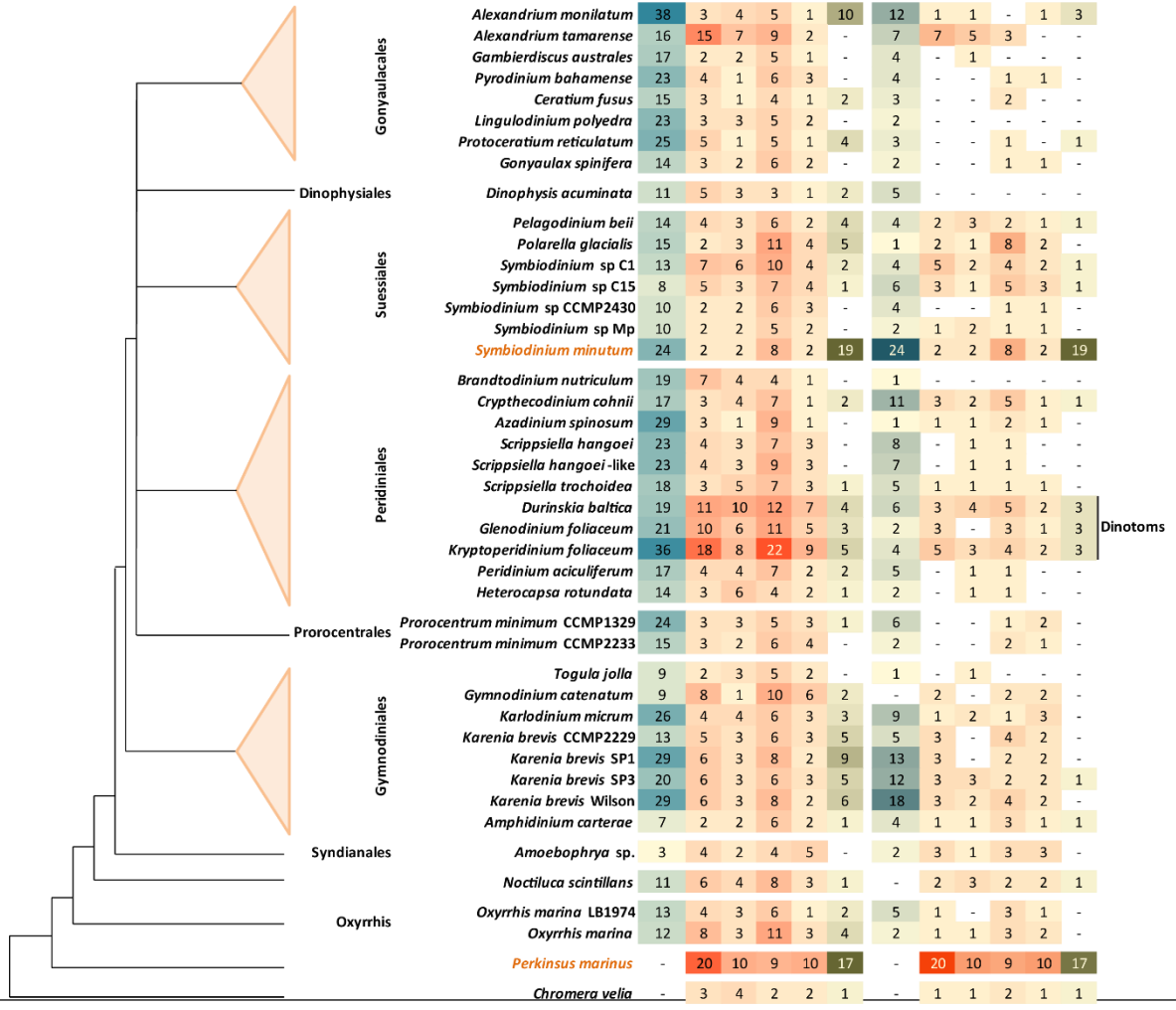
Current Biology 22, 2303–2312, December 18, 2012 ©2012 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2012.10.036

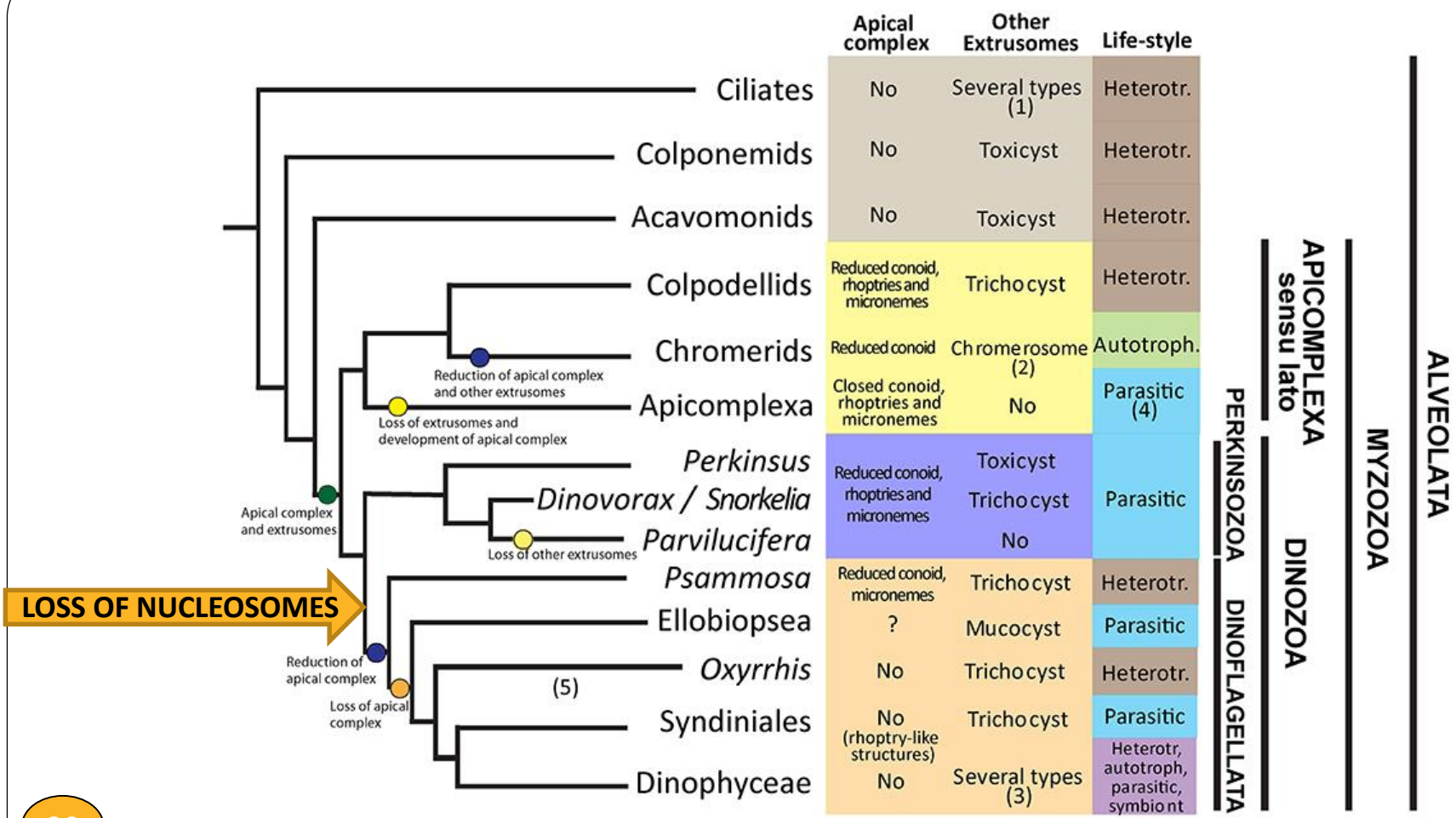
Article

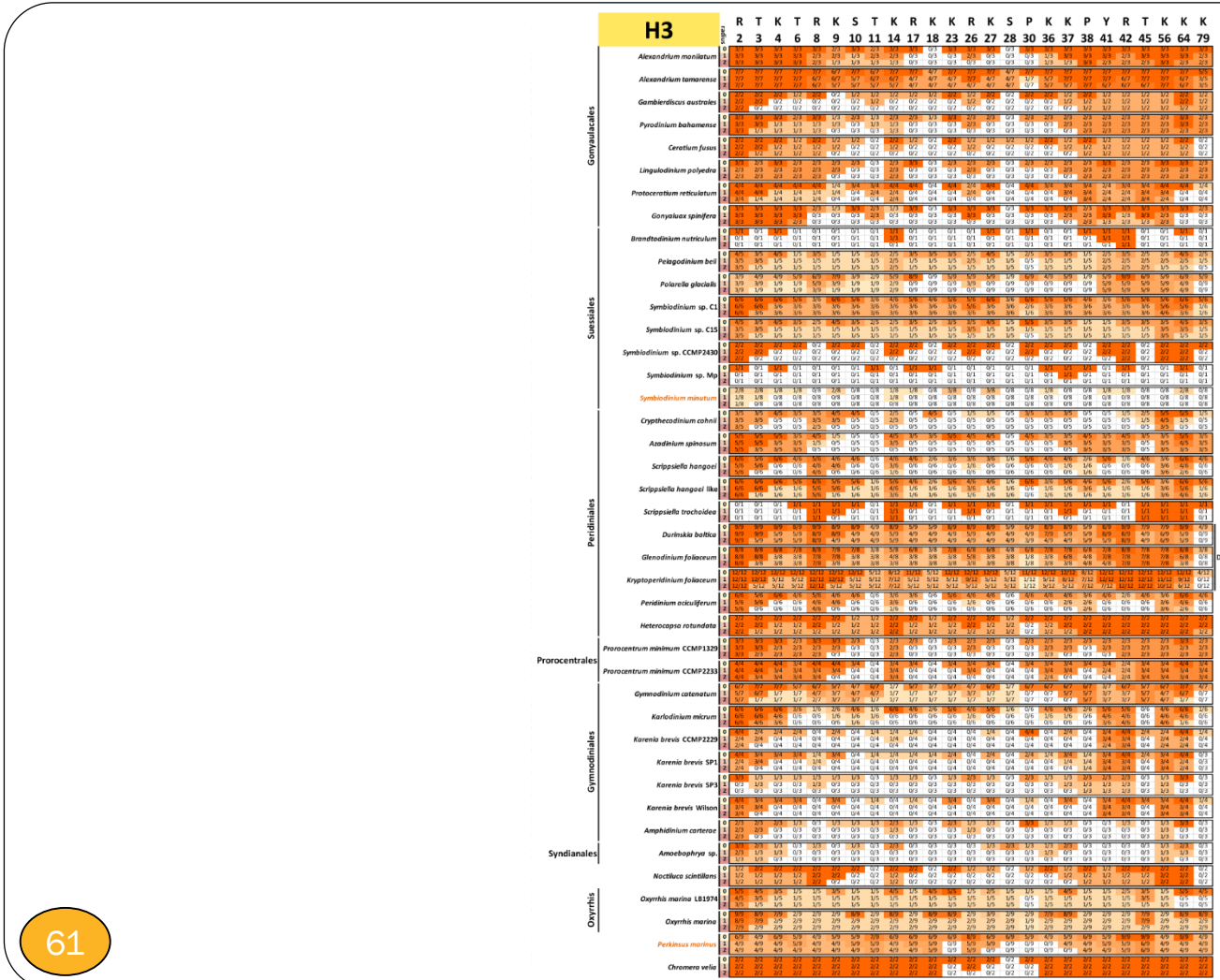
Loss of Nucleosomal DNA Condensation Coincides with Appearance of a Novel Nuclear Protein in Dinoflagellates



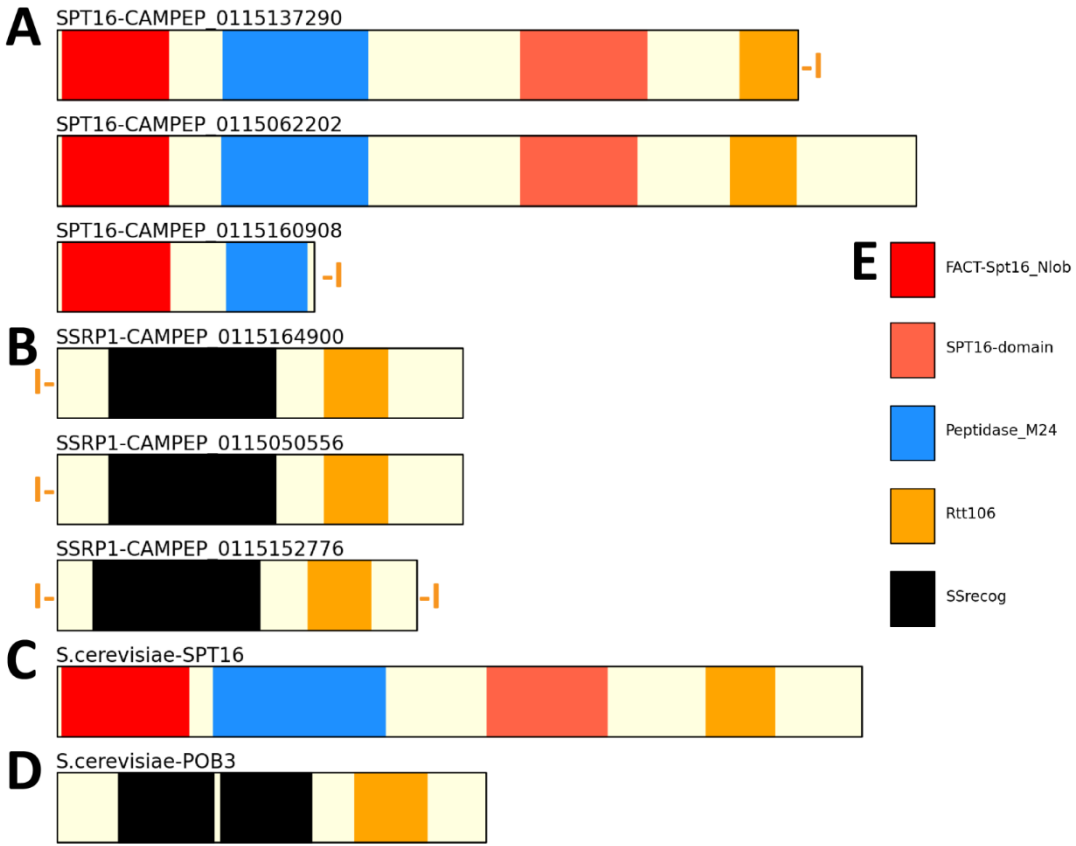
BUT HISTONES ARE PRESENT




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
Dinovorax / Snorkelia
Parvilucifera
Reduced conoid, rhoptries and micronemes
Toxicyst
Trichocyst
No
Parasitic
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
(5)
Reduction of apical complex and other extrusomes
Loss of extrusomes and development of apical complex
Apical complex and extrusomes
Loss of other extrusomes
Reduction of apical complex
Loss of apical complex
LOSS OF NUCLEOSOMES
APICOMPLEXA sensu lato
DINOZOA
PERKINSOZOA
DINOFLAGELLATA
MYZOZOA
ALVEOLATA


H3
Gonyalacales
Suessiales
Peridiniales
Prorocentrales
Gymnodiniales
Syndianales
Oxyrrhis
Dinotoms
Alexandrium monilatum
Alexandrium tamarense
Gambierdiscus australes
Pyrodinium bahamense
Ceratium fusus
Lingulodinium polyedra
Protoceratium reticulatum
Gonyaulax spinifera
Brandtodinium nutriculum
Peisgodinium bell
Polarella glacialis
Symbiodinium sp. C1
Symbiodinium sp. C115
Symbiodinium sp. CCMP2430
Symbiodinium sp. Mp
Symbiodinium minutum
Crypthecodinium cohnii
Azadinium spinosum
Scrippsiella hangoei
Scrippsiella hangoei like
Scrippsiella trochoidea
Durinskia baltica
Glenodinium foliaceum
Kryptoperidinium foliaceum
Peridinium aciculiferum
Heterocapsa rotundata
Prorocentrum minimum CCMP1329
Prorocentrum minimum CCMP2233
Gymnodinium catenatum
Karlodinium micrum
Karenia brevis CCMP2229
Karenia brevis SP1
Karenia brevis SP3
Karenia brevis Wilson
Amphidinium carterae
Amoebophrya sp.
Noctiluca scintillens
Oxyrrhis marina LB1974
Oxyrrhis marina
Perkinsus marinus
Chromera velia

THE FACT COMPLEX IN DINOFLAGELLATES



QUESTIONS:

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

CORAL SYMBIOSIS

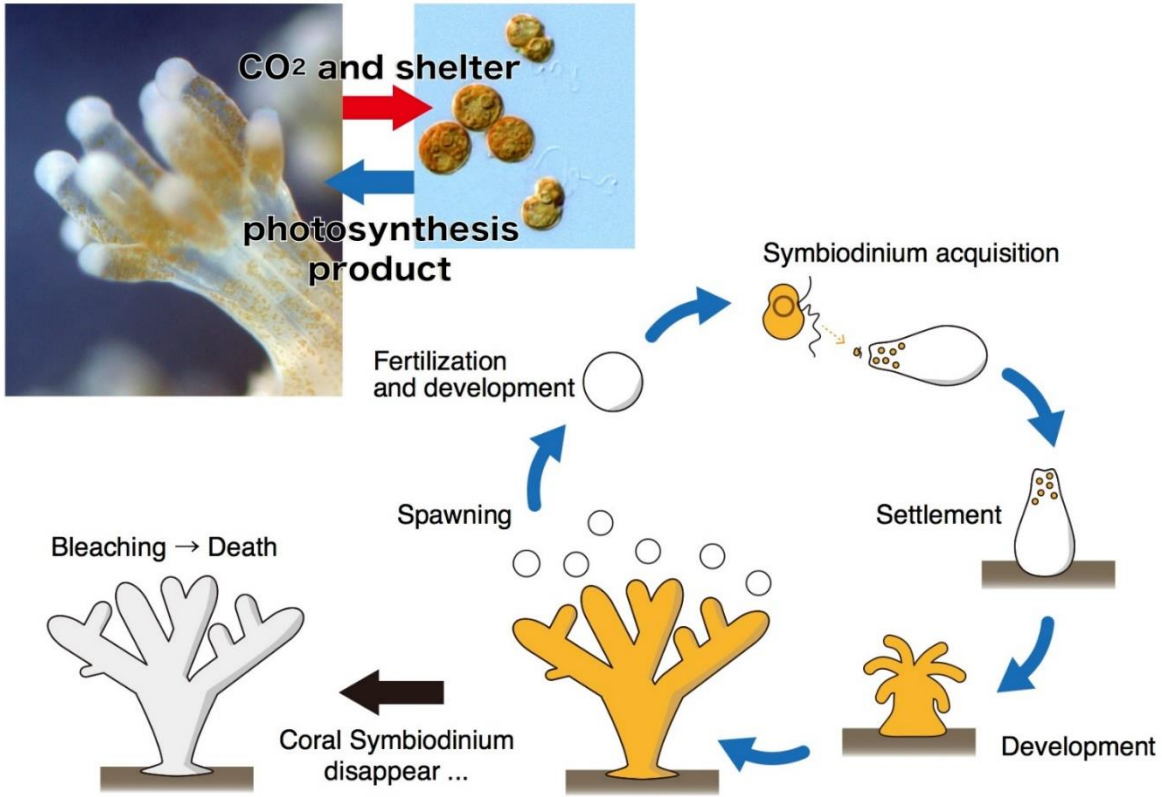
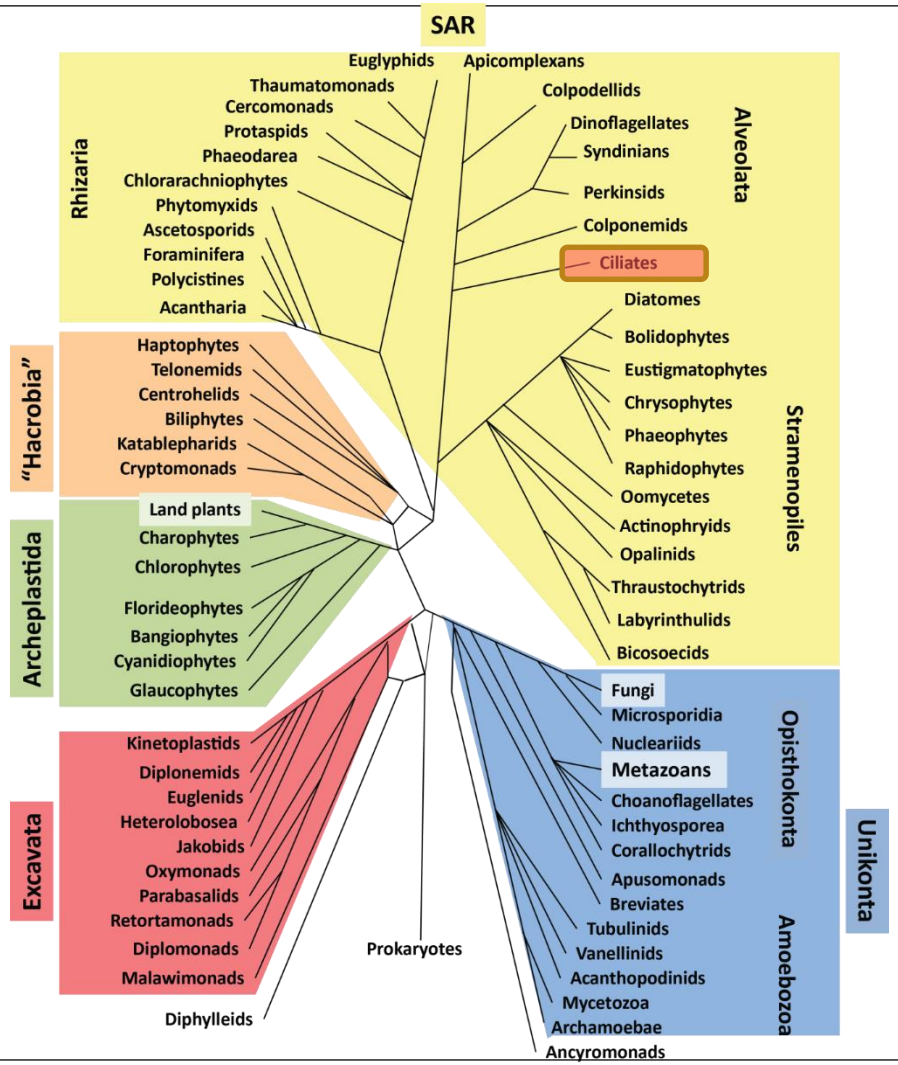


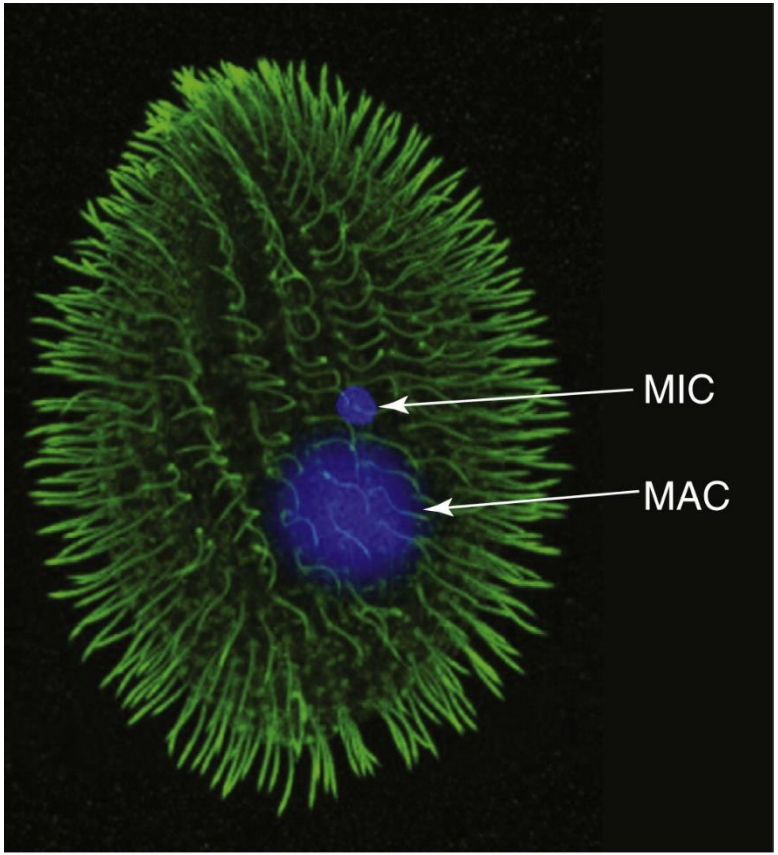
Figure 2. A symbiotic relationship between corals and Symbiodinium

CORAL BLEACHING

CILIATES

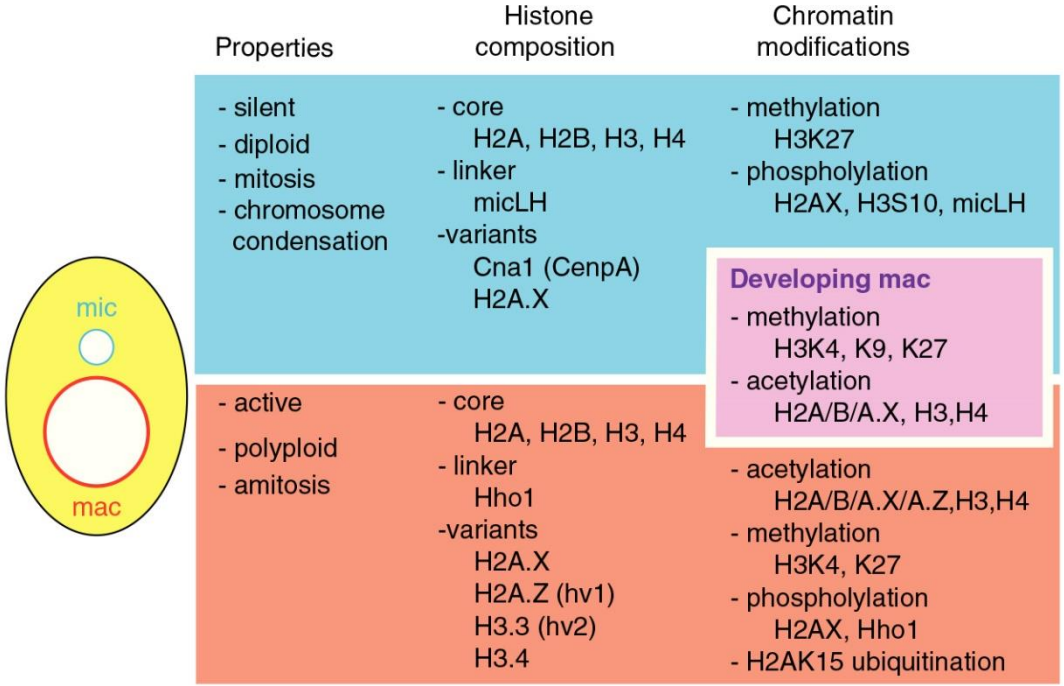

SAR
Rhizaria
Euglyphids
Thaumatomonads
Cercomonads
Protaspids
Phaeodarea
Chlorarachniophytes
Phytomyxids
Ascetosporids
Foraminifera
Polycistines
Acantharia
Alveolata
Apicomplexans
Colpodellids
Dinoflagellates
Syndinians
Perkinsids
Colponemids
Ciliates
Stramenopiles
Diatomes
Bolidophytes
Eustigmatophytes
Chrysophytes
Phaeophytes
Raphidophytes
Oomycetes
Actinophryids
Opalinids
Thraustochytrids
Labyrinthulids
Bicosoecids
"Hacrobia"
Haptophytes
Telonemids
Centrohelids
Biliphytes
Katablepharids
Cryptomonads
Archeplastida
Land plants
Charophytes
Chlorophytes
Florideophytes
Bangiophytes
Cyanidiophytes
Glaucophytes
Excavata
Kinetoplastids
Diplonemids
Euglenids
Heterolobosea
Jakobids
Oxymonads
Parabasalids
Retortamonads
Diplomonads
Malawimonads
Diphylleids
Prokaryotes
Unikonta
Opisthokonta
Fungi
Microsporidia
Nucleariids
Metazoans
Choanoflagellates
Ichthyosporea
Corallochytrids
Amoebozoa
Apusomonads
Breviates
Tubulinids
Vanellinids
Acanthopodinids
Mycetozoa
Archamoebae
Ancyromonads

NUCLEAR DIMORPHISM



Current Biology

NUCLEAR DIMORPHISM

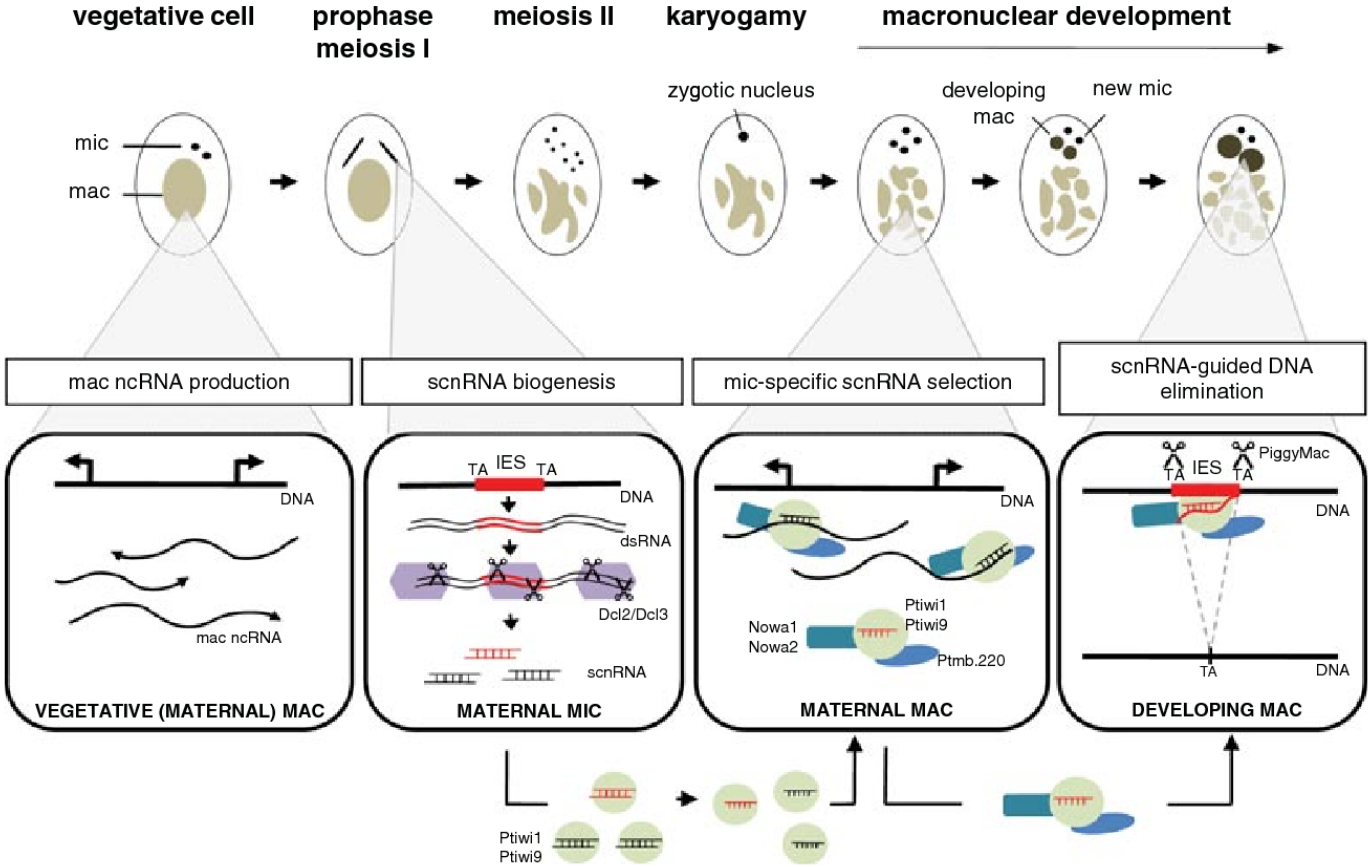


MAC ploidy levels:

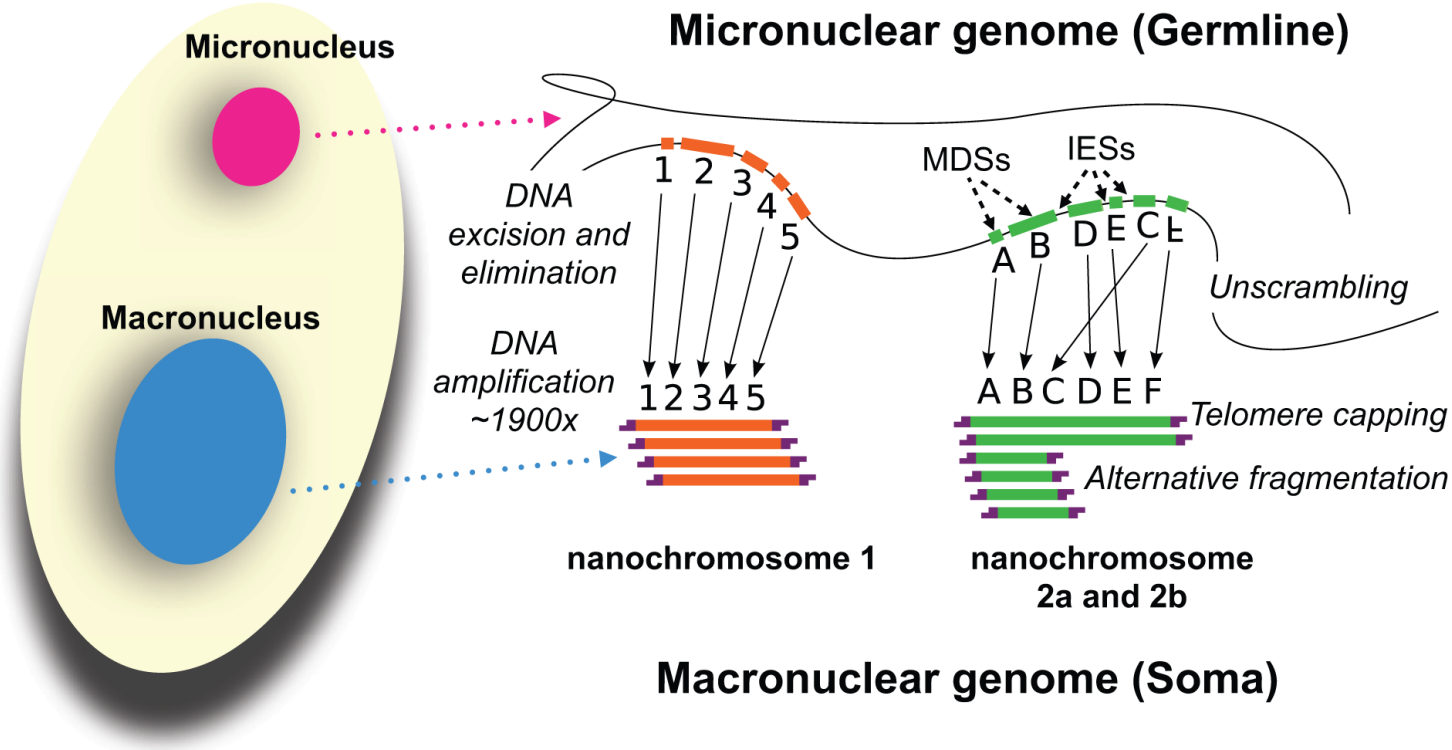
- 45n in *Tetrahymena*
- ~1000n in *Paramecium*
- >10,000n *in Stylonichia*

Figure 2. Nuclear dimorphism of ciliates. The germline micronucleus (mic), the developing macronucleus (mac), and the somatic macronucleus contain different histone complements and modifications. Those known to occur specifically in each or in the developing somatic genome are listed.

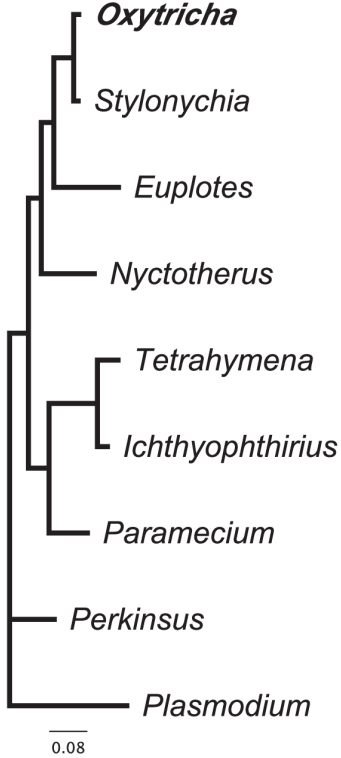
FROM MIC TO MAC, PROGRAMMED DNA ELIMINATION:



PROGRAMMED DNA ELIMINATION WITH UNSCRAMBLING AND NANOCHROMOSOMES:

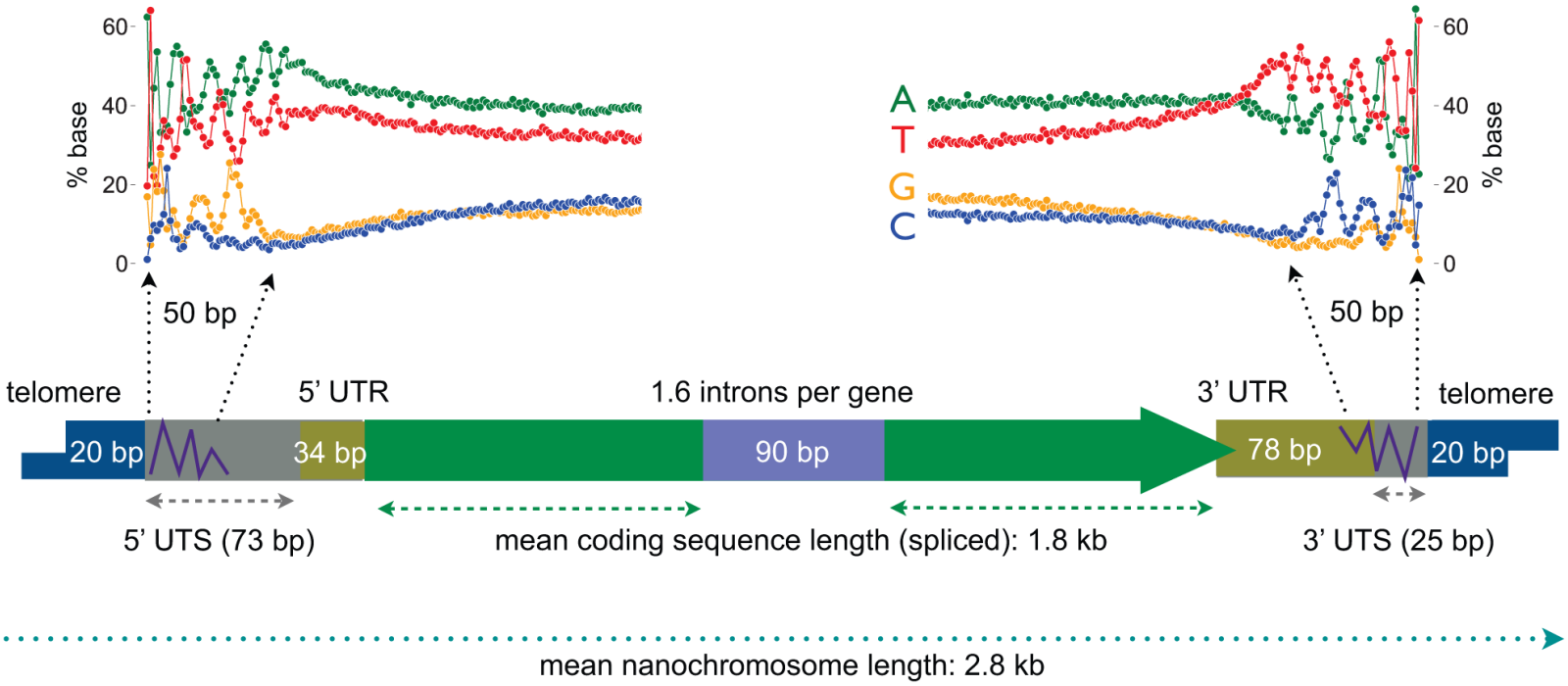


NANOCHROMOSOMES:



	Genome size	Genes	Chromo-somes	Ploidy	Alternative fragmen-tation
Oxytricha	**~50 Mb**	**~18,400**	**~15,600**	**Variable ~1,900[a]**	**Yes**
Stylonychia	~50 Mb[b]	~12,000[c]	~10-15,000[b]	Variable ~15,000[b]	Yes
Euplotes	~50 Mb?	?	nano ?	Variable ~2,000[d]	No?
Nyctotherus	~50 Mb[e]	?	nano ?	Variable[e] ?	?
Tetrahymena	105 Mb[f]	24,700[g]	225[f]	45[f]	limited[h,i]
Ichthyophthirius	49 Mb[j]	8,100[j]	71[j]	~12,000[j]	?
Paramecium	72 Mb[k]	40,000[k]	~200[k]	~800[l]	limited[l]
Perkinsus	87 Mb	23,700	?	1	NA
Plasmodium	23 Mb[m]	5,300[m]	14[m]	1	NA

NANOCHROMOSOMES:



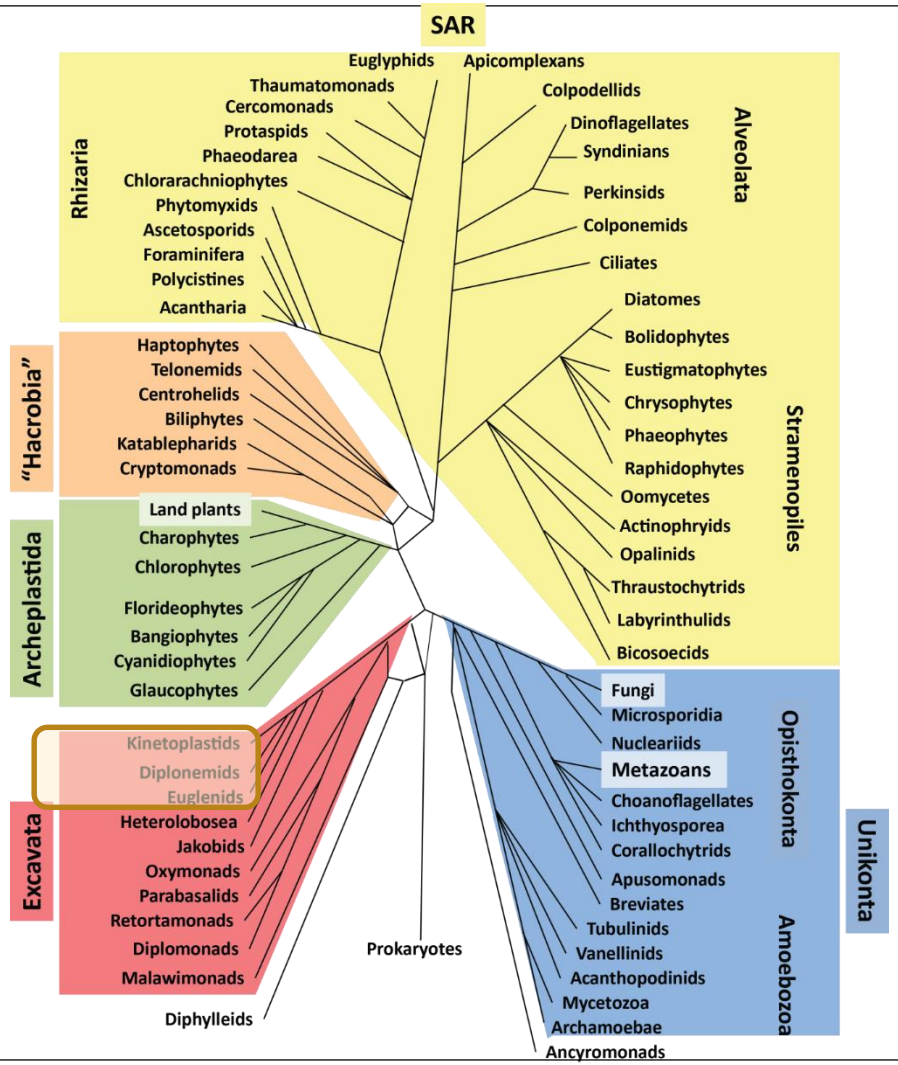
QUESTIONS:

- How is chromatin organized in 3D space when millions of nanochromosomes are present?
- How is transcriptional regulation accomplished with so little regulatory space?

TARGET SPECIES:

- *Tetrahymena*
- *Oxytricha / Euplotes*

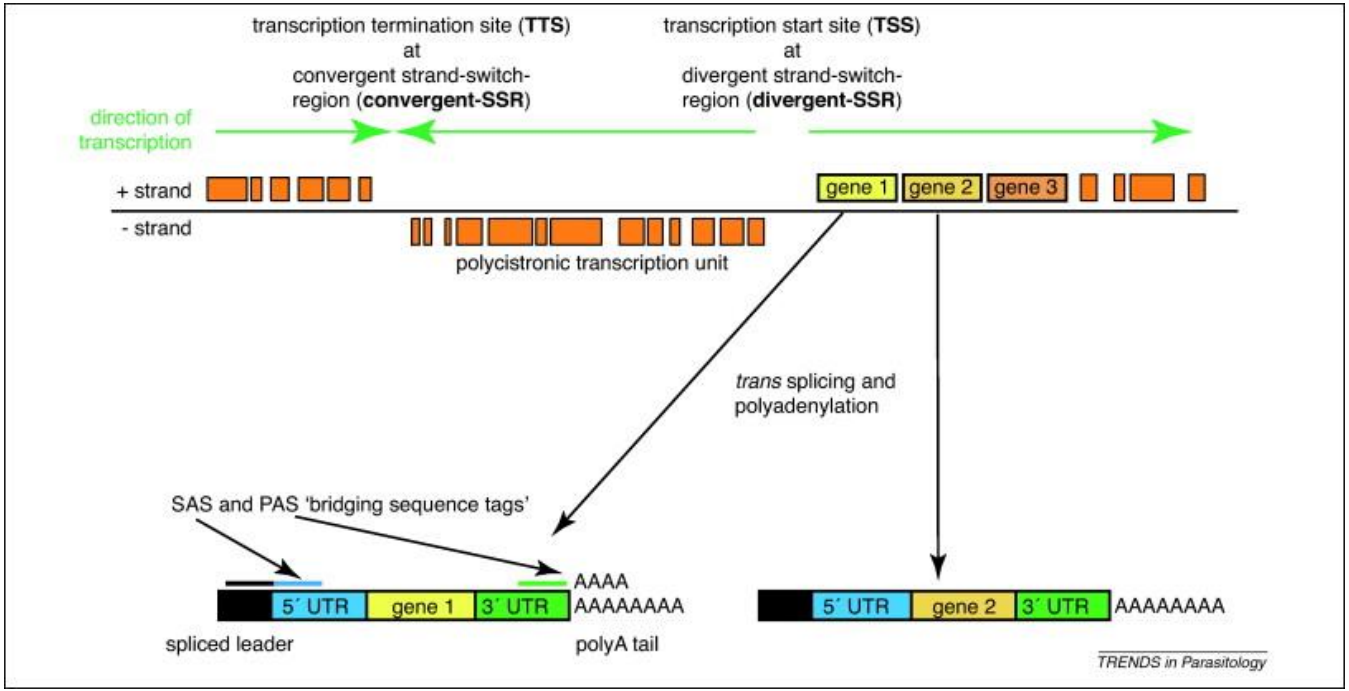
EUGLENIDS


SAR
Rhizaria
Euglyphids
Thaumatomonads
Cercomonads
Protaspids
Phaeodarea
Chlorarachniophytes
Phytomyxids
Ascetosporids
Foraminifera
Polycistines
Acantharia
Alveolata
Apicomplexans
Colpodellids
Dinoflagellates
Syndinians
Perkinsids
Colponemids
Ciliates
Stramenopiles
Diatomes
Bolidophytes
Eustigmatophytes
Chrysophytes
Phaeophytes
Raphidophytes
Oomycetes
Actinophryids
Opalinids
Thraustochytrids
Labyrinthulids
Bicosoecids
"Hacrobia"
Haptophytes
Telonemids
Centrohelids
Biliphytes
Katablepharids
Cryptomonads
Archeplastida
Land plants
Charophytes
Chlorophytes
Florideophytes
Bangiophytes
Cyanidiophytes
Glaucophytes
Excavata
Kinetoplastids
Diplonemids
Euglenids
Heterolobosea
Jakobids
Oxymonads
Parabasalids
Retortamonads
Diplomonads
Malawimonads
Diphylleids
Prokaryotes
Unikonta
Opisthokonta
Fungi
Microsporidia
Nucleariids
Metazoans
Choanoflagellates
Ichthyosporea
Corallochytrids
Apusomonads
Breviates
Amoebozoa
Tubulinids
Vanellinids
Acanthopodinids
Mycetozoa
Archamoebae
Ancyromonads

SPECIAL PROPERTIES:

- Near complete loss of transcriptional regulation

- Polycistronic gene arrays



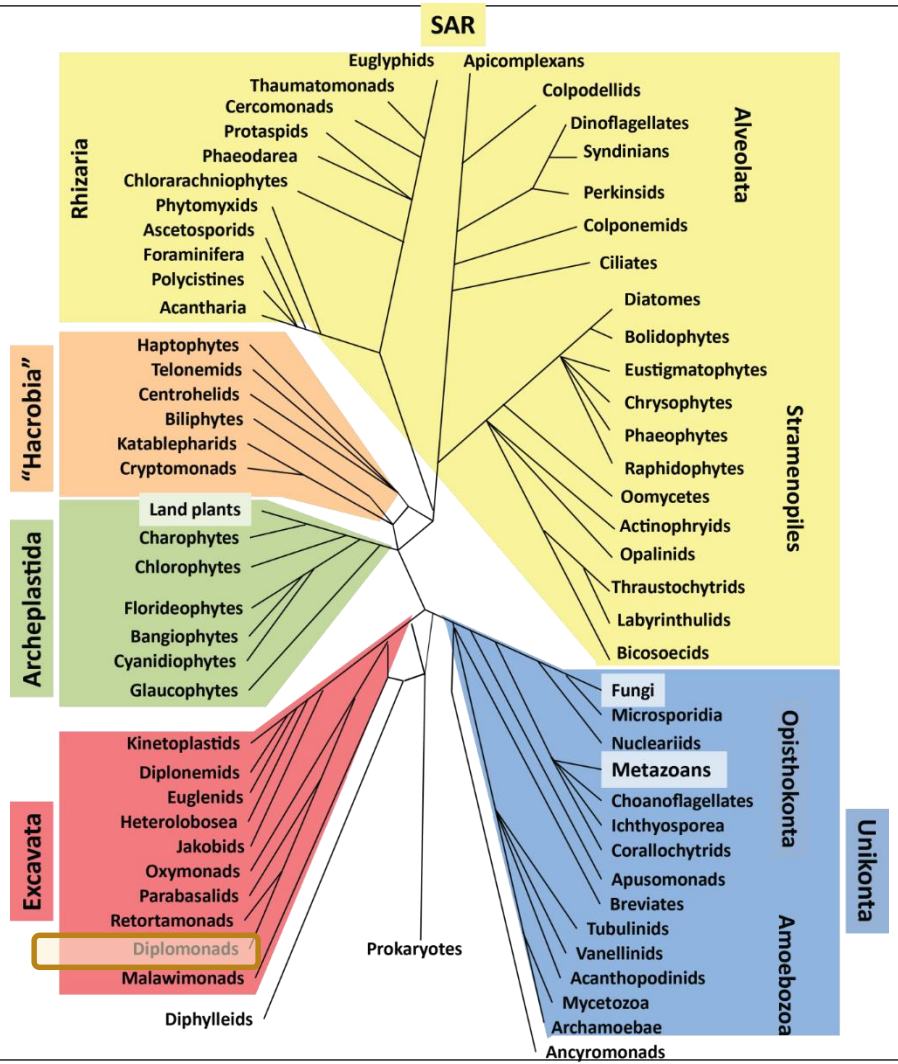
SPECIAL PROPERTIES:

- Near complete loss of transcriptional regulation
- Polycistronic gene arrays
- Permanently condensed chromatin in diplonemids and perhaps in a few other groups

TARGET SPECIES:

- *Trypanosoma / Leishmania*
- *Bodo saltans*
- *Euglena*
- *Diplonema*

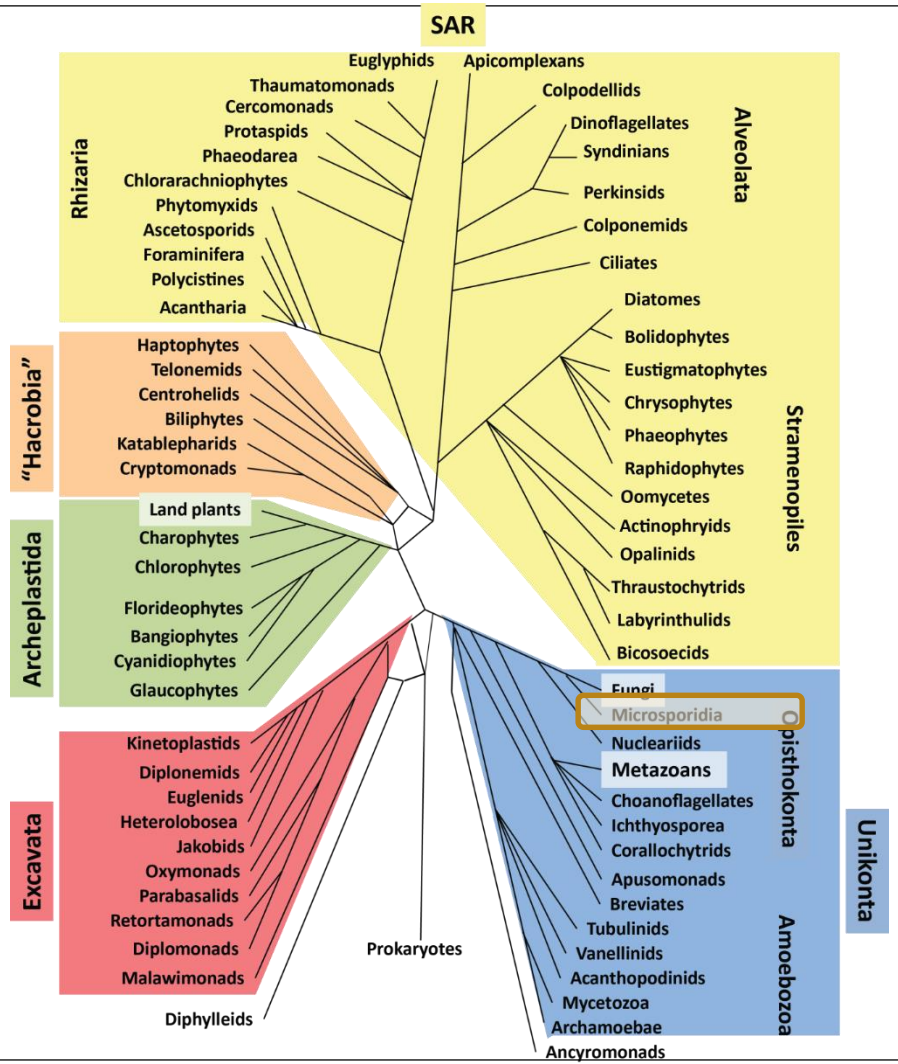
DIPLOMONADS



SPECIAL PROPERTIES:

- Absence of most general transcription factors
- Suspected pervasive bidirectional transcription

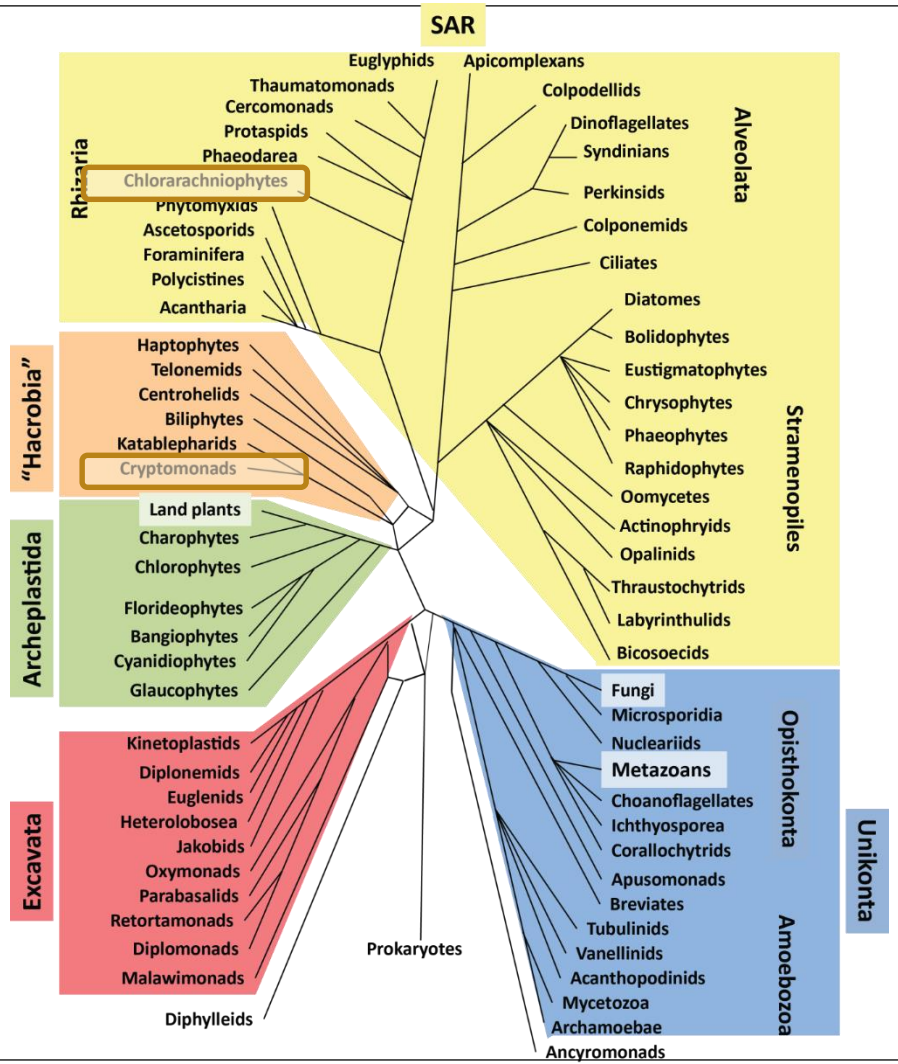
MICROSPORIDIA



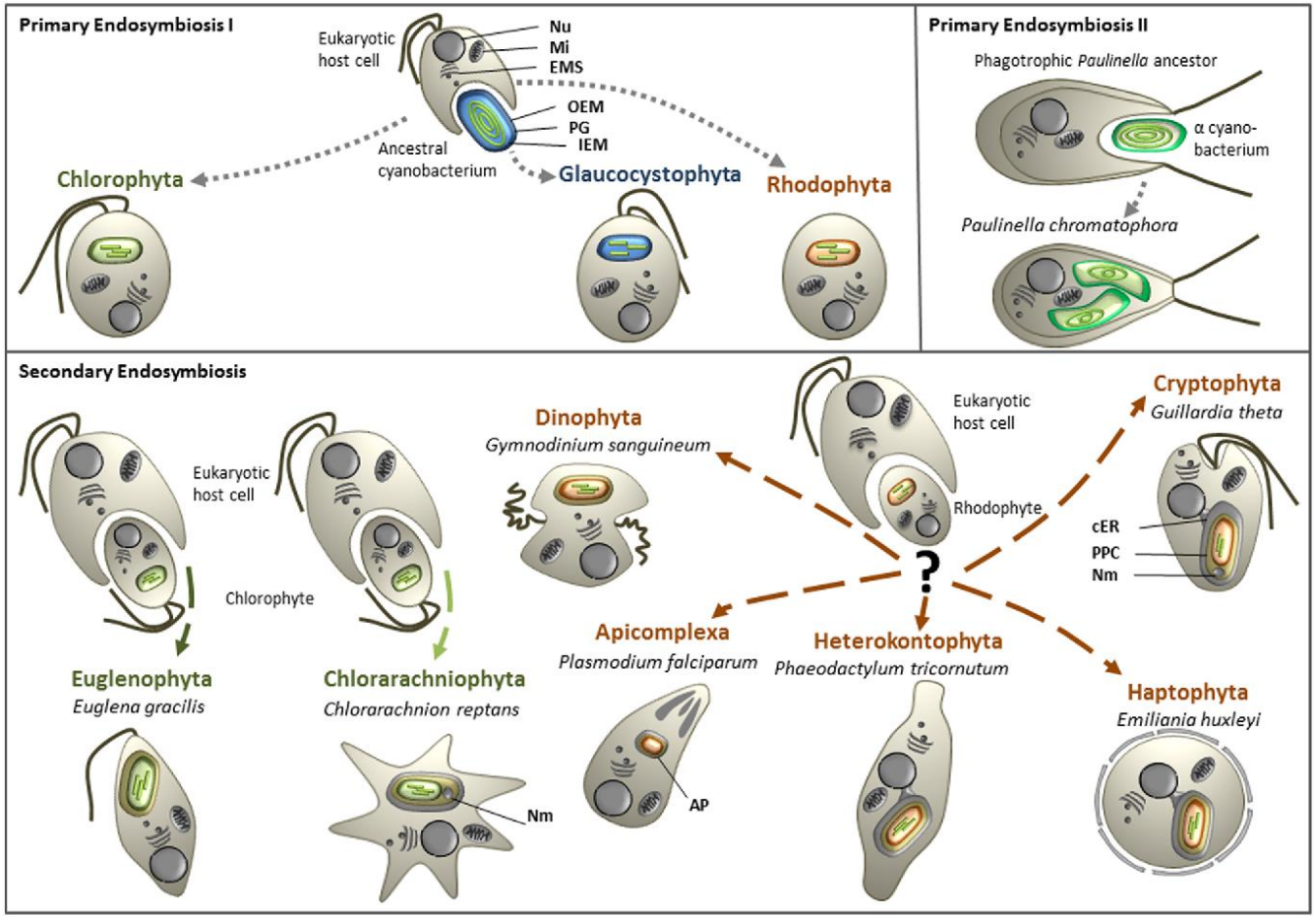
SPECIAL PROPERTIES:

- Obligatory intracellular parasites
- Most reduced autonomous eukaryote genomes known – down to 2.5 Mbp
- Extremely tightly packed genes, loss of some key histone marks

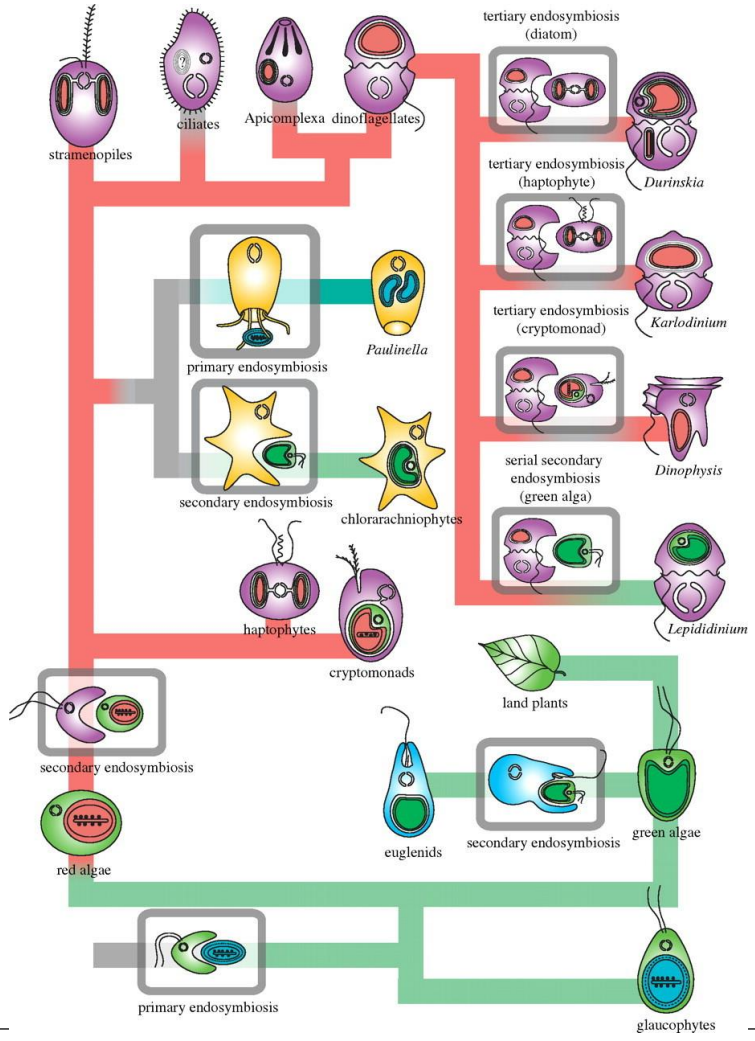
NUCLEOMORPHS



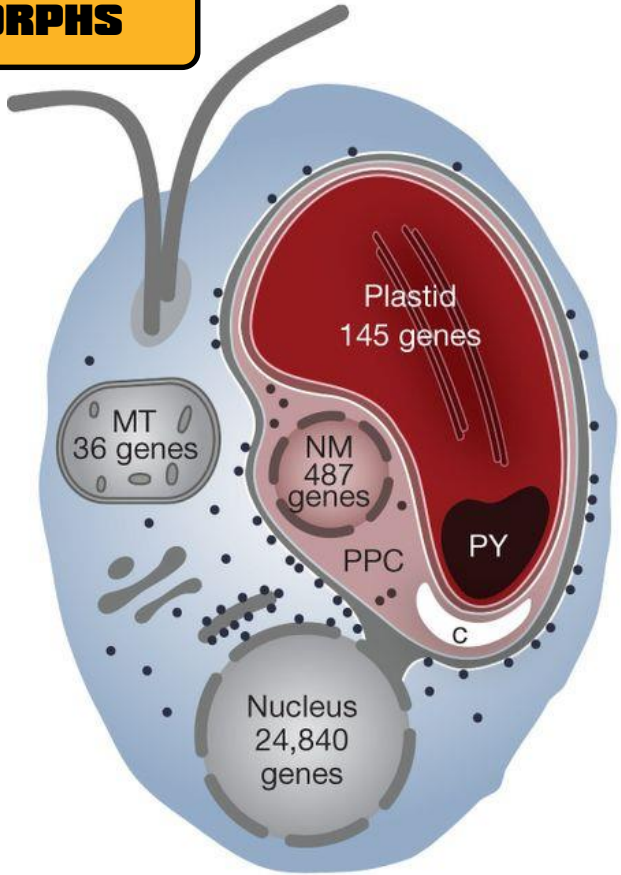
SECONDARY ENDOSYMBIOSIS



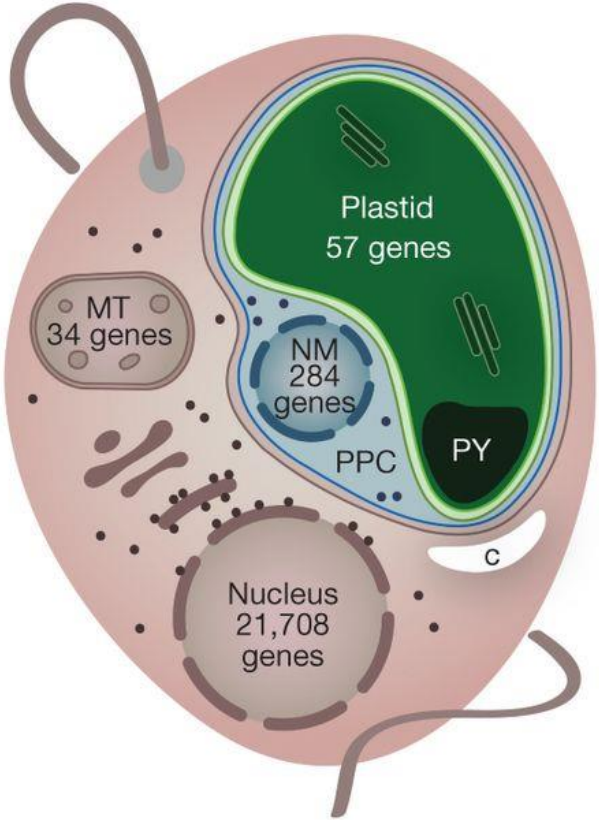
TERTIARY ENDOSYMBIOSIS



NUCLEOMORPHS



Guillardia theta



Bigelowiella natans

SPECIAL PROPERTIES:

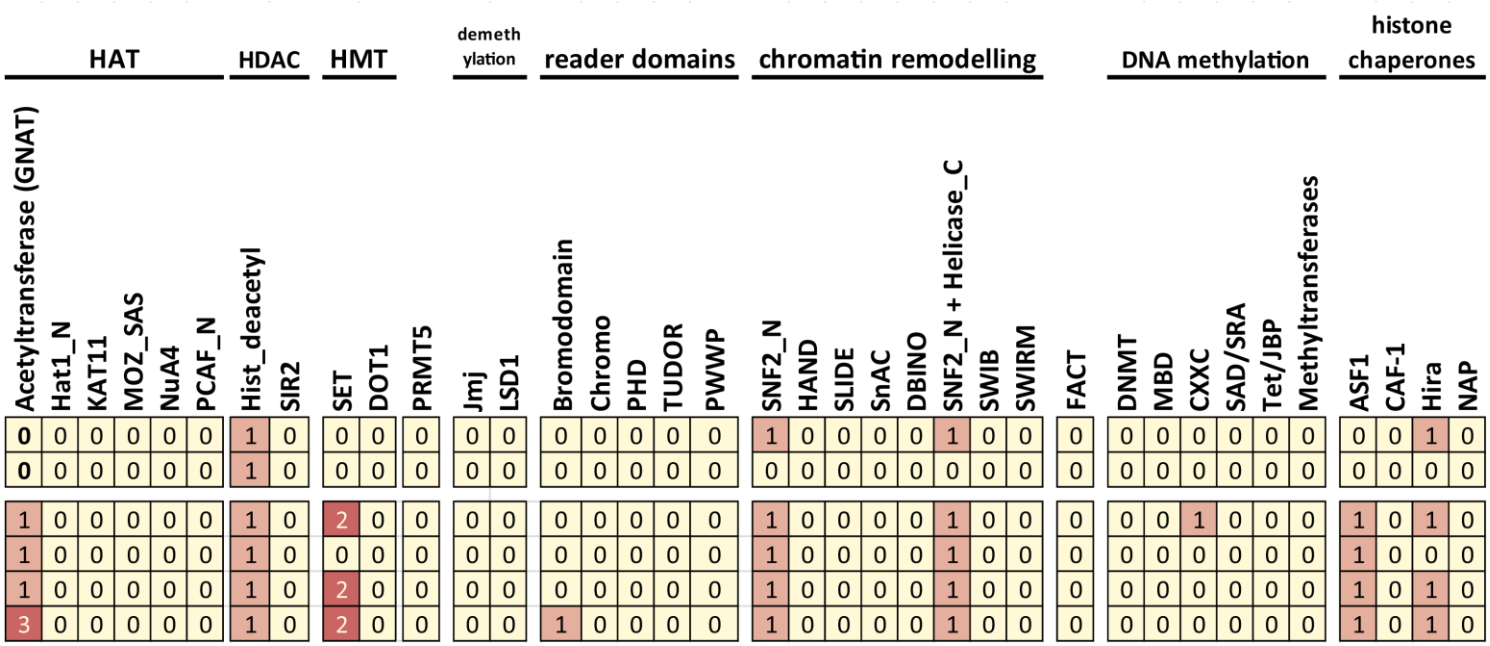
- Most reduced known eukaryote genomes
- Overlapping genes, extremely little intergenic space
- Surprisingly convergent evolution between the two groups – 3 chromosomes with subtelomeric rDNA arrays
- Divergent histone code and RNA Pol2 CTD

A **H3**

radius:	*Bigelowiella natans*			*Lotharella oceanica*			*Chroomonas mesostigmatica*			*Cryptomonas paramecium*			*Guillardia theta*			*Hemiselmis andersenii*		
	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

B **H4**

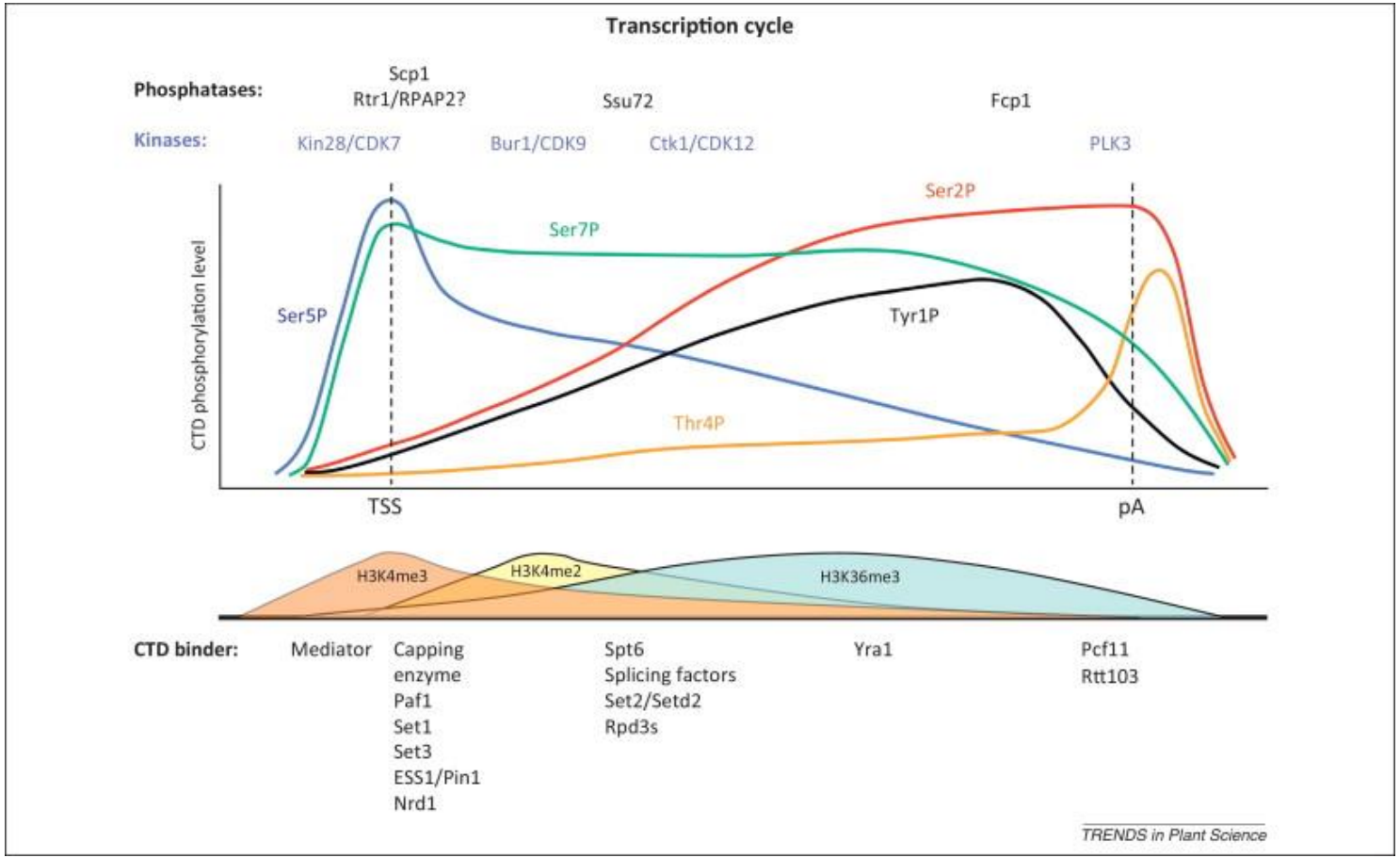
radius:	*Bigelowiella natans*			*Lotharella oceanica*			*Chroomonas mesostigmatica*			*Cryptomonas paramecium*			*Guillardia theta*			*Hemiselmis andersenii*		
	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2
S 1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
R 3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K 5	0	0	0	0	0	0	1	0	0	1	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


HAT
HDAC
HMT
demethylation
reader domains
chromatin remodelling
DNA methylation
histone chaperones
AcetylTransferase (GNAT)
Hat1_N
KAT11
MOZ_SAS
NuA4
PCAF_N
Hist_deacetyl
SIR2
SET
DOT1
PRMT5
Jmj
LSD1
Bromodomain
Chromo
PHD
TUDOR
PWWP
SNF2_N
HAND
SLIDE
SnAC
DBINO
SNF2_N + Helicase_C
SWIB
SWIRM
FACT
DNMT
MBD
CXXC
SAD/SRA
Tet/JBP
Methyltransferases
ASF1
CAF-1
Hira
NAP
Bigelowiella natans 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0
Lotharella oceanica 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Chroomonas mesostigmatica 1 0 0 0 0 0 1 0 2 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 1 0
Cryptomonas paramecium 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0
Guillardia theta 1 0 0 0 0 0 1 0 2 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0
Hemiselmis andersenii 3 0 0 0 0 0 1 0 2 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0

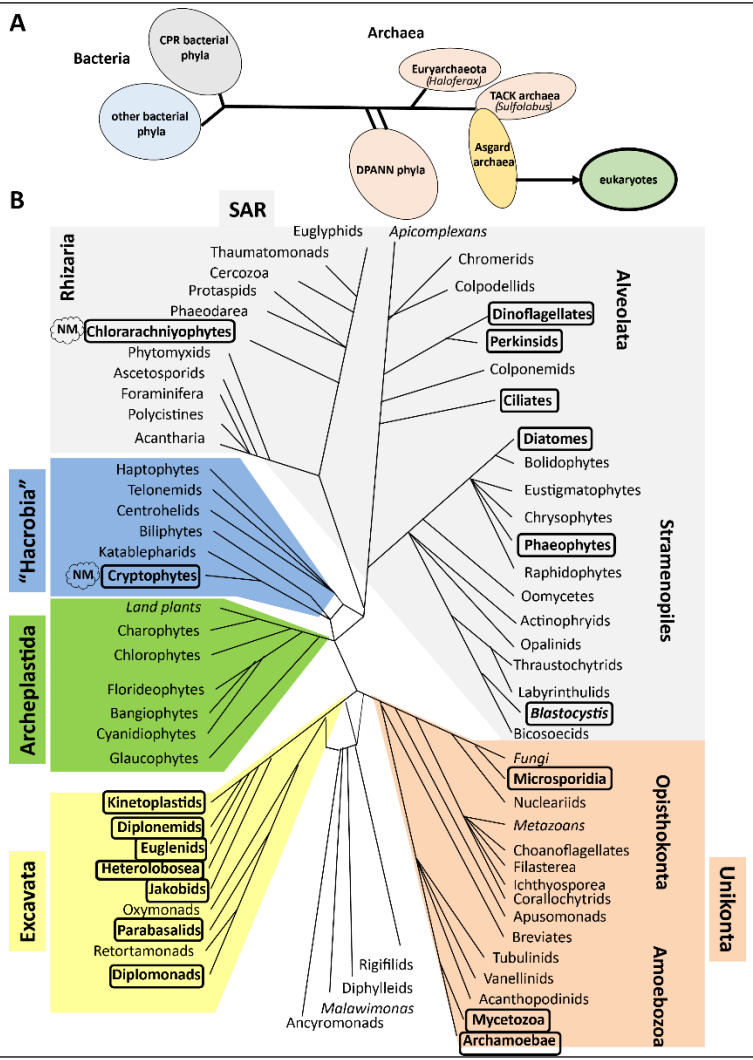

RPB1_Saccharomyces_cerevisiae/1-1733 1366 RHMALLVDVMTTQGGLTSVTRHGFNRSNTGALMRCSFEETVEILFEAGASAELDDCRGVSENVILGQMAPIGTGAFDVMIDEESLVKYMPEQKITEIEDGQDGGVTPYSNES 1477
Bigelowiella_natans/1-1510 1371 HHISCLADYSTNTGRPNAINRYGLKVQSLSTLQECSFEETMKTLLKASVETKVDRFSSITETIIFGKESNFGTGFSNMKKTKFQEVQ------ISKNSFKRSPLIIPSFNQS 1476
Lotharella_oceanica/1-1502 1354 HHILCLTDFMTNLGFLNSINRFGLKYQALSALQQCSFEETAKTFIKASINTKRDELNSITENVILGKLIKLGTGFFNPFVKPQLGNHKIKN--IVKQTLQQTKEINNNSLVI 1463
Guillardia_theta/1-1473 1392 RHLCTLVDLITFRGSLMSINRHGINKCNSGPIAKSTFEETVDIFYESAVFGLNDNLRGVSAQIMTGMVPKIGTGKIDLFVNI---------- 1473
Cryptomonas_paramecium/1-1452 1370 RHIFLLADIMTHGGKLMSITRHGISKIEIGPIAKSSFEETVDIFYQSAVFGIYDNIKGVSANIMVGNLLSVGTGKVDLFVDDF---------- 1452
Chroomonas_mesostigmatica/1-1473 1381 RHLYLLVDIMTHRGKLLSITRHGINKTEIGPIAKCSFEETIDILYQAAAFGISDNLKSISASVLLGQLPLLGTGKIDLFLSEKKLMEKEHTAK---------- 1473
Hemiselmis_andersenii/1-1464 1380 RHLSLLVDVMTHKGKLMSITRHGINRTDIGPIAKCSFEETIEVLYQAAAFGVCDNLKGVSGSILVGQVSSLGTGKIDLFLCEKKF---------- 1464
RPB1_Saccharomyces_cerevisiae/1-1733 1478 GLVNADLDVKDELMFSPLVDSGSNDAMAGGFTAYGGADYGEATSPFGAYGEAPTSPGFGVSSPGFSPTSPTYSPTSPAYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSP 1589
Bigelowiella_natans/1-1510 1477 KSSYKLVSYYK-----PLIKNKKLQYNPYHQKKFLEWEY---------- 1510
Lotharella_oceanica/1-1502 1464 FEQYKLYNLFKFKTVNNILSKKNNKINLNSKKYYYDWDY---------- 1502
Guillardia_theta/1-1473 ----------
Cryptomonas_paramecium/1-1452 ----------
Chroomonas_mesostigmatica/1-1473 ----------
Hemiselmis_andersenii/1-1464 ----------
RPB1_Saccharomyces_cerevisiae/1-1733 1590 SYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPNYSPTSP 1701
Bigelowiella_natans/1-1510 ----------
Lotharella_oceanica/1-1502 ----------
Guillardia_theta/1-1473 ----------
Cryptomonas_paramecium/1-1452 ----------
Chroomonas_mesostigmatica/1-1473 ----------
Hemiselmis_andersenii/1-1464 ----------
RPB1_Saccharomyces_cerevisiae/1-1733 1702 SYSPTSPGYSPGSPAYSPKQDEQKHNENENSR 1733
Bigelowiella_natans/1-1510 ----------
Lotharella_oceanica/1-1502 ----------
Guillardia_theta/1-1473 ----------
Cryptomonas_paramecium/1-1452 ----------
Chroomonas_mesostigmatica/1-1473 ----------
Hemiselmis_andersenii/1-1464 ----------

QUESTIONS:

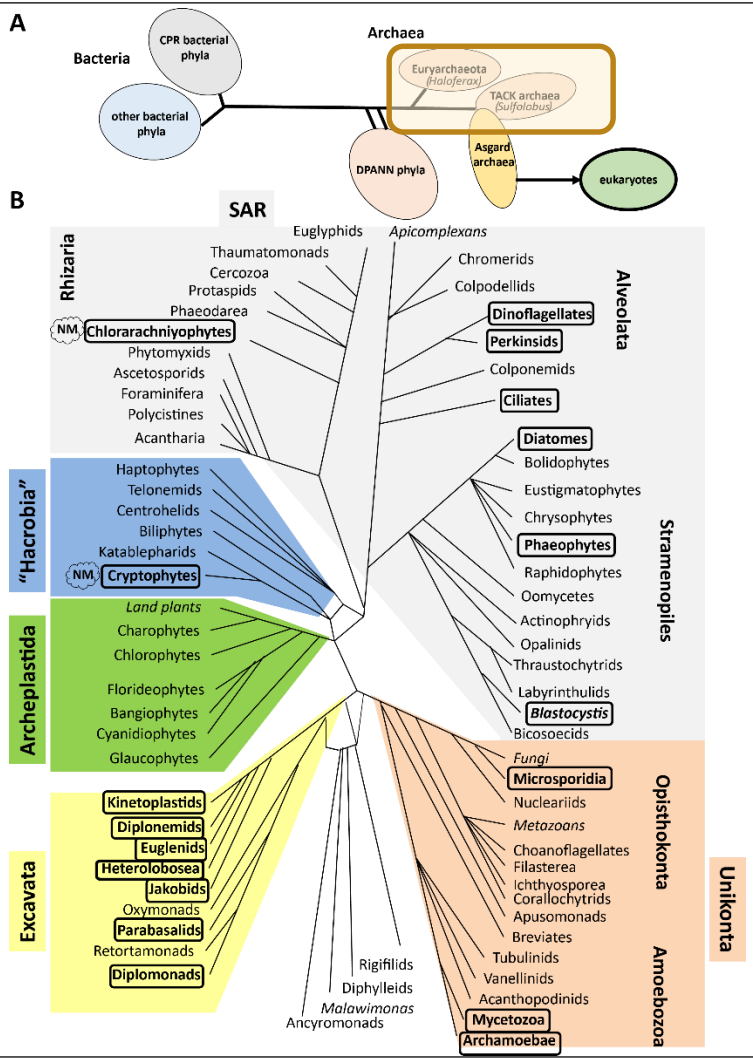
- How has the transcriptional machinery adapted to the context of extreme genomic reduction
- How has the histone code evolved
- How has the transcriptional cycle evolved, on its own and in relation to the histone code?


Transcription cycle
Phosphatases:
Scp1
Rtr1/RPAP2?
Ssu72
Fcp1
Kinases:
Kin28/CDK7
Bur1/CDK9
Ctk1/CDK12
PLK3
Ser2P
Ser7P
Ser5P
Tyr1P
Thr4P
CTD phosphorylation level
TSS
pA
H3K4me3
H3K4me2
H3K36me3
CTD binder:
Mediator
Capping enzyme
Paf1
Set1
Set3
ESS1/Pin1
Nrd1
Spt6
Splicing factors
Set2/Setd2
Rpd3s
Yra1
Pcf11
Rtt103
TRENDS in Plant Science

GENERAL SURVEY OF ALL MAJOR SUPERGROUPS:



ARCHAEA AS THE ANCESTORS



AIMS:

Specific Aim 1. Characterizing candidate cis-regulatory elements (cCREs) and chromatin accessibility states.

- ATAC-seq in each species under multiple conditions
- SMAC-seq for finding coaccessible regulatory elements

Specific Aim 2. Charting the relationship between the linear epigenome and 3D genomic organization.

- deep Hi-C maps in each targeted organism.
- histone mark ChIP-seq.
- DNA methylation

Specific Aim 3. Identification of trans-acting structural and regulatory non-coding RNA species.

- ChAR-seq

Specific Aim 4: Identifying the sequence determinants of CREs using interpretative deep learning methods

OUTCOME:

- Submitted the grant for the February 5th cycle

- Scores:

SUMMARY STATEMENT

PROGRAM CONTACT:
Jennifer Troyer
301.480.3565
troyerj@mail.nih.gov

(Privileged Communication)

Release Date: **06/21/2019**
Revised Date:

Application Number: **1 R01 HG010888-01**

Principal Investigators (Listed Alphabetically):

GREENLEAF, WILLIAM JAMES (Contact)
KUNDAJE, ANSHUL

Applicant Organization: STANFORD UNIVERSITY

Review Group: **GVE**
Genetic Variation and Evolution Study Section

Meeting Date: **06/06/2019**
Council: **OCT 2019**
Requested Start: **09/01/2019**

RFA/PA: **PA19-056**
PCC: **X6JT**

Project Title: **The physical genome across evolution**

SRG Action: **Impact Score:40 Percentile:32**
Next Steps: **Visit https://grants.nih.gov/grants/next_steps.htm**
Human Subjects: **10-No human subjects involved**
Animal Subjects: **10-No live vertebrate animals involved for competing appl.**

OUTCOME:

- Submitted the grant for the February 5th cycle

- Scores:

CRITIQUE 1

Significance: 4
Investigator(s): 3
Innovation: 2
Approach: 5
Environment: 1

CRITIQUE 2

Significance: 2
Investigator(s): 1
Innovation: 2
Approach: 2
Environment: 1

CRITIQUE 3

Significance: 1
Investigator(s): 4
Innovation: 1
Approach: 1
Environment: 1

OUTCOME:

Overall Impact: This proposal aims to study genome organization and gene regulation in a group of understudied but medically relevant organisms sampled among Eukaryotes, trying to cover as much diversity as possible. This need stems from recognizing that physical genome organization can be extremely variable across Eukaryotes and this variability has not been captured so far. This project showcases several innovating technologies and computational analyses, and as such promises to advance the field. PI Greenleaf has pioneered cutting-edge and widely used tools and is now leveraging nanopore-based technologies. Although very innovative, this proposal is extremely ambitious and lists a large effort that is unlikely to be accomplished on time. Together with having to adapt very new technologies to each single organism, high quality genomes and corresponding annotations are not available for most of the species included in the project, suggesting that the lab will need to invest a consistent amount of extra work to improve these genomes in order to make the datasets usable. A major concern about this proposal is that the ultimate product of this project will be large number of annotations and sequencing datasets but there is no plan to further interpret them in the context of human disease and evolution and/or compare them to each other. This might also be due to the lack of experience with these organisms and the paucity of collaborators proficient and interested in the biology of these organisms. As a result, the impact of this project will be greatly reduced.

OUTCOME:

Overall Impact: This is an exciting proposal that can substantially improve our understanding of gene expression and chromosome organization in eukaryotes. The careful study design provides insight into organisms of mayor health impact while at the same time allowing a reasonable assessment of eukaryote diversity. The proposed approach of studying ~20 species in parallel is highly advantageous as it allows comparisons between organisms that are broadly unaffected by batch effects. The study proposes to use cutting edge technologies that are likely going to be refined as part of the proposed work. These refinements add to the impact of the proposed work.

Overall Impact: This is a strong application by outstanding researchers. PI Greenleaf has developed a number of useful assays for genomics, some of which will be employed in this project. Characterization of functional genomic architectures and regulatory mechanisms in diverse eukaryotes (~20 species) across large evolutionary distances will have a significant impact. Furthermore, the genome sequencing and functional genomic data (regarding chromatin accessibility, 3D genomic organization, epigenetic modifications, and non-coding RNAs) produced by this project will be a useful resource to the genomics community. In particular, this project will apply novel single-molecule long-read techniques to probe the mechanisms of gene regulation.

[MORE] PRELIMINARY DATA AND INTERESTING RESULTS

1.2. SYMBIODINIUM HI-C

SYMBIODINIUM GENOME ASSEMBLIES

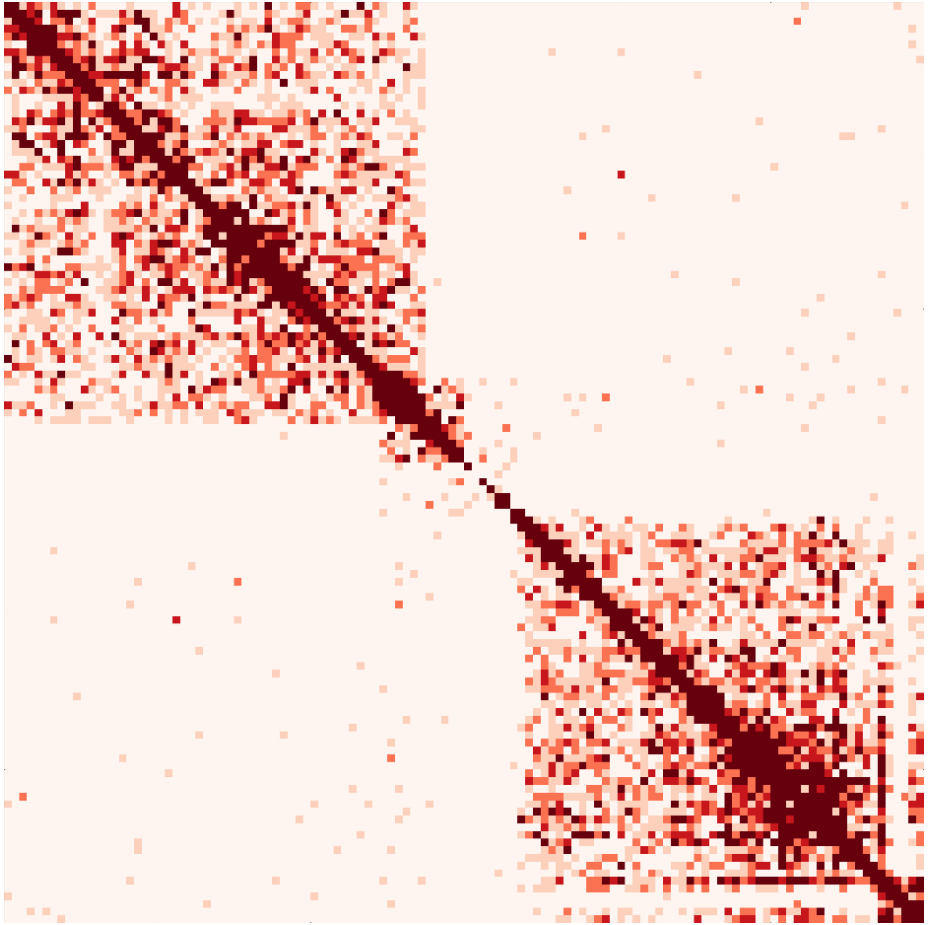
#Contig length	S. kawagutii	S. microadriaticum	S. minutum_Clade_B1
0	23412	136	12051
1,000	2611	6868	2305
10,000	1434	1039	5587
100,000	2534	1516	1956
1,000,000	49	136	0
10,000,000	0	0	0
100,000,000	0	0	0

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

SYMBIODINIUM HI-C

5kb resolution,
uncorrected homemade maps

(Juicer takes forever on so fragmented assemblies)

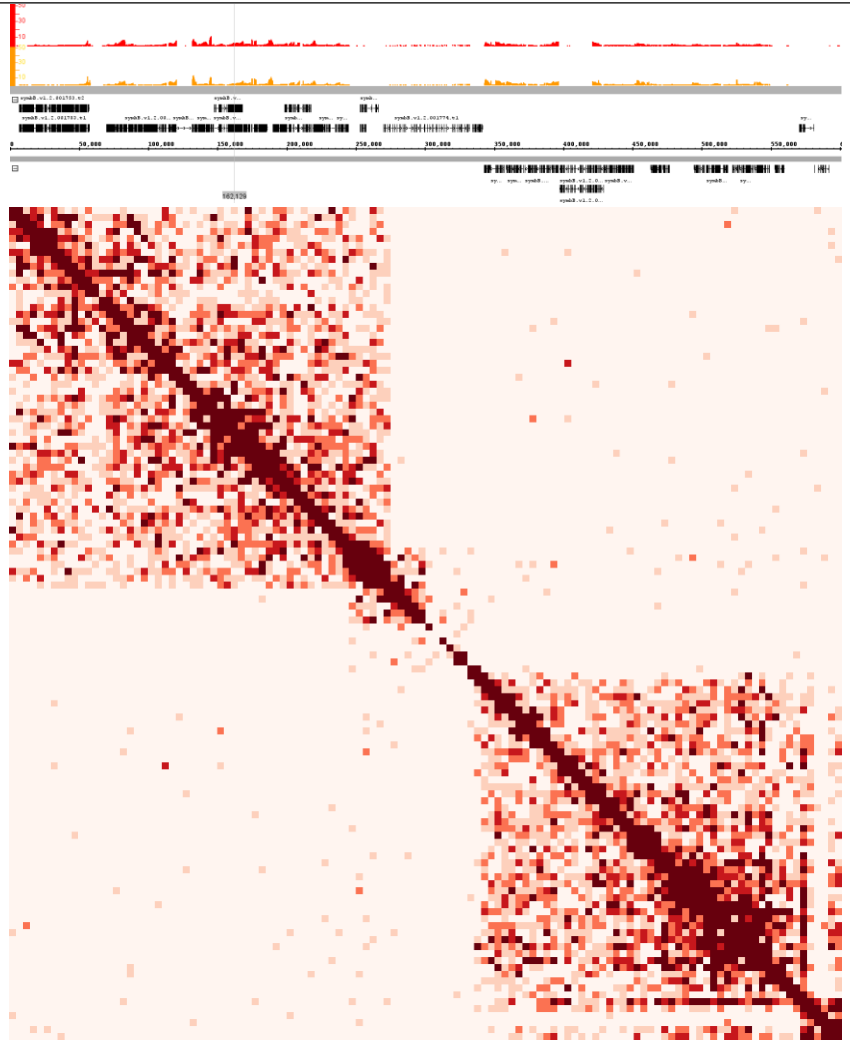


34C 8d RNA-seq

34C 0d RNA-seq

+ strand genes

- strand genes

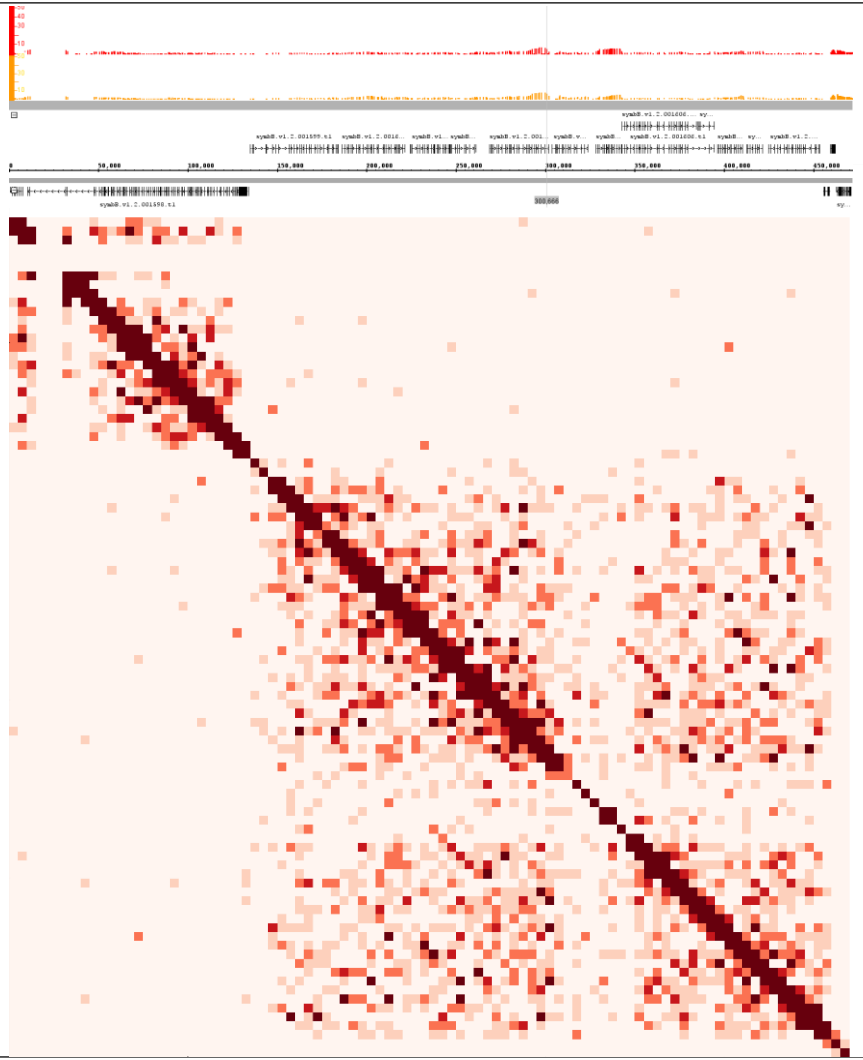


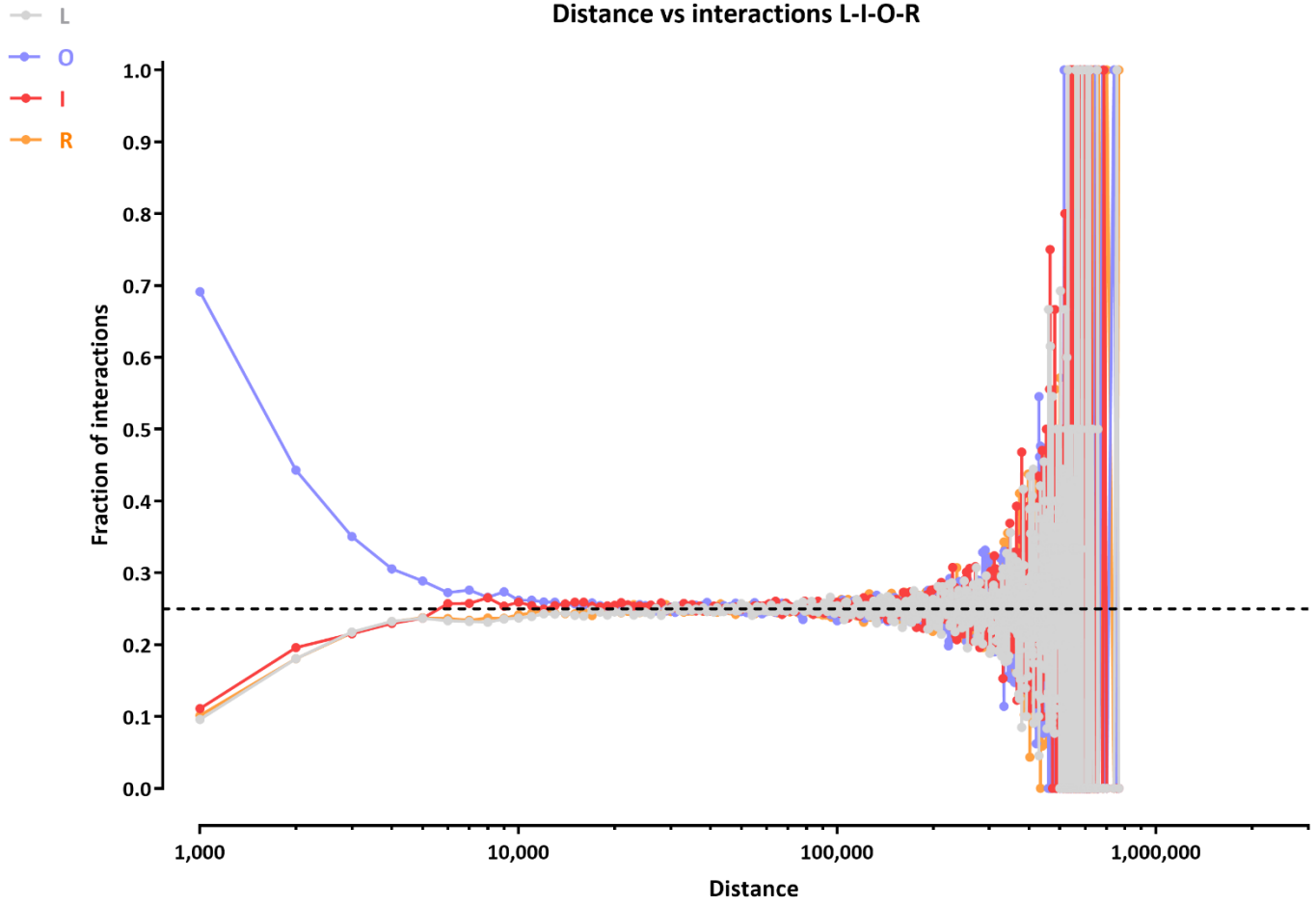
34C 8d RNA-seq

34C 0d RNA-seq

+ strand genes

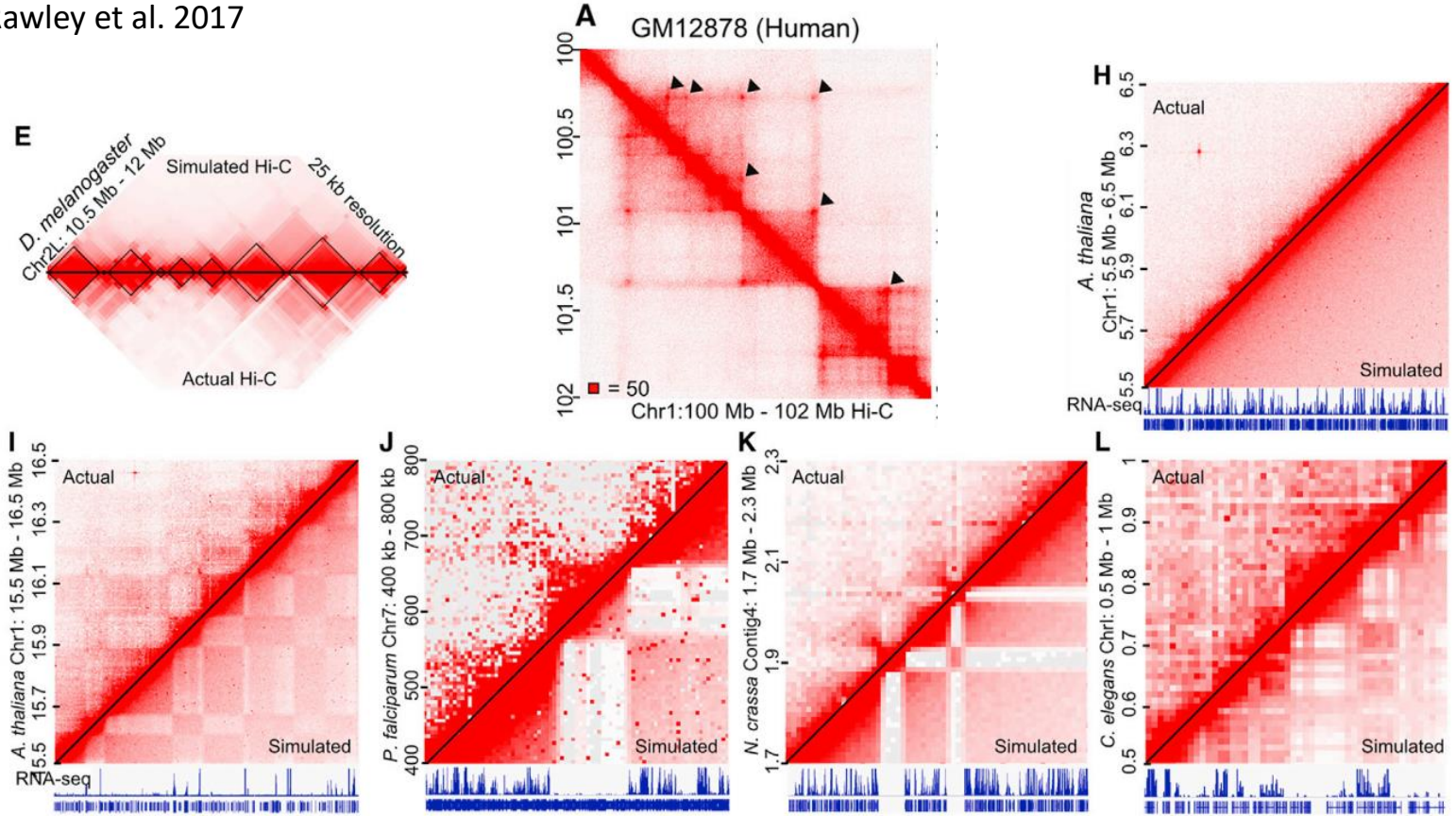
- strand genes



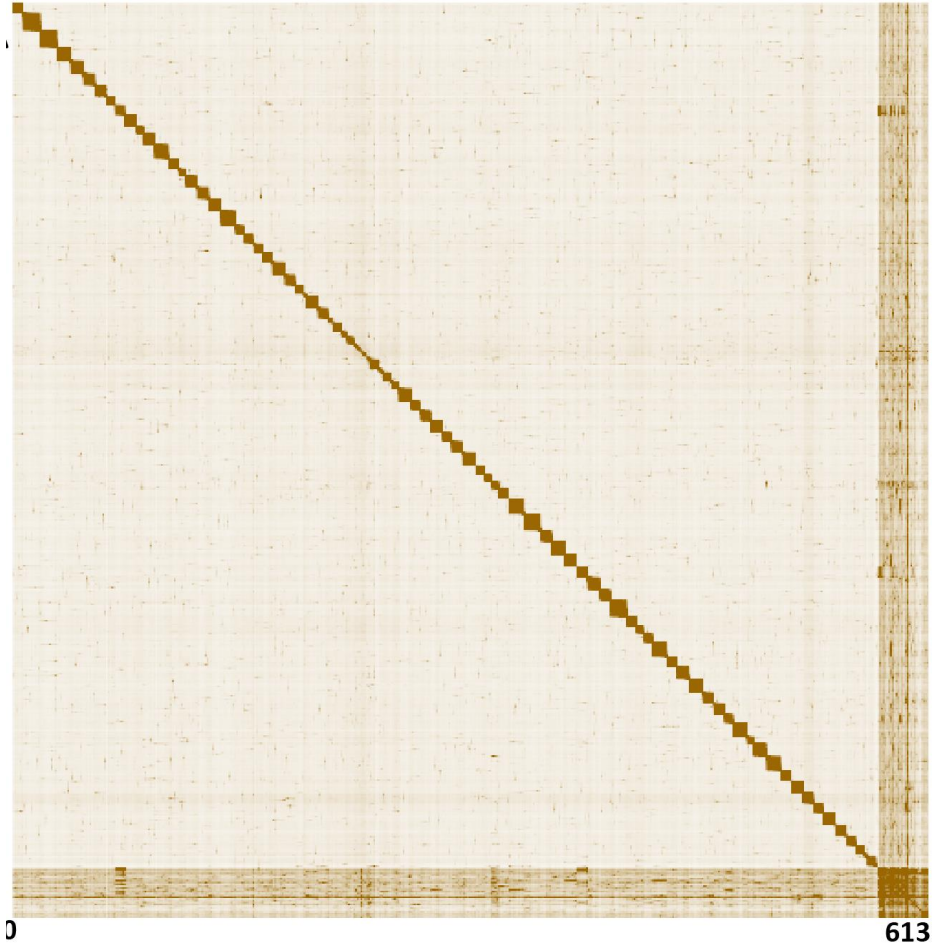

Distance vs interactions L-I-O-R
L
O
I
R
Fraction of interactions
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1,000
10,000
100,000
1,000,000
Distance

COMPARISON WITH OTHER EUKARYOTES:

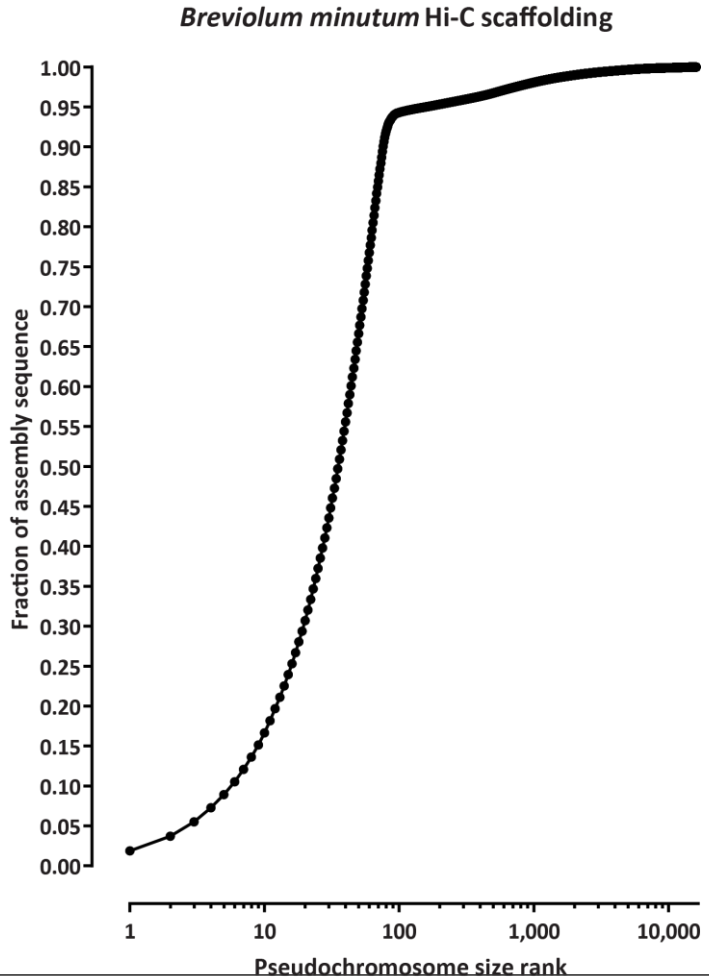
from Rawley et al. 2017



HI-C SCAFFOLDING



HI-C SCAFFOLDING



ONE PROBLEM:

Current Biology 23, 1399–1408, August 5, 2013 ©2013 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2013.05.062

Article

Draft Assembly of the *Symbiodinium minutum* Nuclear Genome

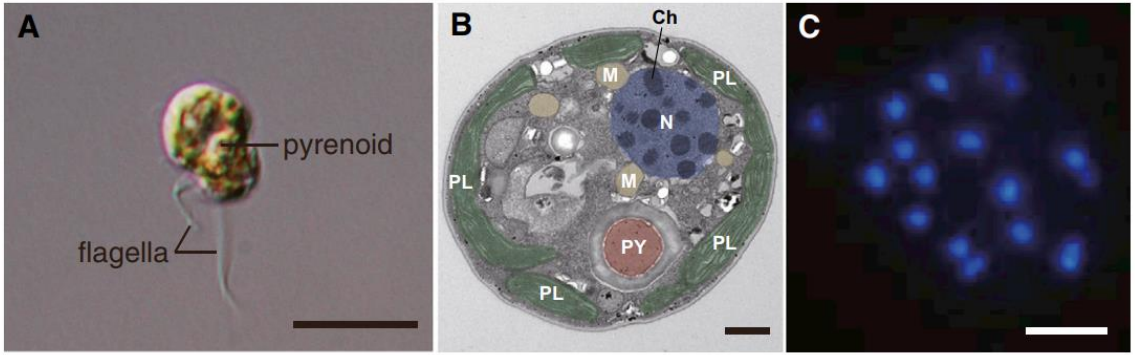


Figure 1. A Dinoflagellate, *Symbiodinium minutum*

(A) *S. minutum* zoospore. A short, transverse flagellum originating from the cingulum and a long, longitudinal flagellum originating from the sulcus are evident in the zoospore. A pyrenoid is also visible. The scale bar represents 10 μm.

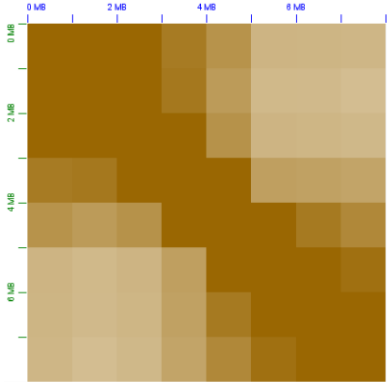
(B) Electron micrograph showing permanently condensed chromosomes (Ch) of *S. minutum*. The nucleus (N) is shown in purple, plastids (PL) in green, mitochondria (M) in orange, and pyrenoid (PY) in brown. The scale bar represents 1 μm.

(C) DAPI staining of the nucleus showing permanently condensed chromosomes of *S. minutum*. The scale bar represents 1 μm.

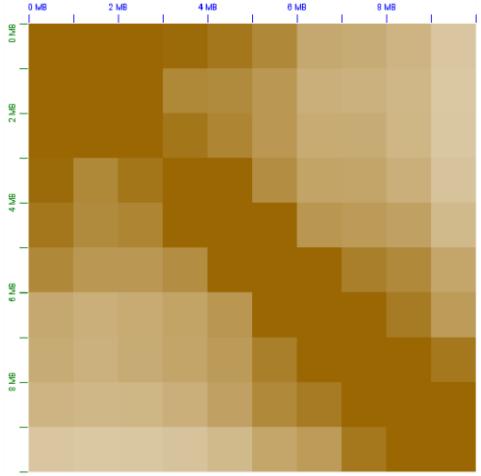
18 chromosomes, but we get ~90

BROAD STRUCTURE:

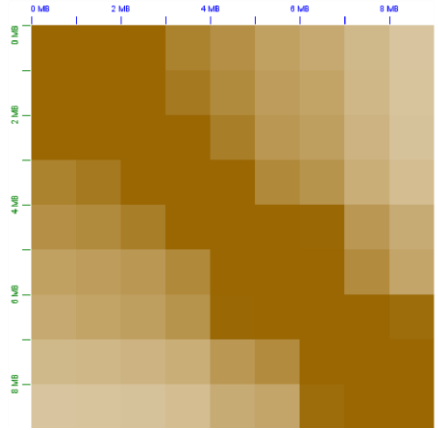
pseudochromosome 10



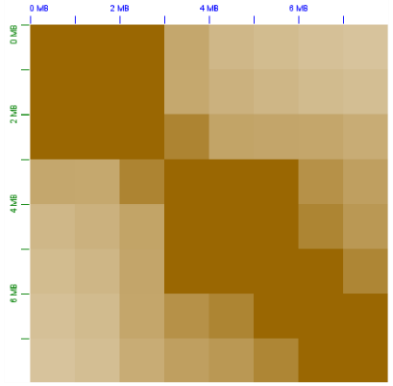
pseudochromosome 4



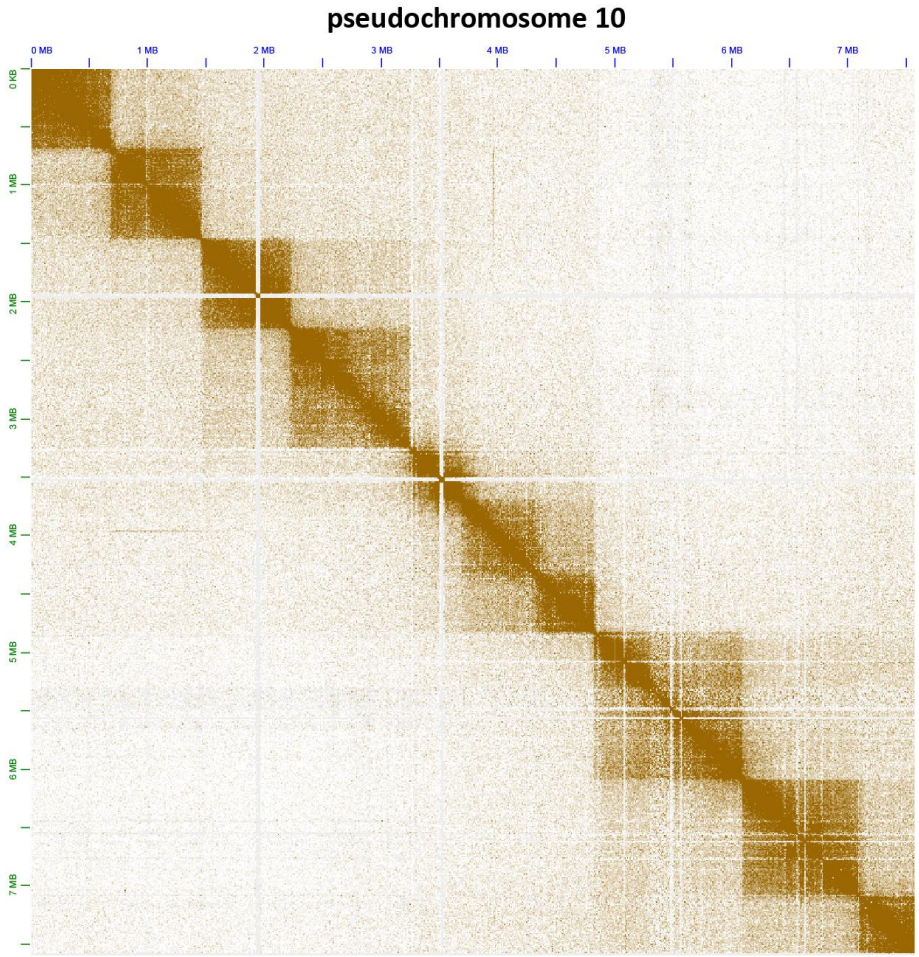
pseudochromosome 5

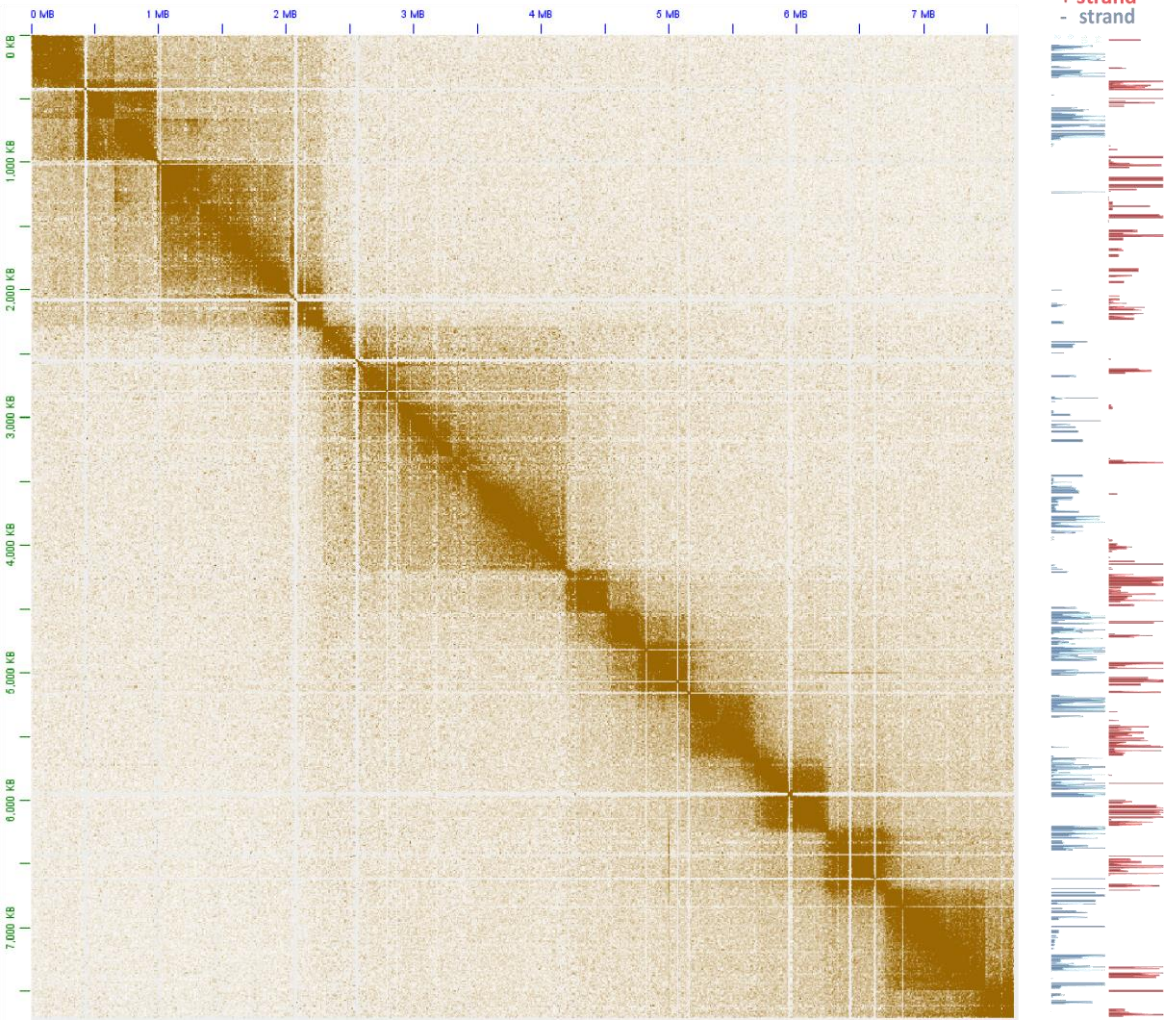


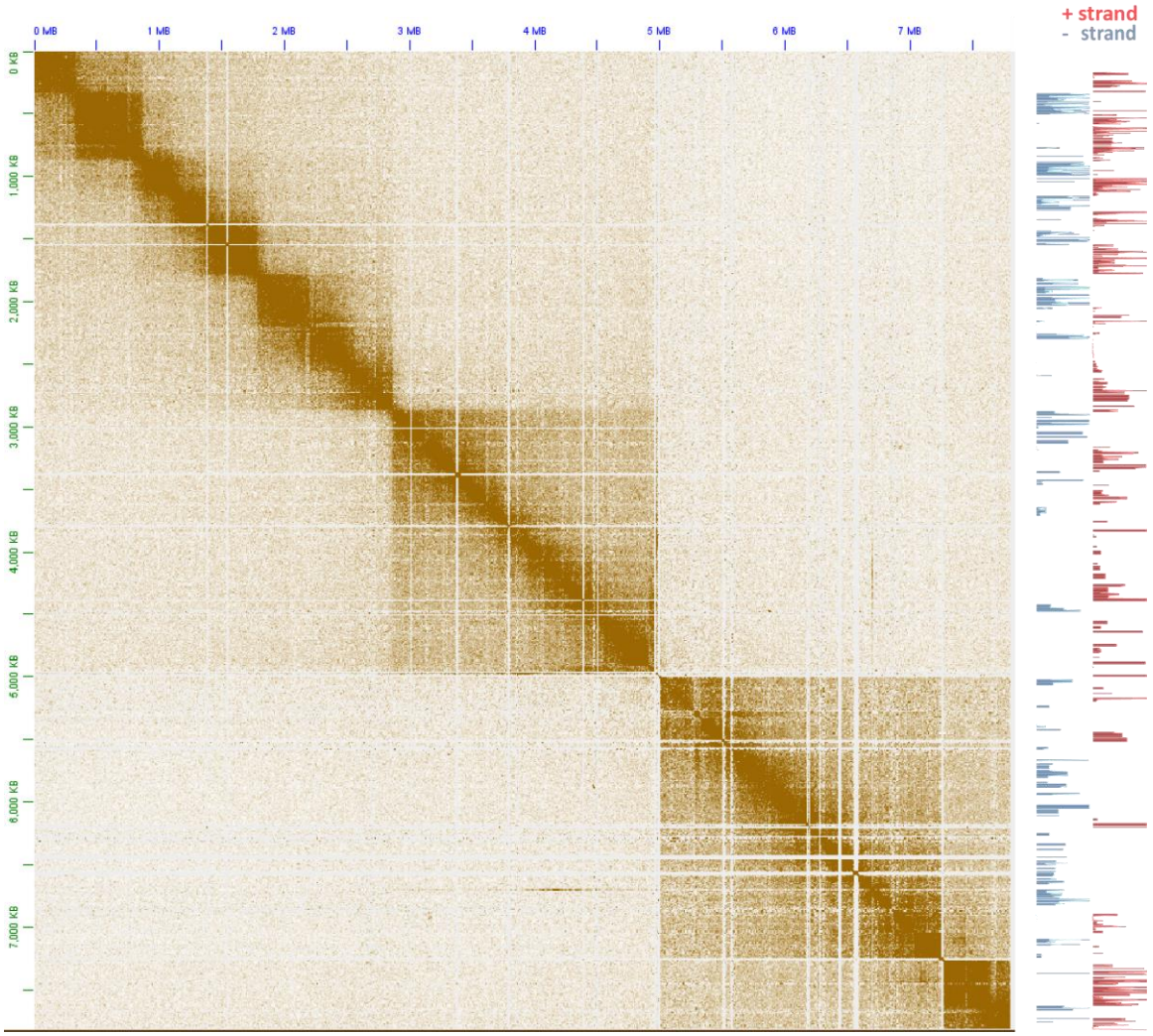
pseudochromosome 6

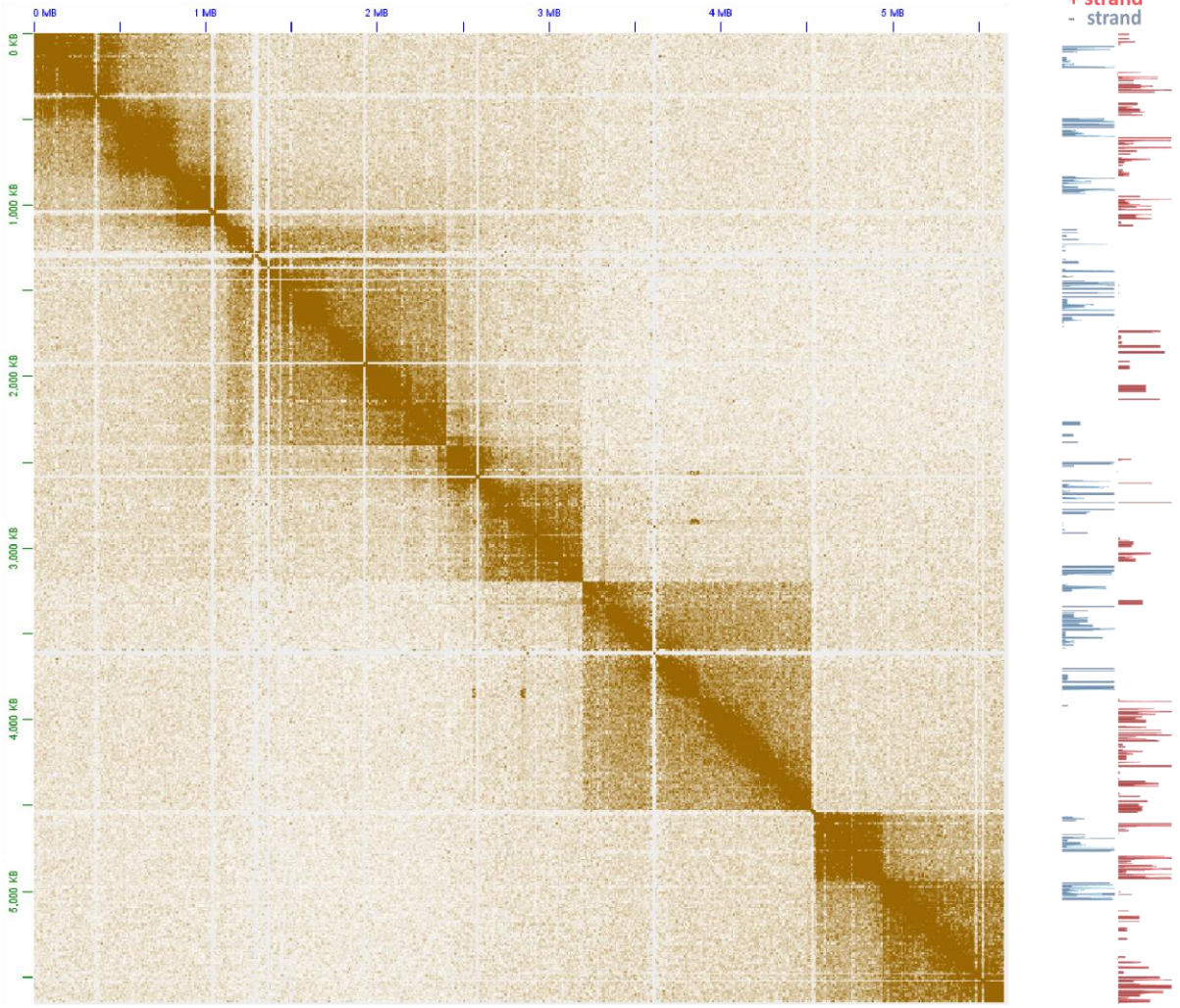


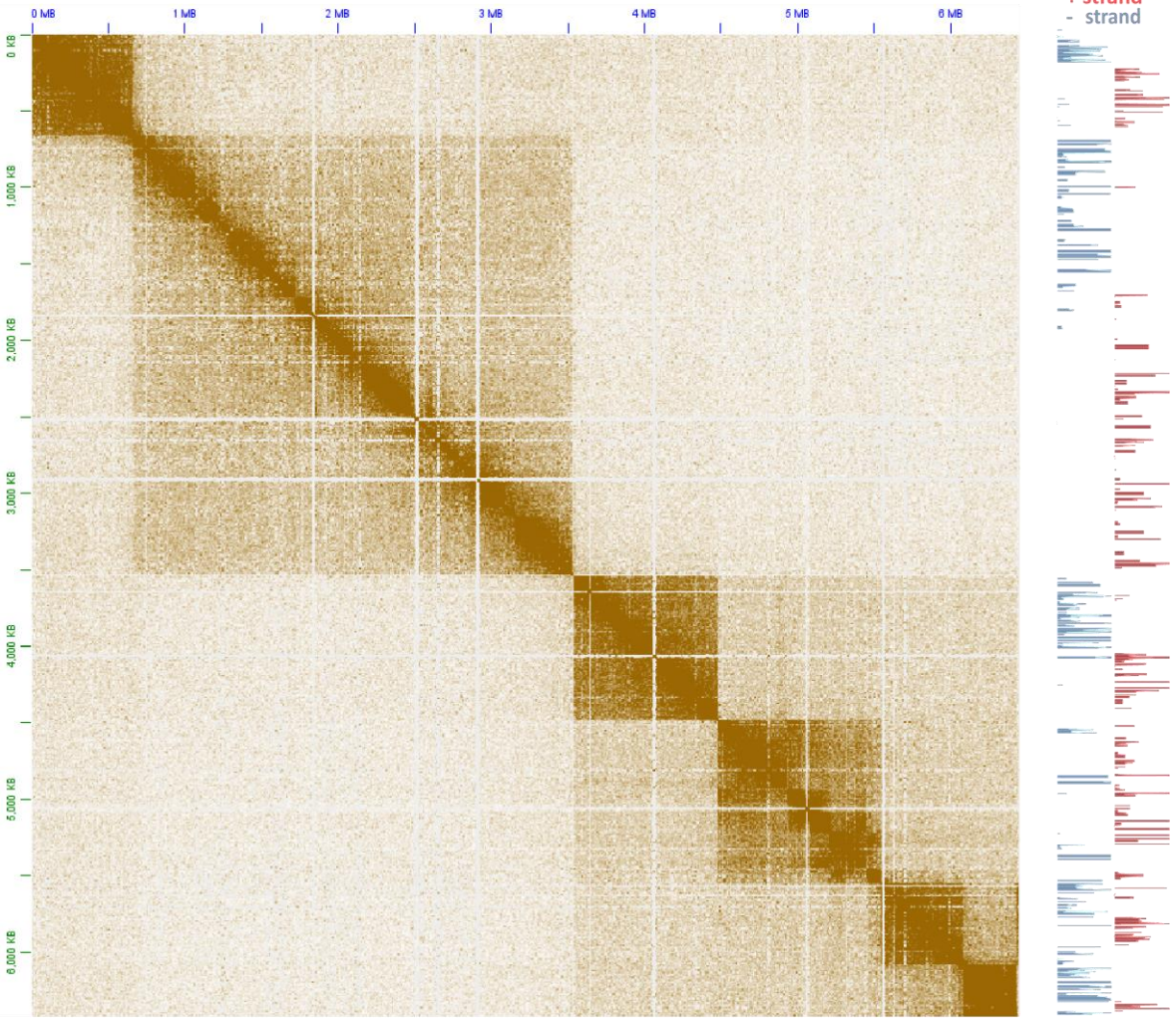
DOMAIN STRUCTURE:

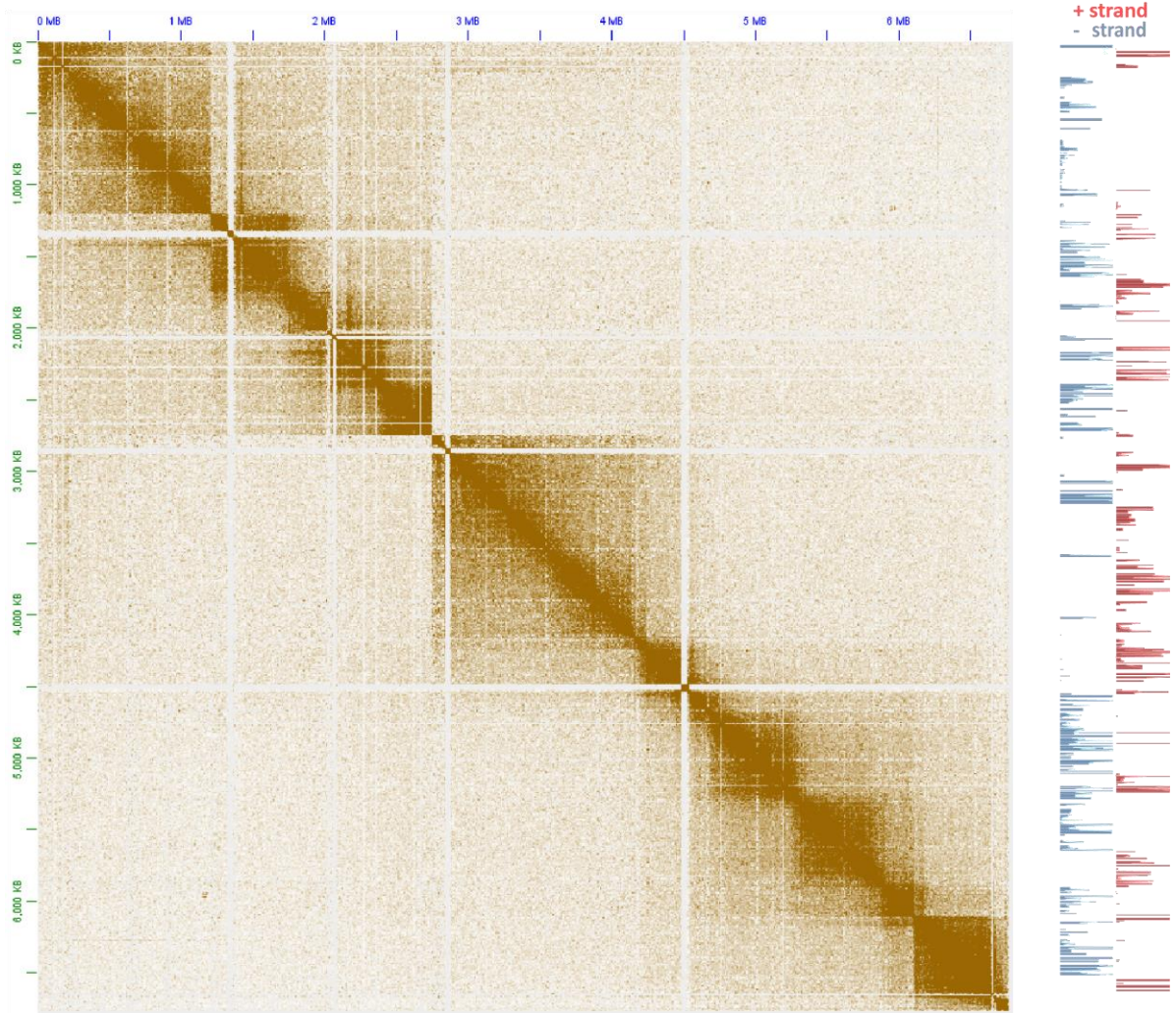


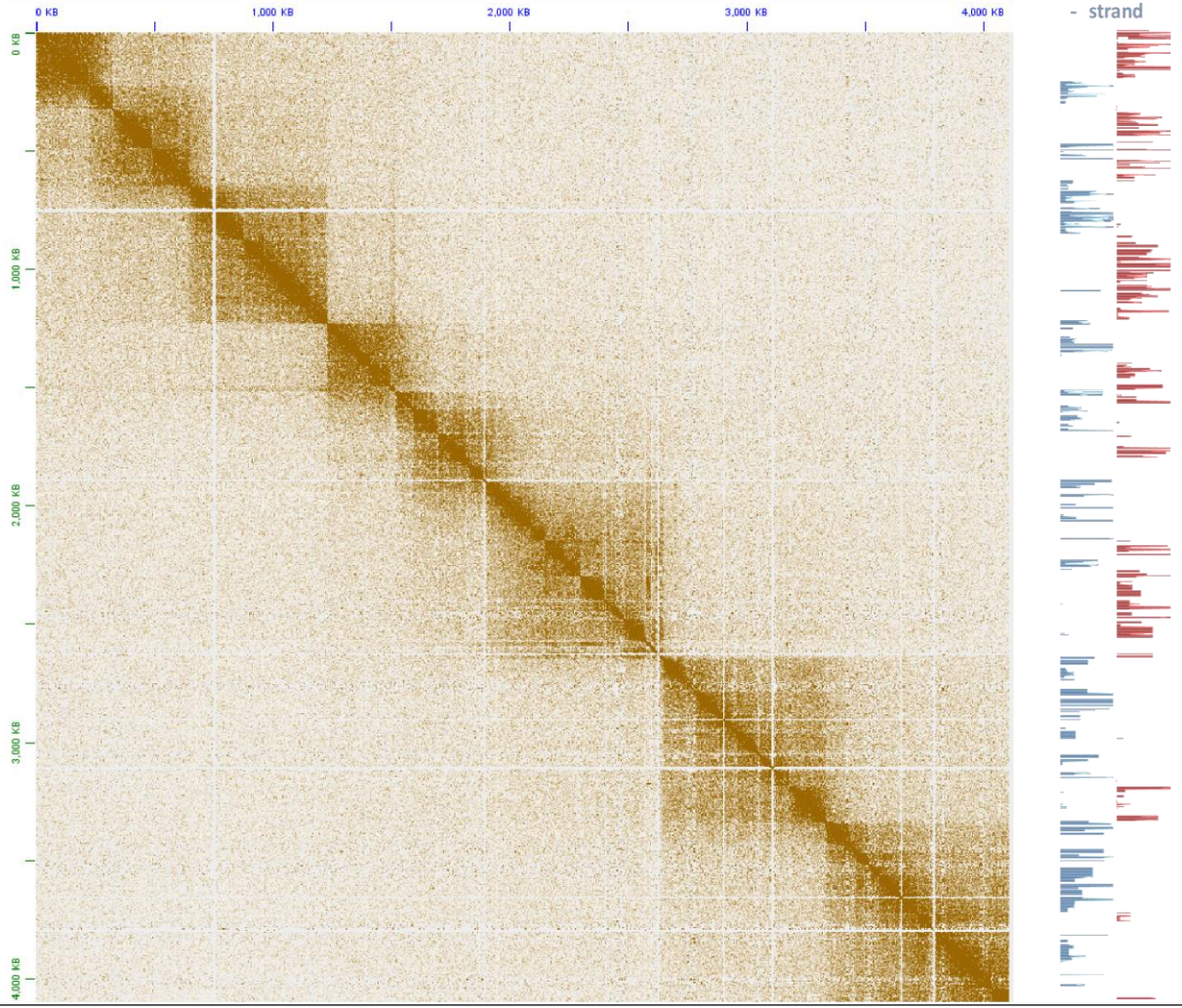

0 MB
1 MB
2 MB
3 MB
4 MB
5 MB
6 MB
7 MB
0 KB
1,000 KB
2,000 KB
3,000 KB
4,000 KB
5,000 KB
6,000 KB
7,000 KB
+ strand
- strand

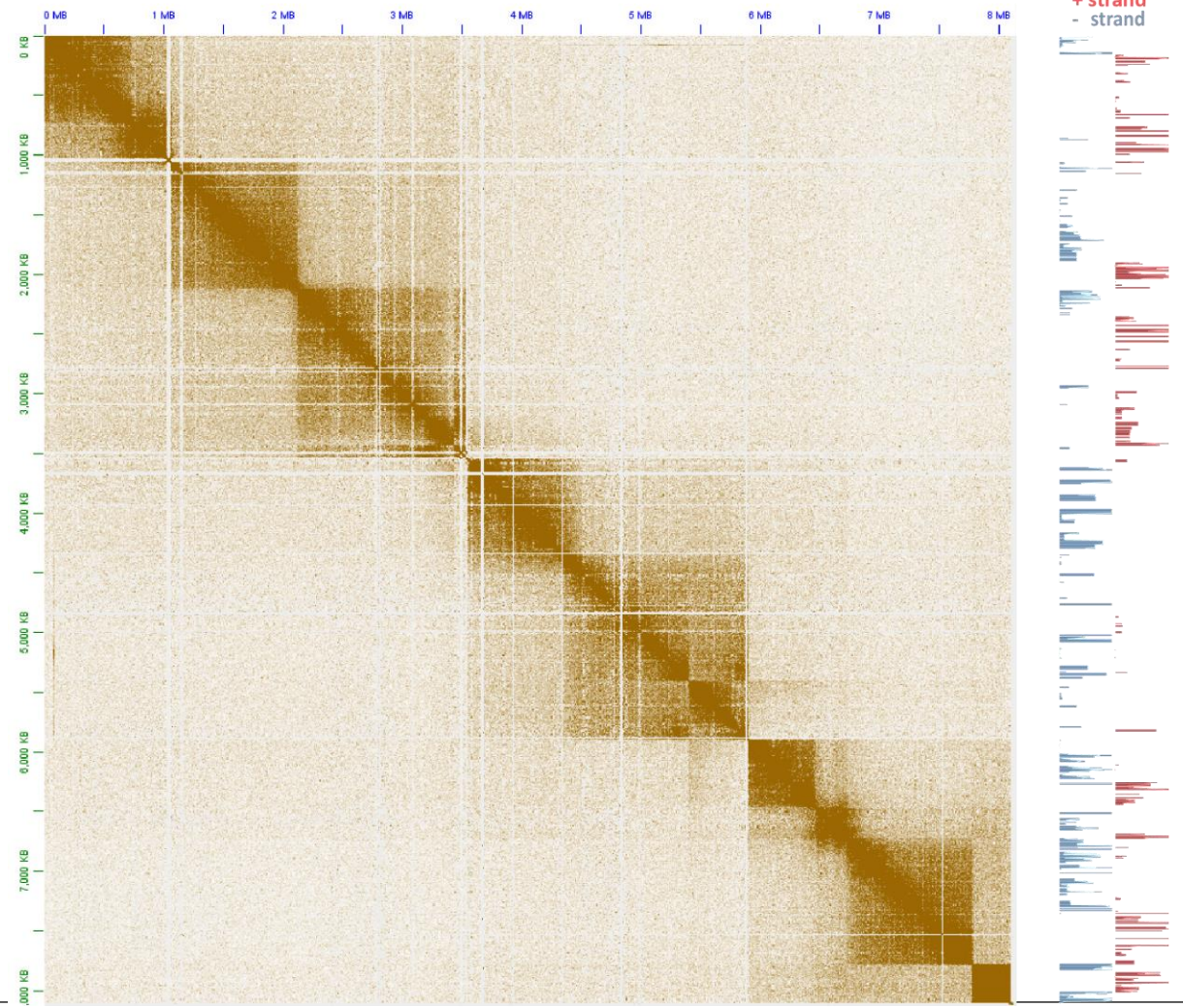

0 MB
1 MB
2 MB
3 MB
4 MB
5 MB
6 MB
7 MB
0 KB
1,000 KB
2,000 KB
3,000 KB
4,000 KB
5,000 KB
6,000 KB
7,000 KB
+ strand
- strand


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

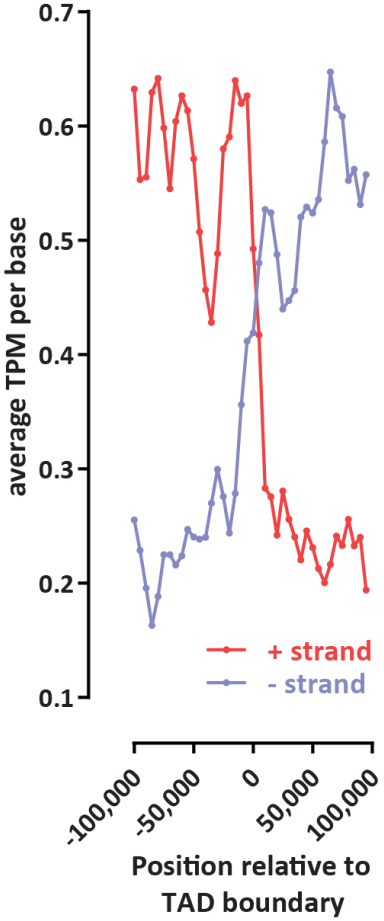

0 MB
1 MB
2 MB
3 MB
4 MB
5 MB
6 MB
0 KB
1,000 KB
2,000 KB
3,000 KB
4,000 KB
5,000 KB
6,000 KB
+ strand
- strand


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


0 KB
1,000 KB
2,000 KB
3,000 KB
4,000 KB
+ strand
- strand

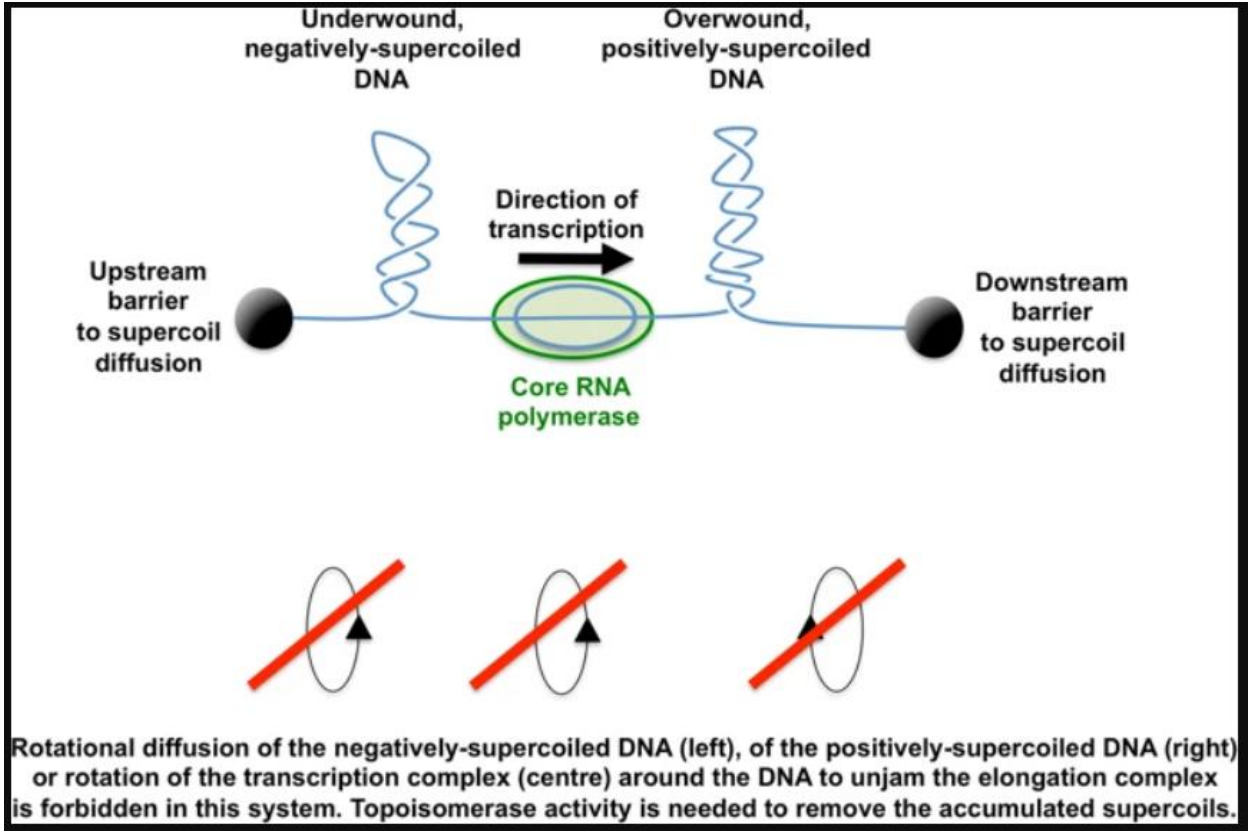

0 MB
1 MB
2 MB
3 MB
4 MB
5 MB
6 MB
7 MB
8 MB
0 KB
1,000 KB
2,000 KB
3,000 KB
4,000 KB
5,000 KB
6,000 KB
7,000 KB
+ strand
- strand

DOMAIN BOUNDARIES



WHAT IS THE MECHANISM?

- Transcription-induced supercoiling is one possibilitiy



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

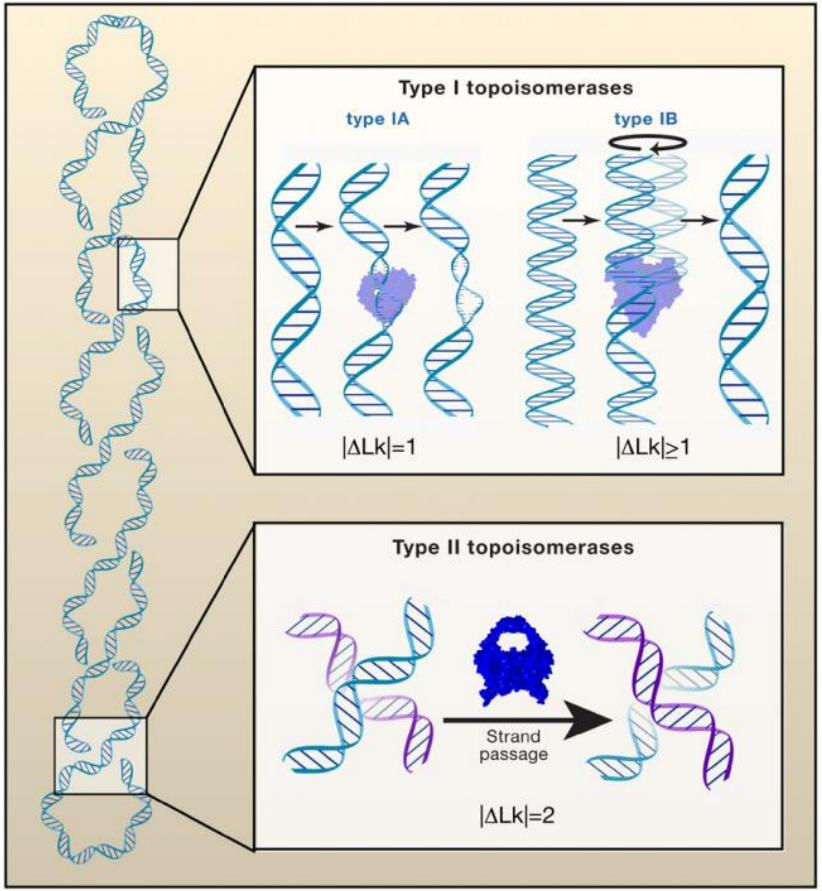


Figure 4. Types of Topoisomerases

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

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

EXPANSION OF THE TOPISOMERASE GENE REPERTOIRE IN DINOFLAGELLATES

	PCNA	MCM	RPA1	RPA2	RPA3	RFC1	DNA Topoisomerase I	DNA Topoisomerase II	DNA Topoisomerase III	DNA primase	DNA ligase	DNA Polymerase A/D	DNA Polymerase A/E
Alexandrium monilatum	8	19	9	0	1	4	1	11	9	1	3	2	3
Alexandrium tamarense	18	45	10	3	4	2	3	18	12	6	2	5	7
Alexandrium-fundyense	18	0	0	0	0	0	0	0	0	1	0	0	0
Gambierdiscus australes	14	9	6	0	0	2	1	5	0	1	3	2	3
Pyrodinium bahamense	19	29	11	0	0	3	1	21	8	1	4	2	4
Ceratium fusus	12	18	9	1	1	3	1	15	10	1	1	4	7
Lingulodinium polyedra	11	19	11	1	0	3	1	17	8	1	2	5	5
Protoceratium reticulatum	11	18	10	0	0	2	2	20	9	2	1	3	5
Gonyaulax spinifera	10	10	8	1	1	1	1	2	0	2	1	1	1
Dinophysis acuminata	13	29	8	0	0	2	4	15	9	1	2	5	5
Pelagodinium beii	11	12	4	0	0	4	1	8	2	1	1	3	4
Polarella glacialis	5	23	5	0	0	8	1	28	5	3	2	10	7
Symbiodinium sp C1	6	9	4	0	0	3	1	9	4	1	3	2	3
Symbiodinium sp C15	3	12	4	1	0	3	1	7	2	1	2	3	3
Symbiodinium sp CCMP2430	7	10	4	0	0	3	1	7	2	1	1	3	3
Symbiodinium sp Mp	6	13	3	0	0	3	1	7	3	1	2	3	3
Brandtodinium nutriculum	21	30	4	0	0	3	1	13	9	1	2	3	4
Crypthecodinium cohnii	2	15	4	0	0	3	1	6	5	1	4	4	5
Azadinium spinosum	11	35	6	0	0	3	1	12	13	1	2	2	5
Scrippsiella hangoei	3	14	5	0	0	2	1	8	11	1	5	5	5
Scrippsiella hangoei -like	6	22	16	0	2	2	1	8	6	1	3	5	5
Scrippsiella trochoidea	12	38	8	1	1	3	1	27	10	1	3	5	5
Durinskia baltica	9	18	8	0	0	4	2	12	9	3	7	12	6
Glenodinium foliaceum	7	23	6	0	1	4	2	9	3	2	6	1	4
Kryptoperidinium foliaceum	16	64	10	1	0	7	4	14	11	3	5	9	15
Peridinium aciculiferum	6	11	5	0	0	3	1	7	5	1	2	2	5
Heterocapsa triquestra	5	13	4	0	0	3	1	8	5	1	2	3	2
Heterocapsa rotundata	6	12	4	0	0	6	2	19	4	1	1	6	3
Prorocentrum minimum CCMP1329	13	29	6	0	0	3	1	15	6	1	4	3	4
Prorocentrum minimum CCMP2233	12	29	5	0	0	3	1	14	4	1	4	3	6
Togula jolla	3	21	6	0	0	4	1	17	3	1	1	5	4
Karlodinium micrum	13	46	31	2	0	5	2	9	7	6	2	6	3
Karenia brevis CCMP2229	8	10	7	0	1	4	1	14	8	1	2	5	4
Karenia brevis SP1	6	16	8	0	1	4	1	14	13	1	3	5	4
Karenia brevis SP3	8	13	10	0	1	4	1	12	9	1	2	4	4
Karenia brevis Wilson	9	14	8	0	2	5	1	14	7	1	3	5	4
Amphidinium carterae	2	8	4	0	0	3	2	5	5	1	2	5	3
Amoebophrya sp.	0	13	1	0	0	0	2	8	1	1	0	3	2
Noctiluca scintillans	1	9	6	0	1	1	1	7	3	1	1	5	5
Oxyrrhis marina LB1974	4	10	2	0	0	2	1	2	4	1	1	4	3
Oxyrrhis marina	7	9	3	0	1	2	1	2	5	1	2	3	3
Perkinsus marinus ATCC50439	2	1	0	0	0	0	1	0	0	0	0	0	0
Perkinsus chesapeaki ATCC_PRA_65	0	0	0	0	0	0	2	0	0	0	0	0	0
Vitrella brassicaformis CCMP3346	2	9	1	0	0	1	1	1	2	1	3	2	2
Pseudonitzschia australis 10249_10_AB	1	8	1	1	0	1	1	3	4	1	2	2	2
Pseudonitzschia fradulenta WWA7	4	24	5	0	0	3	2	11	6	2	2	2	6
Thalassiosira antarctica CCMP982	1	12	1	3	1	1	1	4	2	1	1	3	3
Thalassiosira gravida GMp14c1	1	13	1	2	1	1	1	1	3	0	1	0	2
Thalassiosira miniscula CCMP1093	1	10	1	2	1	1	1	13	6	1	2	3	3
Thalassiosira oceanica CCMP1005	1	10	1	0	0	1	1	10	1	0	1	3	3
Mesodinium pulex SPMC105	9	16	4	0	0	6	2	13	2	1	1	3	2
Micromonas sp RCC472	1	7	2	1	0	1	1	2	2	1	2	3	3
Hemiselmis andersenii CCMP644	1	12	2	0	1	1	1	2	5	1	3	5	3

CONCLUSIONS

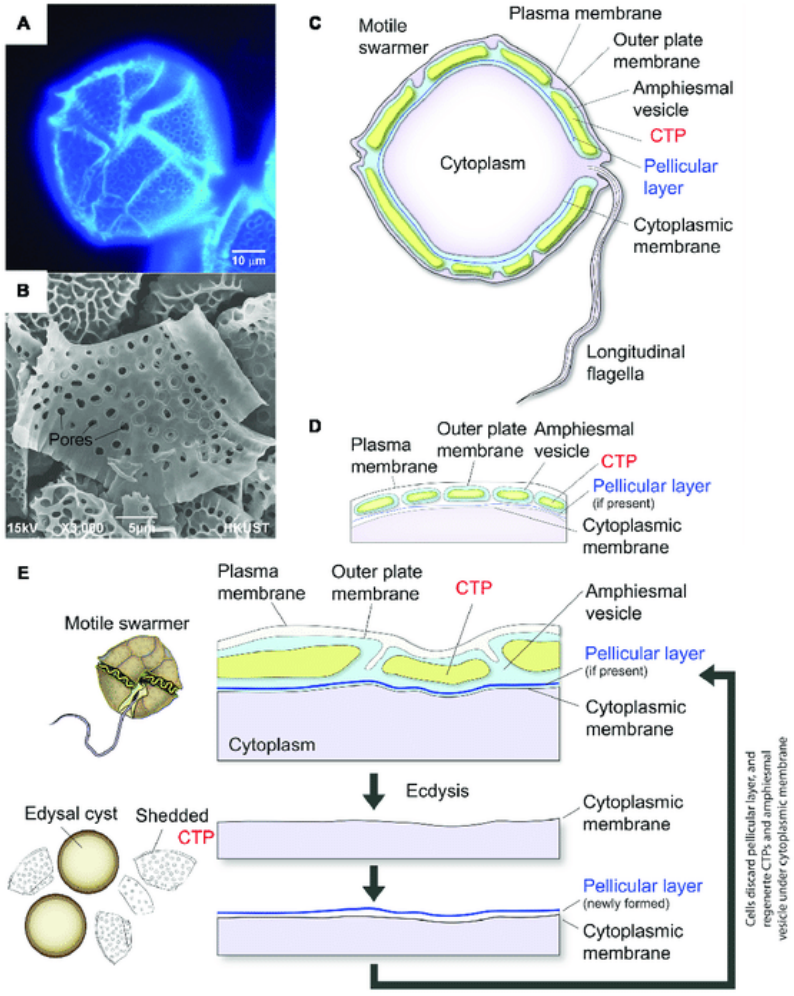
- Strong insulation domains are observed in the *Symbiodinium* genome
- These appear to correspond to individual multicistronic gene arrays (with some exceptions)
- Possibly related to transcription-induced supercoiling
- Clear loop contacts are not immediately obvious

IN PROGRESS (WITH ALEX)

- Inhibition of transcription (DRB, amanitin)
  - Prediction: domains should disappear as topoisomerases resolve topological stress

- Inhibition of topoisomerases
  - Prediction: domains should become stronger

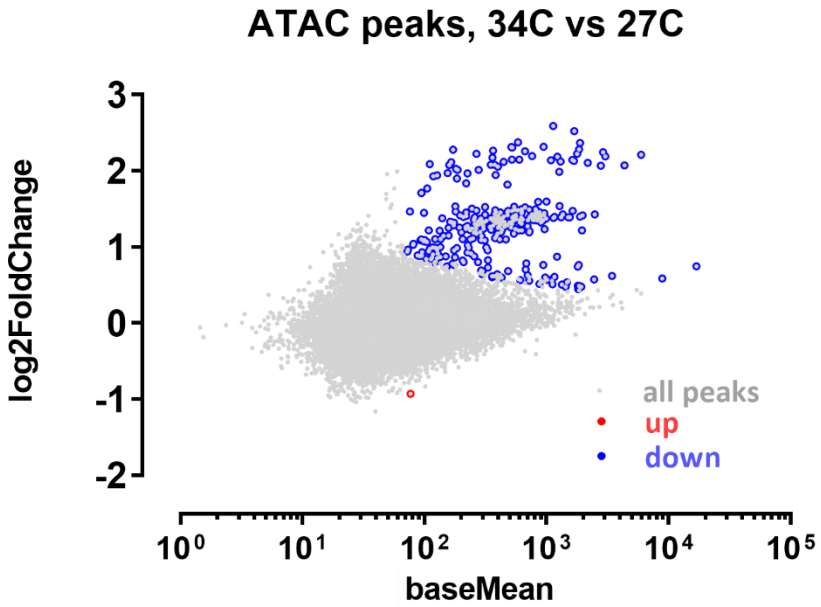
- Problem: nobody has ever done such experiments on *Symbiodinium* and we have no idea:
  - Whether the inhibitors will actually get inside the nucleus
  - How quickly cells will die as a result of the treatment
  - There isn't even a way to measure cell death


A
10 µm
B
Pores
15kV X3,000 5µm HKUST
C
Motile swarmer
Plasma membrane
Outer plate membrane
Amphiesmal vesicle
CTP
Pellicular layer
Cytoplasm
Cytoplasmic membrane
Longitudinal flagella
D
Plasma membrane
Outer plate membrane
Amphiesmal vesicle
CTP
Pellicular layer
(if present)
Cytoplasmic membrane
E
Motile swarmer
Plasma membrane
Outer plate membrane
CTP
Amphiesmal vesicle
Pellicular layer
(if present)
Cytoplasmic membrane
Cytoplasm
Ecdysis
Edysal cyst
Shedded CTP
Cytoplasmic membrane
Pellicular layer
(newly formed)
Cytoplasmic membrane
Cells discard pellicular layer, and regenerate CTPs and amphiesmal vesicle under cytoplasmic membrane

1.3. DVNPs AND ACCESSIBILITY

ATAC-SEQ DATA

A	B	p-val		p-adj	
		up in B vs A	down in B vs A	up in B vs A	down in B vs A
27C	34C	372	1,395	2	317



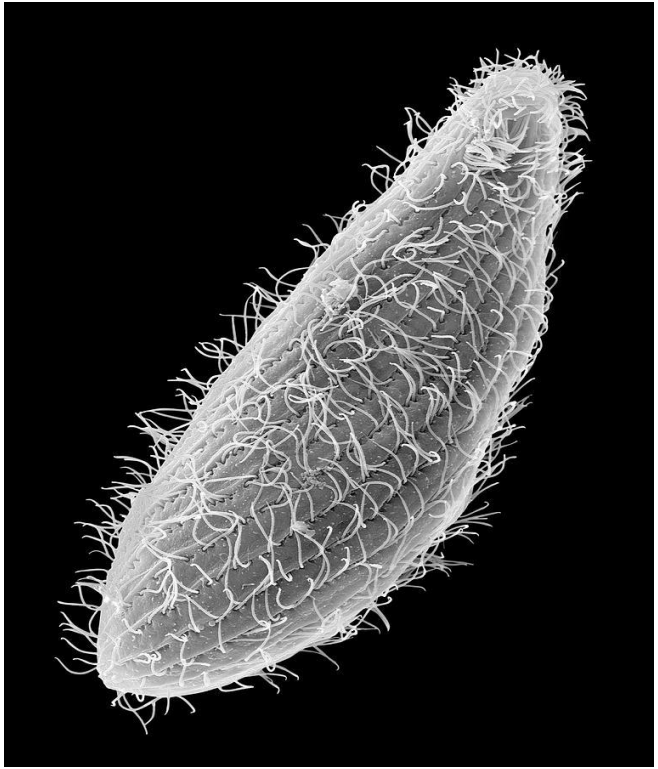
Obtained ATAC-seq data but is inconclusive in terms of whether there are clear regions of enrichment

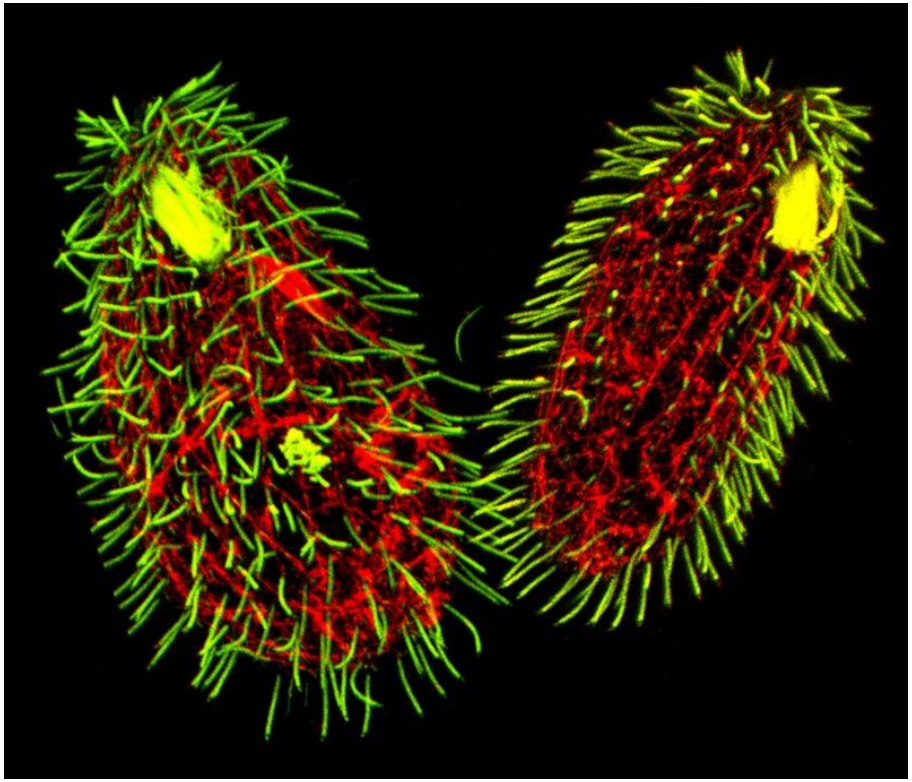
CURRENT EXPERIMENTS:

- ATAC-see
- Heterologous expression of DVNPs in yeast, then ChIP-seq, ATAC-seq and dSMF
- dSMF on Symbiodinium

1.4. TETRAHYMENA HI-C

TETRAHYMENA THERMOPHILA





NOTE ON EXISTING ASSEMBLY:

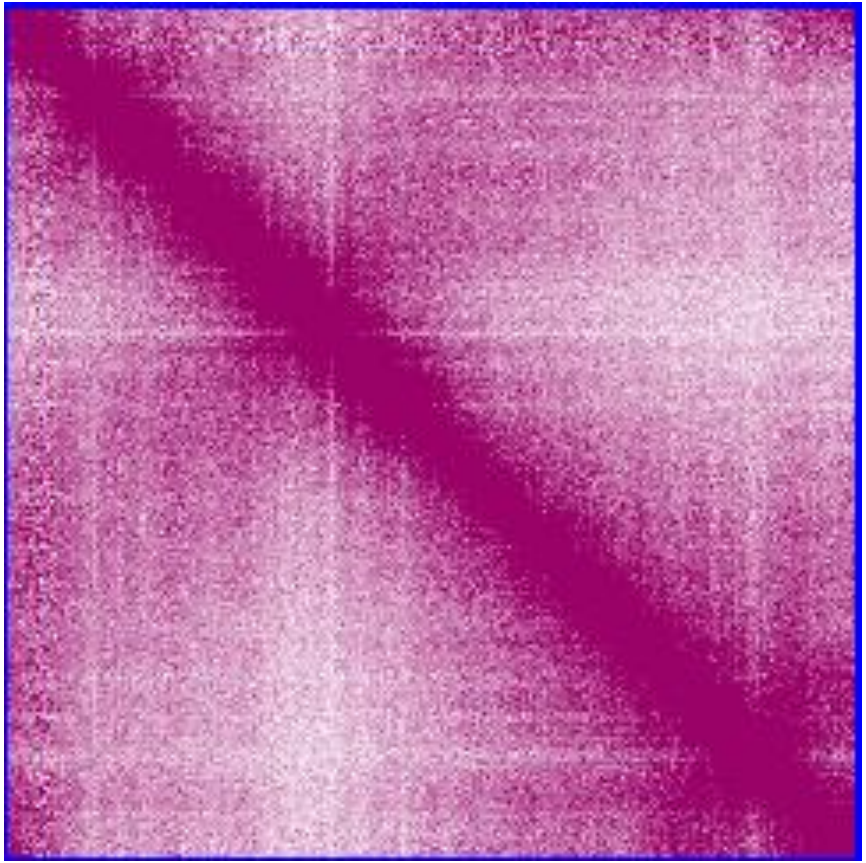
- The *Tetrahymena thermophila* genome was sequenced back in 2006 using Sanger sequencing and is of quite high quality (and I ordered exactly the same strain from the Cornell stock center)

- ~100 Mbp in total

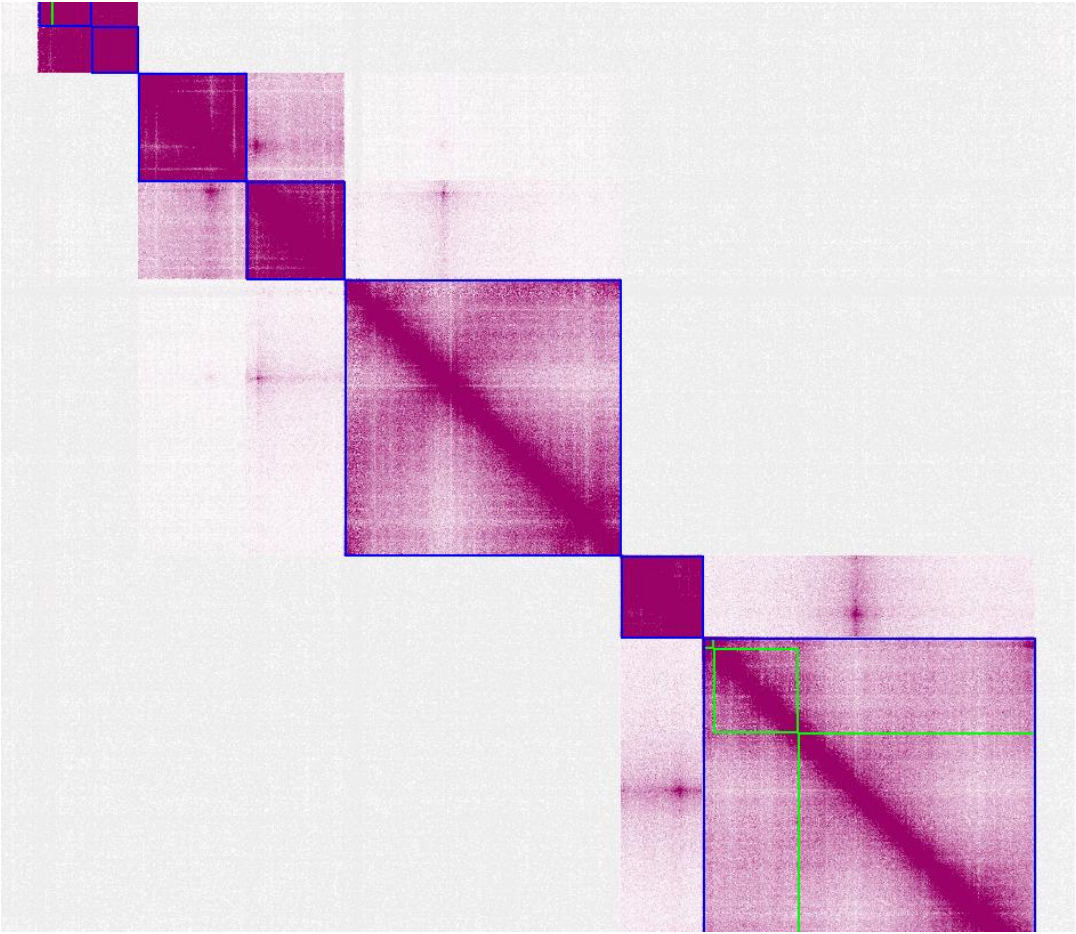
- However, it is not complete – it has some 400 scaffolds, of which only 185 have telomeres at their ends

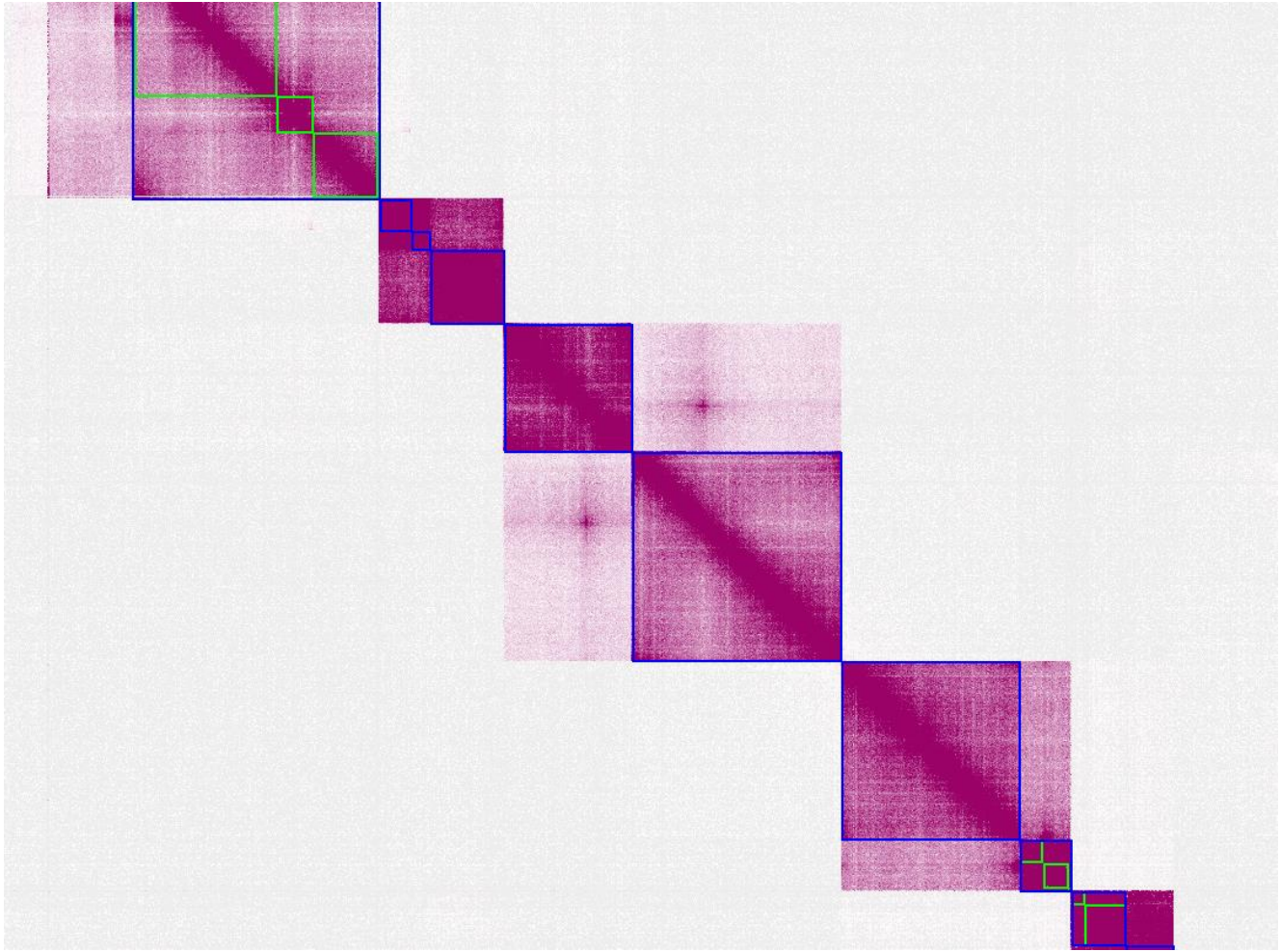
- Another problem: the genome is 75% AT-rich.
  - MboI, the restriction enzyme usually used for Hi-C recognizes GATC
  - Most experiments were done with MluCI, which recognizes AATT

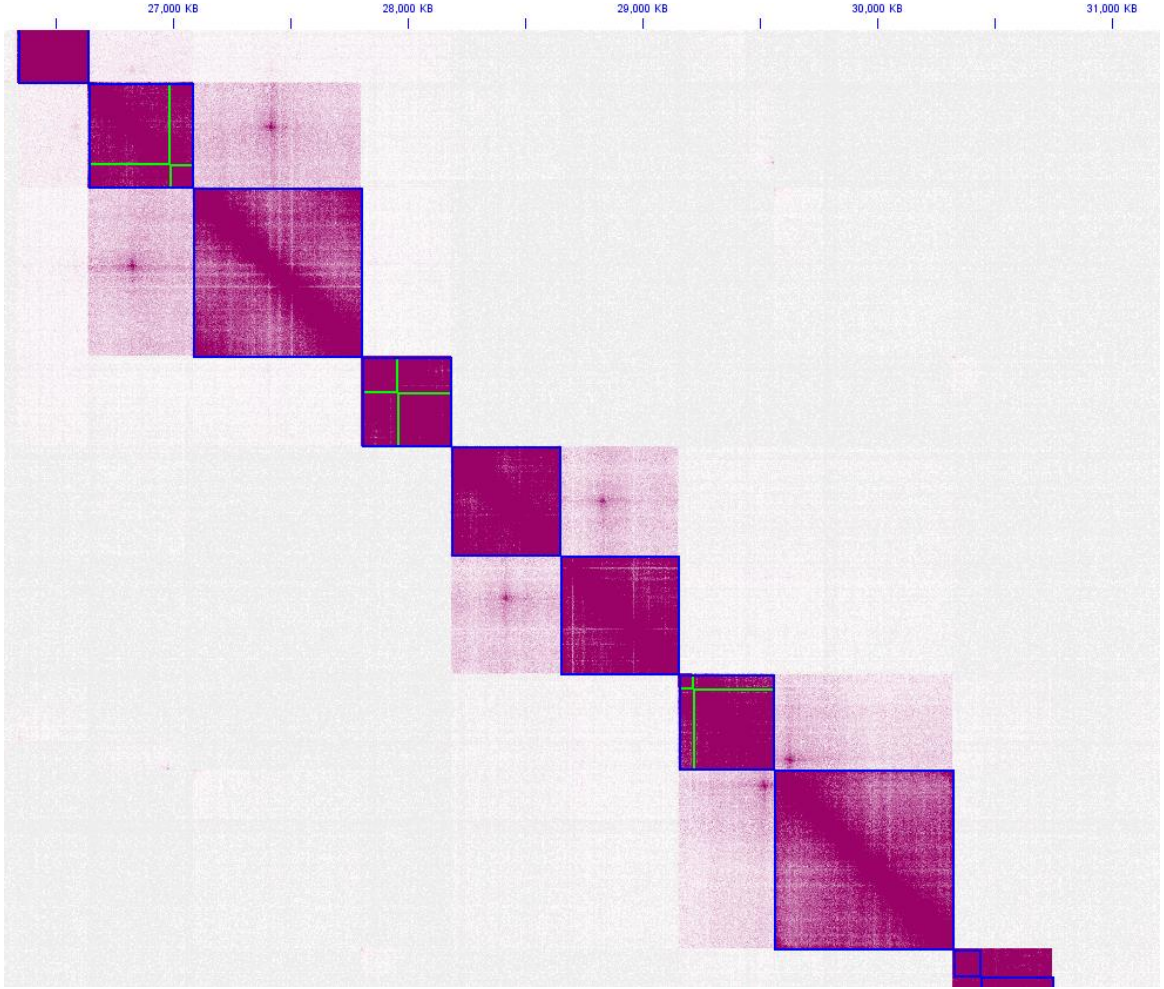
- Initially there was not a lot of structure when looking at individual complete chromosomes:

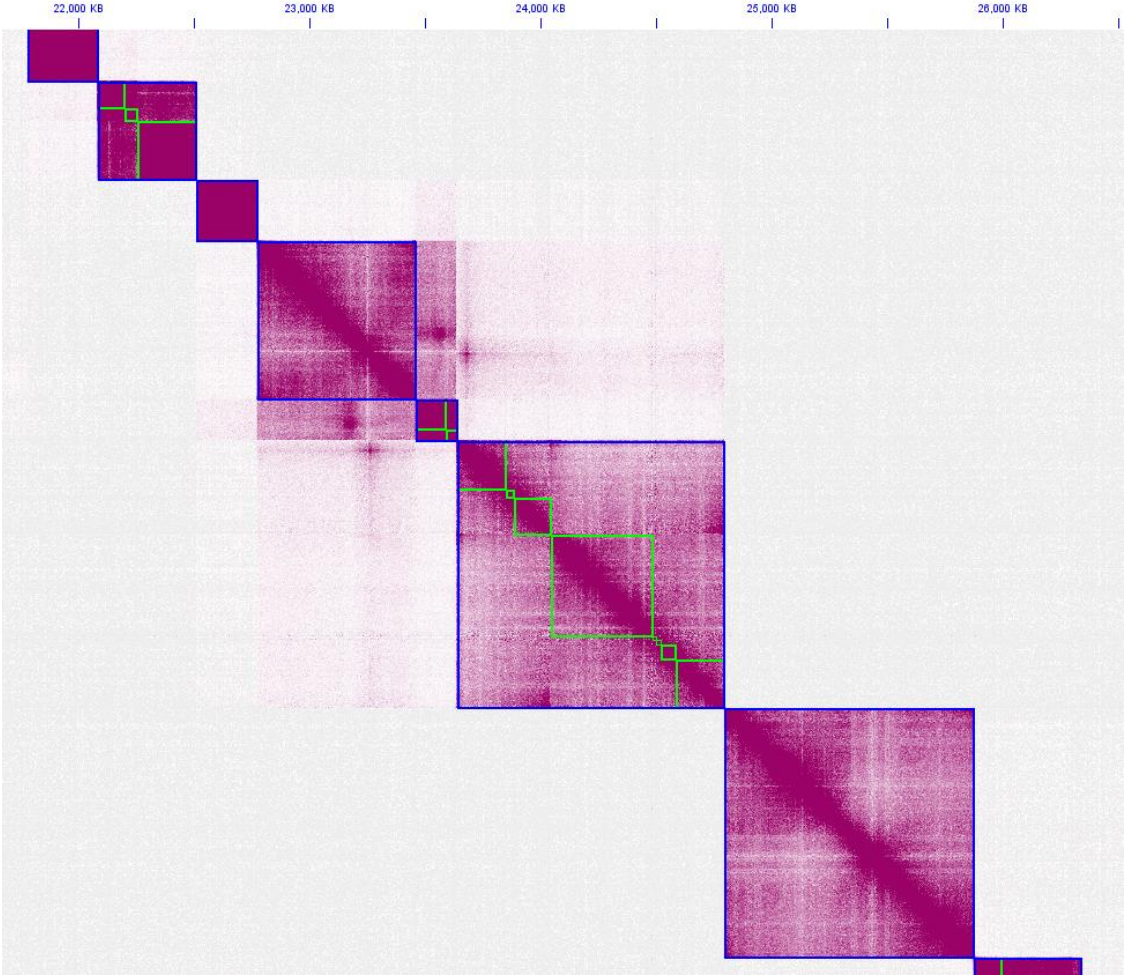


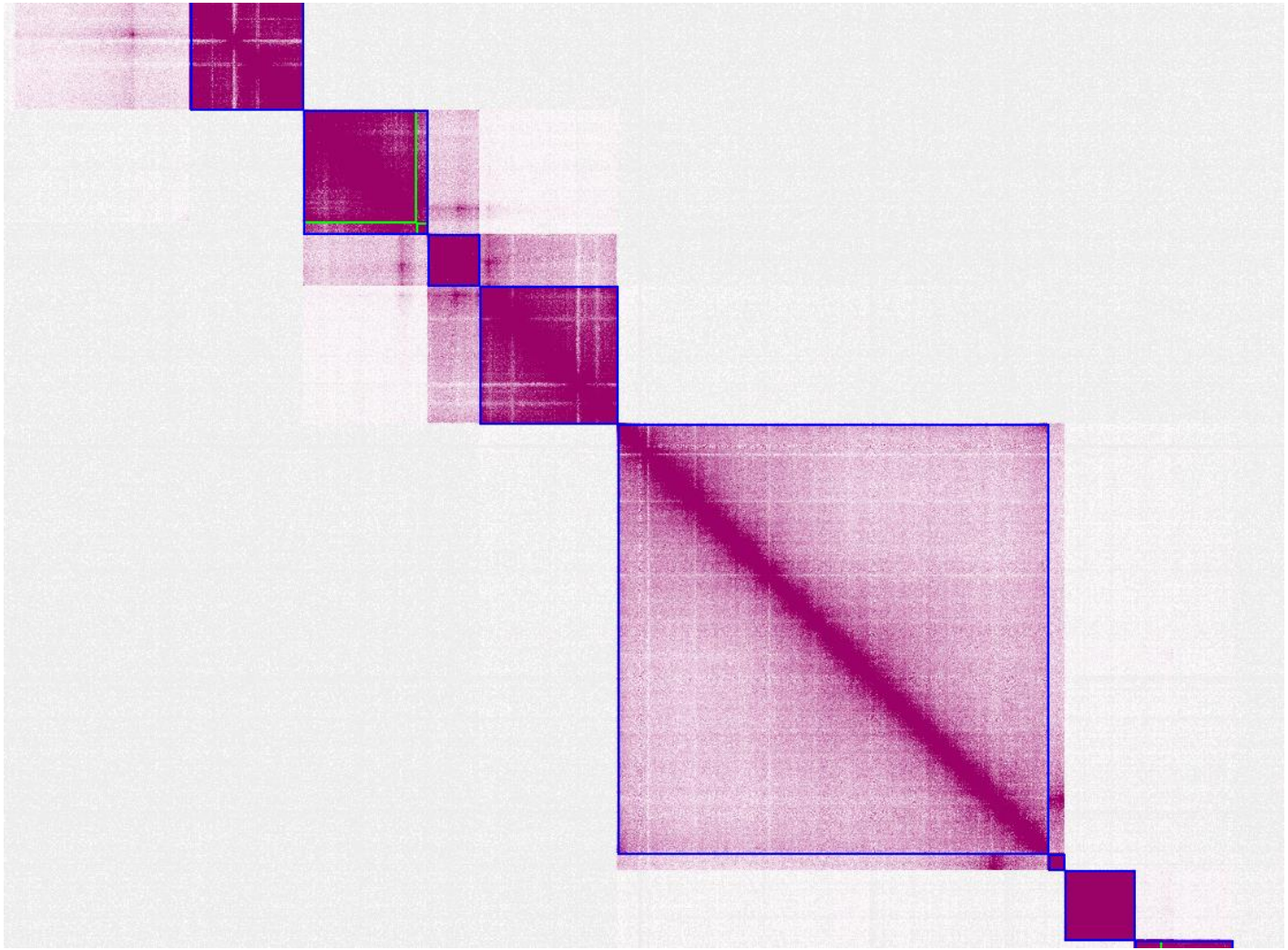
BUT AFTER 3D DNA SCAFFOLDING...

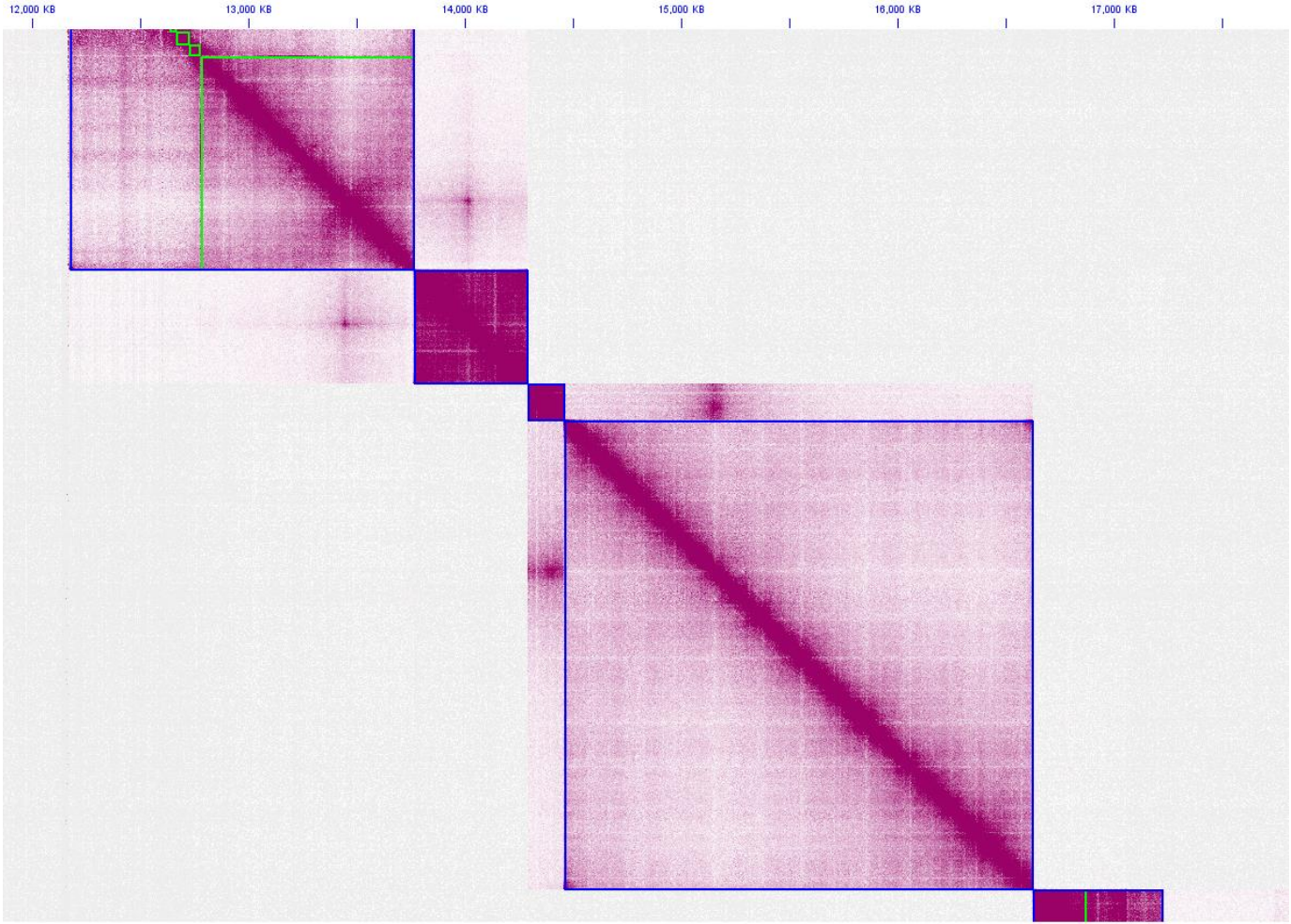





27,000 KB
28,000 KB
29,000 KB
30,000 KB
31,000 KB


22,000 KB
23,000 KB
24,000 KB
25,000 KB
26,000 KB

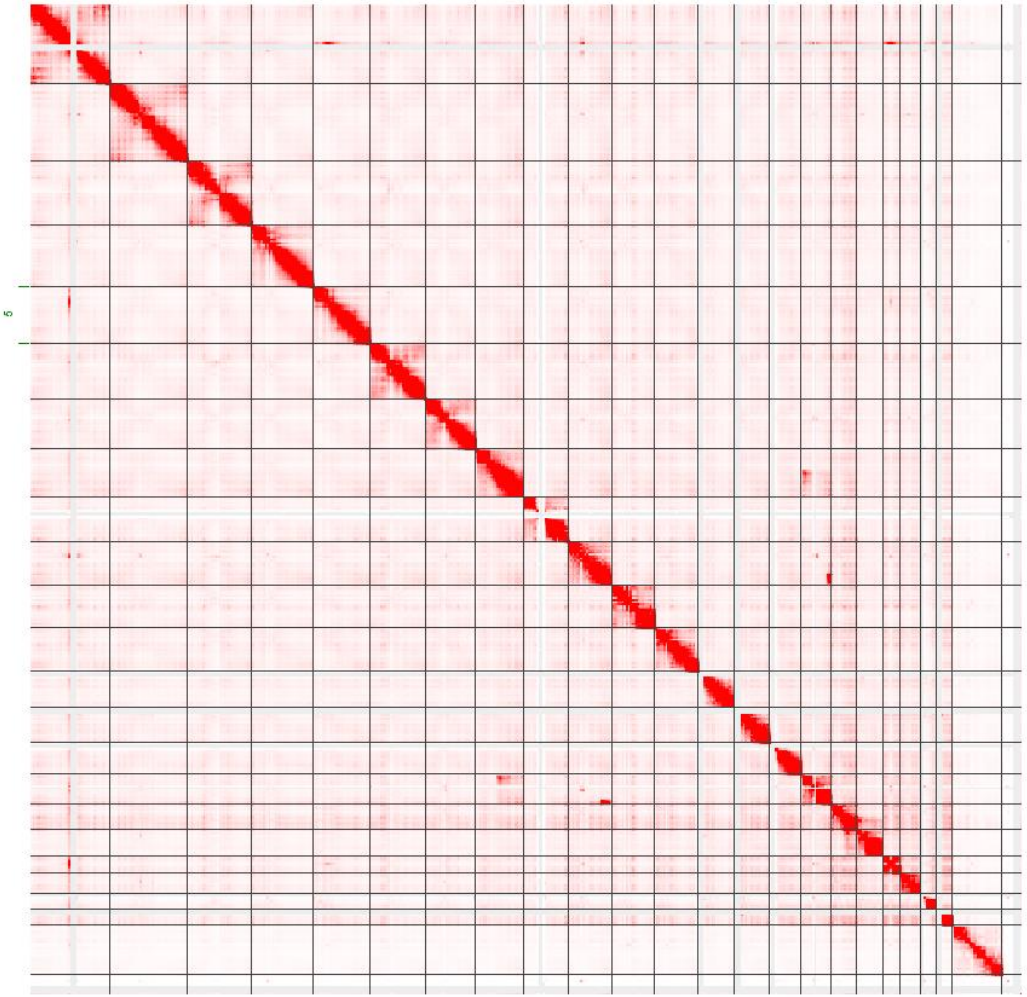


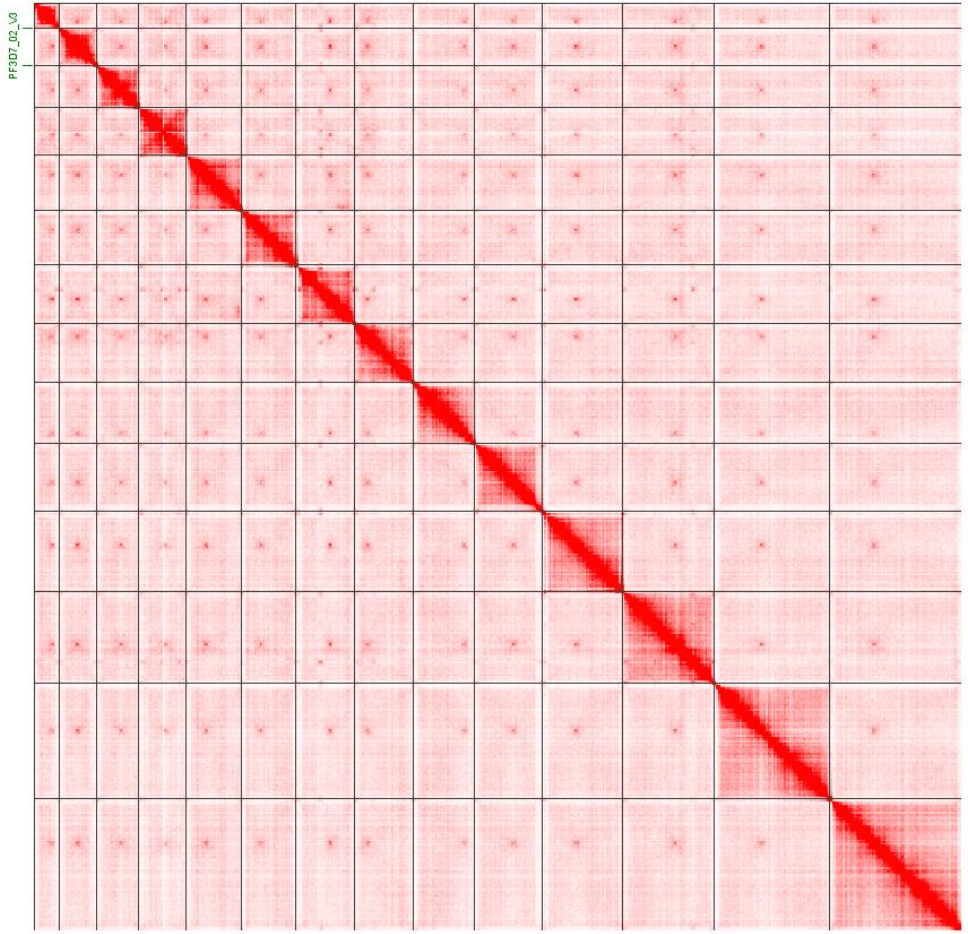

12,000 KB
13,000 KB
14,000 KB
15,000 KB
16,000 KB
17,000 KB

- These sorts of interchromosomal interactions are actually quite common.

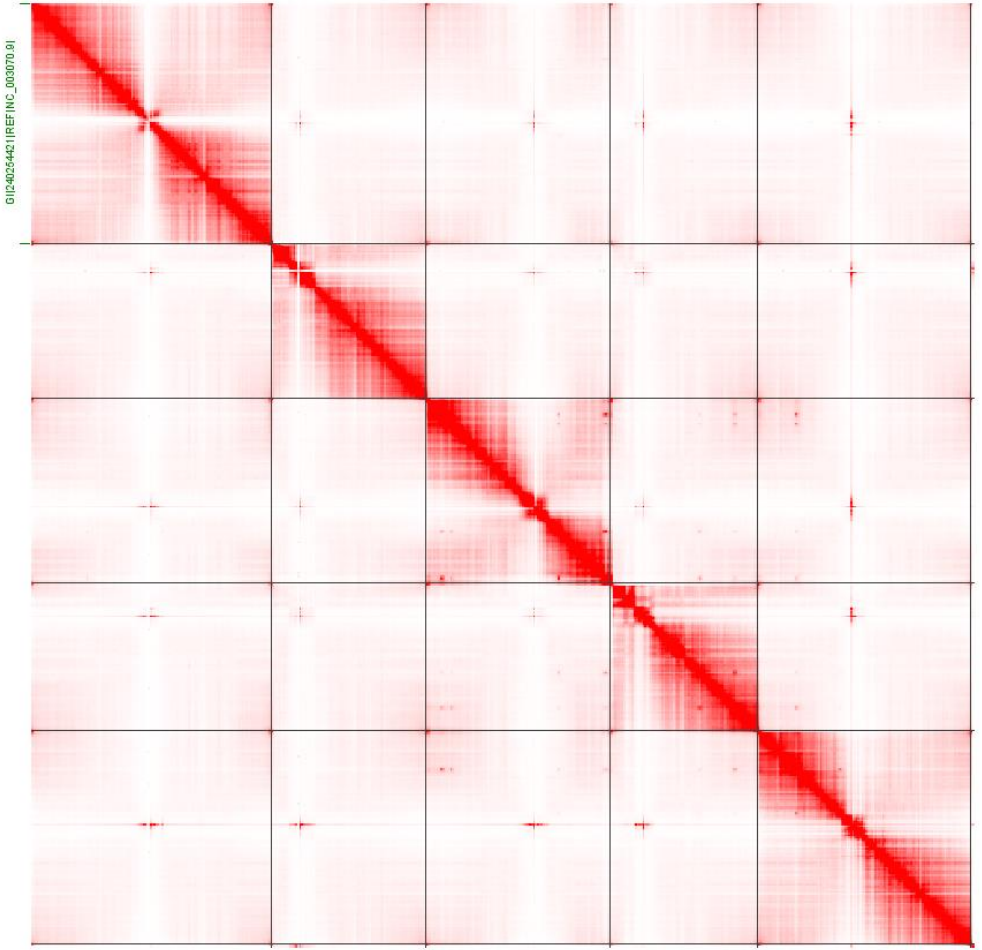
- Though not in mammals.

HUMAN




PF3D7_02_v3

ARABIDOPSIS



- These sorts of interchromosomal interactions are actually quite common.

- Though not in mammals.

- Usually they involve centromeres

- However, previously observed centromere-to-centromere interactions are always all-vs-all, while here we have pairwise interchromosomal loops (and sometimes three-way)

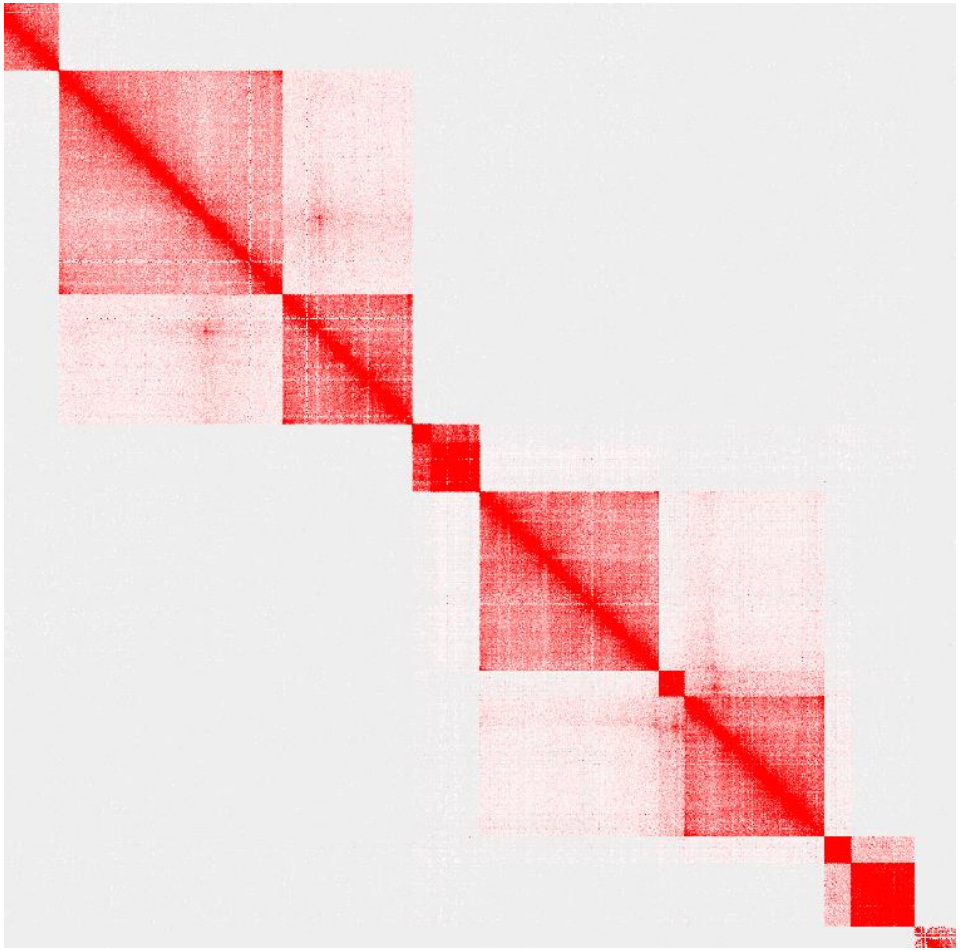
- MIC contamination cannot explain the patterns because the interactions are internal to chromosomes and there is no scrambling in *Tetrahymena*

- The data suggests that chromosomes exist in very specific compartments in the MAC, preferentially associating with one or two other chromosomes. Also, keep in mind the 45n ploidy

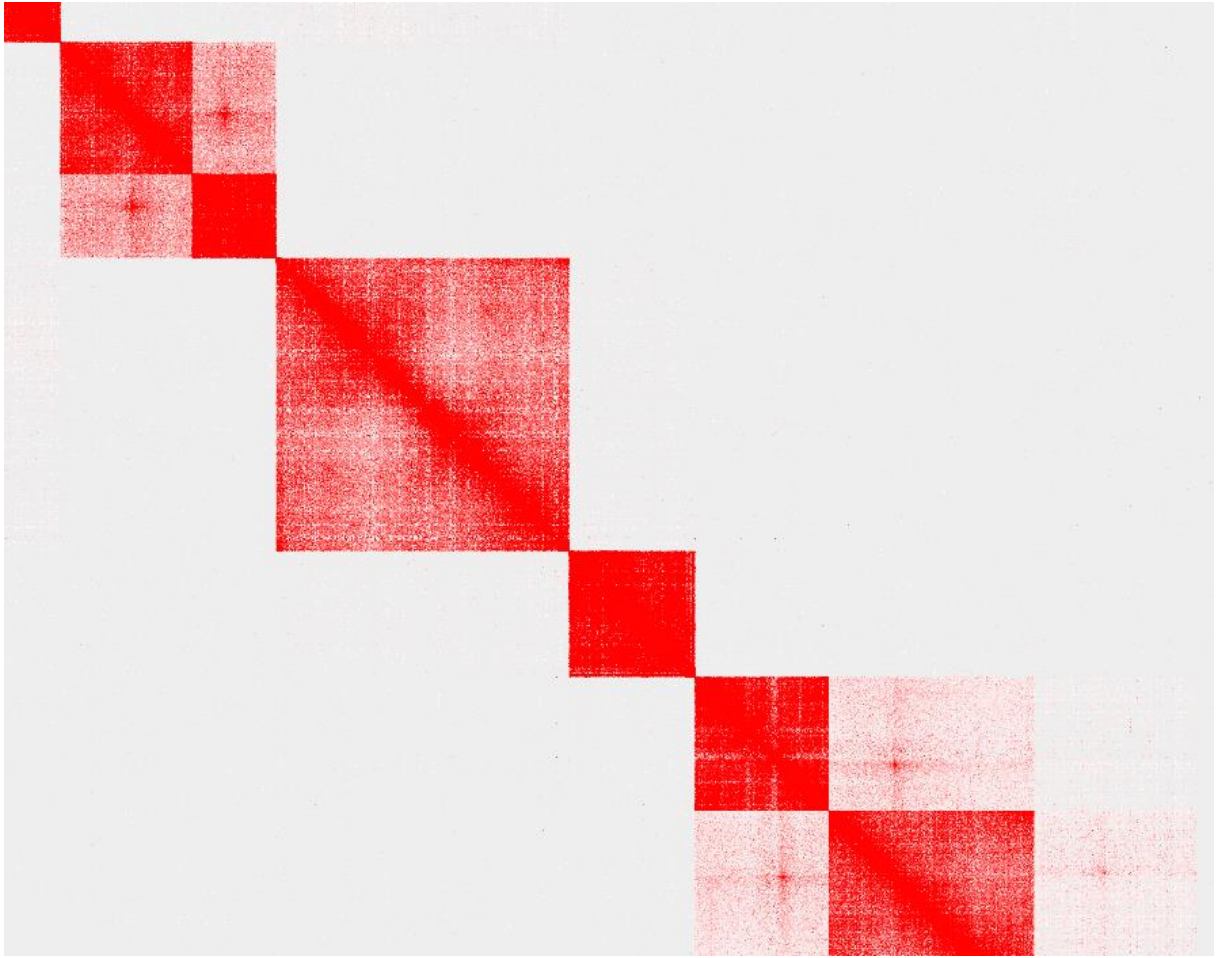
ONE WORRY

- Before ligation, the restriction enzyme needs to be heat-inactivated.
- For MboI, that is done at 62C for 20 minutes
- MluCI is heat-inactivated at 80C 30 minutes
- Tried Hi-C at 80C but that failed completely (likely due to crosslinks being reversed)

- Fortunately, the same patterns appear on maps generated with MboI, so they are not an artifact:



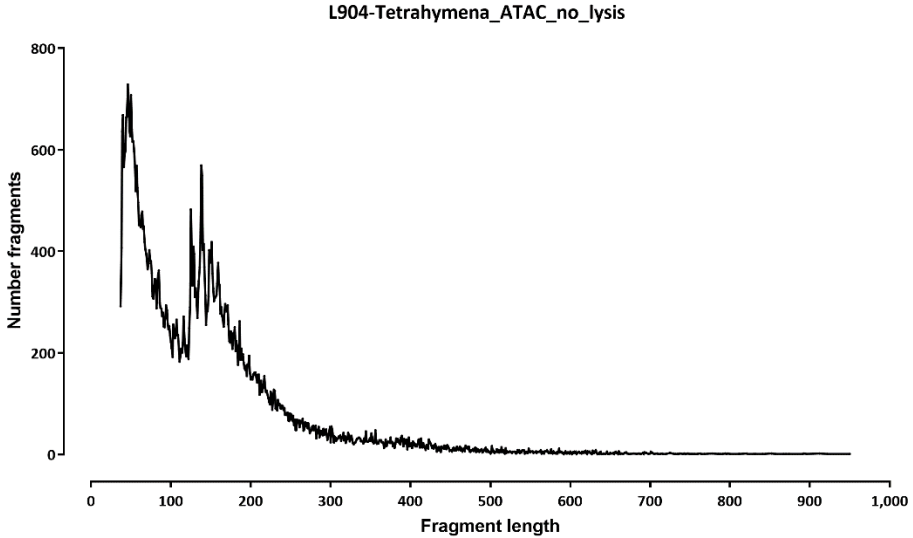
- Fortunately, the same patterns appear on maps generated with MboI, so they are not an artifact:

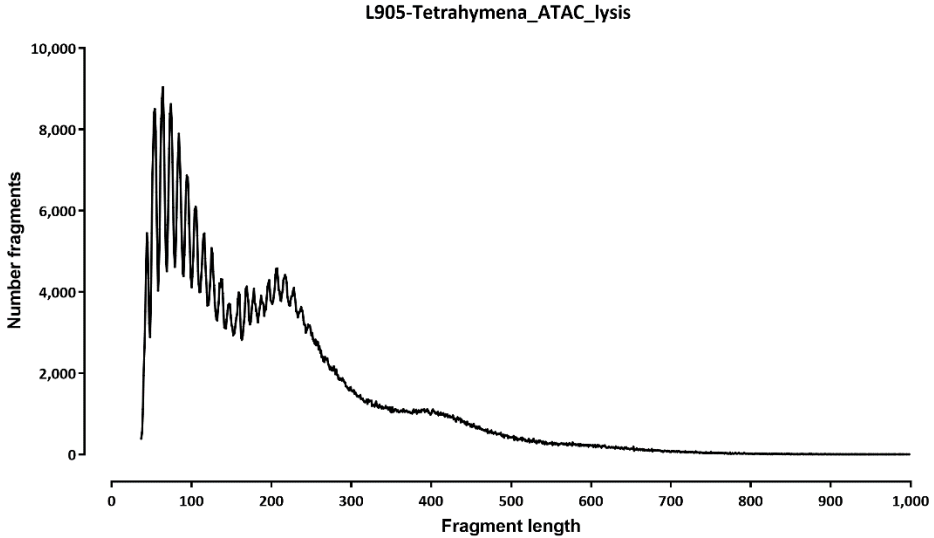


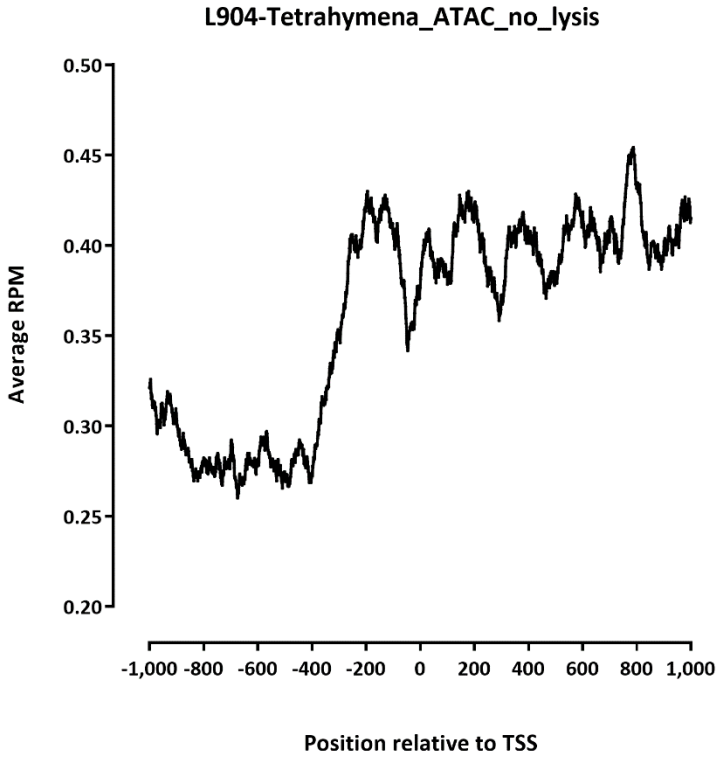
- We have done a MinION flowcell for validation of scaffolding (analysis in progress)

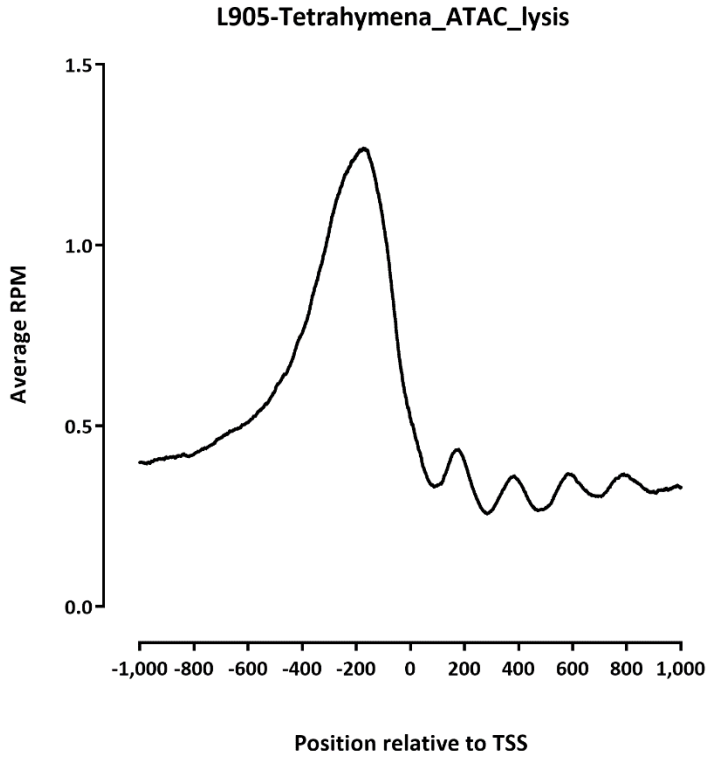
- Desperately need FISH validation of interchromosomal pairsings

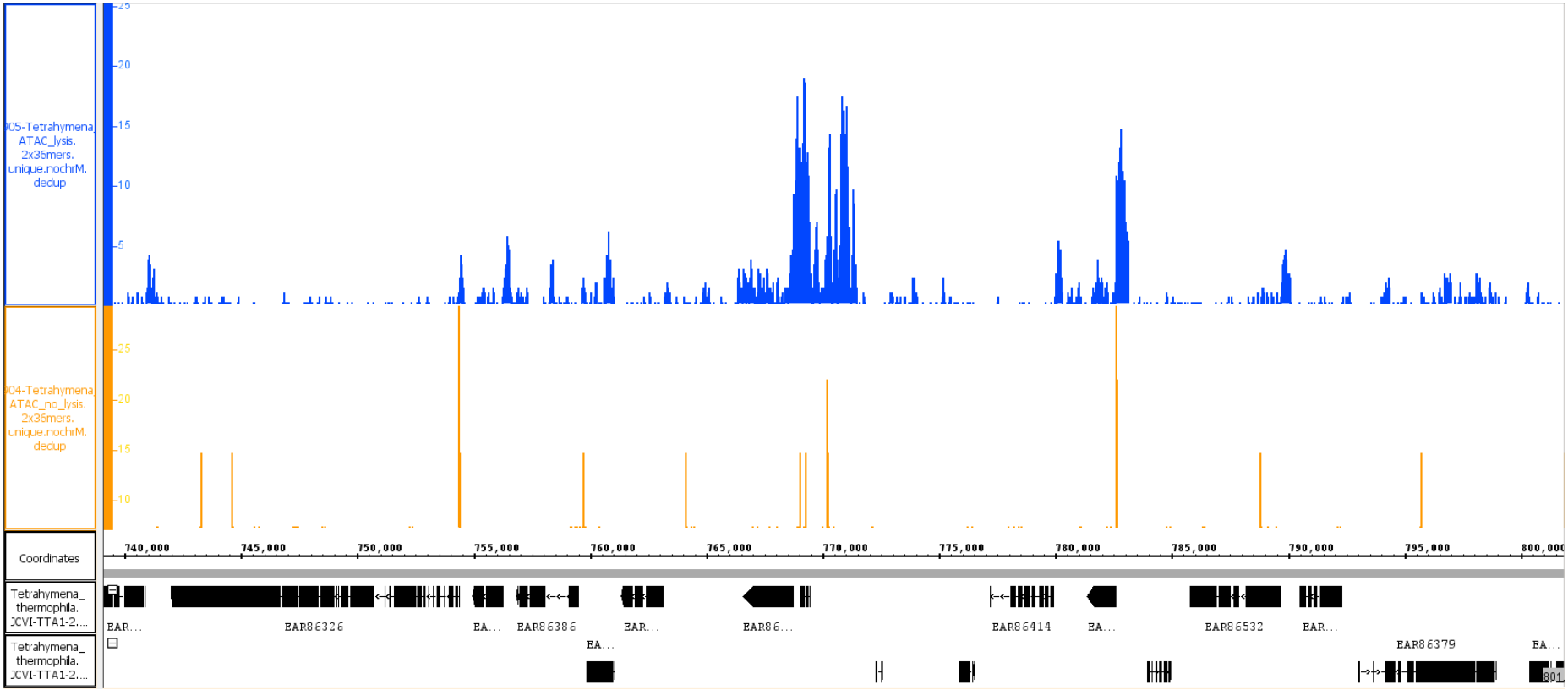
1.5. TETRAHYMENA ATAC-SEQ

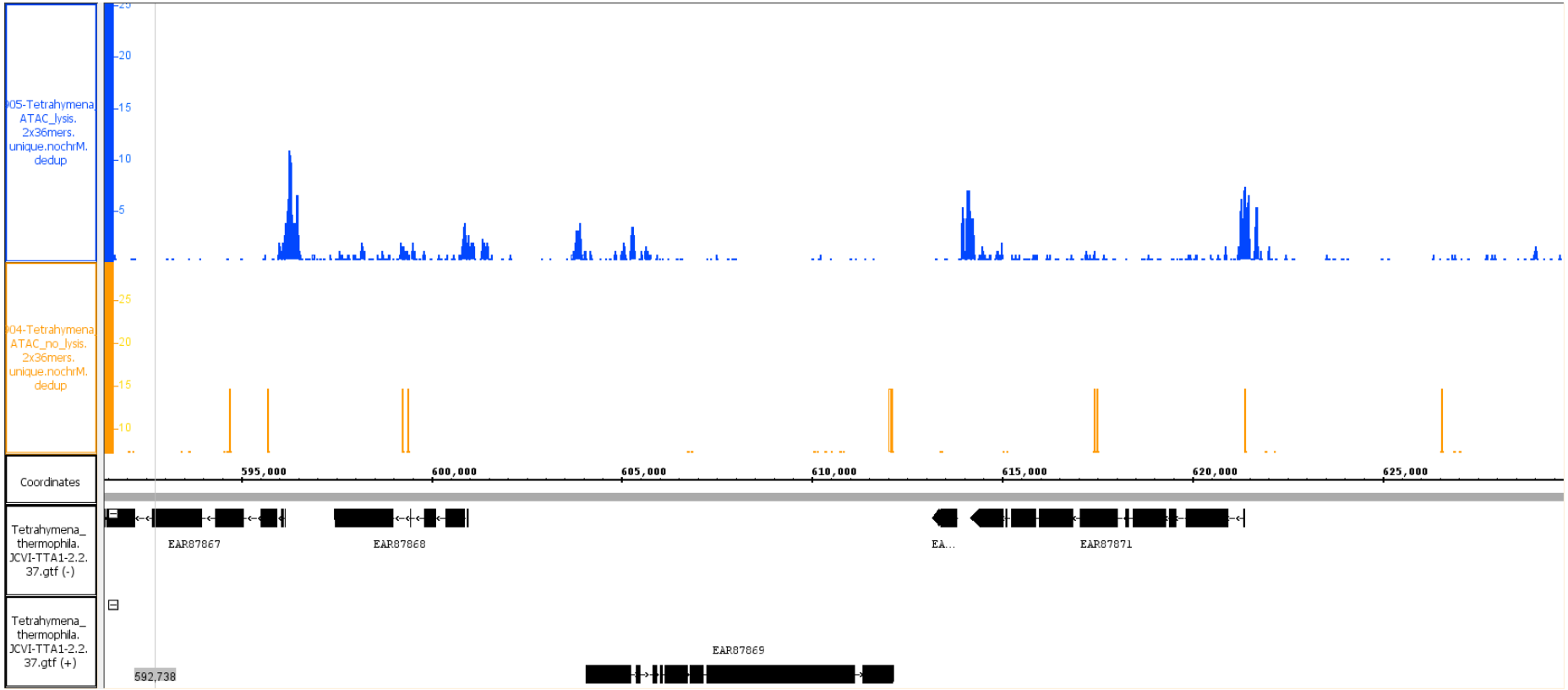

L904-Tetrahymena_ATAC_no_lysis
Number fragments
800
600
400
200
0
0
100
200
300
400
500
600
700
800
900
1,000
Fragment length

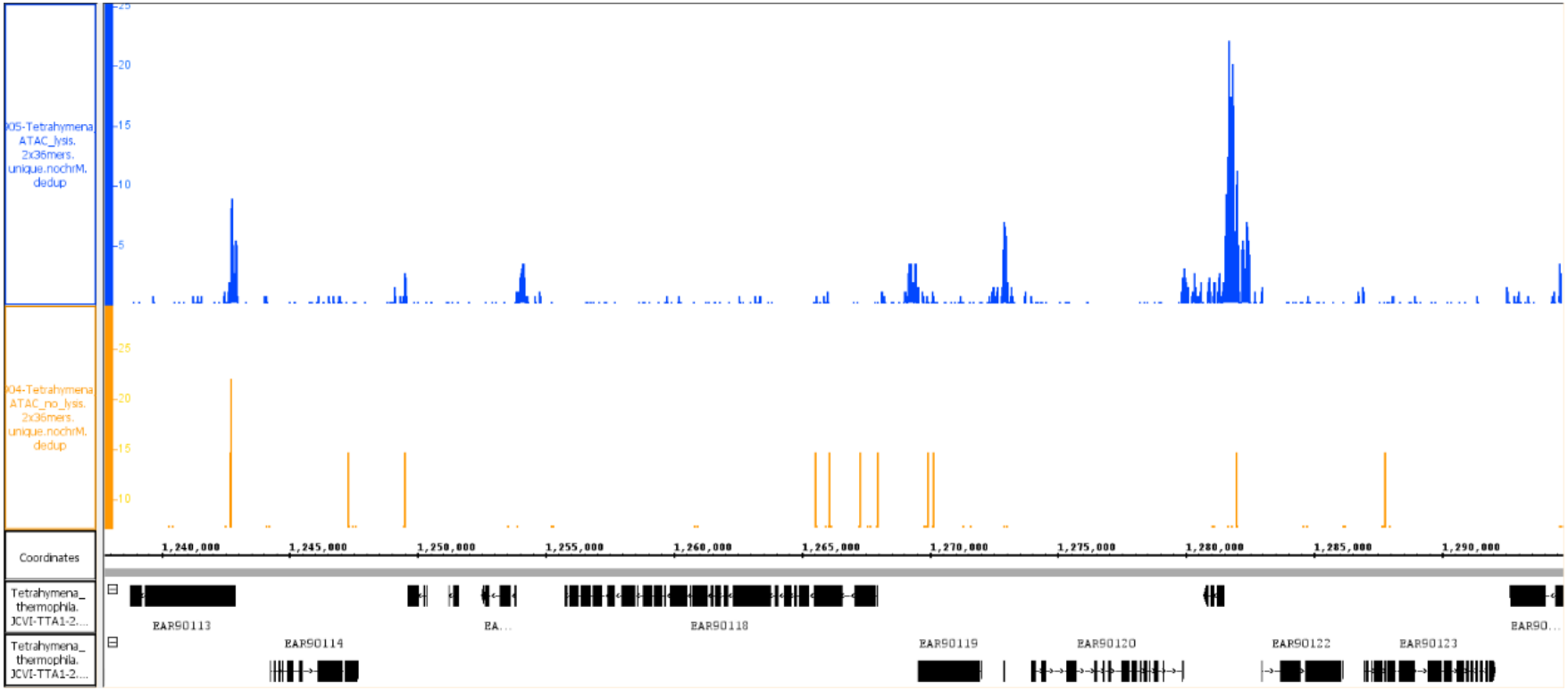

L905-Tetrahymena_ATAC_lysis
Number fragments
10,000
8,000
6,000
4,000
2,000
0
0
100
200
300
400
500
600
700
800
900
1,000
Fragment length


L904-Tetrahymena_ATAC_no_lysis
Average RPM
0.50
0.45
0.40
0.35
0.30
0.25
0.20
-1,000 -800 -600 -400 -200 0 200 400 600 800 1,000
Position relative to TSS


L905-Tetrahymena_ATAC_lysis
Average RPM
1.5
1.0
0.5
0.0
-1,000 -800 -600 -400 -200 0 200 400 600 800 1,000
Position relative to TSS

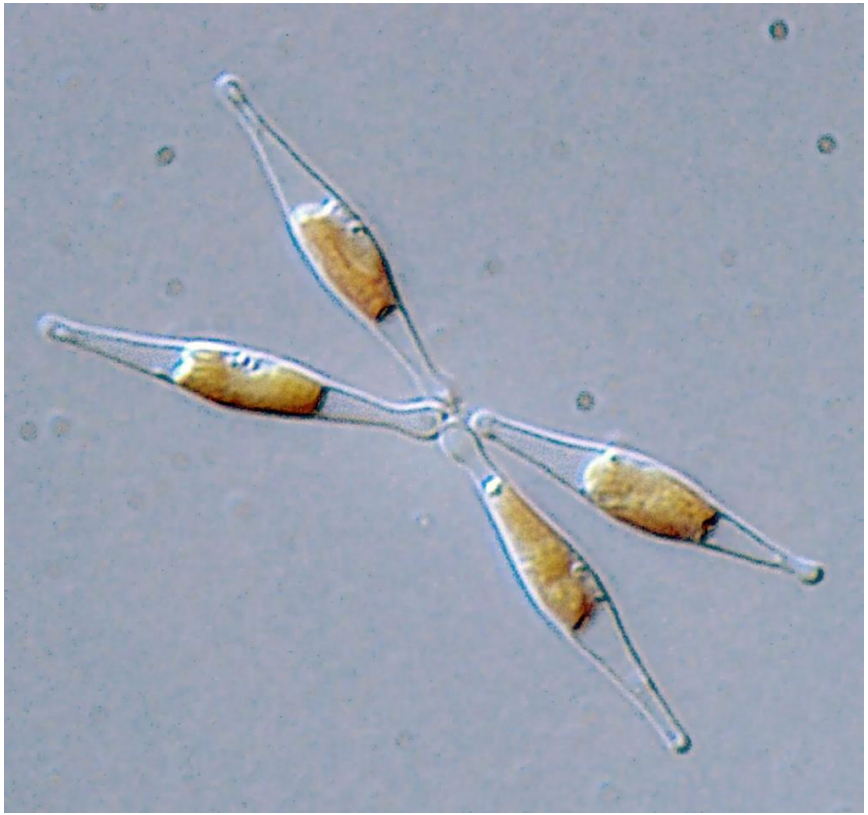

05-Tetrahymena_ATAC_lysis.2x36mers.unique.nochrM.dedup
04-Tetrahymena_ATAC_no_lysis.2x36mers.unique.nochrM.dedup
Coordinates
Tetrahymena_thermophila.JCVI-TTA1-2....
Tetrahymena_thermophila.JCVI-TTA1-2....
740,000
745,000
750,000
755,000
760,000
765,000
770,000
775,000
780,000
785,000
790,000
795,000
800,000
EAR86326
EAR86386
EAR86414
EAR86532
EAR86379


05-Tetrahymena ATAC_lysis. 2x36mers. unique.nochrM. dedup
04-Tetrahymena ATAC_no_lysis. 2x36mers. unique.nochrM. dedup
Coordinates
595,000
600,000
605,000
610,000
615,000
620,000
625,000
Tetrahymena_ thermophila. JCVI-TTA1-2.2. 37.gtf (-)
EAR87867
EAR87868
EA...
EAR87871
Tetrahymena_ thermophila. JCVI-TTA1-2.2. 37.gtf (+)
EAR87869
592,738


005-Tetrahymena ATAC_lysis. 2x35mers. unique.nochrM. dedup
004-Tetrahymena ATAC_no_lysis. 2x35mers. unique.nochrM. dedup
Coordinates
1,240,000
1,245,000
1,250,000
1,255,000
1,260,000
1,265,000
1,270,000
1,275,000
1,280,000
1,285,000
1,290,000
Tetrahymena_ thermophila. JCVI-TTA1-2....
Tetrahymena_ thermophila. JCVI-TTA1-2....
EAR90113
EAR90114
EAR90118
EAR90119
EAR90120
EAR90122
EAR90123

1.6. DIATOM HI-C

- *Phaeodactylum tricoturnum* (diatom)



NOTE ON EXISTING ASSEMBLY:

- The *T Phaeodactylum* genome was also sequenced using Sanger sequencing and is of good quality (though it did turn out it had a couple missasmblies)

- ~28 Mbp in total

- It was also not complete – it had 33 chromosomes plus a few quite sizable scaffolds

BRIEF REVIEW OF FOLDING MECHANISMS IN EUKARYOTES:

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

BRIEF REVIEW OF FOLDING MECHANISMS IN EUKARYOTES:

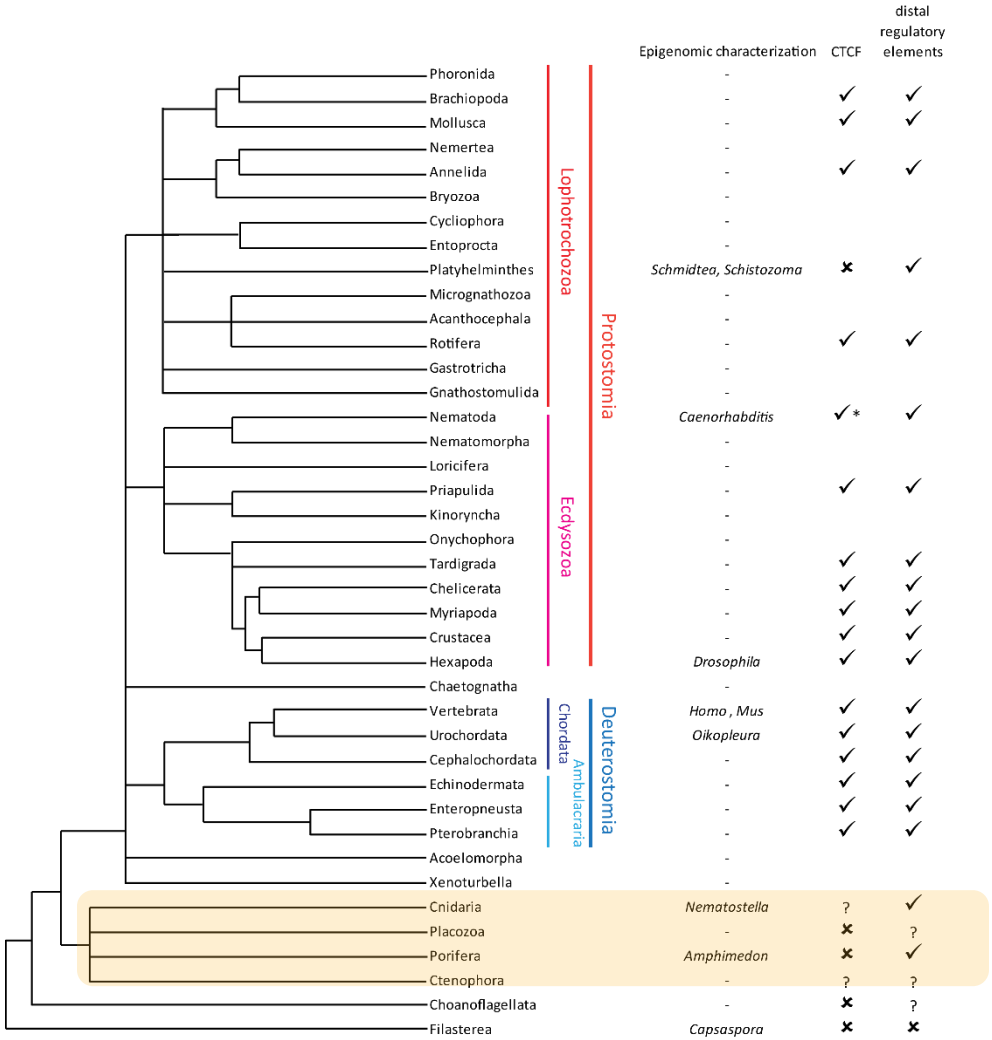
- Remarkably, TADs constrained by CTCF or CTCF-like loops are actually very rare

- Vertebrates have them

- *Drosophila* does not, even though it has CTCF plus additional 6 different insulator proteins (*Drosophila* has Polycomb loops, but those are few in number and very different in nature). It does, however, exhibit TAD-like patterns, but those do not seem to be formed by loop extrusion

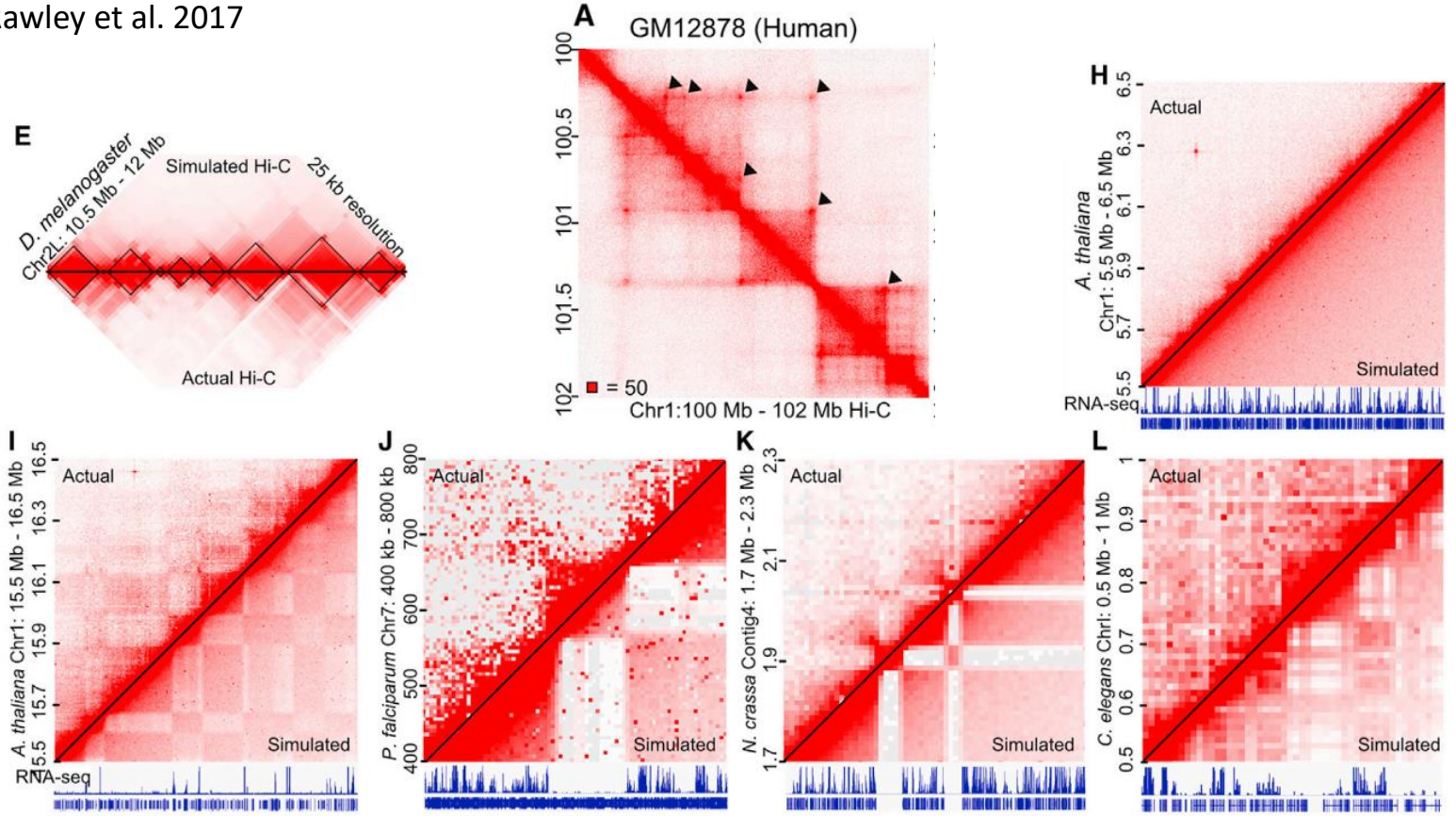
- *C. elegans* has no CTCF or other insulator proteins, and no TADs on its autosomes. But on the X chromosome the Dosage Compensation Complex (DCC) seems to play a similar role to CTCF and to form (weak) TADs

- There have been no reports of loop extrusion and TADs in other organisms


Epigenomic characterization
CTCF
distal regulatory elements
Phoronida
Brachiopoda
Mollusca
Nemertea
Annelida
Bryozoa
Cycliophora
Entoprocta
Platyhelminthes
Micrognathozoa
Acanthocephala
Rotifera
Gastrotricha
Gnathostomulida
Nematoda
Nematomorpha
Loricifera
Priapulida
Kinoryncha
Onychophora
Tardigrada
Chelicerata
Myriapoda
Crustacea
Hexapoda
Chaetognatha
Vertebrata
Urochordata
Cephalochordata
Echinodermata
Enteropneusta
Pterobranchia
Acoelomorpha
Xenoturbella
Cnidaria
Placozoa
Porifera
Ctenophora
Choanoflagellata
Filasterea
Lophotrochozoa
Ecdysozoa
Protostomia
Chordata
Ambulacraria
Deuterostomia
Schmidtea, Schistozoma
Caenorhabditis
Drosophila
Homo , Mus
Oikopleura
Nematostella
Amphimedon
Capsaspora

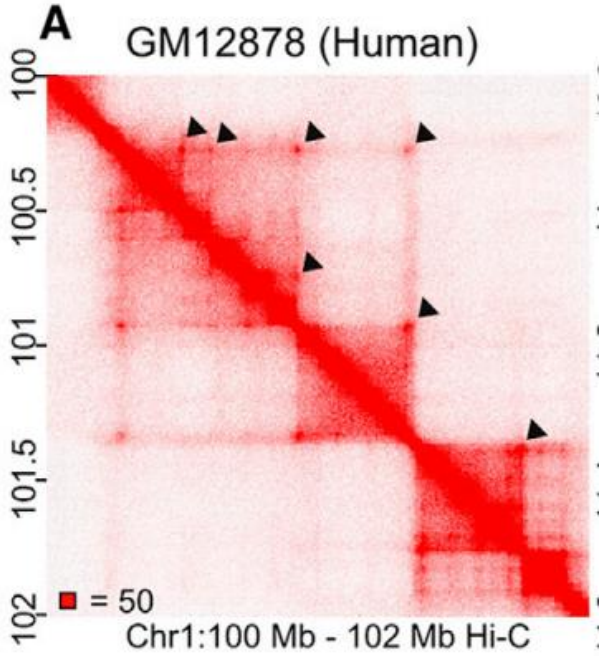
HI-C PATTERNS IN EUKARYOTES STUDIED SO FAR:

from Rawley et al. 2017

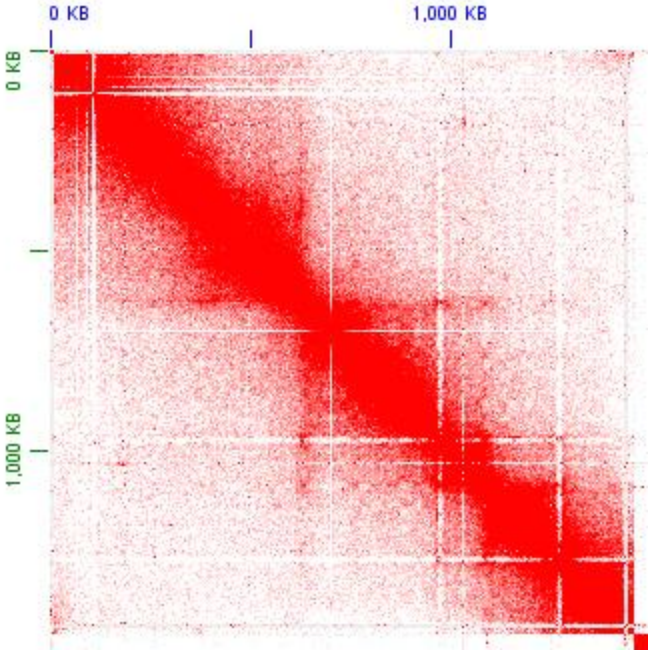


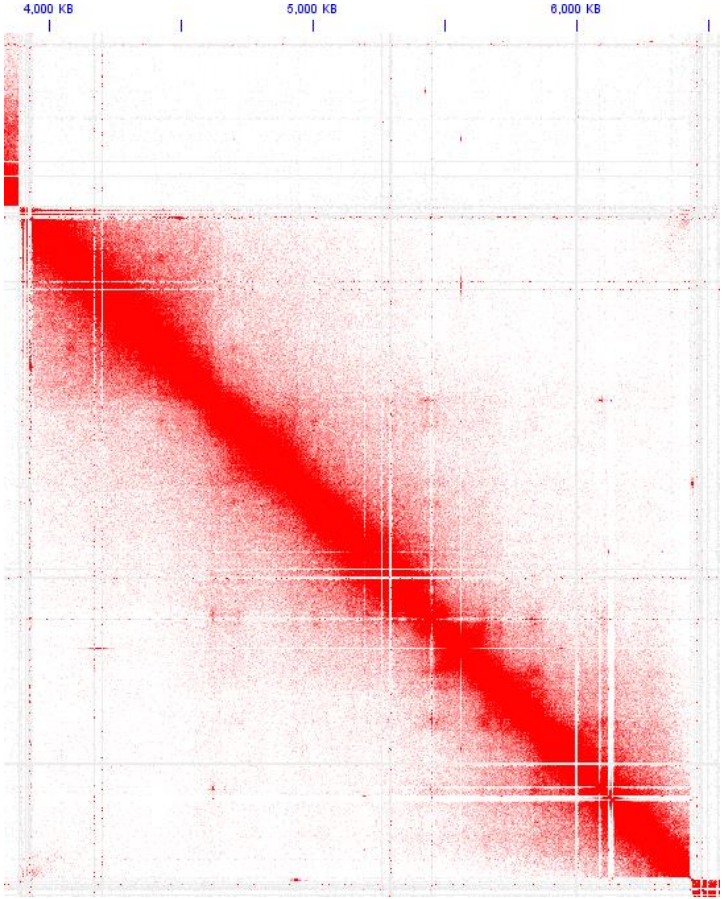
SIGNATURES OF LOOP EXTRUSION:

- TAD domains
- Loops
- Topological stripes

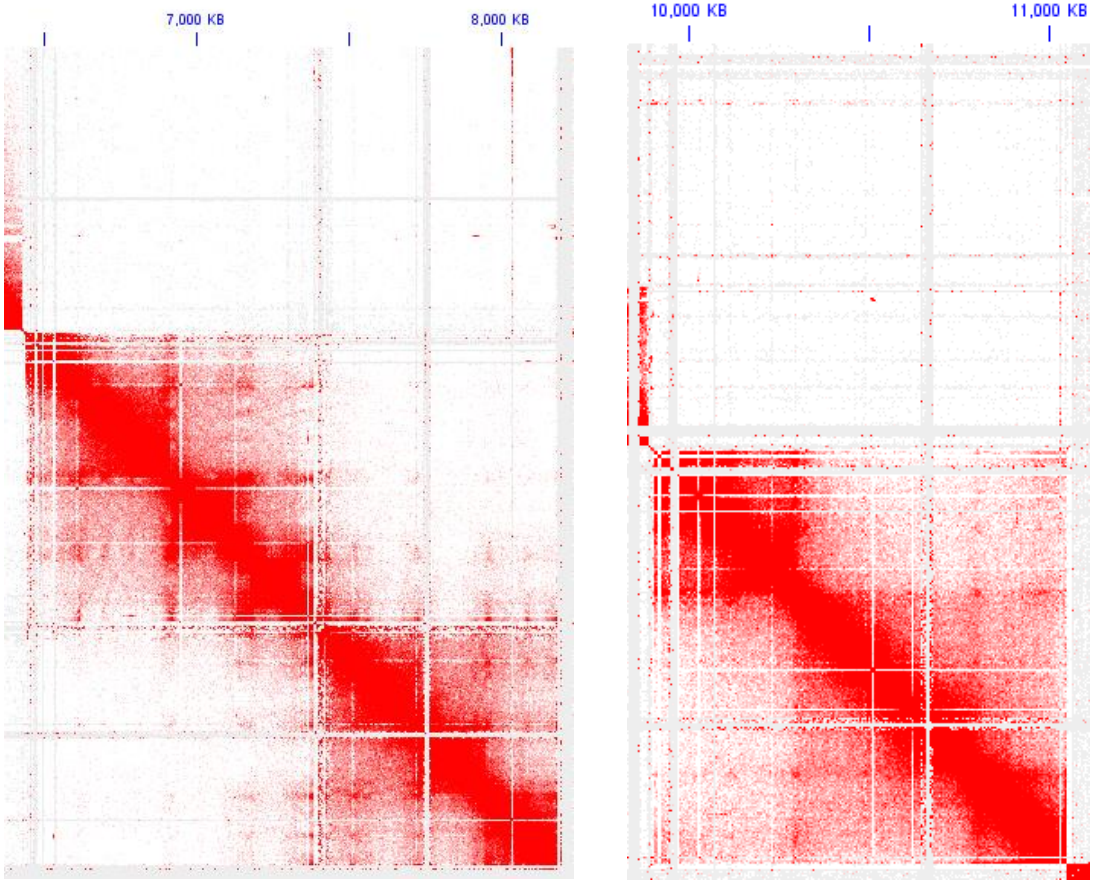


IN DIATOMS:

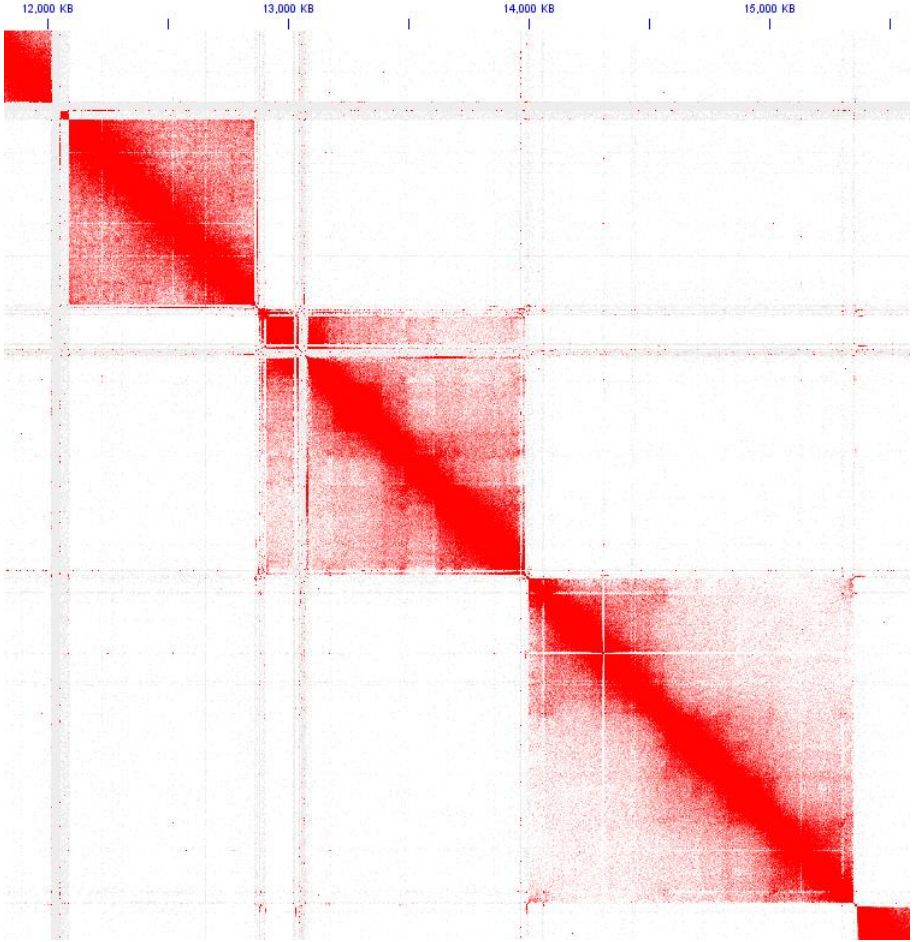


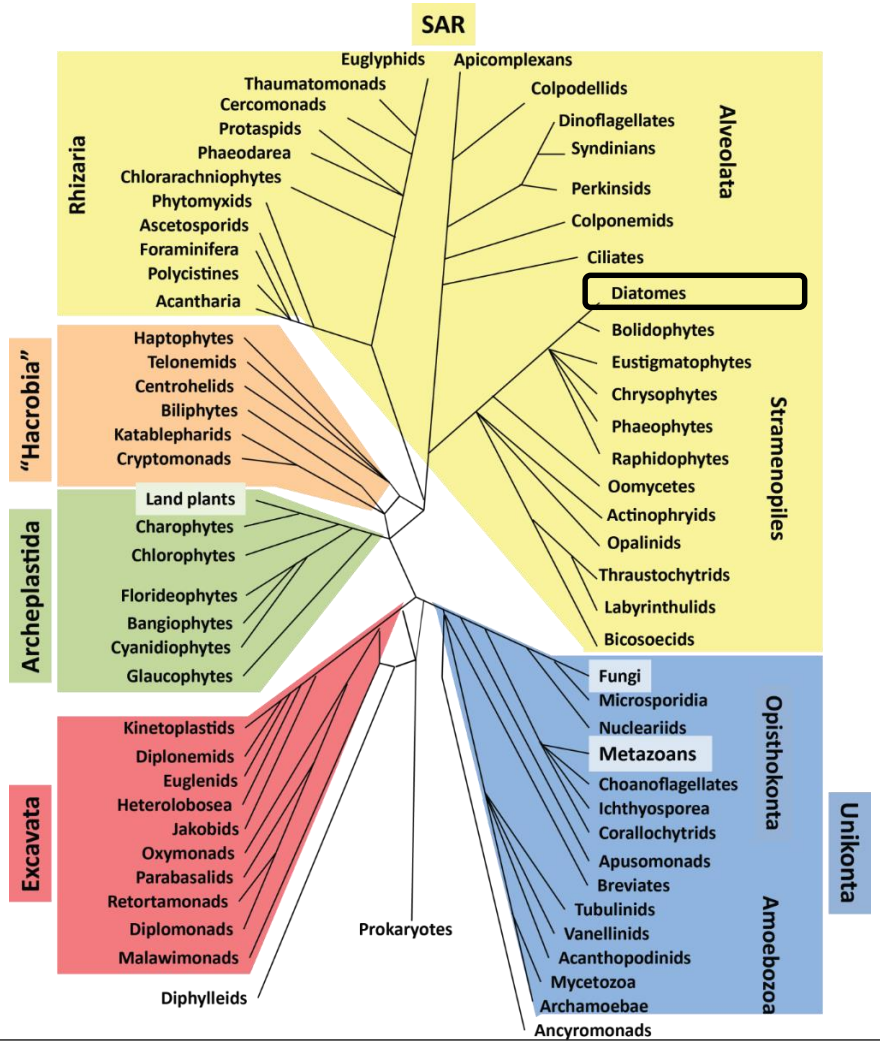


IN DIATOMS:

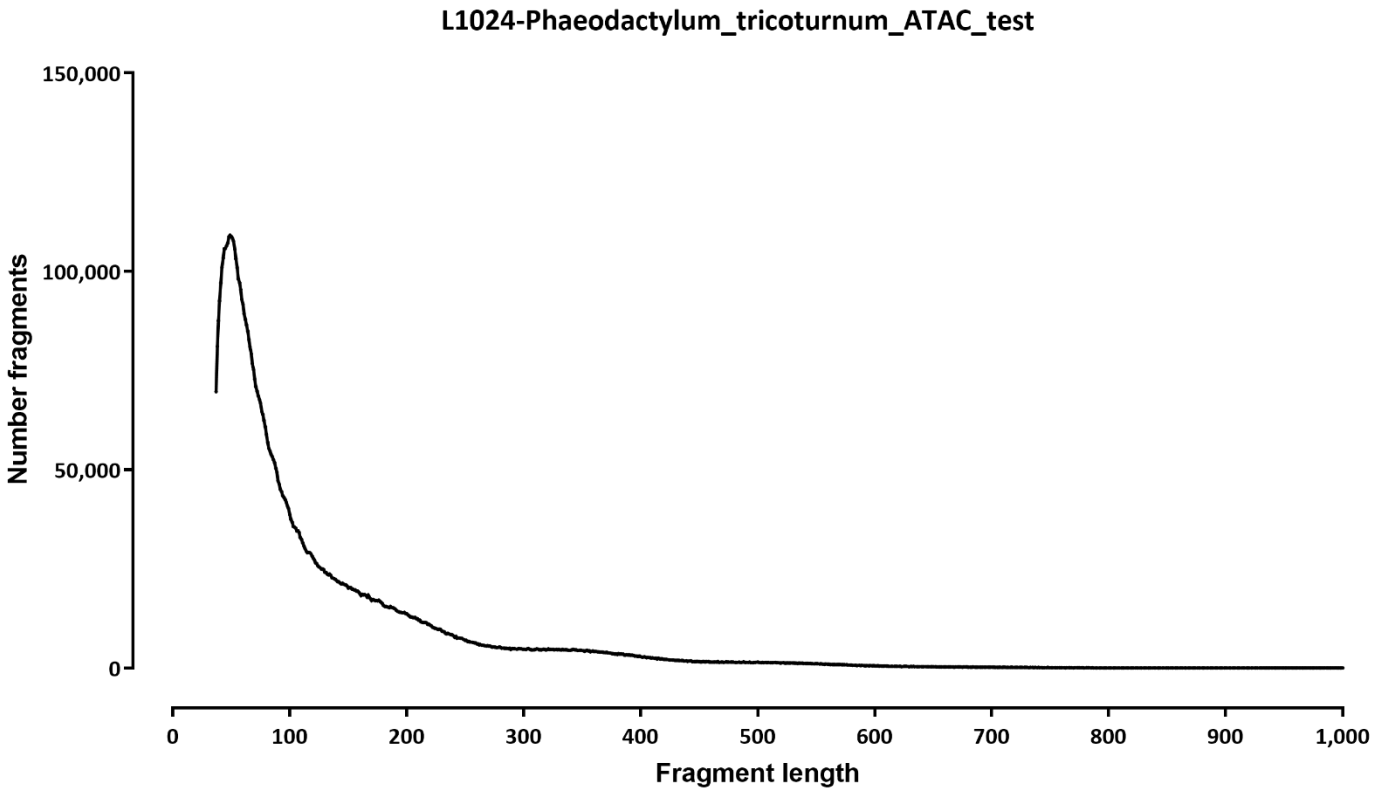


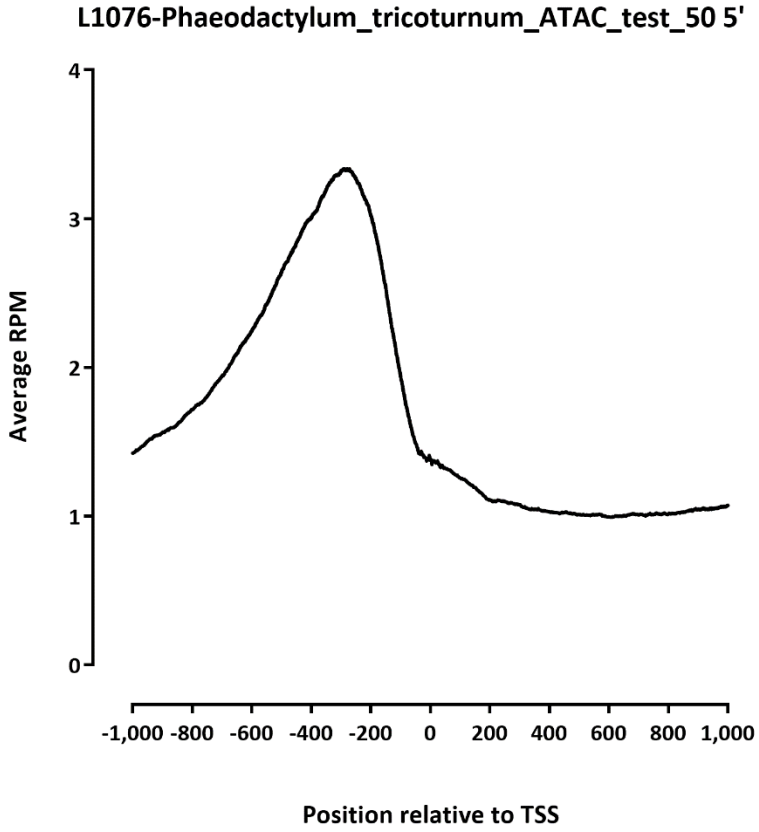
IN DIATOMS:

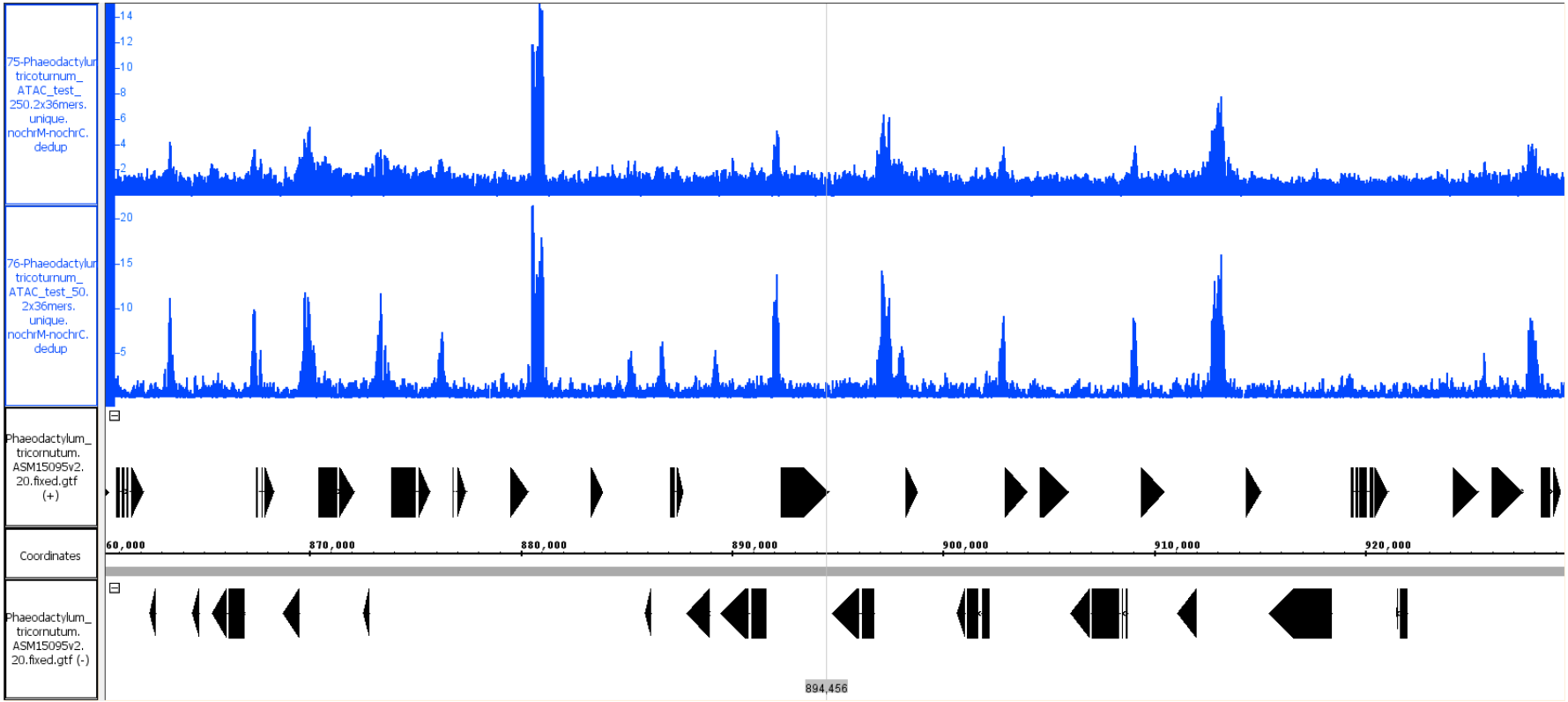


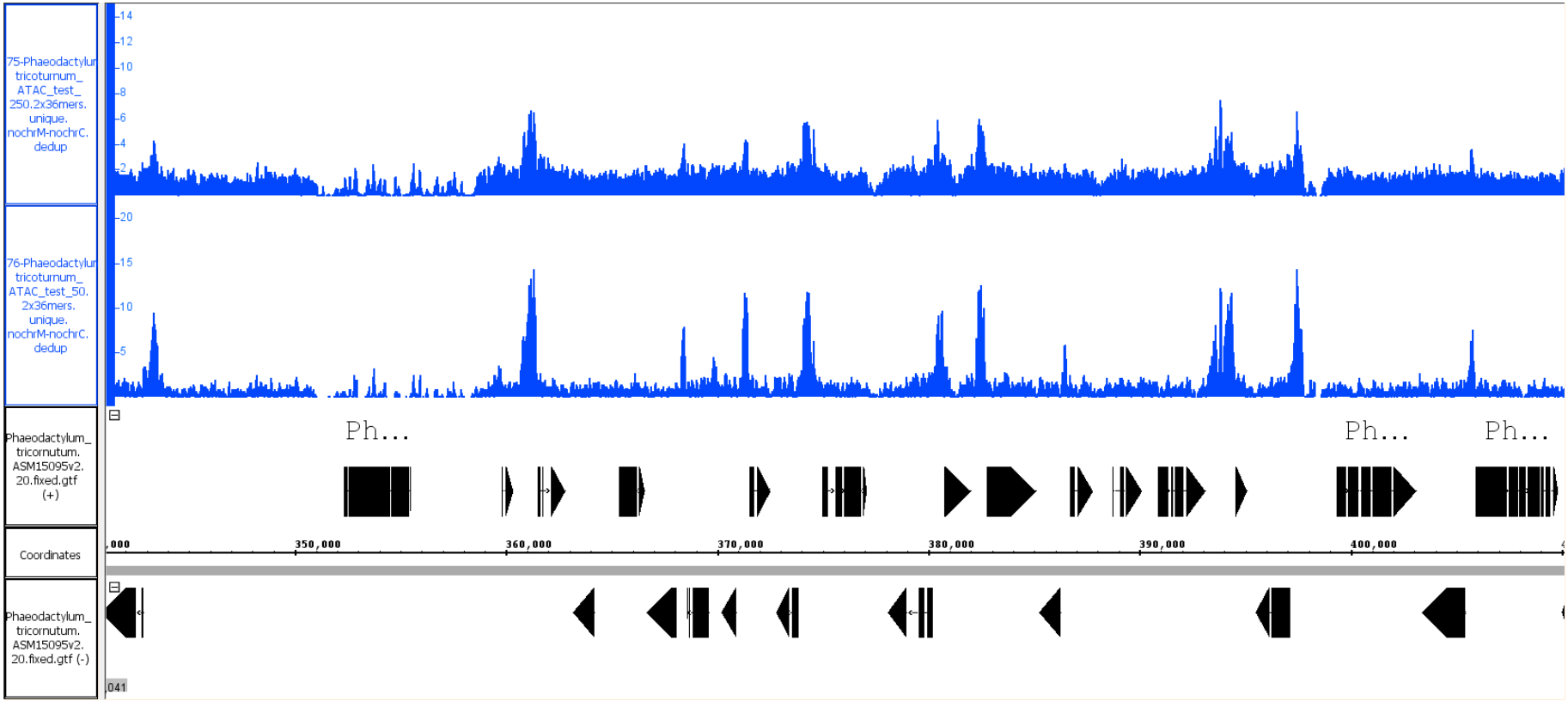

SAR
Rhizaria
Euglyphids
Thaumatomonads
Cercomonads
Protaspids
Phaeodarea
Chlorarachniophytes
Phytomyxids
Ascetosporids
Foraminifera
Polycistines
Acantharia
Alveolata
Apicomplexans
Colpodellids
Dinoflagellates
Syndinians
Perkinsids
Colponemids
Ciliates
Stramenopiles
Diatomes
Bolidophytes
Eustigmatophytes
Chrysophytes
Phaeophytes
Raphidophytes
Oomycetes
Actinophryids
Opalinids
Thraustochytrids
Labyrinthulids
Bicosoecids
"Hacrobia"
Haptophytes
Telonemids
Centrohelids
Biliphytes
Katablepharids
Cryptomonads
Archeplastida
Land plants
Charophytes
Chlorophytes
Florideophytes
Bangiophytes
Cyanidiophytes
Glaucophytes
Excavata
Kinetoplastids
Diplonemids
Euglenids
Heterolobosea
Jakobids
Oxymonads
Parabasalids
Retortamonads
Diplomonads
Malawimonads
Diphylleids
Prokaryotes
Unikonta
Opisthokonta
Fungi
Microsporidia
Nucleariids
Metazoans
Choanoflagellates
Ichthyosporea
Corallochytrids
Amoebozoa
Apusomonads
Breviates
Tubulinids
Vanellinids
Acanthopodinids
Mycetozoa
Archamoebae
Ancyromonads

1.7. DIATOM ATAC-SEQ


L1024-Phaeodactylum_tricoturnum_ATAC_test
150,000
100,000
50,000
0
Number fragments
0
100
200
300
400
500
600
700
800
900
1,000
Fragment length

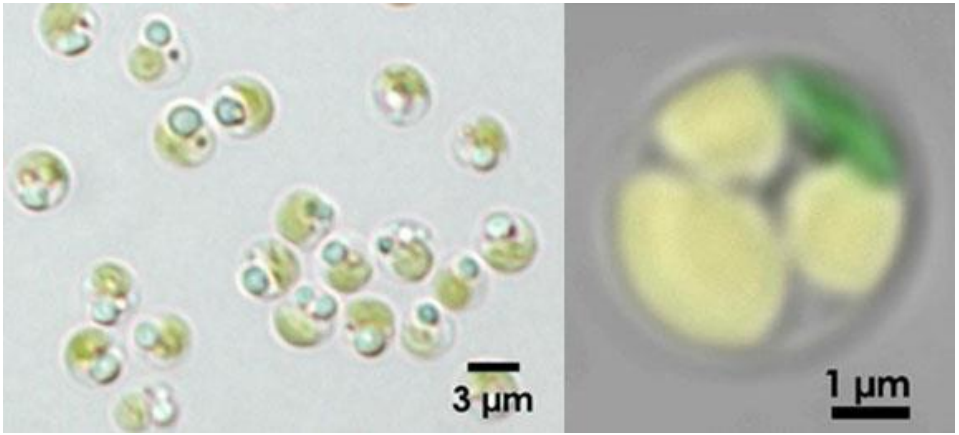

L1076-Phaeodactylum_tricoturnum_ATAC_test_50 5'
Average RPM
0
1
2
3
4
-1,000 -800 -600 -400 -200 0 200 400 600 800 1,000
Position relative to TSS

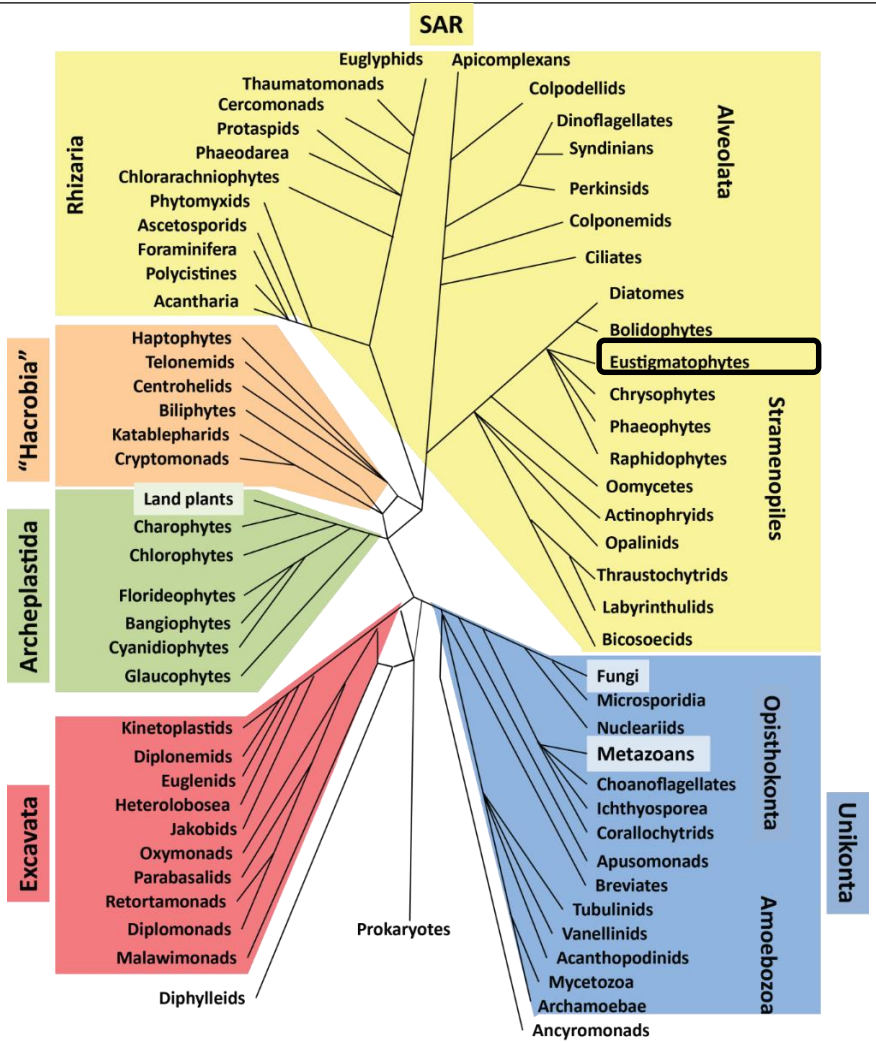

75-Phaeodactylum_tricoturnum_ATAC_test_250.2x36mers.unique.nochrM-nochrC.dedup
14
12
10
8
6
4
2
76-Phaeodactylum_tricoturnum_ATAC_test_50.2x36mers.unique.nochrM-nochrC.dedup
20
15
10
5
Phaeodactylum_tricornutum.ASM15095v2.20.fixed.gtf (+)
Coordinates
60,000
870,000
880,000
890,000
900,000
910,000
920,000
Phaeodactylum_tricornutum.ASM15095v2.20.fixed.gtf (-)
894,456


75-Phaeodactylur tricoturnum_ ATAC_test_ 250.2x36mers. unique. nochrM-nochrC. dedup
14
12
10
8
6
4
2
76-Phaeodactylur tricoturnum_ ATAC_test_50. 2x36mers. unique. nochrM-nochrC. dedup
20
15
10
5
Phaeodactylum_ tricornutum. ASM15095v2. 20.fixed.gtf (+)
Ph...
Ph...
Ph...
Coordinates
,000
350,000
360,000
370,000
380,000
390,000
400,000
Phaeodactylum_ tricornutum. ASM15095v2. 20.fixed.gtf (-)
041

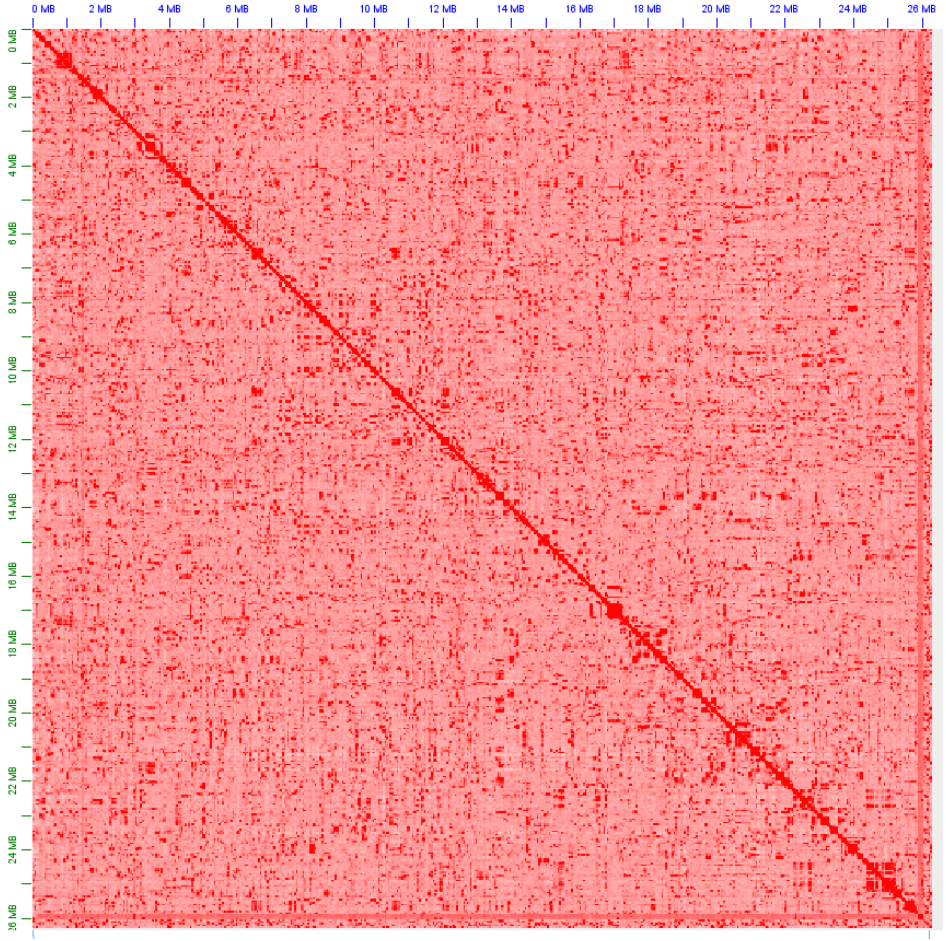
1.8 OTHER PROTOZOANS IN PROGRESS:

- *Nannochloropsis occulata* (eustygmatophyte)



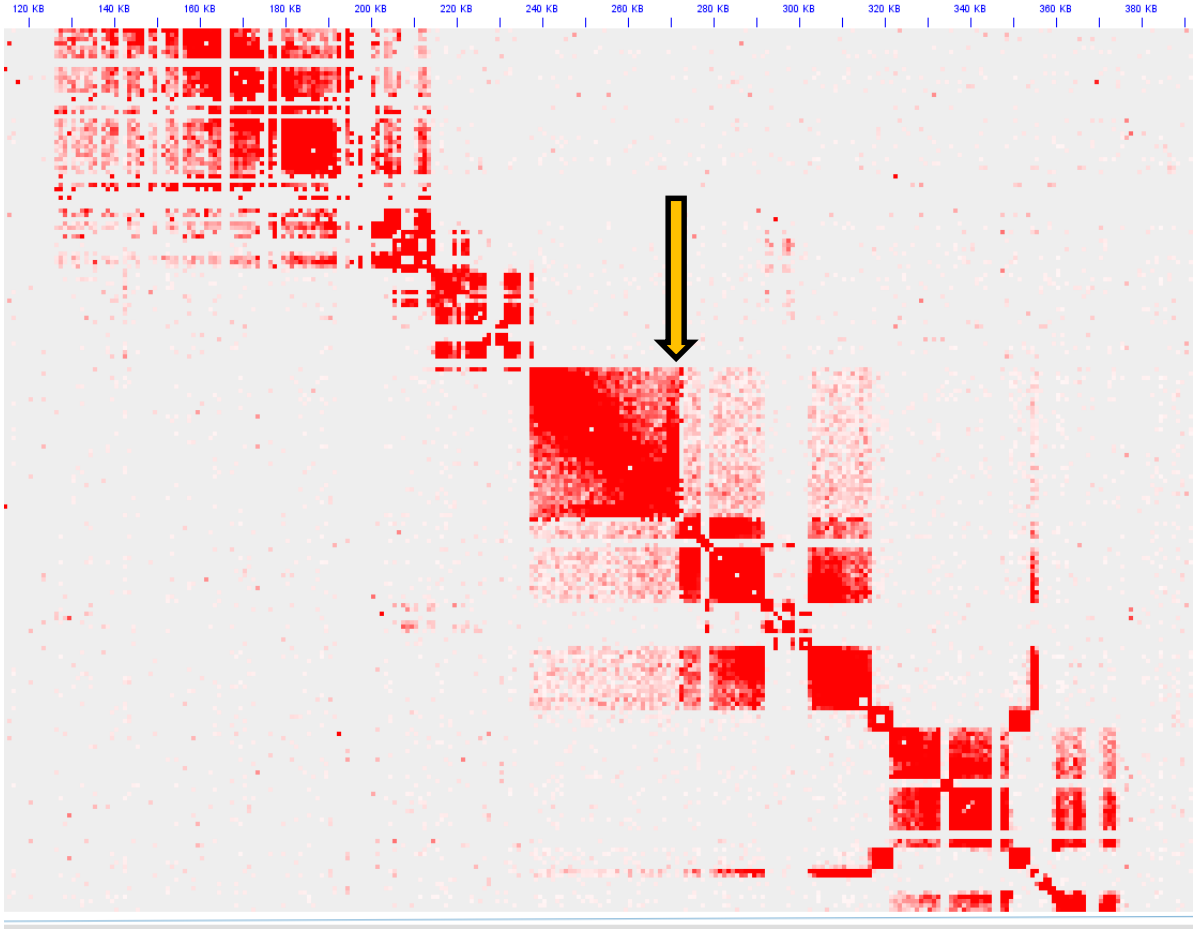

SAR
Euglyphids
Apicomplexans
Thaumatomonads
Colpodellids
Cercomonads
Dinoflagellates
Protaspids
Syndinians
Phaeodarea
Perkinsids
Chlorarachniophytes
Colponemids
Phytomyxids
Ciliates
Ascetosporids
Foraminifera
Polycistines
Acantharia
Rhizaria
Alveolata
Diatomes
Bolidophytes
Eustigmatophytes
Chrysophytes
Phaeophytes
Raphidophytes
Oomycetes
Actinophryids
Opalinids
Thraustochytrids
Labyrinthulids
Bicosoecids
Stramenopiles
Haptophytes
Telonemids
Centrohelids
Biliphytes
Katablepharids
Cryptomonads
"Hacrobia"
Land plants
Charophytes
Chlorophytes
Florideophytes
Bangiophytes
Cyanidiophytes
Glaucophytes
Archeplastida
Fungi
Microsporidia
Nucleariids
Metazoans
Choanoflagellates
Ichthyosporea
Corallochytrids
Apusomonads
Breviates
Tubulinids
Vanellinids
Acanthopodinids
Mycetozoa
Archamoebae
Opisthokonta
Amoebozoa
Unikonta
Kinetoplastids
Diplonemids
Euglenids
Heterolobosea
Jakobids
Oxymonads
Parabasalids
Retortamonads
Diplomonads
Malawimonads
Excavata
Prokaryotes
Diphylleids
Ancyromonads

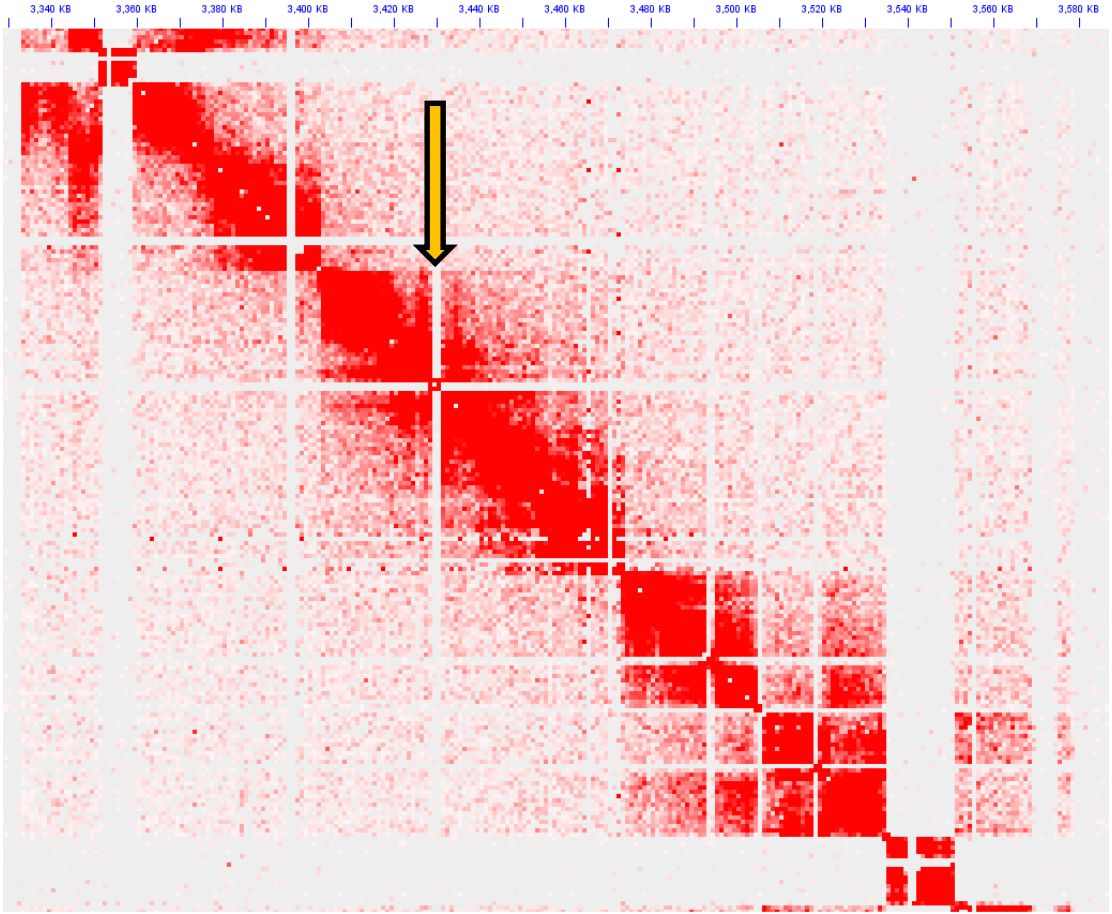
HI-C SCAFFOLDING



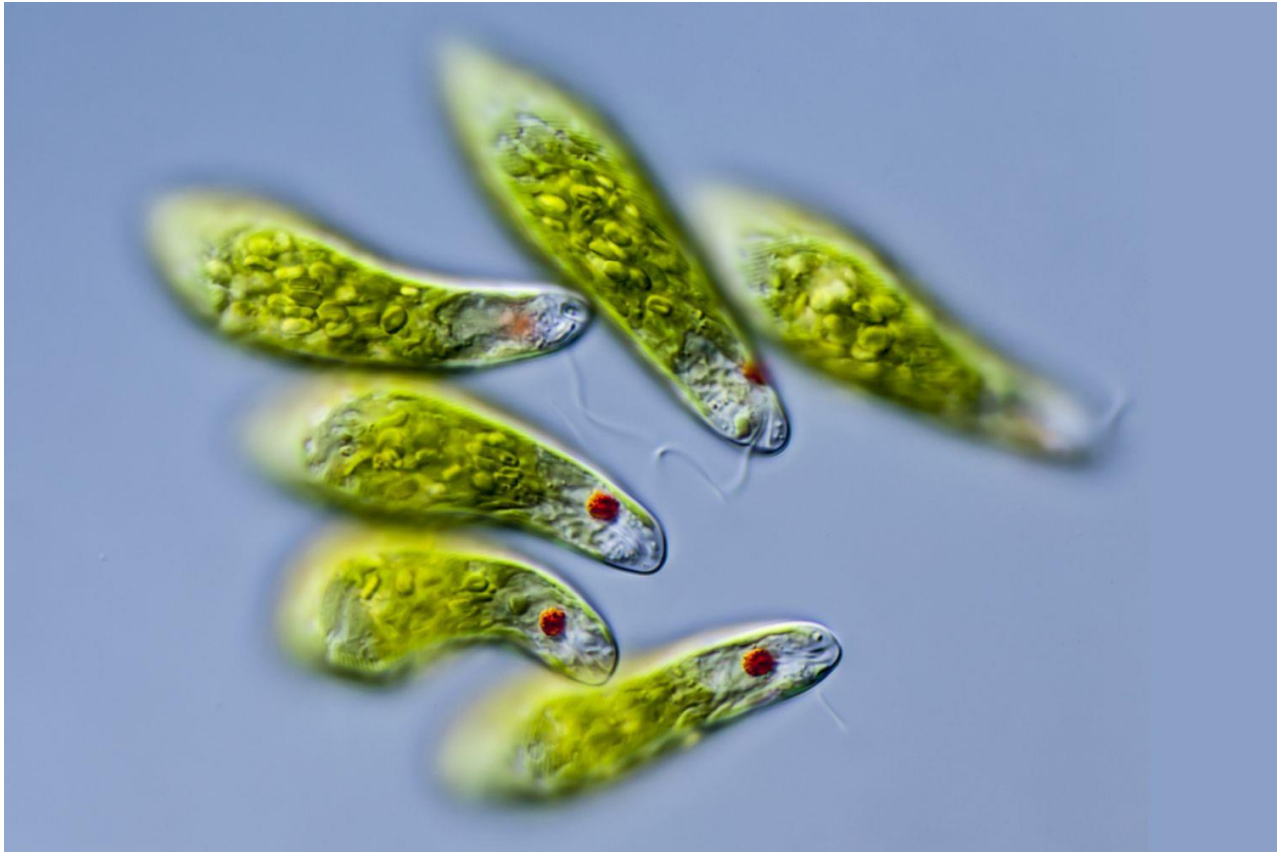
It appears that the strain we have is not a match of the strain from which the public assembly was generated

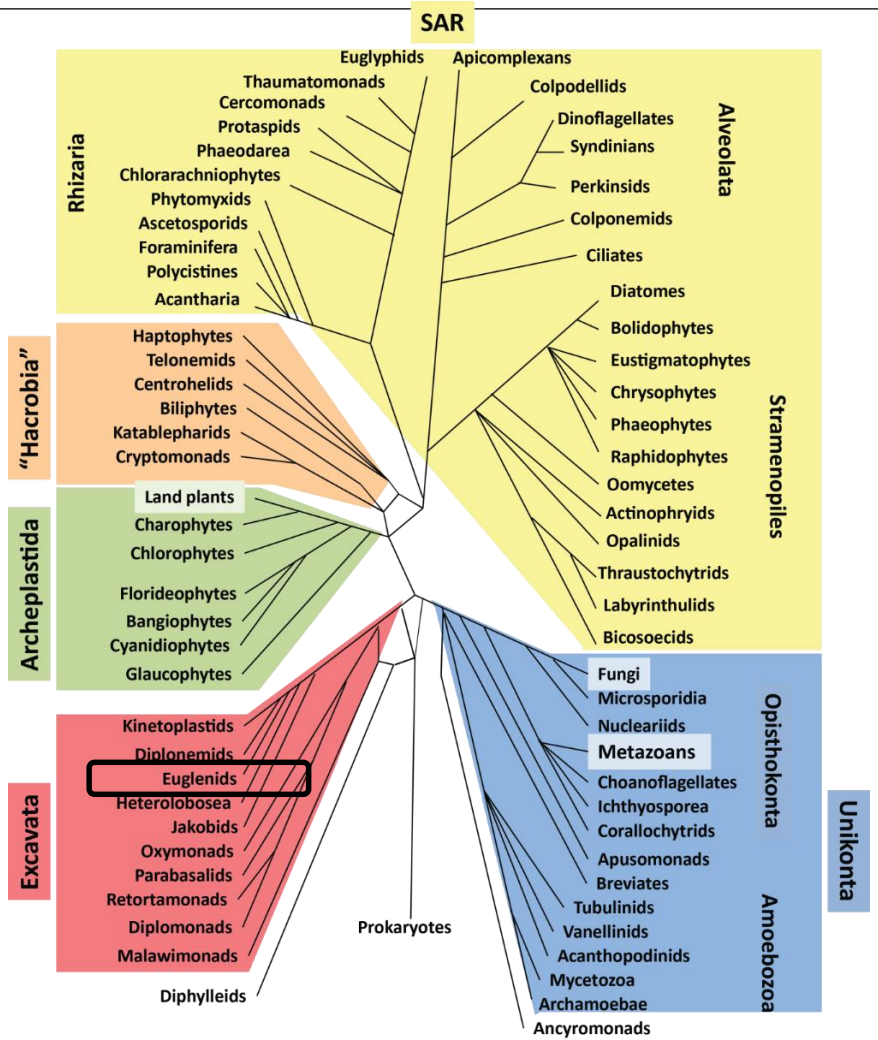
Will have to reassemble at the DNA level (genome is 25Mb)


120 KB
140 KB
160 KB
180 KB
200 KB
220 KB
240 KB
260 KB
280 KB
300 KB
320 KB
340 KB
360 KB
380 KB


3,340 KB
3,360 KB
3,380 KB
3,400 KB
3,420 KB
3,440 KB
3,460 KB
3,480 KB
3,500 KB
3,520 KB
3,540 KB
3,560 KB
3,580 KB

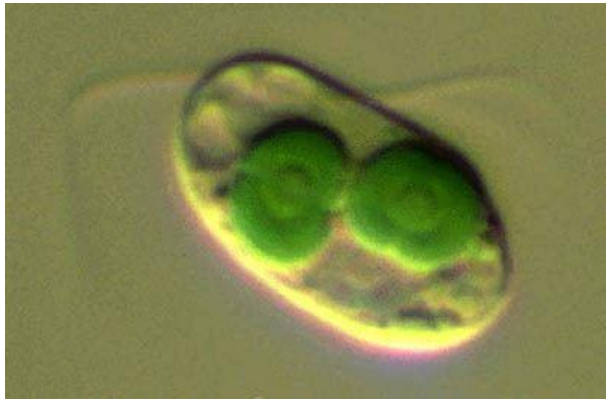
- *Euglena gracilis* (euglenid)



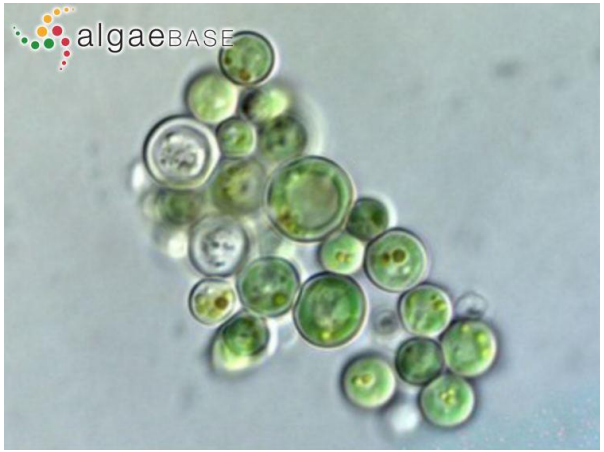


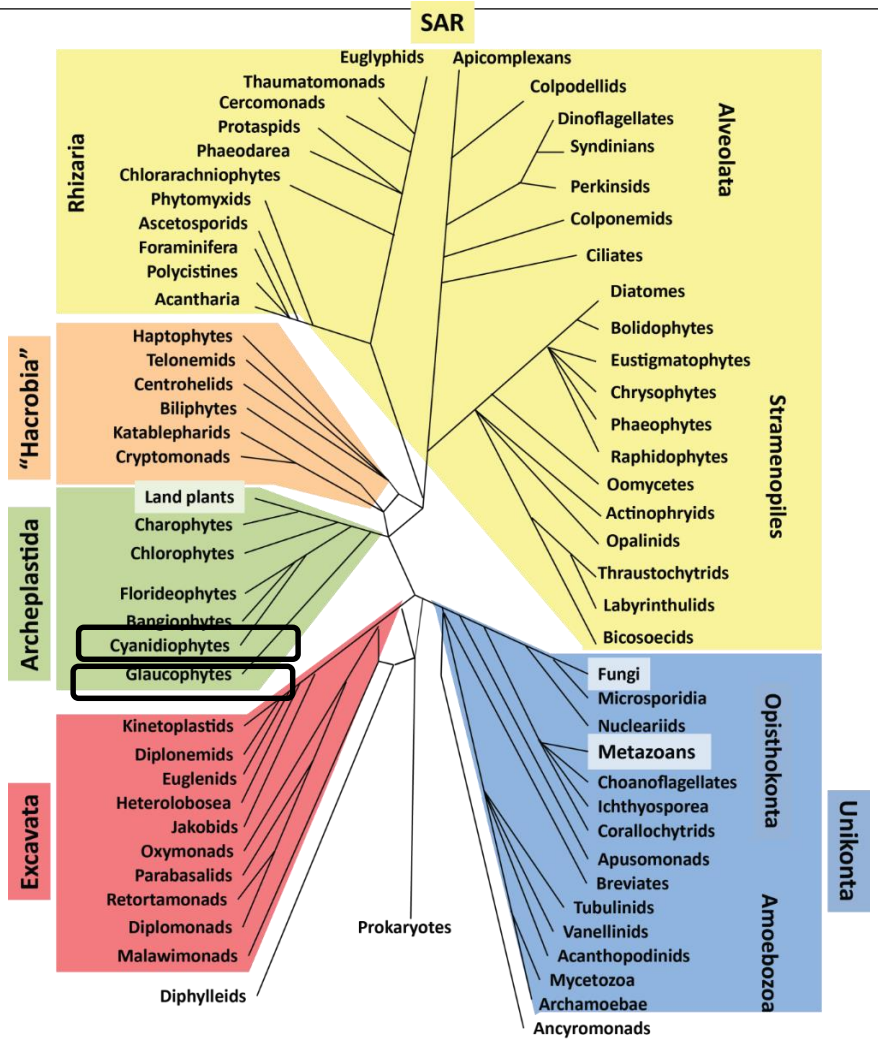
- *Euglena gracilis* (euglenid)
- Had a lot of contamination issues
- Genome is 1.5Gb and is in a horrible state – 250X PacBio assembly, yet still in 2M contigs
- Analysis has been in progress for a long time

- *Cyanophora paradoxa* (glaucophyte)



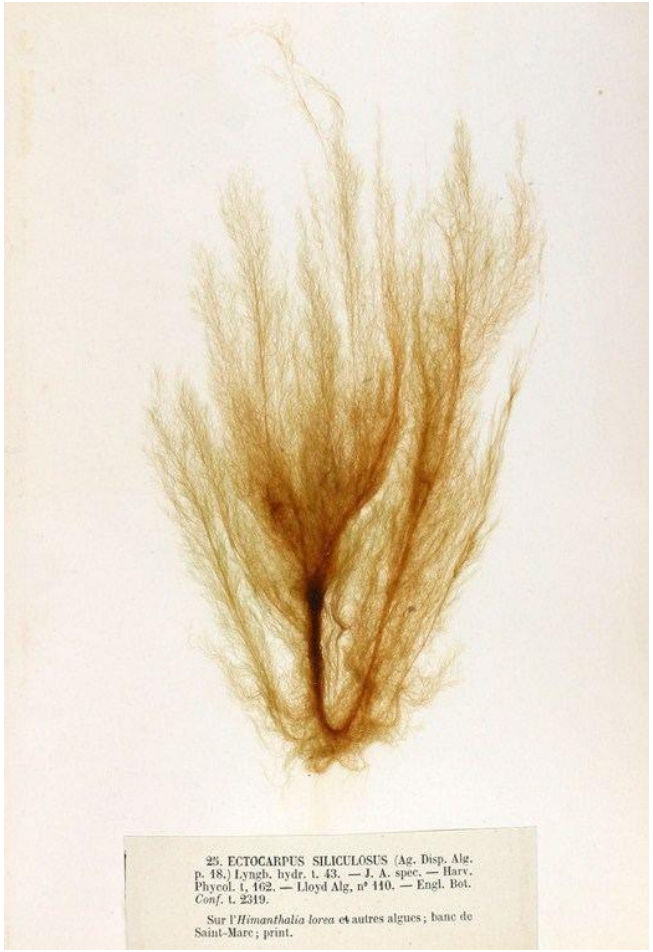
- *Galdieria sulphuraria* (rhodophyte)

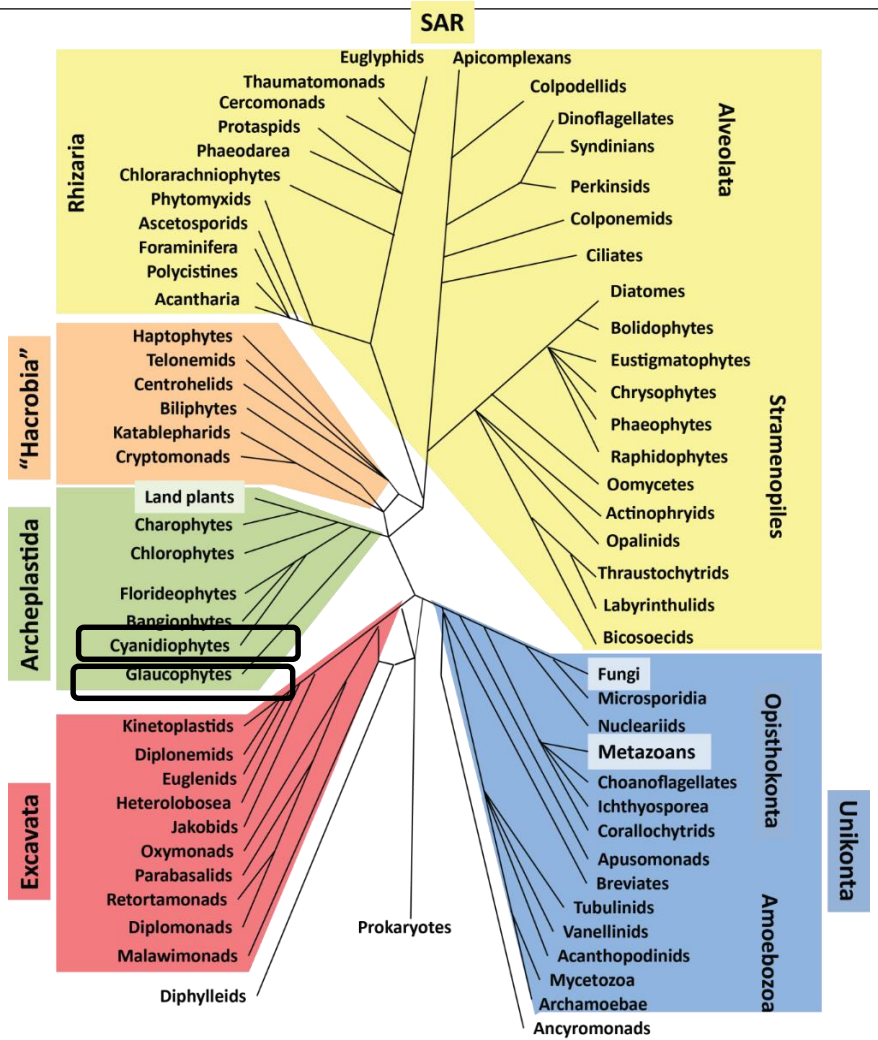




- Generated what seems to be good Hi-C for both of these
- Unfortunately, once again it turned out the strains were not what they were supposed to be
- These are a 14 Mbp and a 70 Mbp genomes so they should be fairly easily rescuable

- *Ectocarpus siliculosus* (multicellular phaeophyte)

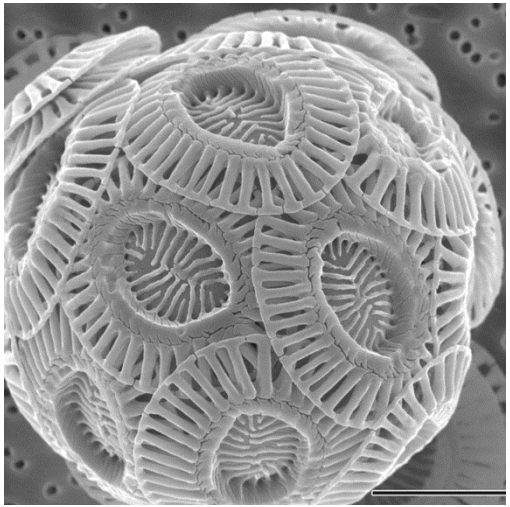


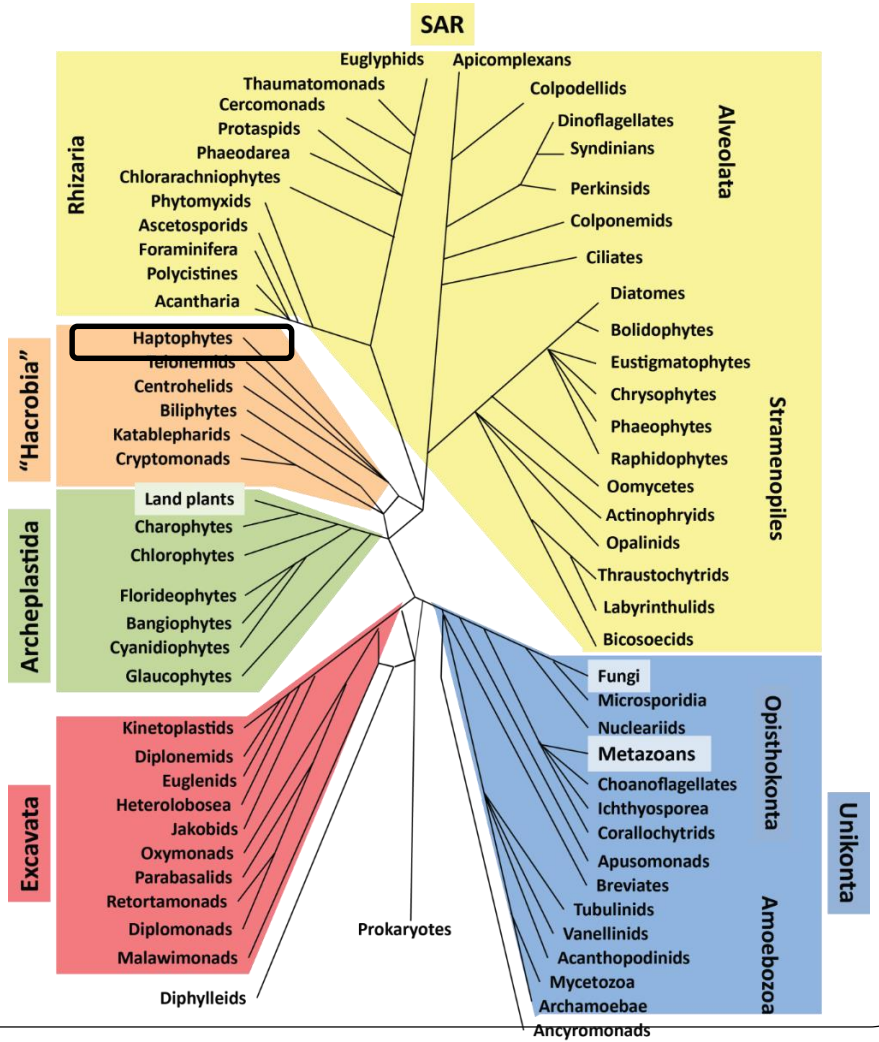


- Generated what seems to be good Hi-C
- The strain is also not the genomic one

- *Emiliania huxley* (haptophyte/coccolithophoride)

- - did not grow ☹





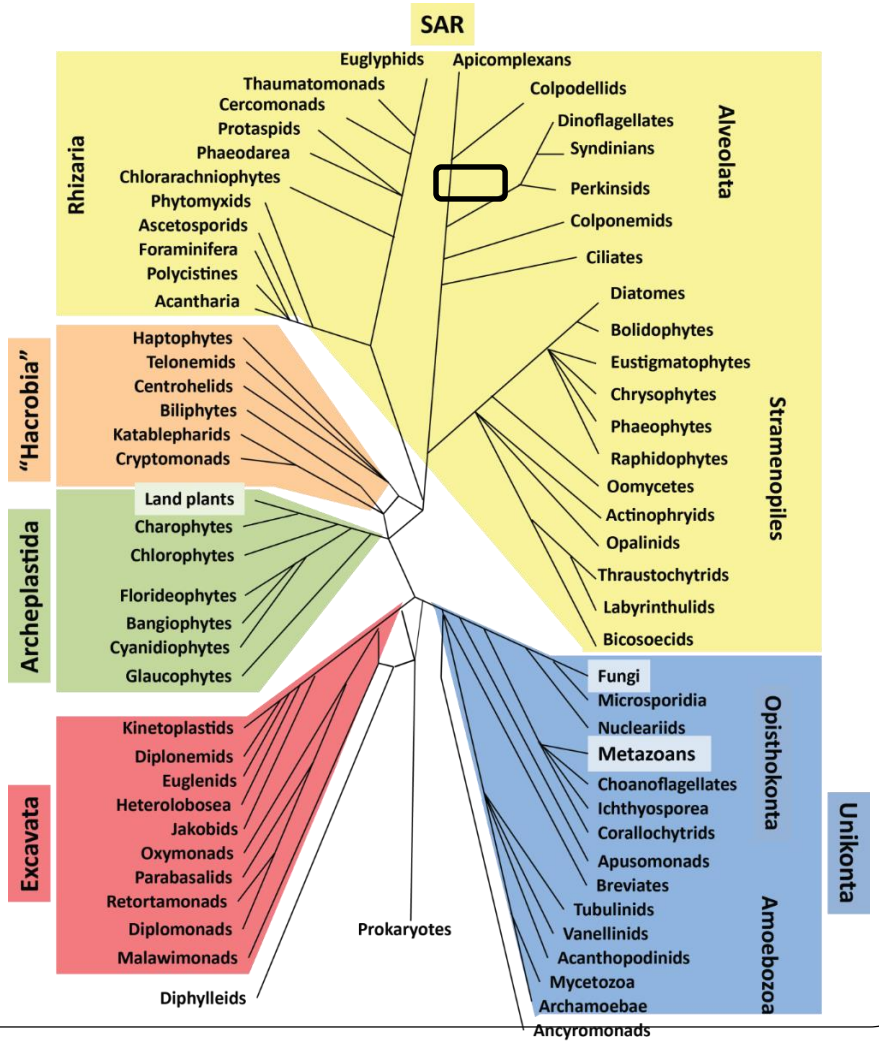
- *Euplotes* (ciliate with nanochromosomes!)

- Also did not grow well, and it appears it never will in axenic condition

- Will have to bite the bullet and eventually do *Oxytricha*

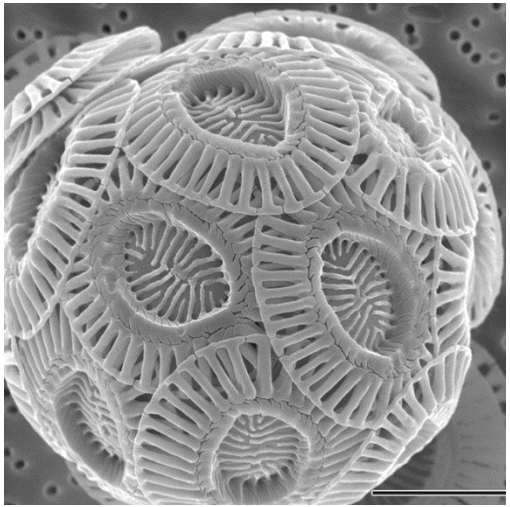
- *Chromera*

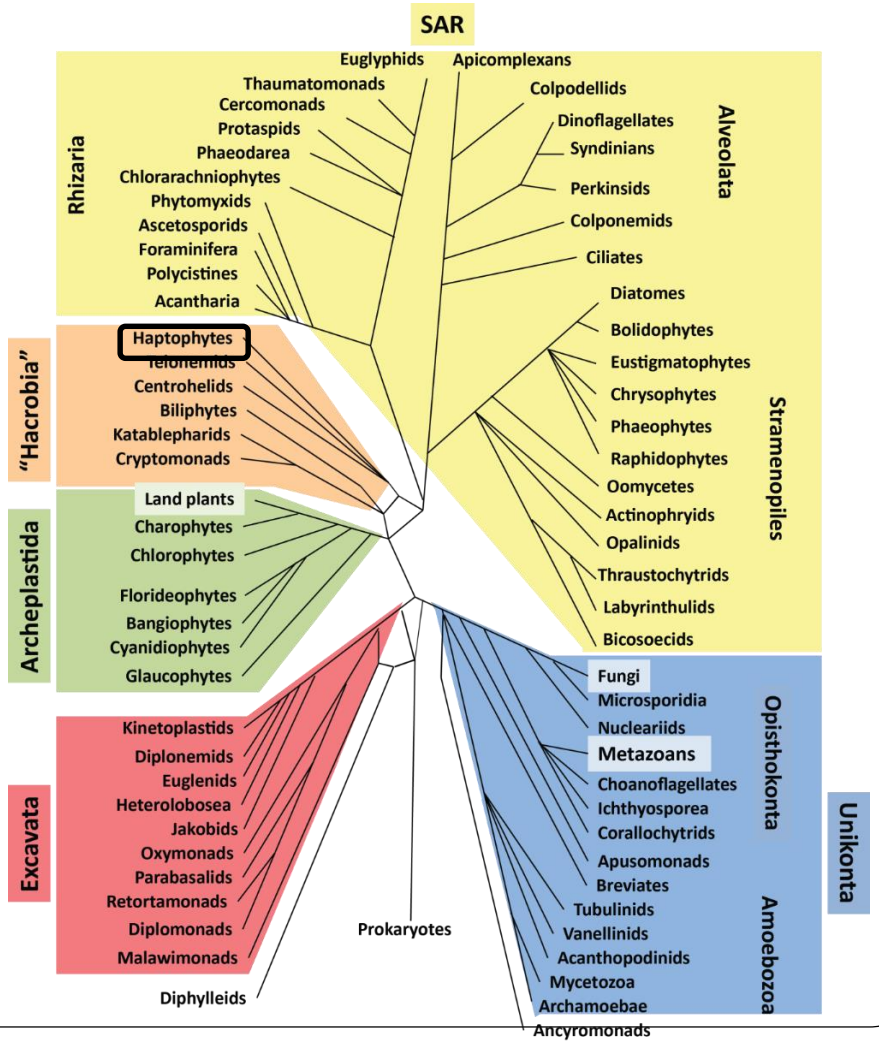
- ordered genomic strain from CCMP



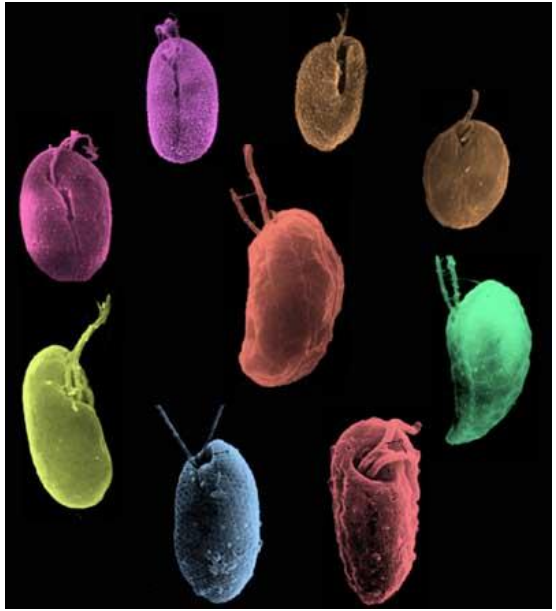
- *Emiliania huxley* (haptophyte/coccolithophoride)

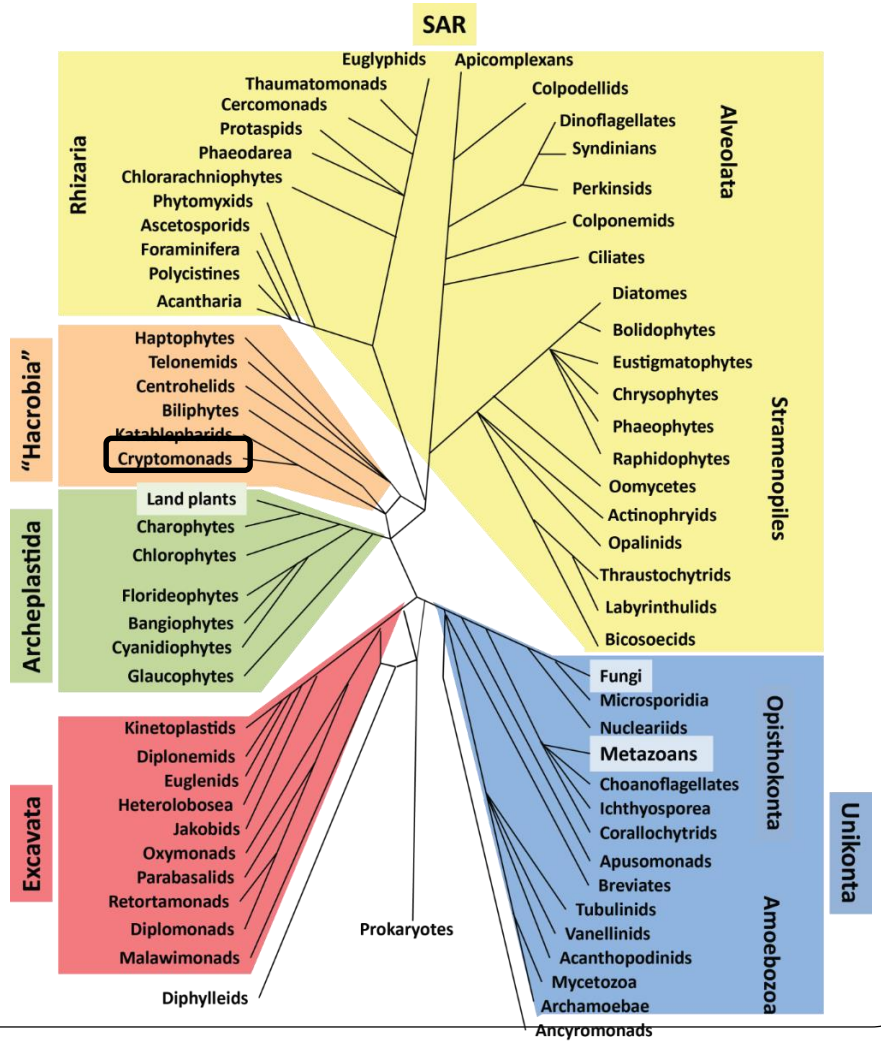
- reordered genomic strain from CCMP (previously sourced it from UTEX)





- *Guillardia theta*
- Has nucleomorph!
- <u>ordered genomic strain from CCMP</u>

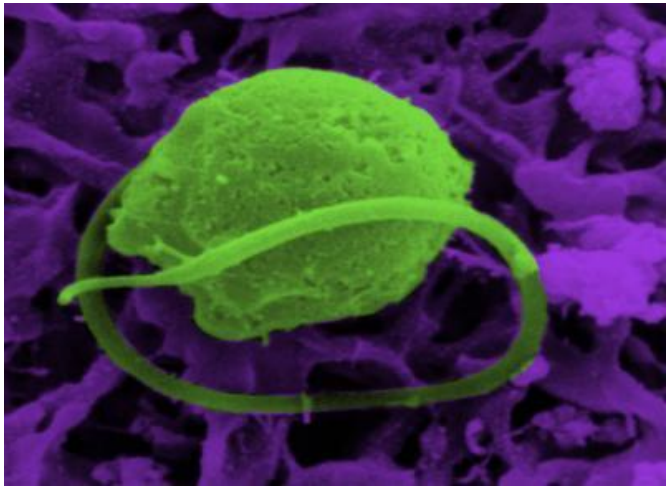


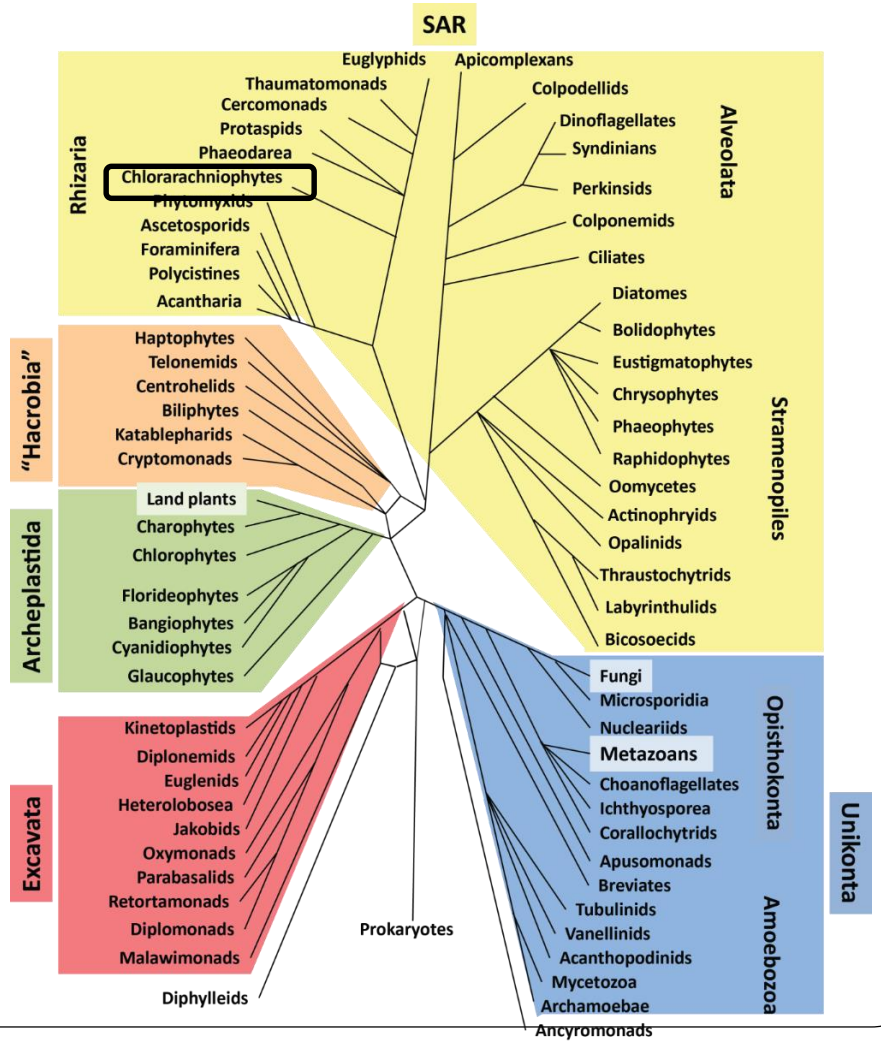


- *Bigelowiella natans*

- Has nucleomorph!

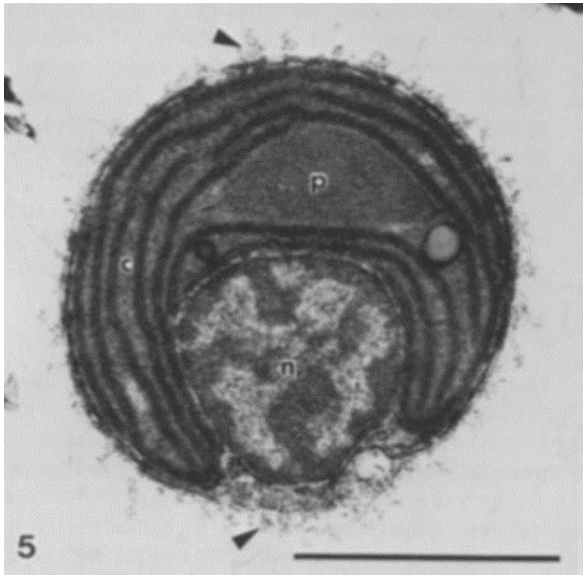
- ordered genomic strain from CCMP
- (will only arrive in January)

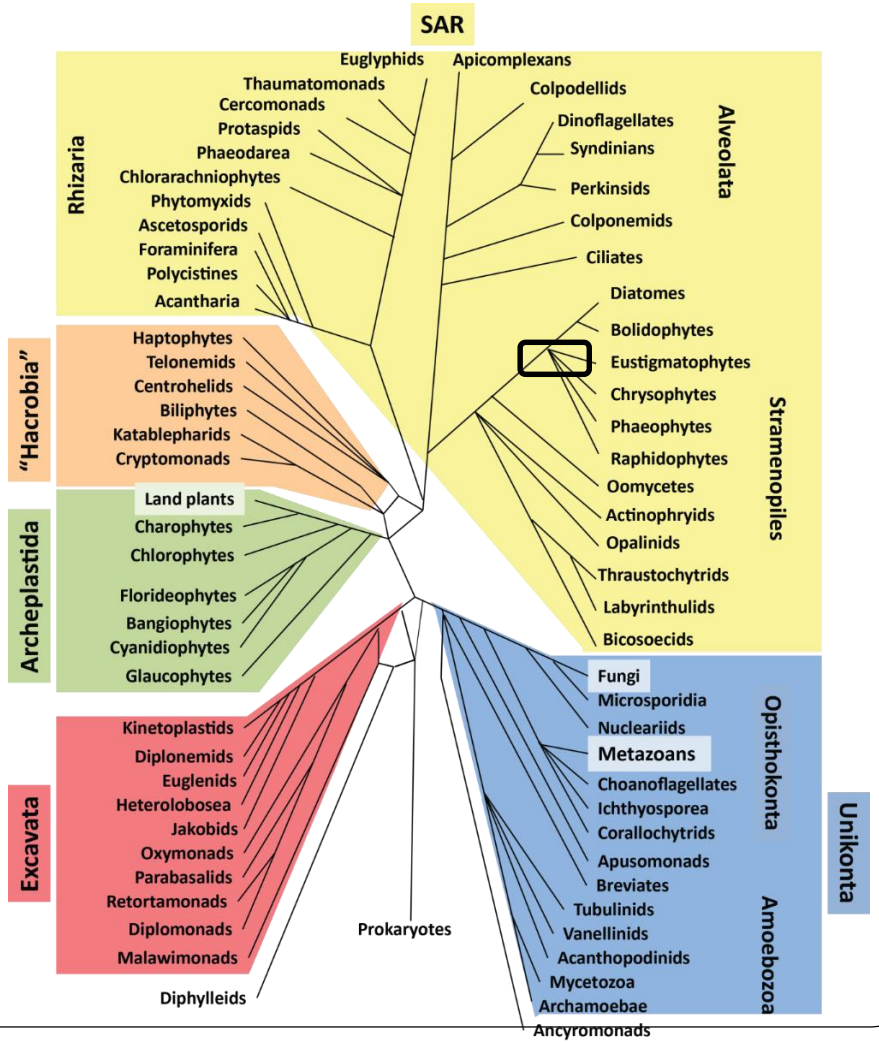




- *Aureococcus*

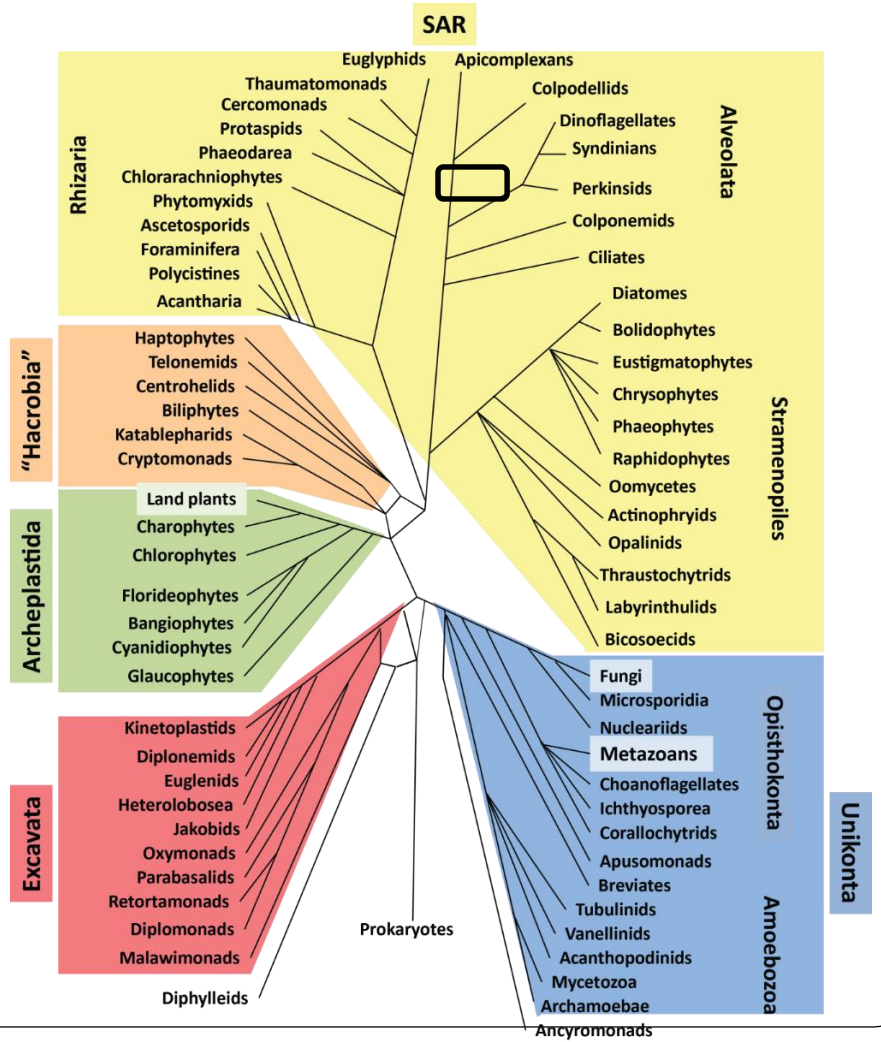
- ordered genomic strain from CCMP





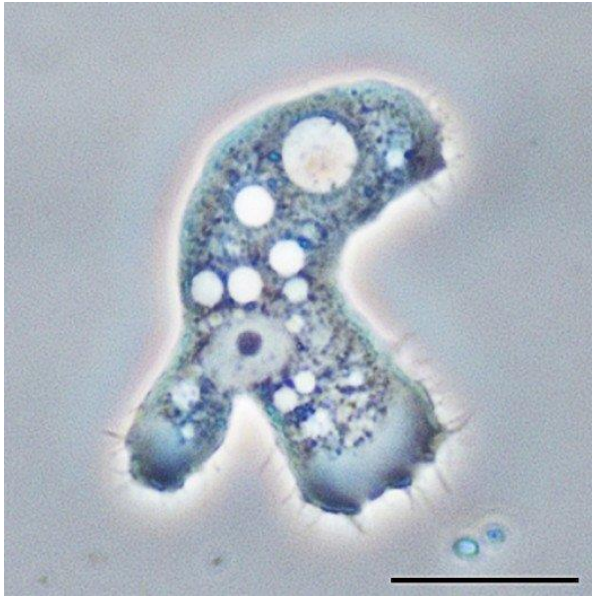
- *Chromera*

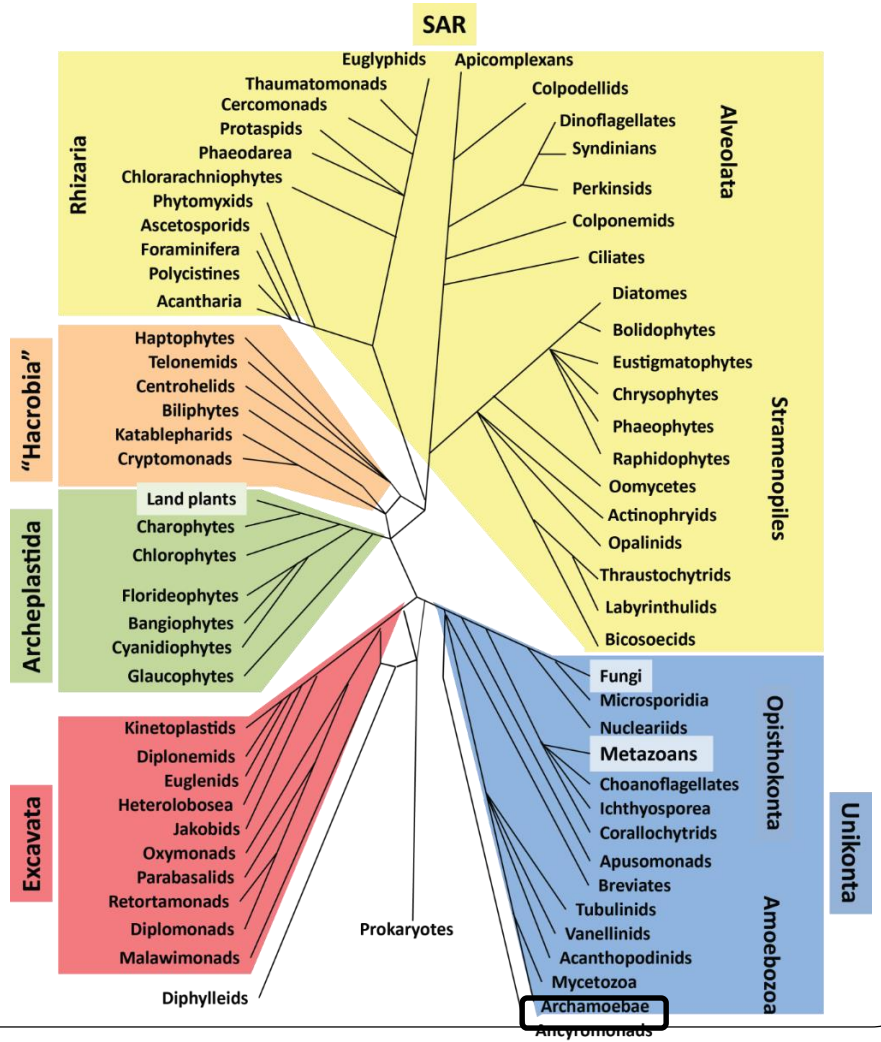
- ordered genomic strain from CCMP



- *Acanthamoeba*

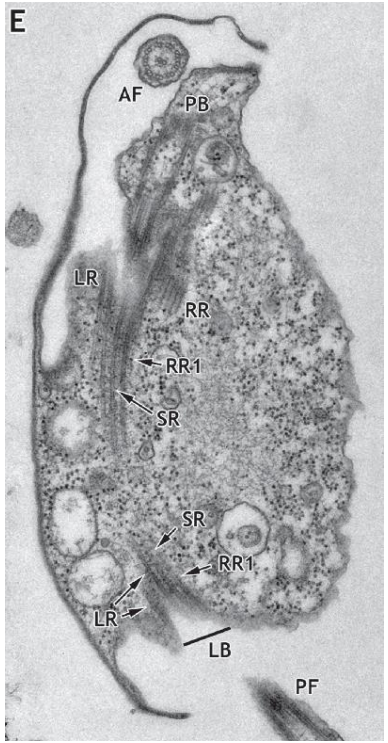
- To be ordered

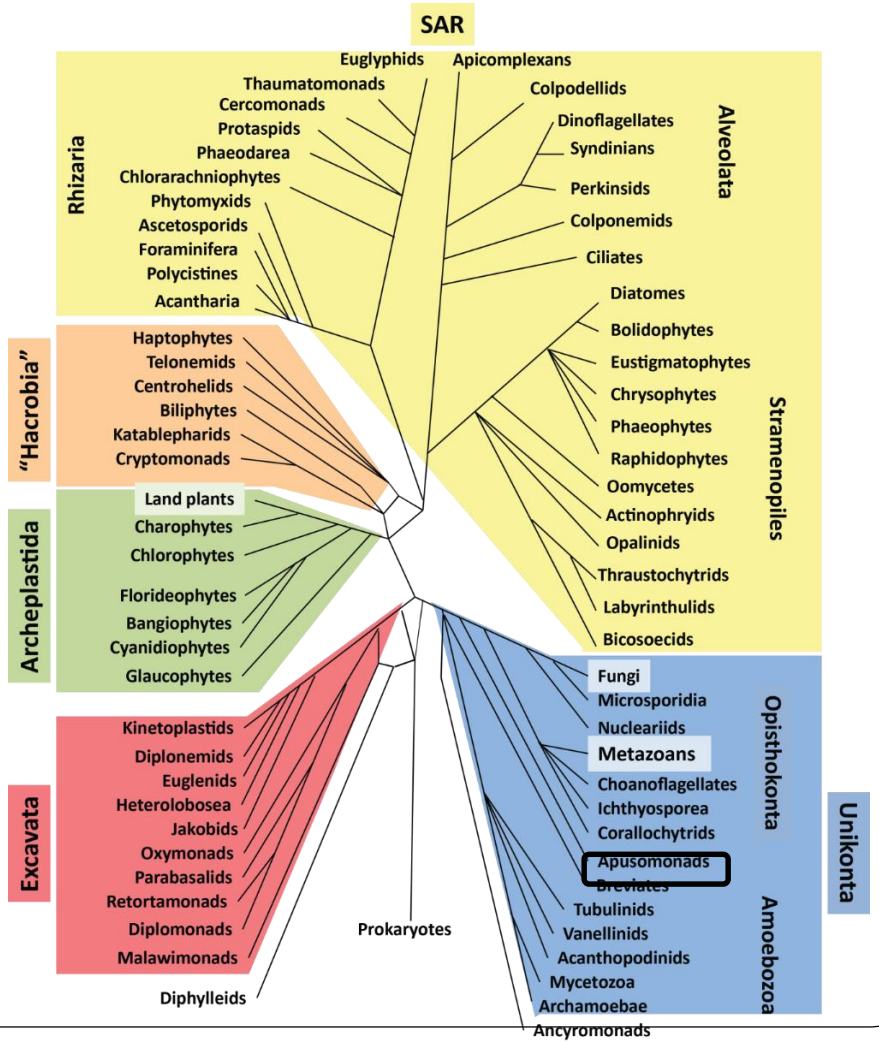




- *Thecamonas*

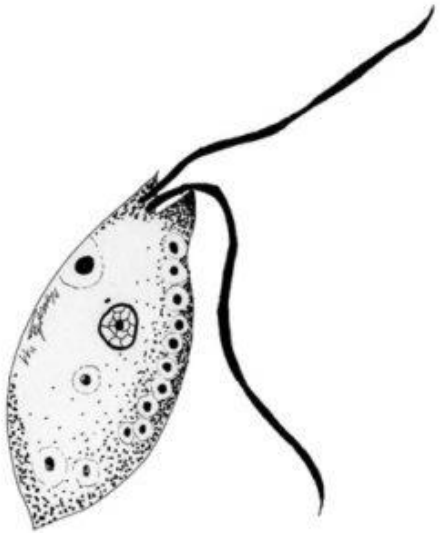
- To be ordered

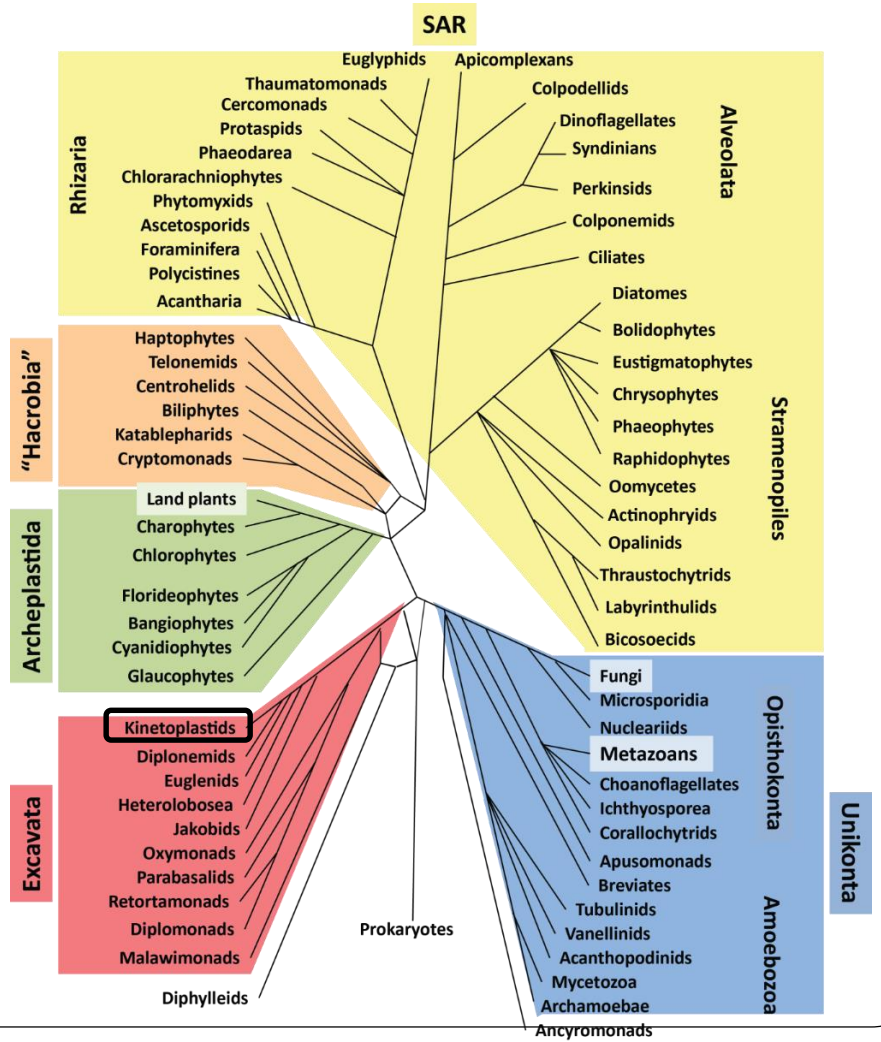




- *Bodo*

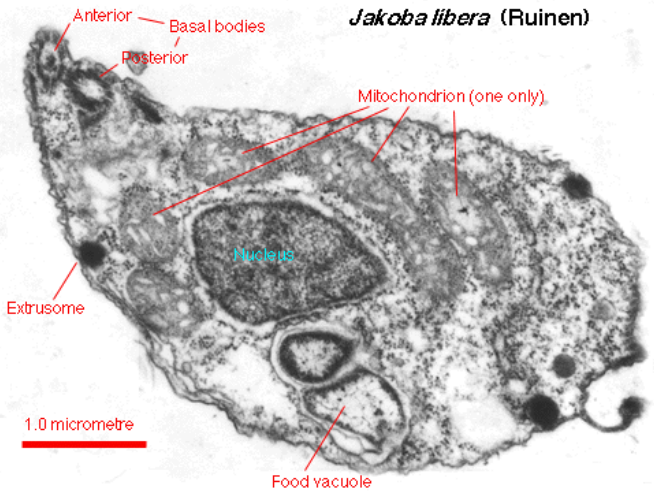
- To be ordered

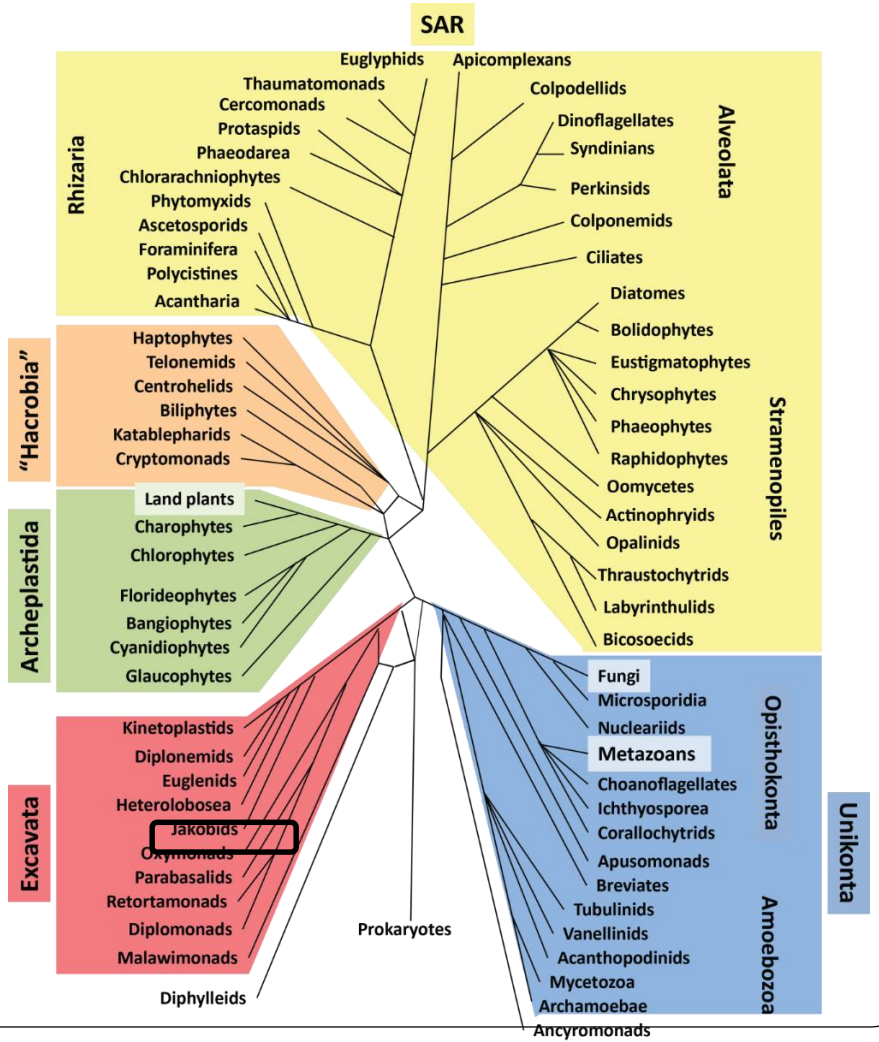




- *Jakoba*

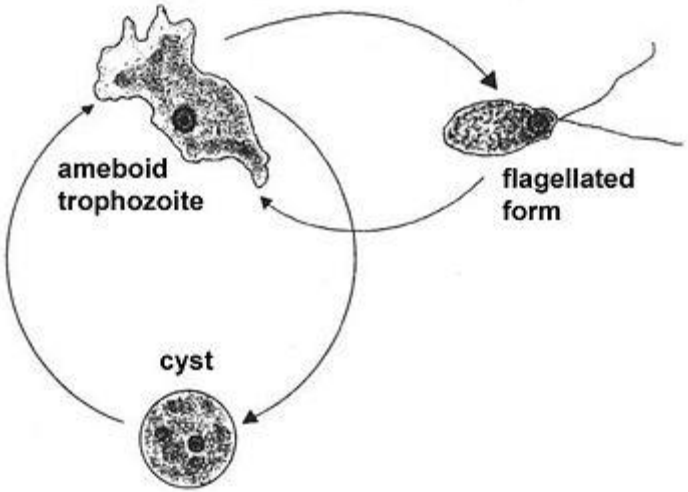
- To be ordered

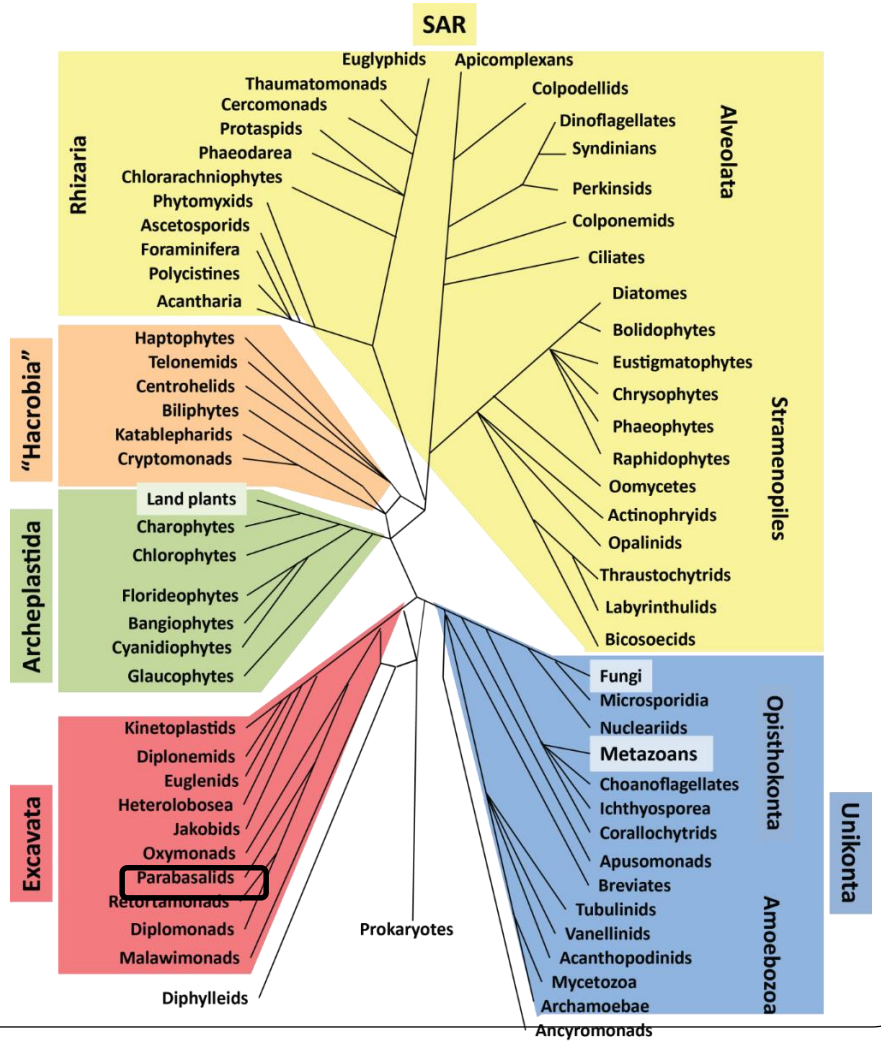




- *Naegleria*

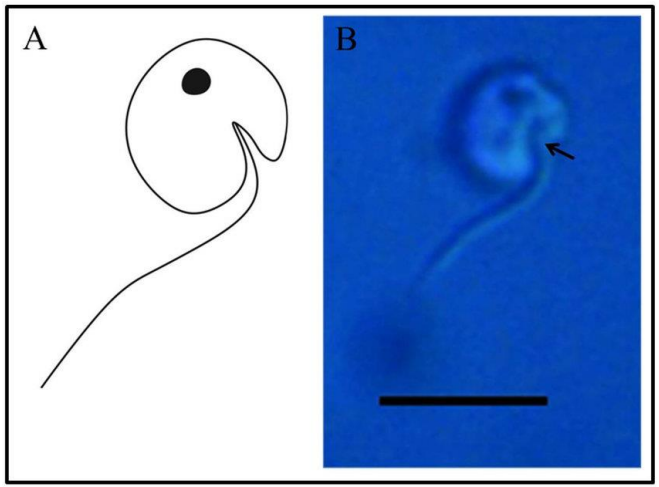
- To be ordered

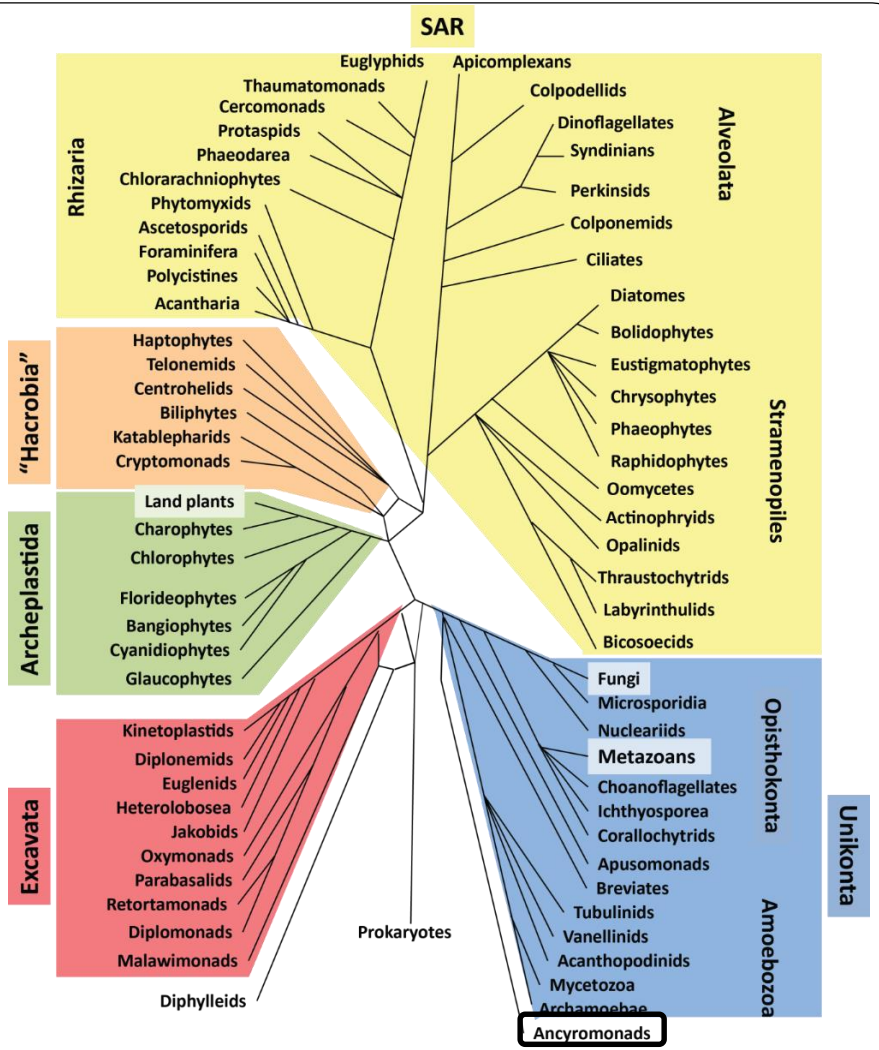




- *Nutomonas*

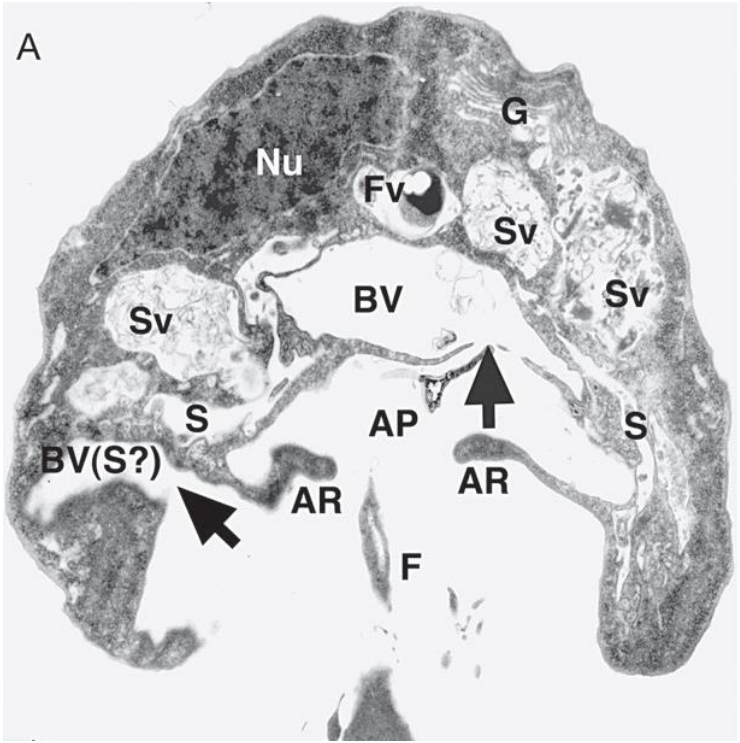
- To be ordered

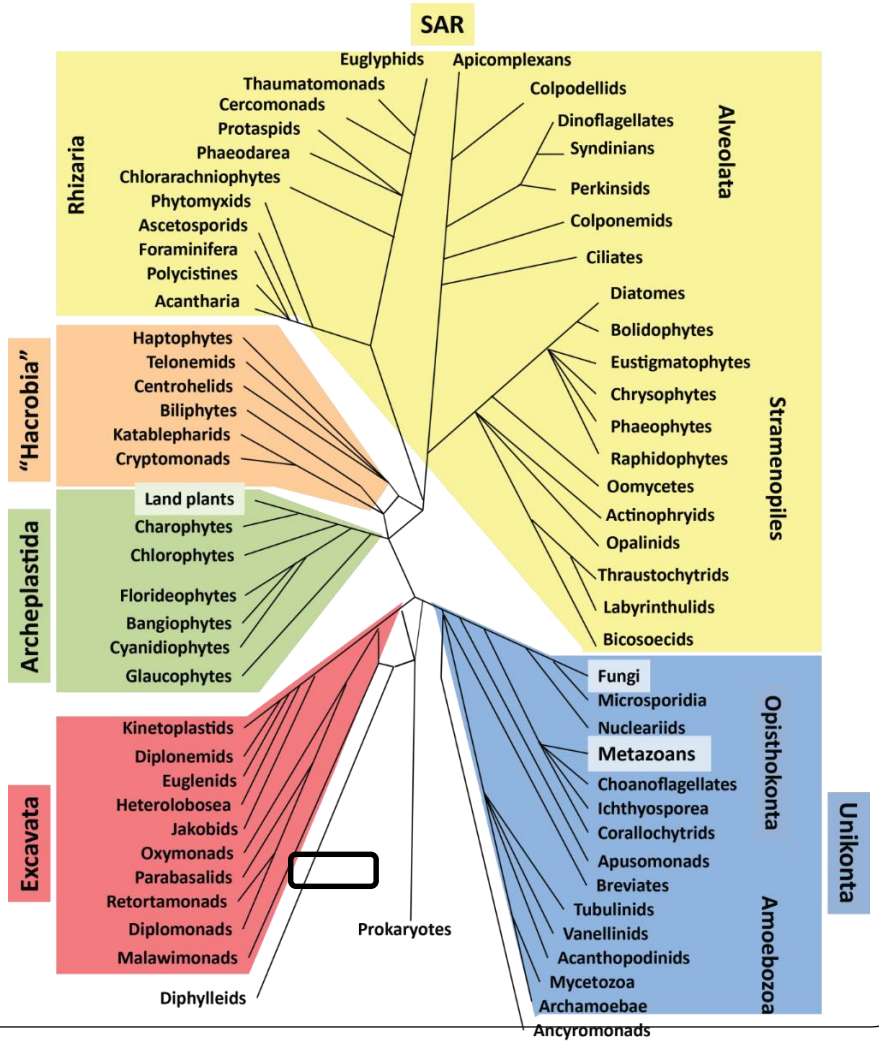




- *Rigifila*

- To be ordered





1.9. ORIGIN OF TADs IN METAZOANS

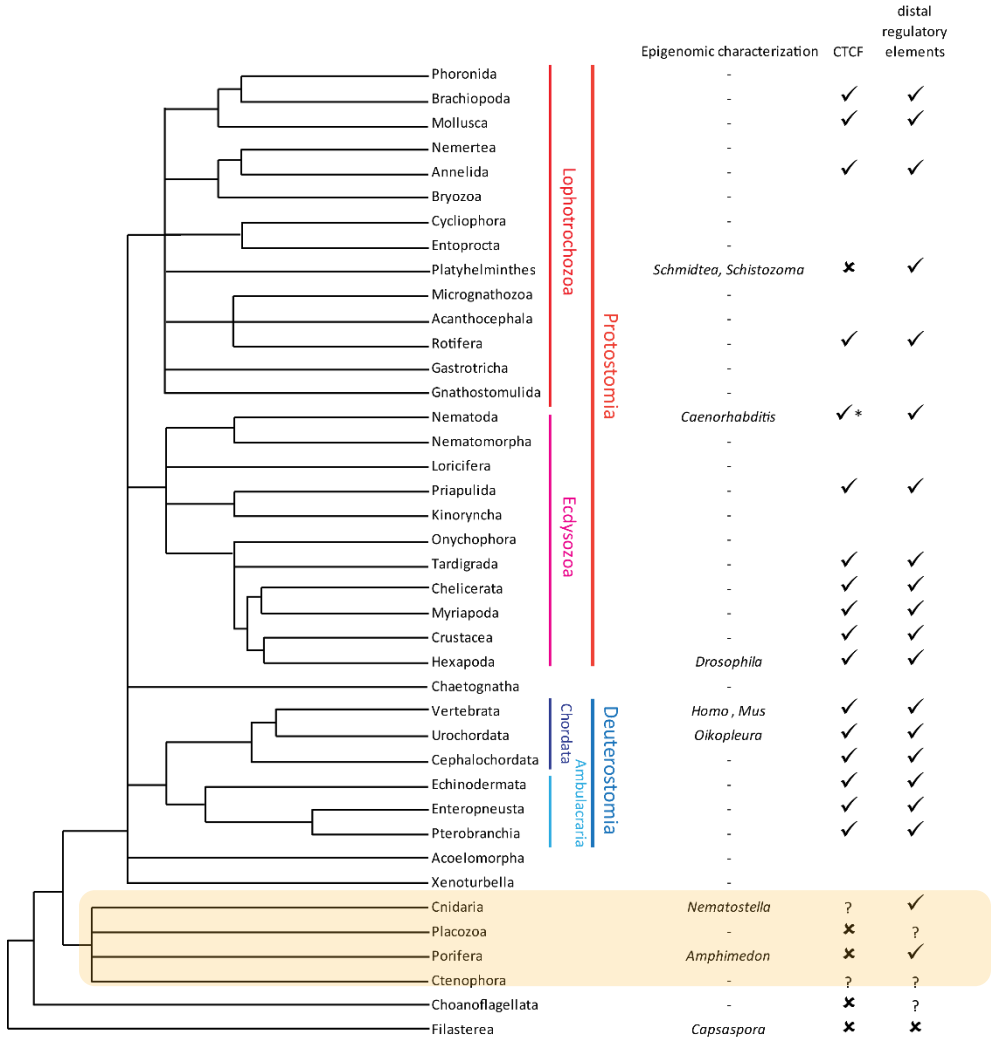
- Compartments (driven by chromatin state) seem to be the fundamental component of folding and Hi-C maps

- Mammals also have TADs and CTCF-mediated loops

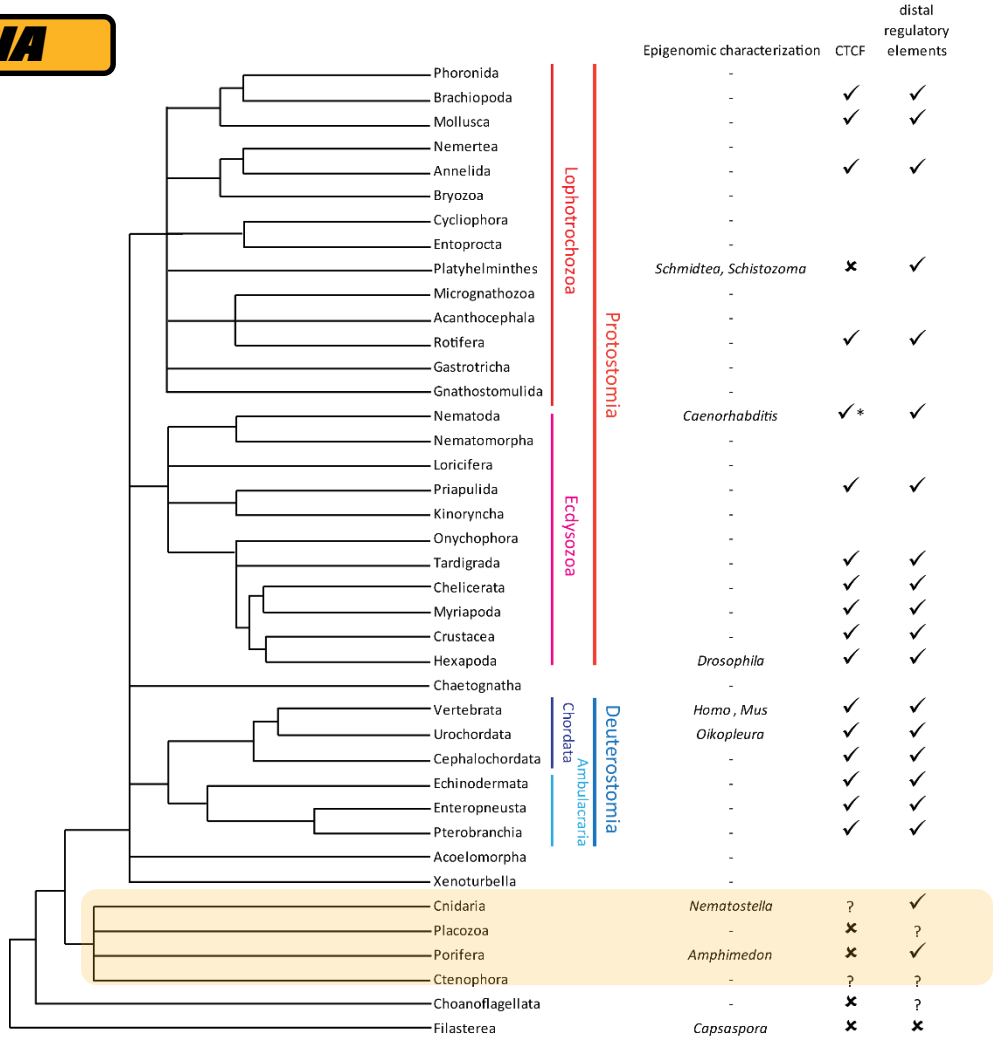
- *Drosophila* has topological domains too, but no loops associated with them (it has other kinds of loops), even though it has multiple insulator proteins

- *C. elegans* has no CTCF and no TADs on the autosomes; on chrX it has TAD-like domains, the boundaries of which are specified by the DCC (dosage compensation complex)

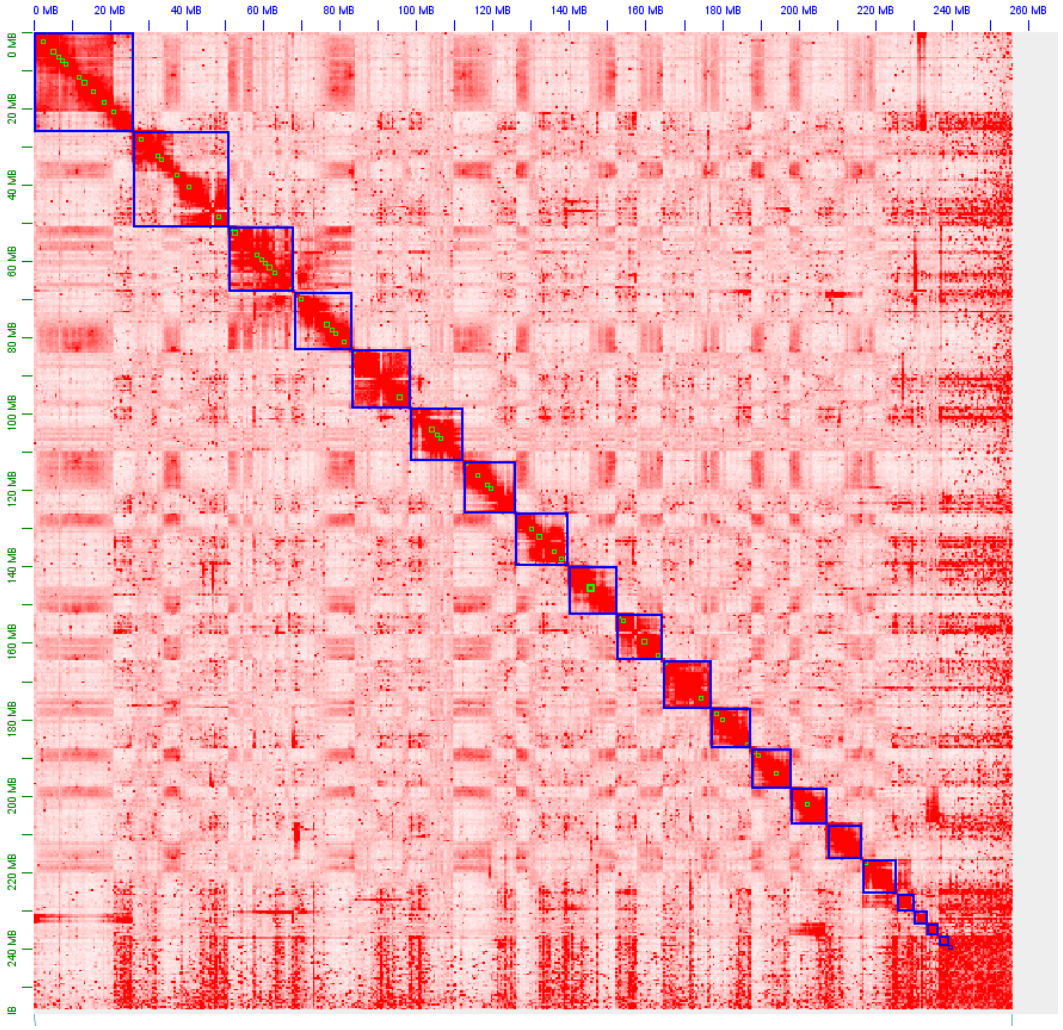
- What about other metazoans?


Epigenomic characterization
CTCF
distal regulatory elements
Phoronida
Brachiopoda
Mollusca
Nemertea
Annelida
Bryozoa
Cycliophora
Entoprocta
Platyhelminthes
Micrognathozoa
Acanthocephala
Rotifera
Gastrotricha
Gnathostomulida
Nematoda
Nematomorpha
Loricifera
Priapulida
Kinoryncha
Onychophora
Tardigrada
Chelicerata
Myriapoda
Crustacea
Hexapoda
Chaetognatha
Vertebrata
Urochordata
Cephalochordata
Echinodermata
Enteropneusta
Pterobranchia
Acoelomorpha
Xenoturbella
Cnidaria
Placozoa
Porifera
Ctenophora
Choanoflagellata
Filasterea
Lophotrochozoa
Ecdysozoa
Protostomia
Chordata
Ambulacraria
Deuterostomia
Schmidtea, Schistozoma
Caenorhabditis
Drosophila
Homo , Mus
Oikopleura
Nematostella
Amphimedon
Capsaspora

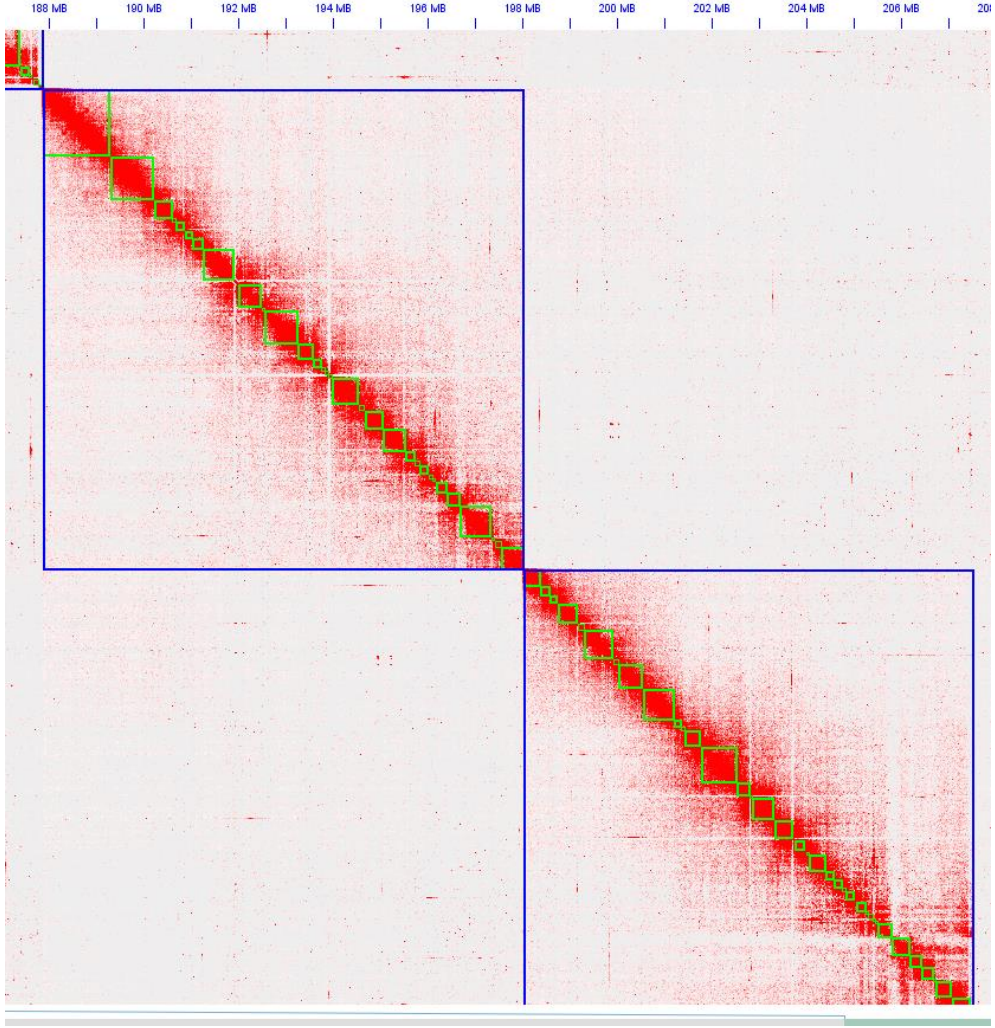
HI-C IN *AIPTASIA*

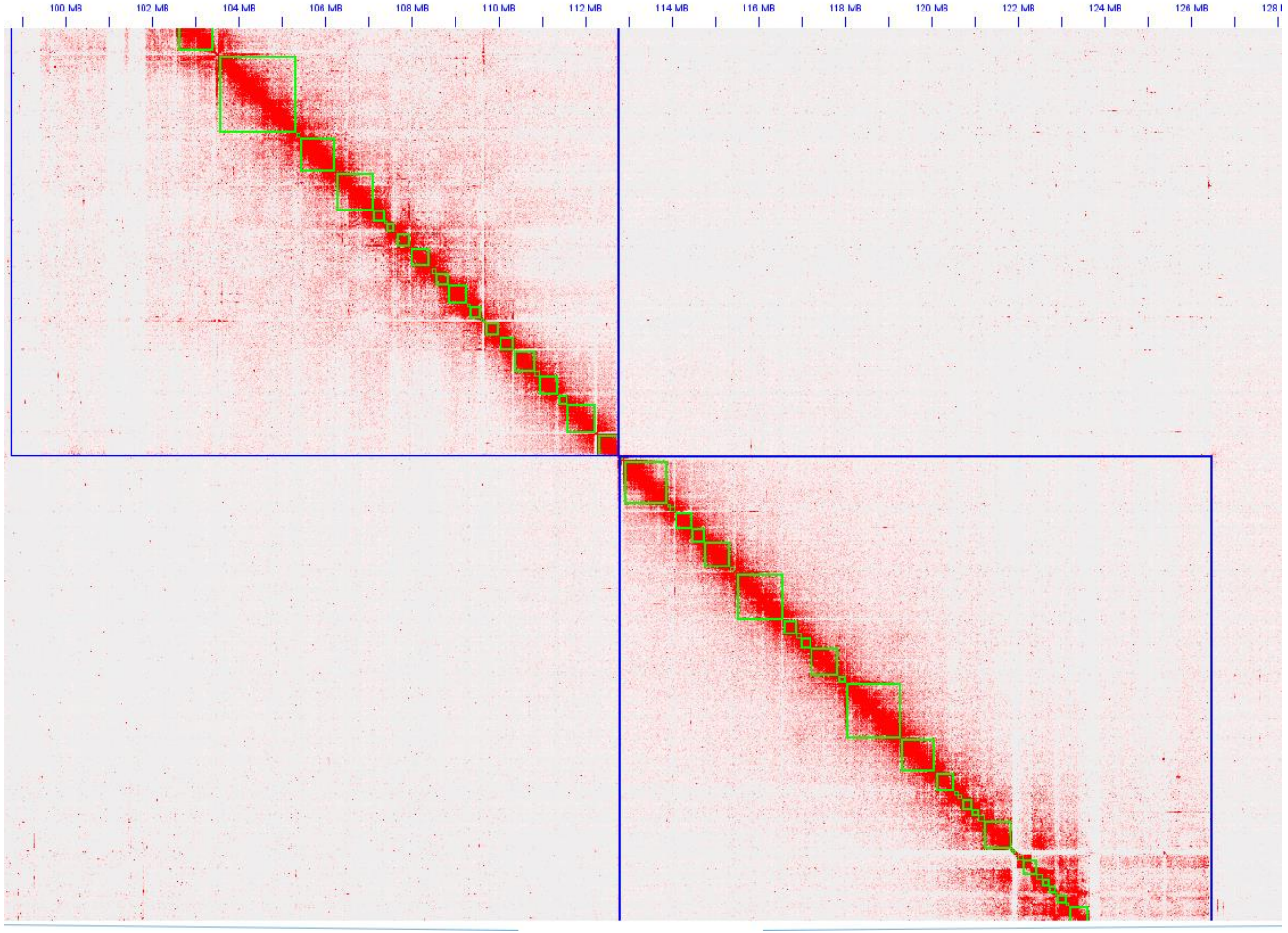


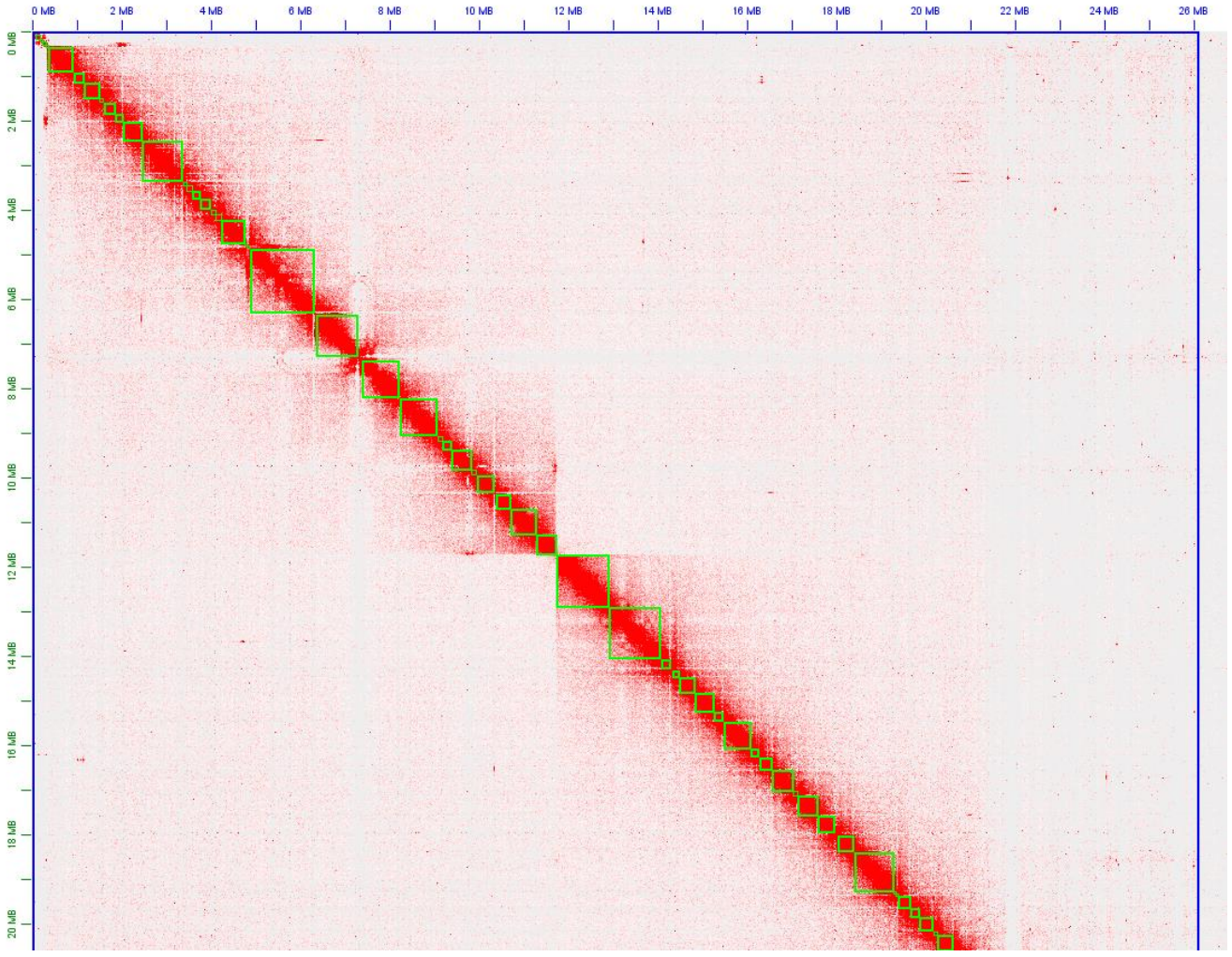
Scaffolding (200 rounds)



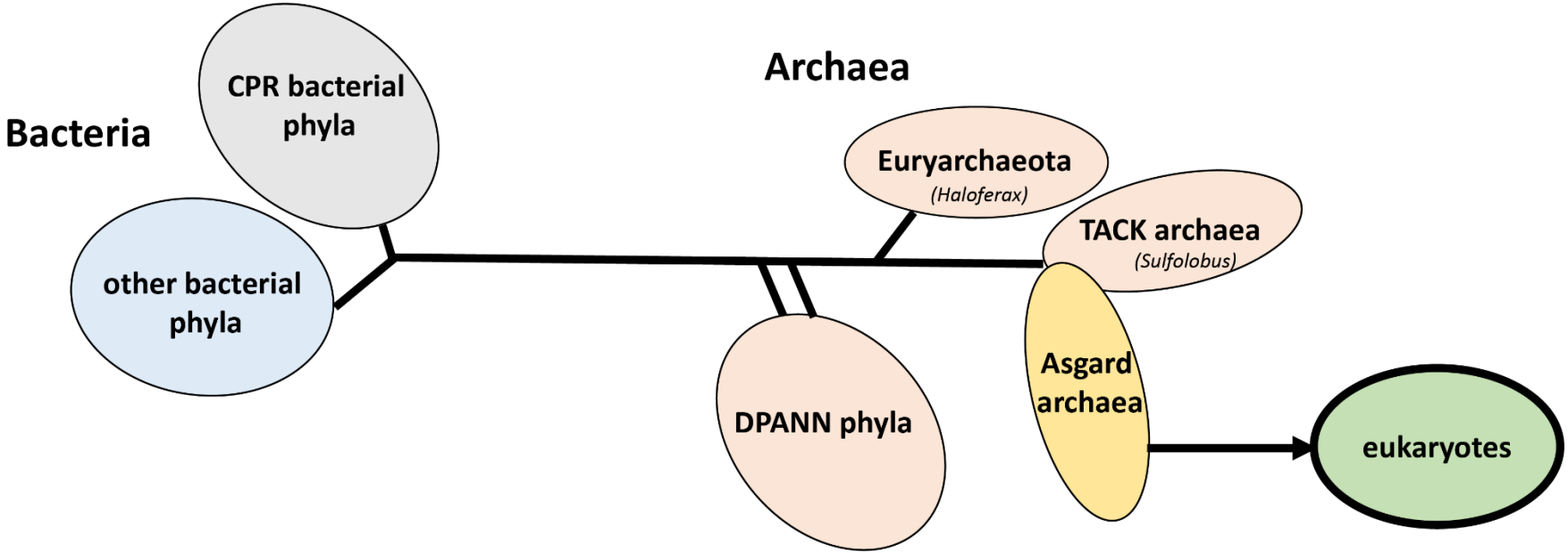
Local look:


188 MB
190 MB
192 MB
194 MB
196 MB
198 MB
200 MB
202 MB
204 MB
206 MB


100 MB
102 MB
104 MB
106 MB
108 MB
110 MB
112 MB
114 MB
116 MB
118 MB
120 MB
122 MB
124 MB
126 MB


0 MB
2 MB
4 MB
6 MB
8 MB
10 MB
12 MB
14 MB
16 MB
18 MB
20 MB
22 MB
24 MB
26 MB

1.10. ARCHAEAL CHROMATIN


Bacteria
CPR bacterial phyla
other bacterial phyla
Archaea
Euryarchaeota
(Haloferax)
TACK archaea
(Sulfolobus)
DPANN phyla
Asgard archaea
eukaryotes

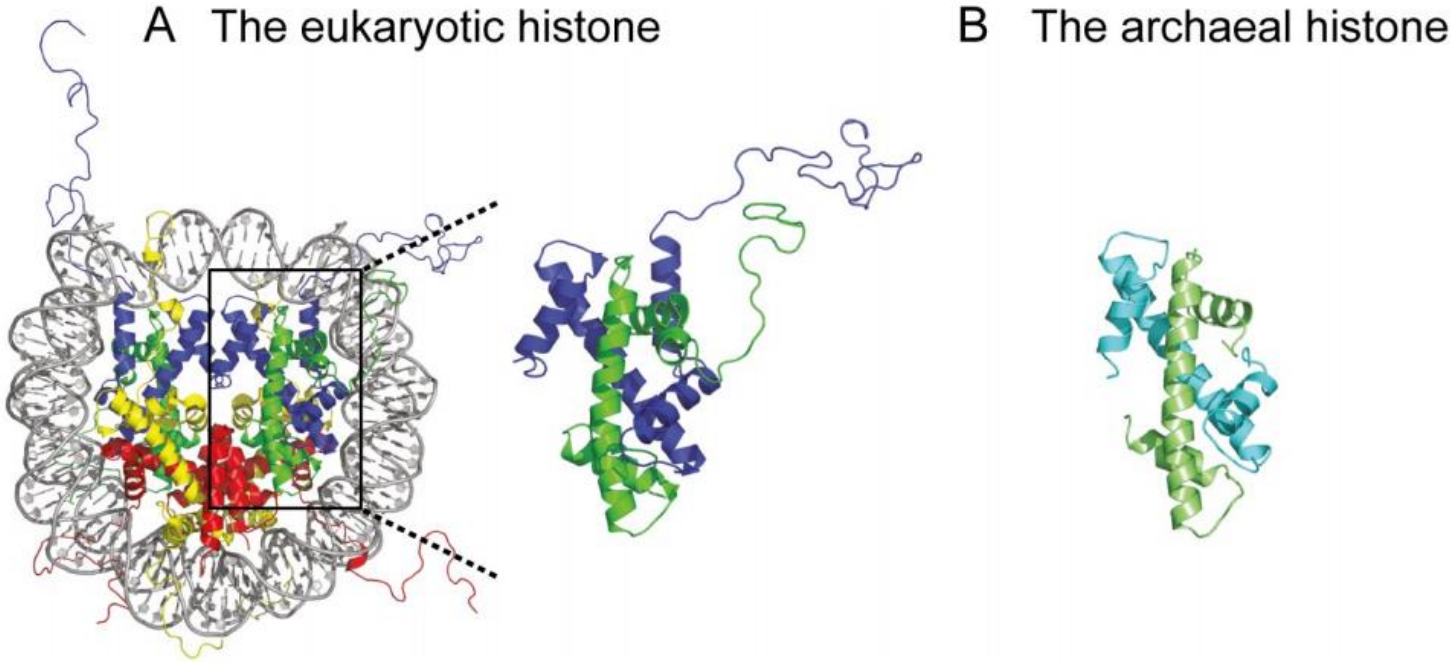


Fig 1. Eukaryotic and archaeal histones. (A) Eukaryotic nucleosome consisting of DNA wrapped around a core of a $(H3\text{-}H4)_2$ tetramer and two H2A-H2B dimers. Yellow, H2A; red, H2B; blue, H3; green, H4. (B) Archaeal histone homodimer of HMfB. HMfB, Histone B from *Methanothermus fervidus*.

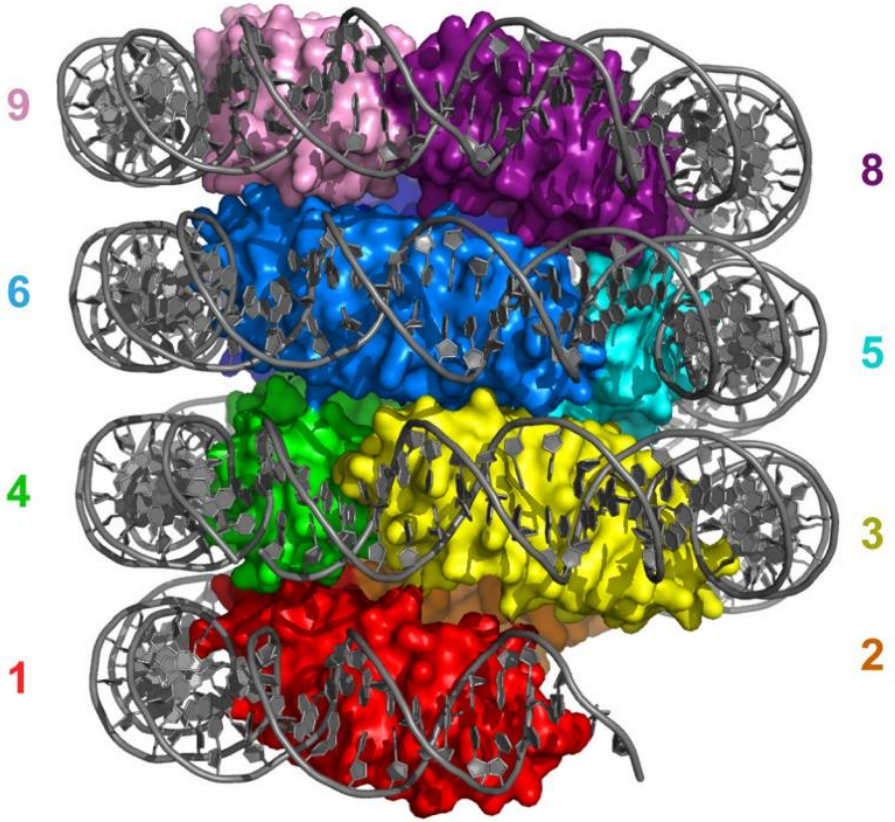
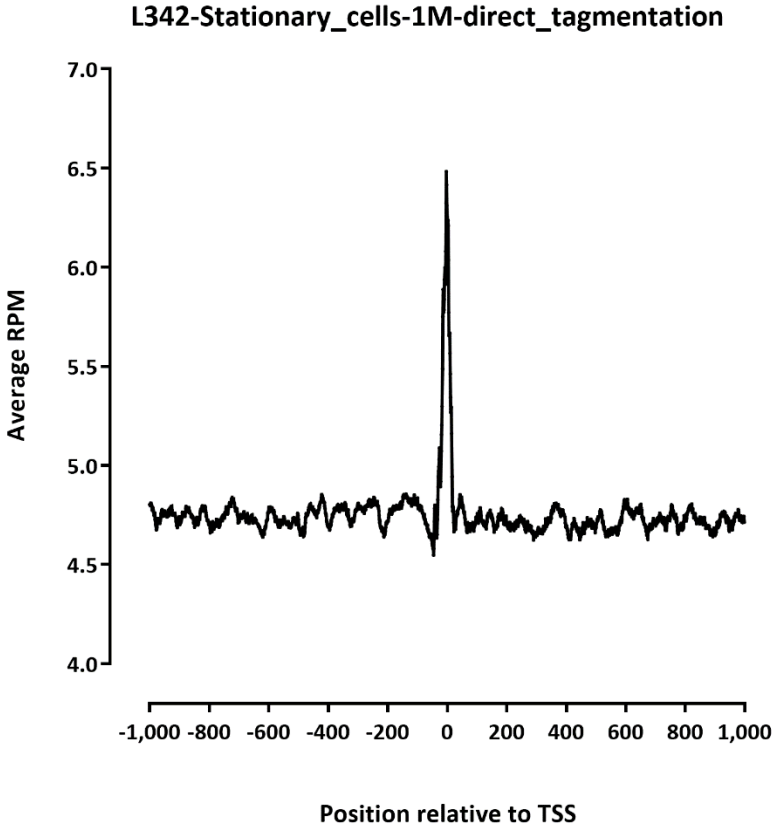
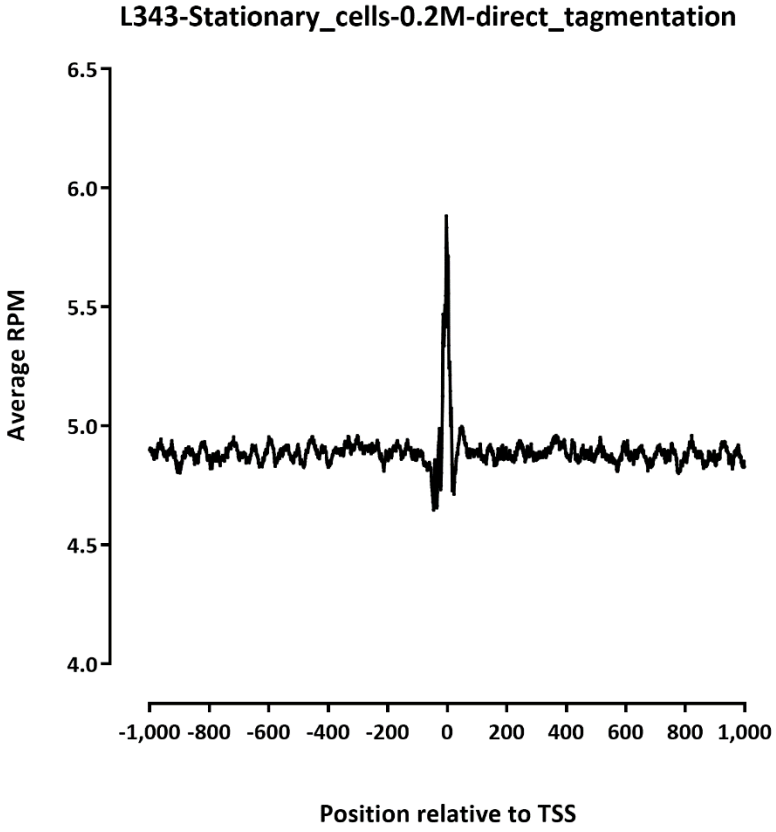


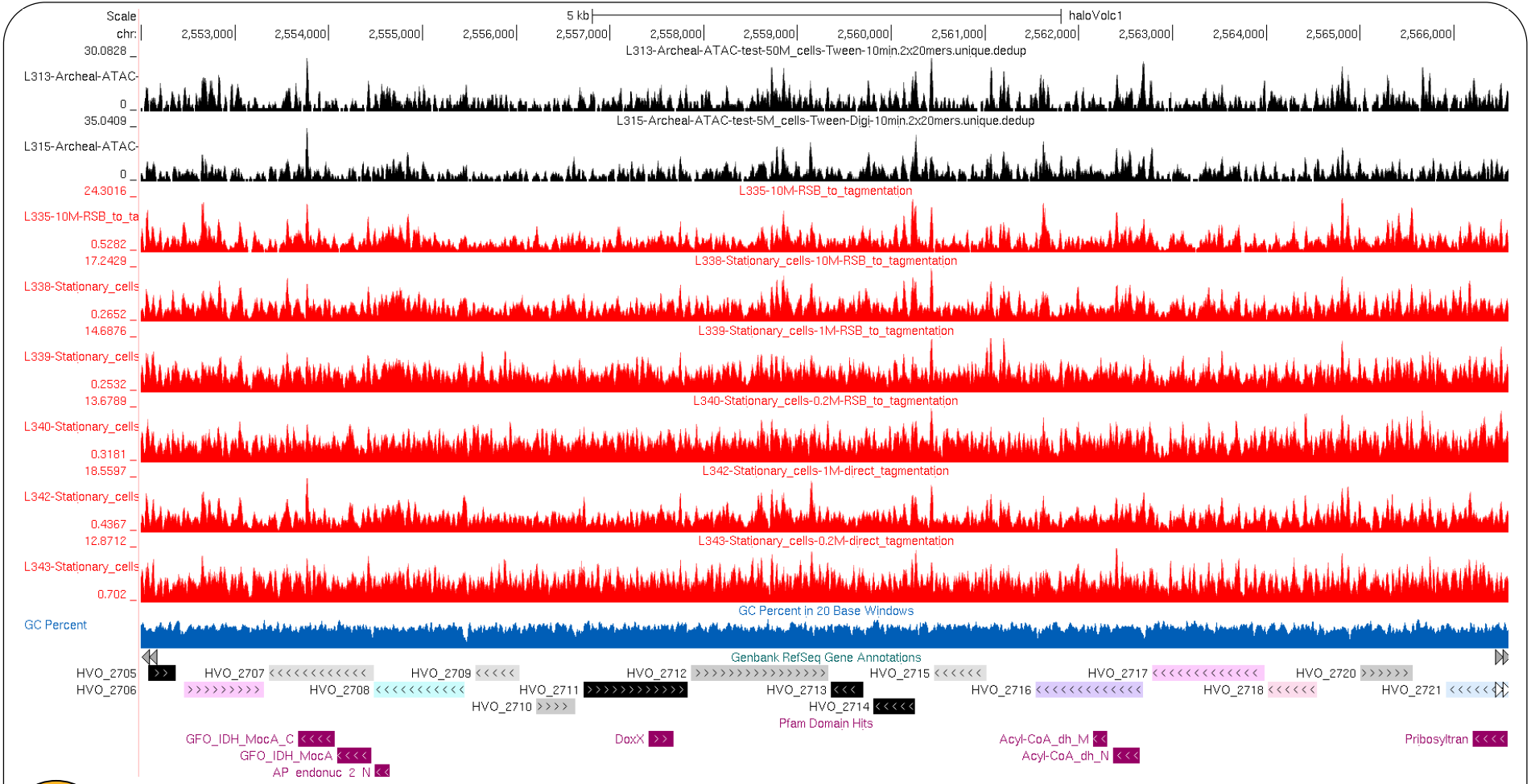
Fig 2. Overview of the hypernucleosome structure. HMfB dimers stack to form a continuous, central protein core that wraps the DNA in a left-handed superhelix. Nine HMfB dimers are shown, each dimer in surface mode and in rainbow colors. Numbering indicates position of the nine histone dimers; note that dimer 5 and 6 occlude the view of dimer 7. DNA is in gray and shown as cartoon. *Image generated using PDB entry 5T5K* [64]. HMfB, Histone B from

ATAC-SEQ

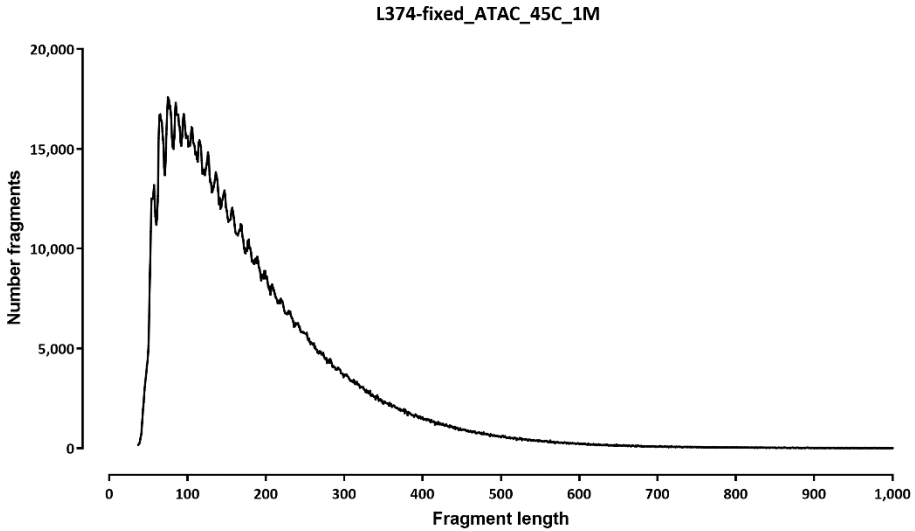
INITITAL EXPERIMENTS (NOT VERY SUCCESSFUL)

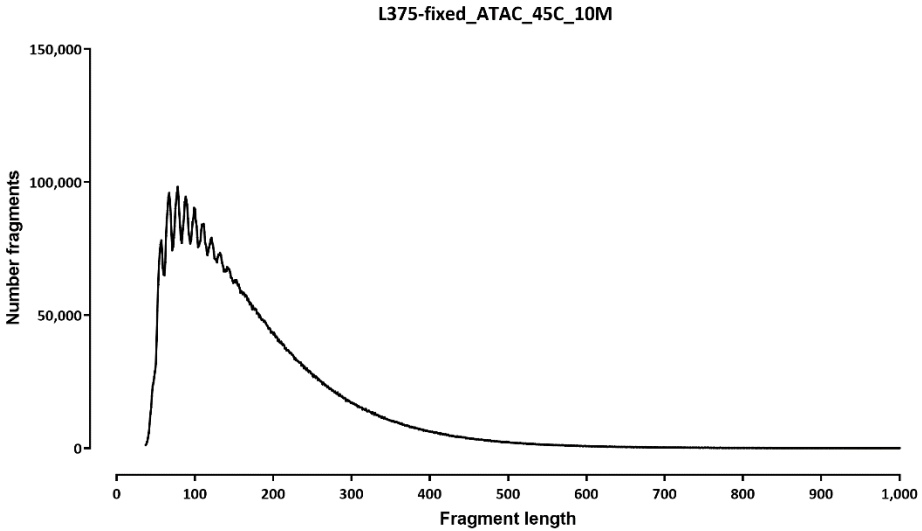




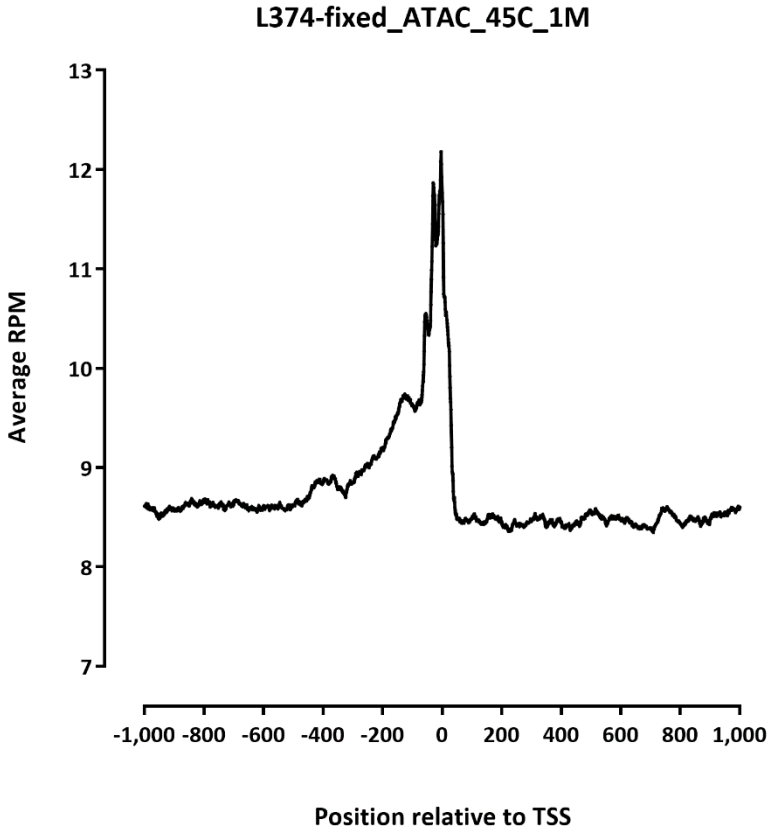

Scale
5 kb
haloVolc1
chr:
2,553,000
2,554,000
2,555,000
2,556,000
2,557,000
2,558,000
2,559,000
2,560,000
2,561,000
2,562,000
2,563,000
2,564,000
2,565,000
2,566,000
30.0828
L313-Archeal-ATAC-test-50M_cells-Tween-10min.2x20mers.unique.dedup
L313-Archeal-ATAC-
0
35.0409
L315-Archeal-ATAC-test-5M_cells-Tween-Digi-10min.2x20mers.unique.dedup
L315-Archeal-ATAC-
0
24.3016
L335-10M-RSB_to_tagmentation
L335-10M-RSB_to_ta
0.5282
17.2429
L338-Stationary_cells-10M-RSB_to_tagmentation
L338-Stationary_cells
0.2652
14.6876
L339-Stationary_cells-1M-RSB_to_tagmentation
L339-Stationary_cells
0.2532
13.6789
L340-Stationary_cells-0.2M-RSB_to_tagmentation
L340-Stationary_cells
0.3181
18.5597
L342-Stationary_cells-1M-direct_tagmentation
L342-Stationary_cells
0.4367
12.8712
L343-Stationary_cells-0.2M-direct_tagmentation
L343-Stationary_cells
0.702
GC Percent in 20 Base Windows
GC Percent
Genbank RefSeq Gene Annotations
HVO_2705
HVO_2706
HVO_2707
HVO_2708
HVO_2709
HVO_2710
HVO_2711
HVO_2712
HVO_2713
HVO_2714
HVO_2715
HVO_2716
HVO_2717
HVO_2718
HVO_2720
HVO_2721
Pfam Domain Hits
GFO_IDH_MocA_C
GFO_IDH_MocA
AP_endonuc_2_N
DoxX
Acyl-CoA_dh_M
Acyl-CoA_dh_N
Pribosyltran

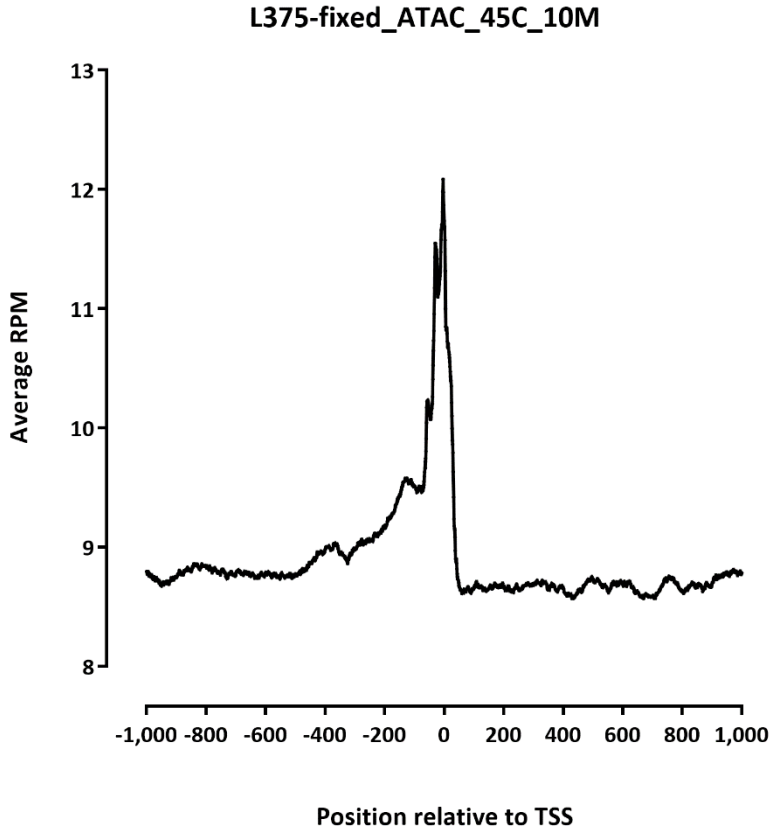
MOVING TO CROSSLINKED ATAC

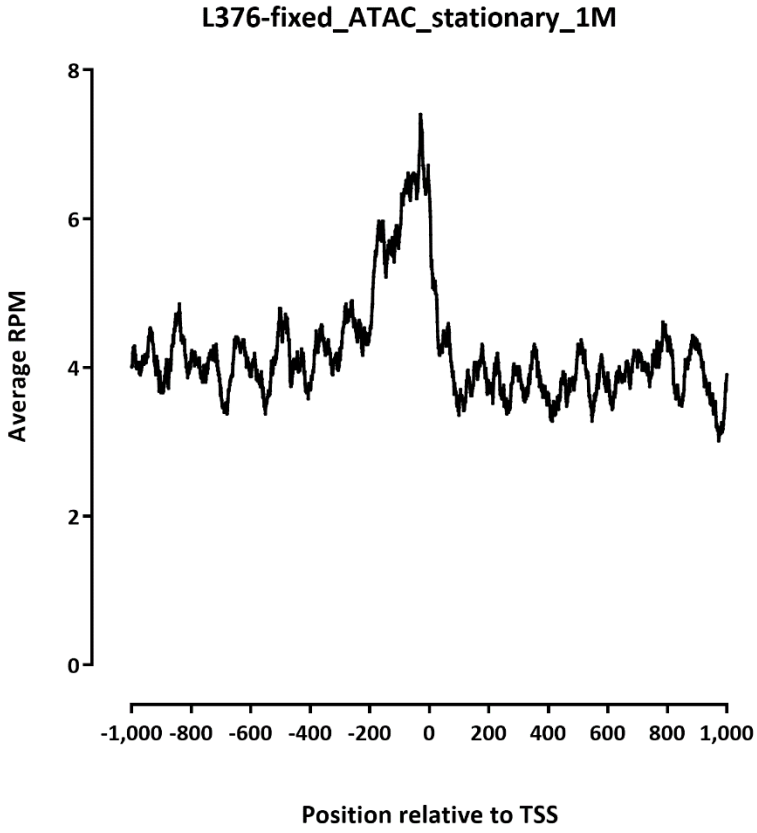


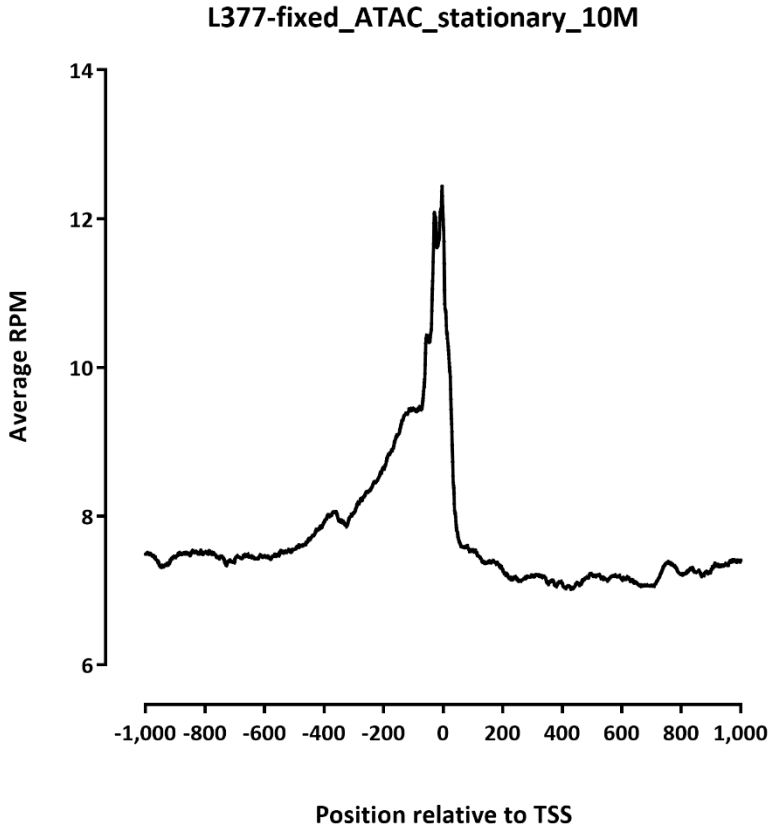


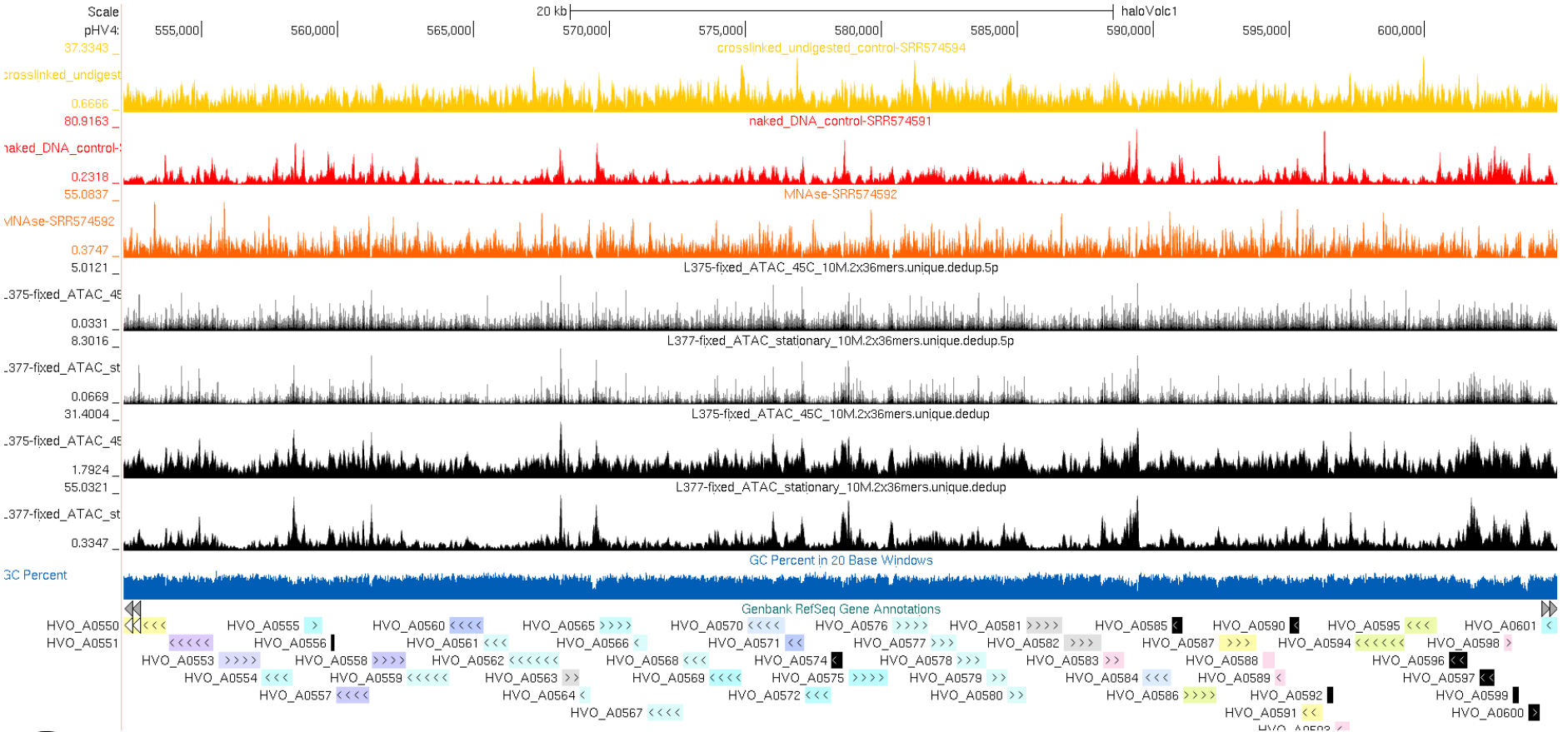
TSS PROFILE

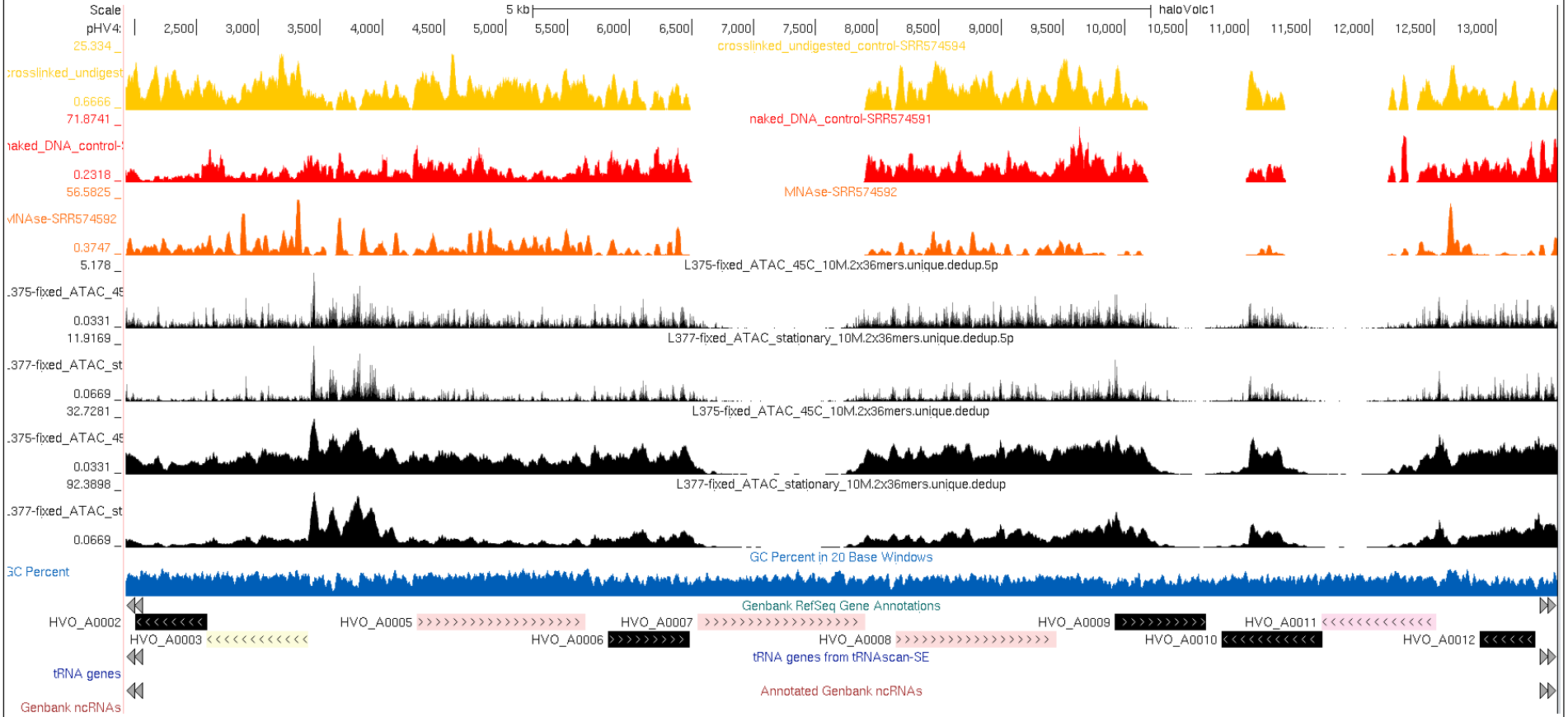

L374-fixed_ATAC_45C_1M
Average RPM
13
12
11
10
9
8
7
-1,000 -800 -600 -400 -200 0 200 400 600 800 1,000
Position relative to TSS


L375-fixed_ATAC_45C_10M
Average RPM
13
12
11
10
9
8
-1,000 -800 -600 -400 -200 0 200 400 600 800 1,000
Position relative to TSS

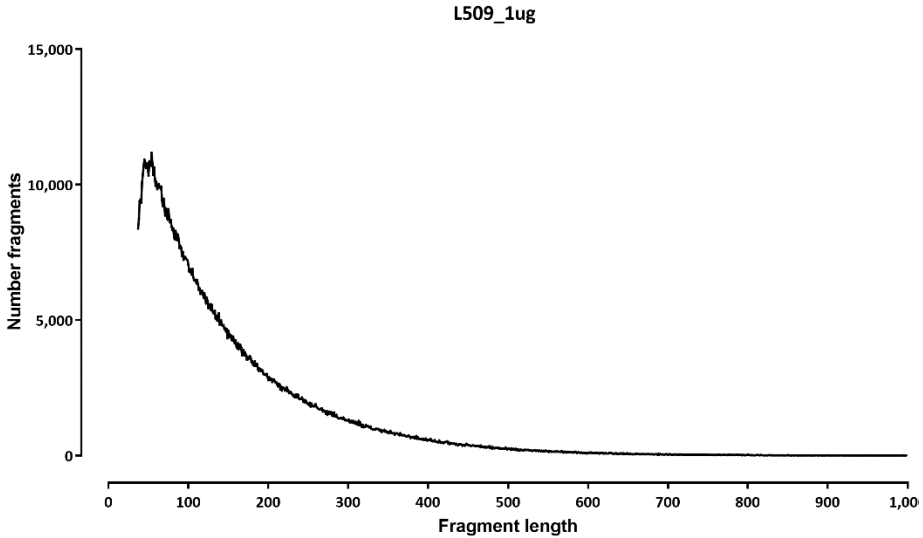

L376-fixed_ATAC_stationary_1M
Average RPM
8
6
4
2
0
-1,000 -800 -600 -400 -200 0 200 400 600 800 1,000
Position relative to TSS

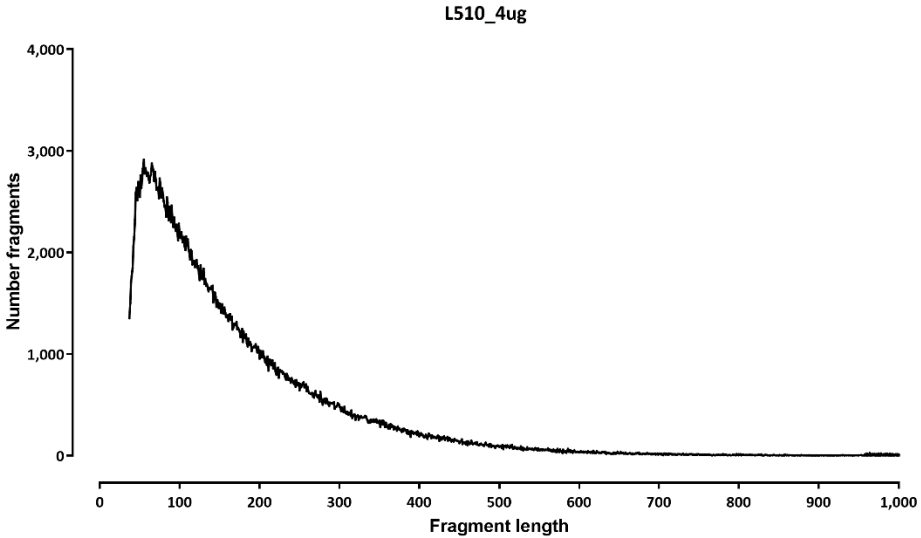

L377-fixed_ATAC_stationary_10M
Average RPM
14
12
10
8
6
-1,000 -800 -600 -400 -200 0 200 400 600 800 1,000
Position relative to TSS

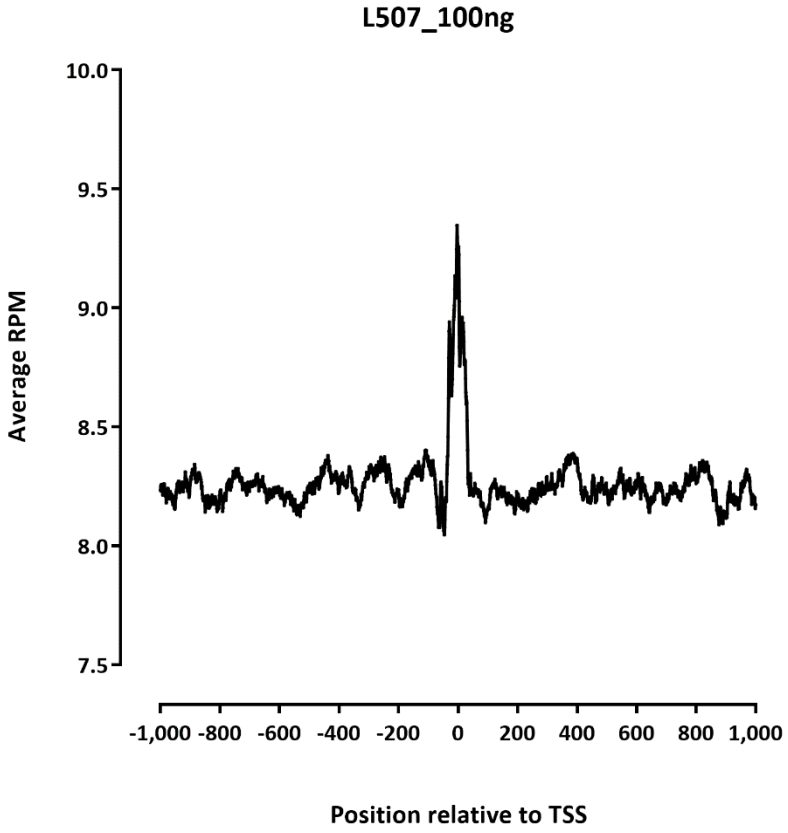

Scale
20 kb
haloVolc1
555,000
560,000
565,000
570,000
575,000
580,000
585,000
590,000
595,000
600,000
crosslinked_undigested_control-SRR574594
naked_DNA_control-SRR574591
MNAse-SRR574592
L375-fixed_ATAC_45C_10M.2x36mers.unique.dedup.5p
L377-fixed_ATAC_stationary_10M.2x36mers.unique.dedup.5p
L375-fixed_ATAC_45C_10M.2x36mers.unique.dedup
L377-fixed_ATAC_stationary_10M.2x36mers.unique.dedup
GC Percent in 20 Base Windows
Genbank RefSeq Gene Annotations
HVO_A0550
HVO_A0551
HVO_A0553
HVO_A0554
HVO_A0557
HVO_A0555
HVO_A0556
HVO_A0558
HVO_A0559
HVO_A0560
HVO_A0561
HVO_A0562
HVO_A0563
HVO_A0564
HVO_A0565
HVO_A0566
HVO_A0567
HVO_A0568
HVO_A0569
HVO_A0570
HVO_A0571
HVO_A0572
HVO_A0574
HVO_A0575
HVO_A0576
HVO_A0577
HVO_A0578
HVO_A0579
HVO_A0580
HVO_A0581
HVO_A0582
HVO_A0583
HVO_A0584
HVO_A0585
HVO_A0586
HVO_A0587
HVO_A0588
HVO_A0589
HVO_A0590
HVO_A0591
HVO_A0592
HVO_A0594
HVO_A0595
HVO_A0596
HVO_A0597
HVO_A0598
HVO_A0599
HVO_A0600
HVO_A0601

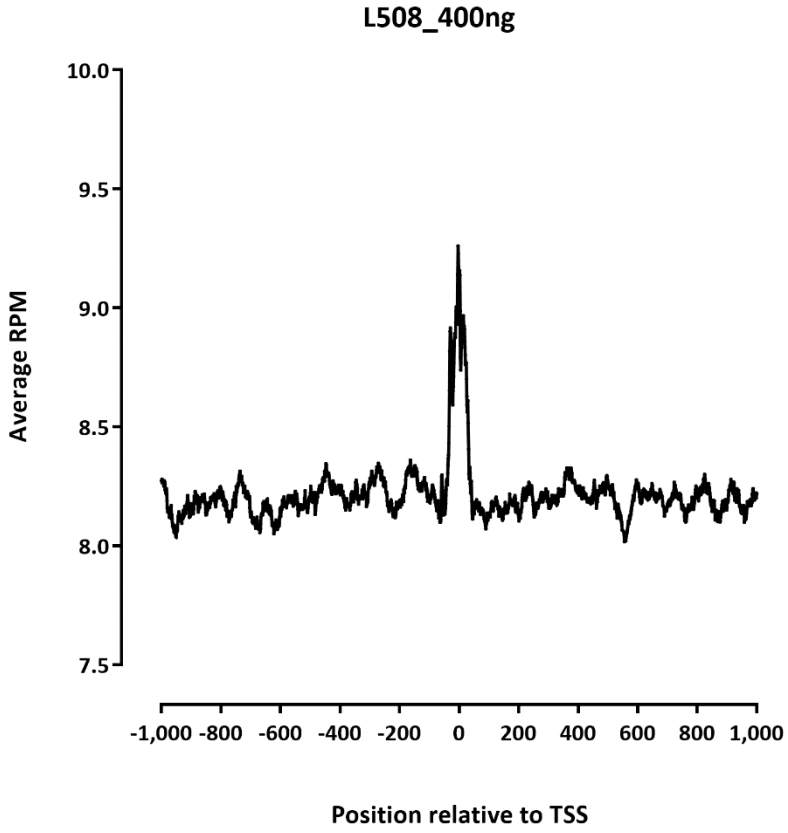

Scale
5 kb
haloVolc1
pHV4:
2,500
3,000
3,500
4,000
4,500
5,000
5,500
6,000
6,500
7,000
7,500
8,000
8,500
9,000
9,500
10,000
10,500
11,000
11,500
12,000
12,500
13,000
25.334
crosslinked_undigested_control-SRR574594
0.6666
71.8741
naked_DNA_control-SRR574591
0.2318
56.5825
MNAse-SRR574592
0.3747
5.178
L375-fixed_ATAC_45C_10M.2x36mers.unique.dedup.5p
0.0331
11.9169
L377-fixed_ATAC_stationary_10M.2x36mers.unique.dedup.5p
0.0669
32.7281
L375-fixed_ATAC_45C_10M.2x36mers.unique.dedup
0.0331
92.3898
L377-fixed_ATAC_stationary_10M.2x36mers.unique.dedup
0.0669
GC Percent in 20 Base Windows
GC Percent
Genbank RefSeq Gene Annotations
HVO_A0002
HVO_A0003
HVO_A0005
HVO_A0006
HVO_A0007
HVO_A0008
HVO_A0009
HVO_A0010
HVO_A0011
HVO_A0012
tRNA genes
tRNA genes from tRNAscan-SE
Genbank ncRNAs
Annotated Genbank ncRNAs

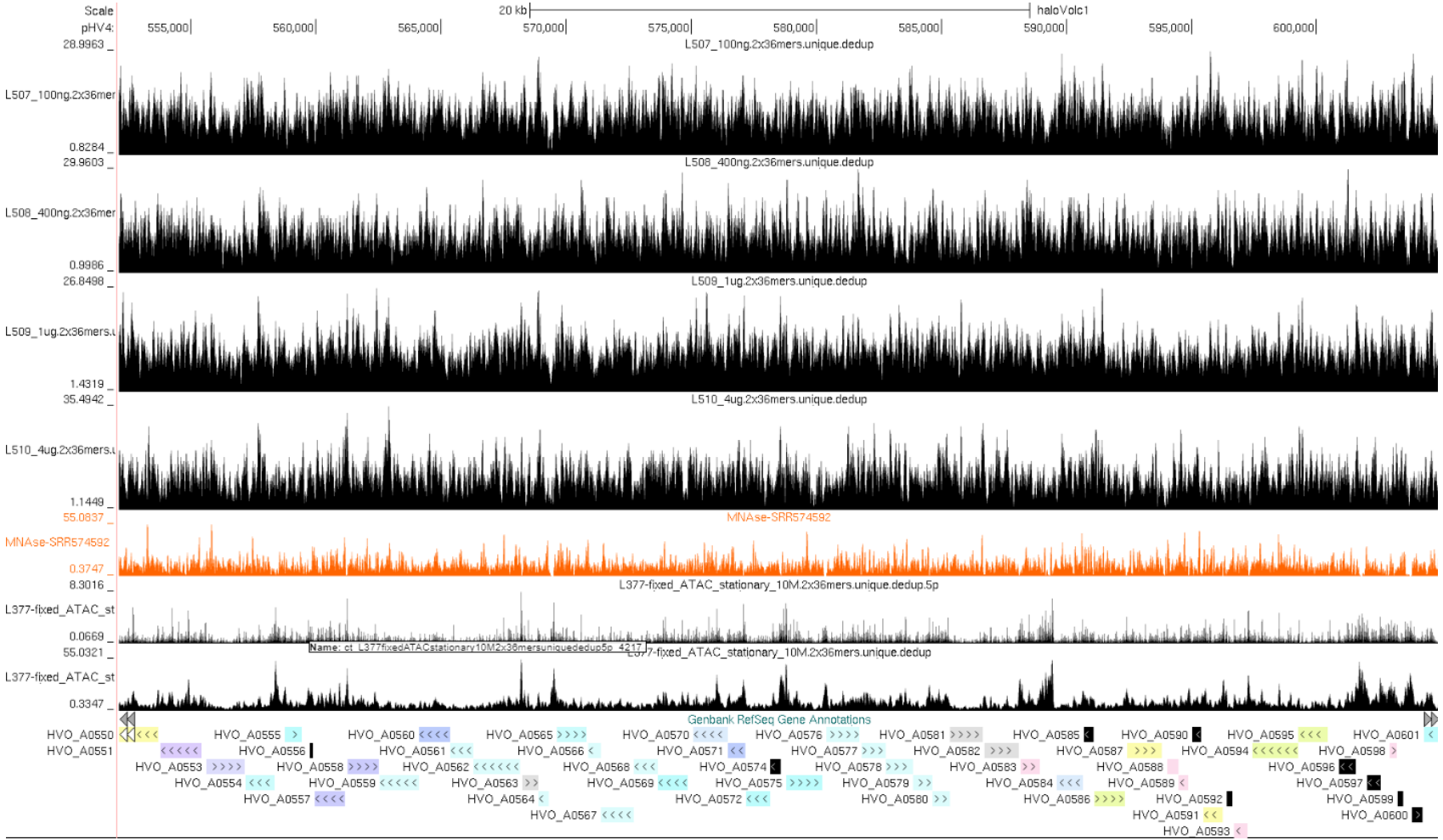
WITH NAKED DNA CONTROL:



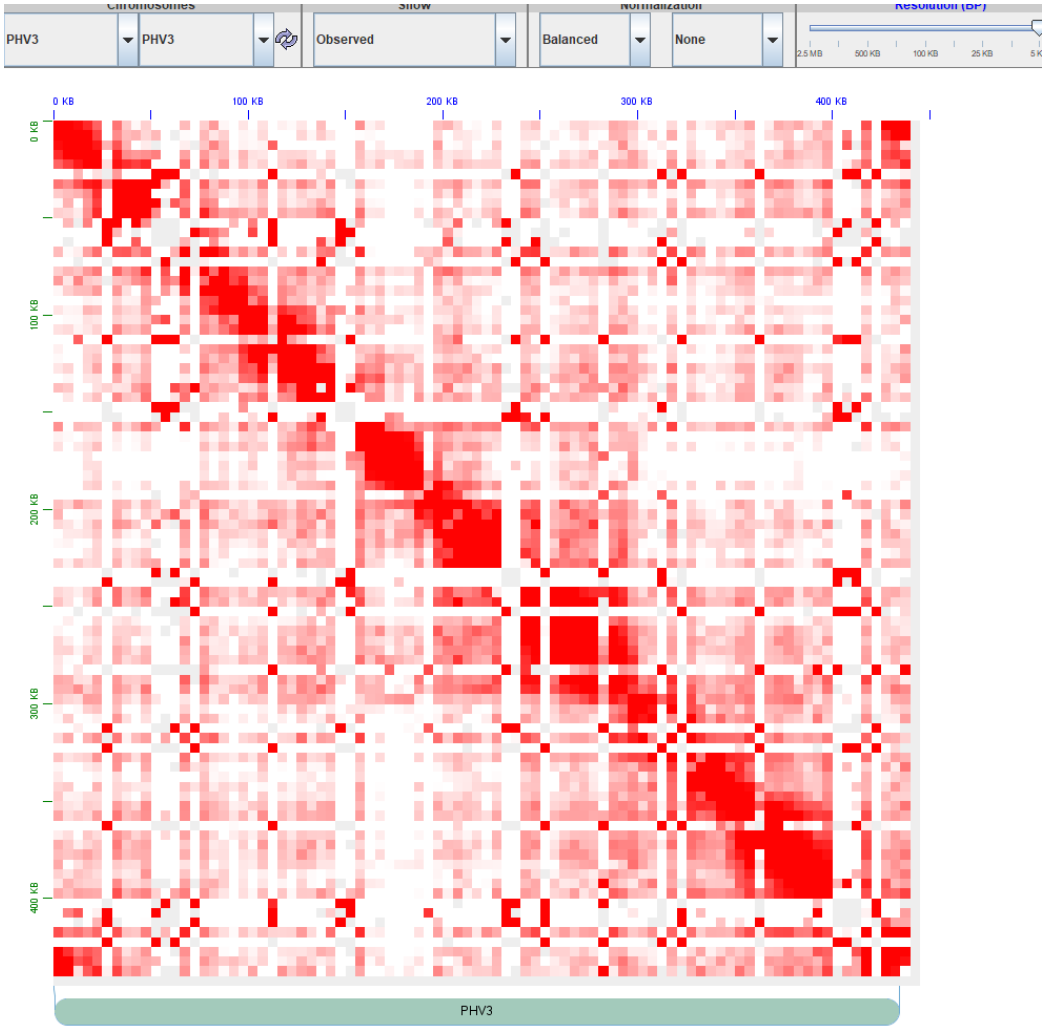


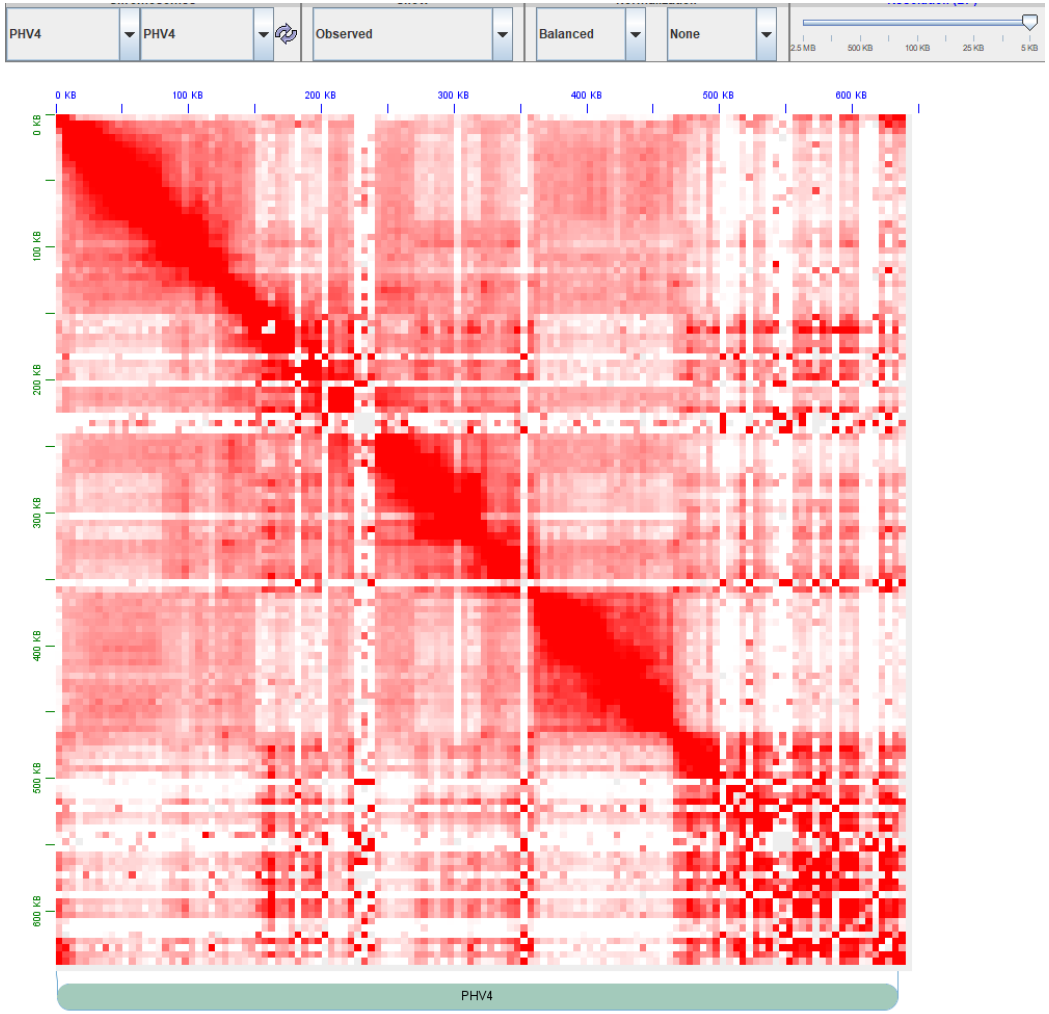

L507_100ng
Average RPM
10.0
9.5
9.0
8.5
8.0
7.5
-1,000 -800 -600 -400 -200 0 200 400 600 800 1,000
Position relative to TSS


L508_400ng
Average RPM
10.0
9.5
9.0
8.5
8.0
7.5
-1,000 -800 -600 -400 -200 0 200 400 600 800 1,000
Position relative to TSS


Scale
pHV4:
20 kb
haloVolc1
555,000
560,000
565,000
570,000
575,000
580,000
585,000
590,000
595,000
600,000
28.9963
L507_100ng.2x36mer
0.8284
29.9603
L508_400ng.2x36mer
0.9986
26.8498
L509_1ug.2x36mers.u
1.4319
35.4942
L510_4ug.2x36mers.u
1.1449
55.0837
MNAse-SRR574592
0.3747
8.3016
L377-fixed_ATAC_st
0.0669
55.0321
L377-fixed_ATAC_st
0.3347
L507_100ng.2x36mers.unique.dedup
L508_400ng.2x36mers.unique.dedup
L509_1ug.2x36mers.unique.dedup
L510_4ug.2x36mers.unique.dedup
MNAse-SRR574592
L377-fixed_ATAC_stationary_10M.2x36mers.unique.dedup.5p
Name: ct_L377fixedATACstationary10M2x36mersuniquededup5p_4217
L377-fixed_ATAC_stationary_10M.2x36mers.unique.dedup
Genbank RefSeq Gene Annotations
HVO_A0550
HVO_A0551
HVO_A0555
HVO_A0556
HVO_A0553
HVO_A0554
HVO_A0557
HVO_A0558
HVO_A0559
HVO_A0560
HVO_A0561
HVO_A0562
HVO_A0563
HVO_A0564
HVO_A0565
HVO_A0566
HVO_A0567
HVO_A0568
HVO_A0569
HVO_A0570
HVO_A0571
HVO_A0572
HVO_A0574
HVO_A0575
HVO_A0576
HVO_A0577
HVO_A0578
HVO_A0579
HVO_A0580
HVO_A0581
HVO_A0582
HVO_A0583
HVO_A0584
HVO_A0585
HVO_A0586
HVO_A0587
HVO_A0588
HVO_A0589
HVO_A0590
HVO_A0591
HVO_A0592
HVO_A0593
HVO_A0594
HVO_A0595
HVO_A0596
HVO_A0597
HVO_A0598
HVO_A0599
HVO_A0600
HVO_A0601

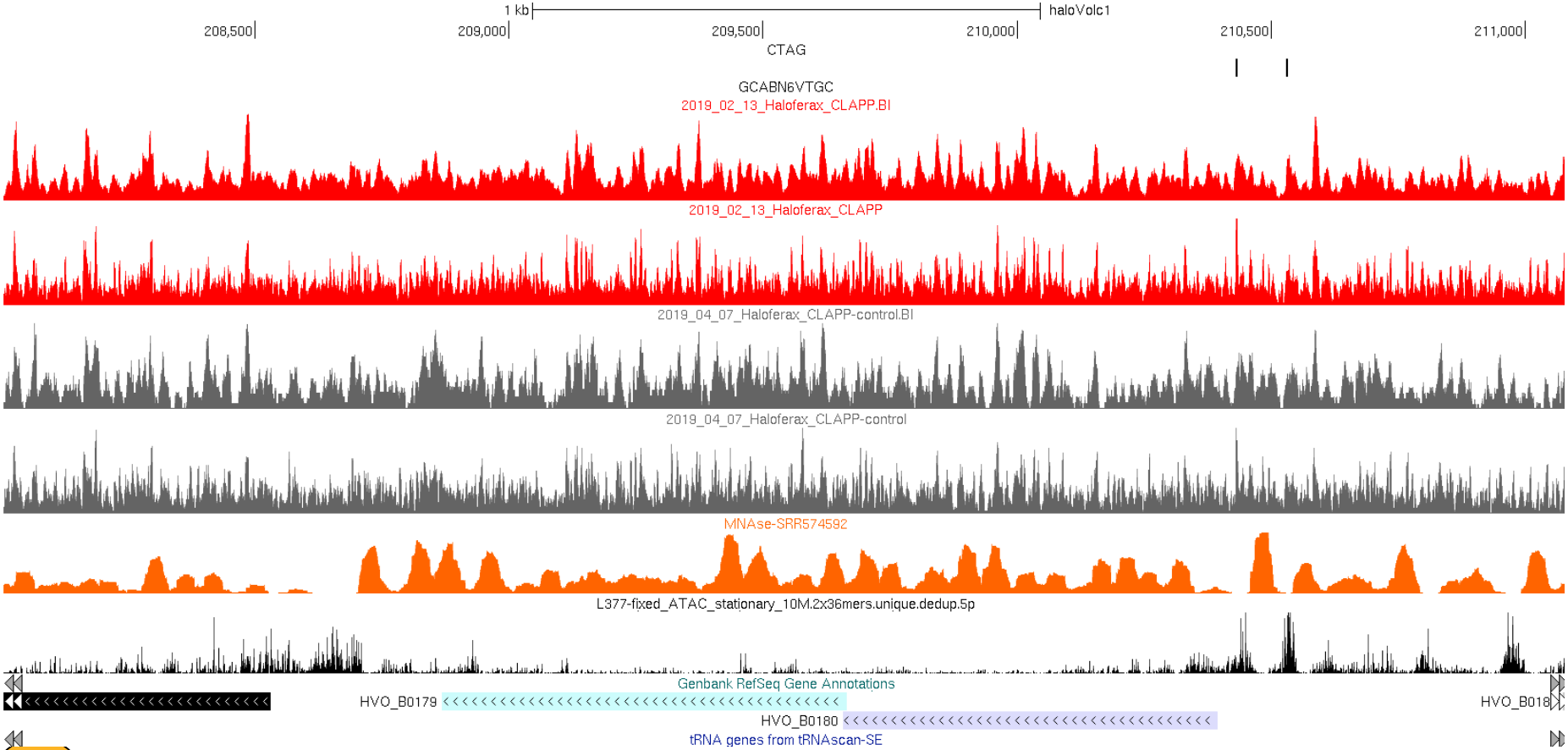
HI-C


Chromosomes
PHV3
PHV3
Show
Observed
Normalization
Balanced
None
Resolution (BP)
2.5 MB
500 KB
100 KB
25 KB
5 K
0 KB
100 KB
200 KB
300 KB
400 KB
PHV3


PHV4
PHV4
Observed
Balanced
None
2.5 MB
500 KB
100 KB
25 KB
5 KB
0 KB
100 KB
200 KB
300 KB
400 KB
500 KB
600 KB
PHV4

ARCHAEAL DNA MODIFICATIONS

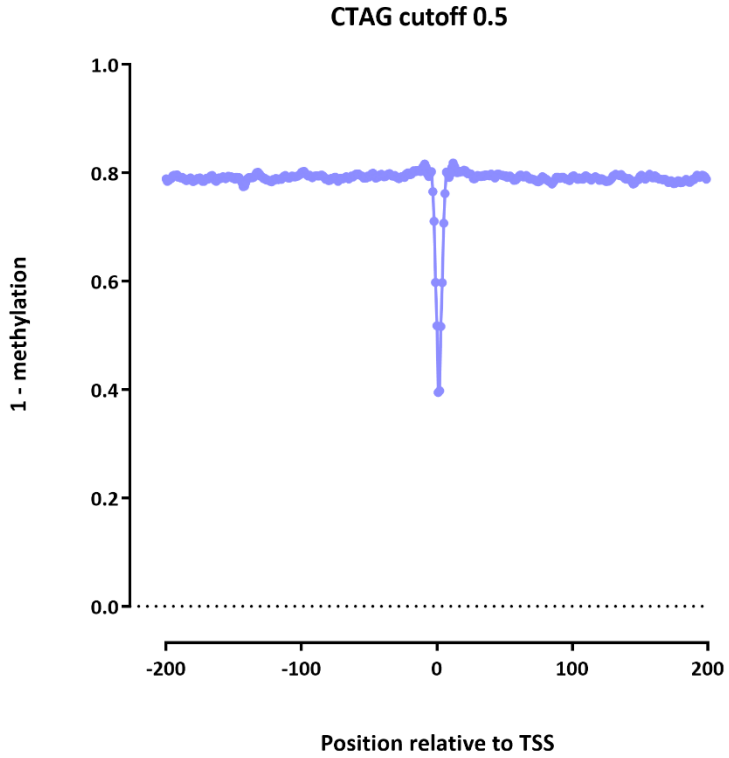
NANOPORE SEQUENCING OF HALOFERAX DNA:

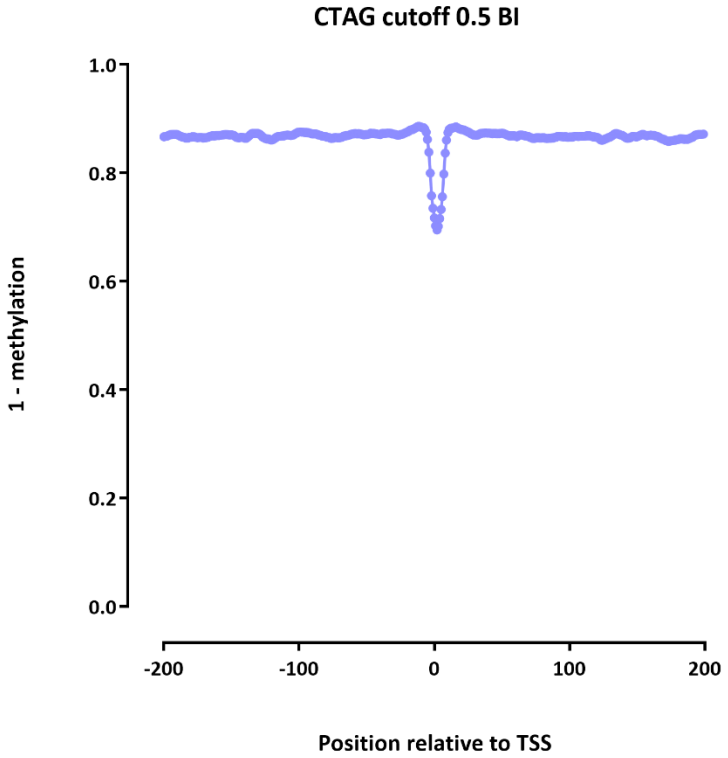


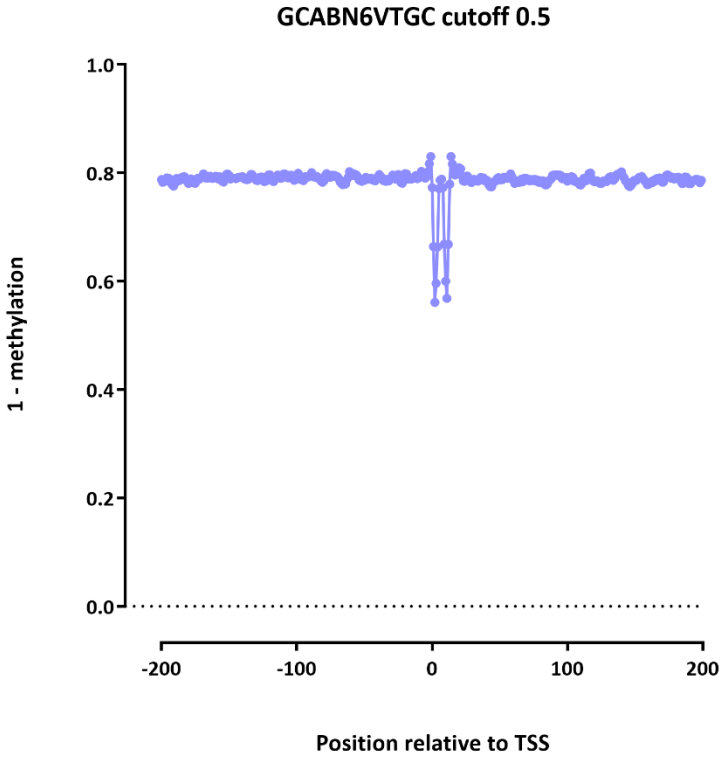
NANOPORE SEQUENCING OF HALOFERAX DNA:

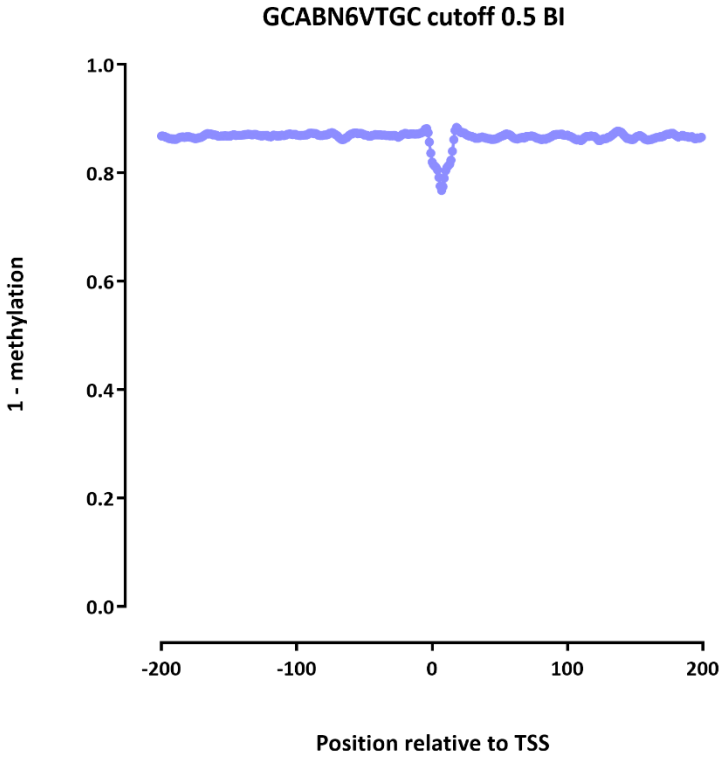
- Strong modifications observed in a lot of places.

- *Haloferax* has two known RM sites:


CTAG cutoff 0.5
1 - methylation
1.0
0.8
0.6
0.4
0.2
0.0
-200
-100
0
100
200
Position relative to TSS


CTAG cutoff 0.5 BI
1 - methylation
1.0
0.8
0.6
0.4
0.2
0.0
-200
-100
0
100
200
Position relative to TSS


GCABN6VTGC cutoff 0.5
1.0
0.8
0.6
0.4
0.2
0.0
1 - methylation
-200
-100
0
100
200
Position relative to TSS


GCABN6VTGC cutoff 0.5 BI
1.0
0.8
0.6
0.4
0.2
0.0
1 - methylation
-200
-100
0
100
200
Position relative to TSS

The RM sites do not explain the observed patterns

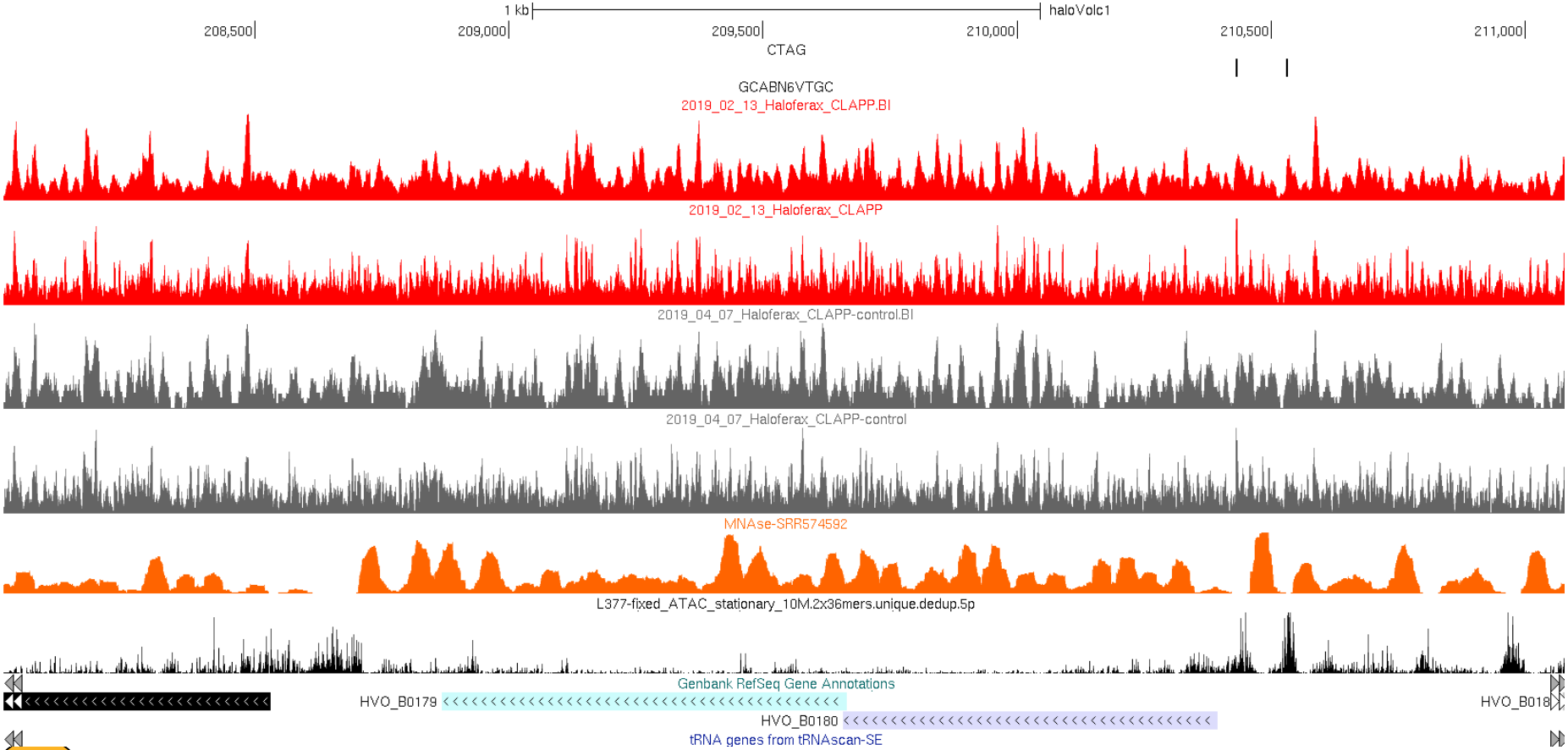


Figure 2. Overview of Sodium Bisulfite Conversion and EM-seq.

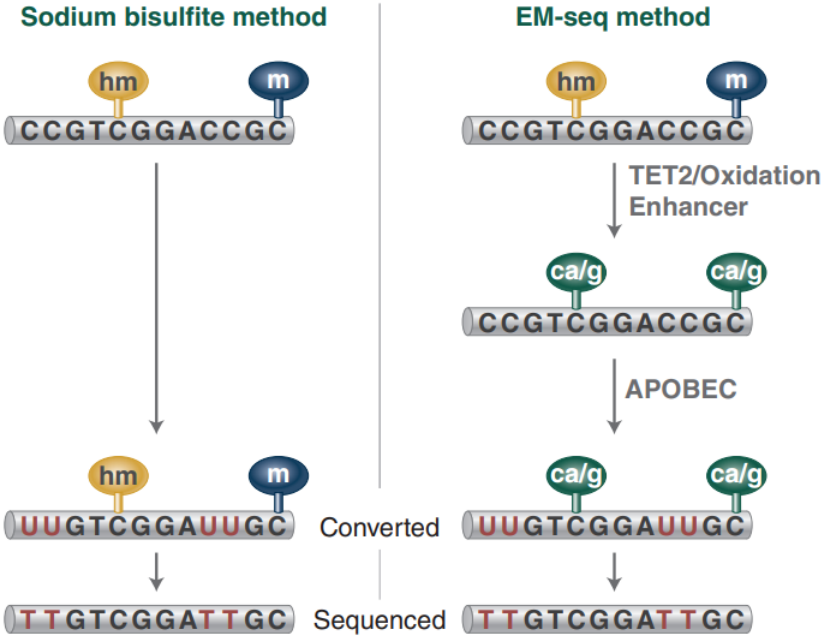


Figure 2 shows a comparison of the sodium bisulfite and EM-seq methods. Sodium bisulfite treatment of DNA results in the deamination of cytosines into uracils, however the modified forms of cytosine (5mC and 5hmC) are not deaminated. Therefore, the preference of bisulfite to chemically deaminate cytosines enables the methylation status of cytosines to be determined. When bisulfite treated DNA is PCR amplified, uracils are replaced by thymines and the 5mC/5hmC are replaced by cytosines. Once sequenced, unmethylated cytosines are represented by thymines and 5mC and 5hmC are represented by cytosines. By comparing sequences to non-converted genomes the appropriate methylation status can be assessed.

EM-SEQ PROTOCOL:

Species	#	CpG %	CHG%	CHH%	Complexity	Read Length	Mean alignment Length	Unique	Unique deduped	Multi	Raw fragments	Alignment fraction
Homo sapiens hg20-female	L575-GM12878_EM-seq	42.80%	10.10%	11.40%	0.97	2x75-trim	74.16	81,747,684			51,390,857	80.00%
Symbiodinium_minutum _Clade_B1_Ver1.120123	L576-SSBO1_EM-seq	63.90%	4.00%	3.80%	0.91	2x75-trim	74.07	174,638,446			158,207,576	55.50%
Haloferax volcanii DS2	L577-Haloferax_volcani_stationary_EM-seq	30.10%	35.70%	31.50%	0.65	2x75-trim	74.29	35,934,240			20,115,037	89.80%

WGBS:

Haloferax volcanii DS2	L625-stationary_2019_03_on_bead_PCR	5.90%	0.50%	0.50%		2x75-trim						1.50%
Haloferax volcanii DS2	L626-42C_2019_07_on_bead_PCR	7.80%	0.80%	1.00%		2x75-trim						2.90%
Haloferax volcanii DS2	L627-stationary_2019_07_on_bead_PCR	7.30%	0.50%	0.60%		2x75-trim						1.60%

- New modification?

ACKNOWLEDGEMENTS

William Greenleaf
Anshul Kundaje

Tingting Xiang
Arthur Grossman

Alex Trevino
Zohar Shipony

Phil Cleves