The chromatin organization of a chlorarachniophyte nucleomorph genome

Georgi K. Marinov[1,#], Xinyi Chen[1], Tong Wu[2], Chuan He[2,3,4], Arthur R. Grossman[5], Anshul Kundaje[1,6], and William J. Greenleaf[1,7,8,9,#]

[1] *Department of Genetics, Stanford University, Stanford, California 94305, USA*
[2] *Department of Chemistry and Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL, 60637, USA*
[3] *Department of Biochemistry and Molecular Biology and Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL, 60637, USA*
[4] *Howard Hughes Medical Institute, The University of Chicago, Chicago, IL, 60637, USA*
[5] *Carnegie Institution for Science, Department of Plant Biology, Stanford, California 94305, USA*
[6] *Department of Computer Science, Stanford University, Stanford, California 94305, USA*
[7] *Center for Personal Dynamic Regulomes, Stanford University, Stanford, California 94305, USA*
[8] *Department of Applied Physics, Stanford University, Stanford, California 94305, USA*
[9] *Chan Zuckerberg Biohub, San Francisco, California, USA*
[#] *Corresponding author*

Abstract

Nucleomoprhs are remnants of secondary endosymbiotic events between two eukaryote cells, in which the endosymbiont has retained its eukaryotic nucleus, unlike the usually observed pattern of its eventual loss. Nucleomoprh genomes are unique in several aspects. They have evolved at least twice independently, in chlorarachniophytes and cryptophytes, yet they have converged on a remarkably similar genomic architecture, characterized by the most extreme compression and minituarization among all known eukaryotes. Previous computational studies have suggested that nucleomorph chromatin likely exhibits a number of divergent features but it has never been studied directly. In this work, we map open chromatin, active transcription, and three-dimensional genome architecture in the nucleomorph of the chlorarachniophyte *Bigelowiella natans*. We find that the *B. natans* nucleomorph genome exists in a highly accessible state, akin to that of ribosomal DNA in some other eukaryotes, and that it is highly transcribed throughout its length, with few signs of polymerase pausing at transcription start sites (TSSs). In the same time, most nucleomorph TSSs show very strong nucleosome positioning. Chromosome conformation analysis shows that nucleomorph chromosomes interact with themselves and with each other at their telomeric regions, and also, surprisingly, that *B. natans* mitochondria, which topologically derive from the host, physically interact more strongly with the endosymbiont-derived plastid and with nucleomorph chromosomes. These results provided a foundation for future understanding of nucleomorph chromatin biology.

Introduction

Endosymbiosis, especially between an eukaryote host and a prokaryote, is a common event in the evolution of eukaryotes, and the consequences of it for the host and the endosymbiont often follow similar general trends. One such trend is the reduction of the endosymbiont's genome due to gene loss and endosymbotic gene transfer[1,2] (EGT) into the host's nucleus, the classic example of which are the extremely reduced genomes of plastids and mitochondria. It is also strongly manifested in the fate of secondary endosymbionts (eukaryotes that become endosymbionts of other eukaryotes). Such endosymbiotic events have occurred on multiple occasions in the evolution of eukaryotes[3], usually resulting in the retention of the plastid of the photosynthetic eukaryotic endosymbiont (as a secondary plastid) while the nucleus of the endosymbiont is lost entirely. There are several notable exceptions to this general rule. One are the dinotoms, the result of the endosymbiosis between a dinoflagellate host and a diatom, in which the diatom has not been substantially reduced[4,5]. More striking are the nucleomorphs, which are classically known from the chlorarachniophytes and the cryptophytes (but may in fact have arisen in other groups too, such as some dinoflagellates[6,7]),

which represent the highly reduced but still retained remnant of the endosymbiont's nucleus and its genome [8,9].

A striking feature of chlorarachniophyte and cryptophyte nucleomorphs is that they have evolved independently, from a green and a red alga, respectively, yet their genomes exhibit remarkably convergent properties [10,11]. In both cases, the genomes of their nucleomorphs are the smallest known among all eukaryotes, usually just a few hundred kilobases in size. All sequenced nucleomorph genomes are organized into three highly AT-rich chromosomes, in which arrays of ribosomal RNA genes form the subtelomeric regions. They are also extremely compressed as there is very little intergenic space between genes, with genes actually even overlapping on occasions, and the genes themselves are also often shortened [12–18].

A number of important questions about nucleomorph biology remain unanswered, including the extent of conservation and divergence of the chromatin organization and transcriptional mechanisms of these extremely reduced nuclei. Previous computational analysis of nucleomorph genome sequences [19] has suggested that a considerable degree of deviation from the conventional eukaryotic state is likely to have developed in nucleomorphs. For example, histone proteins are ancestral to all eukaryotes and the key posttranscriptional modifications (PTMs) that they carry also date back to the last eukaryotic common ancestor (LECA) and are extremely conserved in all branches of the eukaryotic tree [20], with the notable exception of dinoflagellates [21]. This is likely because of the involvement of these PTMs in the so called "histone code" [22], in which different PTMs are deposited in a highly regulated manner on specific residues of histone proteins, and are then read out by various effector proteins. The histone code plays key roles in practically all aspects of chromatin biology, such as the regulation of gene expression, the transcriptional cycle, the formation of repressive heterochromatin, mitotic condensation of chromosomes, DNA repair, and many others.

Nucleomorphs appear to be another exception to this general rule. Inside nucleomorph genomes, in both chlorarachniophytes and cryptophytes, only two histone genes are encoded, one for H3 and for H4, with H2A and H2B apparently imported from the host's nucleus [23]. Sequence analysis of the H3 and H4 proteins shows remarkable divergence from the typical aminoacid sequence in eukaryotes, in particular in chlorarachniophytes, which have lost nearly all key histone code residues [19]. Furthermore, the heptad repeats in the C-terminal domain (CTD) tail of the Rpb1 subunit of RNA Polymerase II, which are also key to the eukaryote transcriptional cycle and mRNA processing [24] are also highly conserved in eukaryotes [25], have been lost.

These observations suggest that the nucleomorph nucleosome may have unconventional properties compared to those of other eukaryotes (e.g. the possibility that it protects DNA differently, perhaps in a looser structure), and raise a number of questions regarding the conservation and divergence of key aspects of transcriptional regulation and the transcription cycle in nucleomorph genomes (e.g. whether nucleomorph promoters display the typical signatures of nucleosome depletion and positioning, histone modifications, and others, and how that relates to transcriptional activity) and their overall organization (e.g. how they are organized in three-dimensional space). However, nucleomorph chromatin has not been directly studied before.

In this work we map chromatin accessibility, active transcription and three-dimensional (3D) genome organization in the chlorarachniophyte *Bigelowiella natans* in order to address these gaps in our knowledge of nucleomorph biology. We find that nucleomorph chromosomes exist in a highly accessible state, reminiscent of what is observed for ribosomal DNA (rDNA) in other eukaryotes, such as budding yeast, where it is thought to exist in a fully nucleosome-free state when actively transcribed [26–28]. However, in the same time, nucleomorph promoters are associated with strongly positioned nucleosomes and exhibit a nucleosome free region upstream of the transcription start site (TSS). Active transcription is nearly uniformly distributed across nucleomorph genomes, with the exception of the subtelomeric rDNA gene, over which it is elevated, together with chromatin accessibility. We find few signs of RNA polymerase pausing over promoters. Nucleomorph chromosomes form a network of telomere-to-telomere interactions in 3D space, and also fold on themselves, but centromeres do not interact preferentially with each other. Curiously, the *B. natans* mitochondrion, though topologically deriving from the host, is more often physically interacting with the endosymbiont compartments (the plastid and the nucleomorph). These results shed light on chlorarachniophyte nucleomorph chromatin structure and provide a framework for future mechanistic studies of transcriptional and regulatory biology in nucleomorphs.

Results

Chromatin accessibility in nucleomorphs

In order to study the chromatin structure of the *B. natans* nucleomorph genome, we carried out ATAC-seq experiments in *B. natans* grown under standard conditions (see Methods). As *B. natans* has four different genomic compartments (Figure 1A) – nucleus, nucleomorph, mitochondrion and plastid – we first examined the fragment length distribution in each (Figure 1B). The nucleus exhibits a subnucleosomal peak at ∼100 bp as well as a second, most likely nucleosomal, peak (or a "shoulder" in the curve) at ∼200 bp. In contrast, the nucleomorph displays two peaks, one at ≤100 bp and another at ∼220 bp, which are tentatively interpreted as subnucleosomal and a nucleosomal one (see further below for details). The mitochondrion and the plastid fragment length distributions are unimodal, consistent with the open DNA structure expected from these compartments.

We then examined the distribution of reads across the compartments (Figure 1C). As expected from the lack of nucleosomal protection over mitochondrial and plastid DNA, *B. natans* ATAC libraries are dominated by reads mapping to those compartments. However, curiously, nucleomorph-mapping reads represent a much larger fraction of mapped reads than expected from the portion of genomic real estate that the nucleomorph genome comprises, and also relative to what is seen in input samples, suggesting that the nucleomorph might exist in a preferentially accessible chromatin state.

We next turned our attention to ATAC-seq profiles in the nucleus, both to characterize accessibility in the *B. natans* host genome, and to verify that the ATAC-seq libraries are of high quality. Figure 1D shows the average ATAC-seq signal over annotated *B. natans* TSSs; it is enriched over promoters, as expected from successful ATAC-seq experiments (we note that the shape of the metaplot is somewhat distorted by the fact that available annotations do not actually include the real TSSs, but only the sites of translation initiation, with most 5'UTR missing). Examination of browser tracks confirmed the enrichment over TSSs (Figure 1E), and did not reveal obvious open chromatin sites outside promoters. We carried out peak calling using MACS2[29], and the distribution of called peaks was also strongly centered on promoters, with almost no open chromatin regions outside the ±2 kbp range around TSSs. Thus *B. natans* appears to have a functional genomic organization typical for an eukaryote with a small compact genome such as yeast, with all regulatory elements located immediately adjacent to TSSs, and few to no distal regulatory elements that exhibit increased accessibility. In addition, in a standard *B. natans* culture the majority of promoters exhibit an open chromatin configuration (Figure 1G).

Genome browser examination of ATAC-seq profiles over the nucleomorph genome (Figure 1H) showed high levels of chromatin accessibility throughout all chromosomes, with numerous localized peaks and generally increased accessibility over the rDNA located near telomeres. Strikingly, the average ATAC-seq profile over nucleomorph TSSs (Figure 1I) showed a strong increase in accessibility around the TSS, but also a clear signature of multiple positioned nucleosomes around each TSS (a clear +1 nucleosome immediately downstream of the TSS, as well as a putative +2 one, together with a –1 nucleosome upstream of the TSS). This phasing is also clearly visible from the individual ATAC-seq profiles over each nucleomorph gene (Figure 1J).

We then quantified the extent of increased accessibility over the nucleomorph genome by calculating the enrichment of ATAC-seq signal over it relative to an input sample. We find that the nucleomorph is ∼10× enriched in ATAC-seq libraries (Figure 1J). Notably, this is comparable to what is observed for rDNA genes in the budding yeast *S. cerevisiae* (Figure 1J), which are known to exist in an almost fully nucleosome-free configuration when actively transcribed, which is thought to be ∼50% of the time[26–28,30].

Thus the nucleomorph apparently exists in a highly accessible, presumably actively transcribed state. Of note, this estimation is not driven by the rDNA genes within it, although those are indeed more accessible than the rest of the nucleomorph genome, as the difference in accessibility between the rDNA arrays and the rest of the genome is on the order of ∼2× and they occupy a minor portion of it.

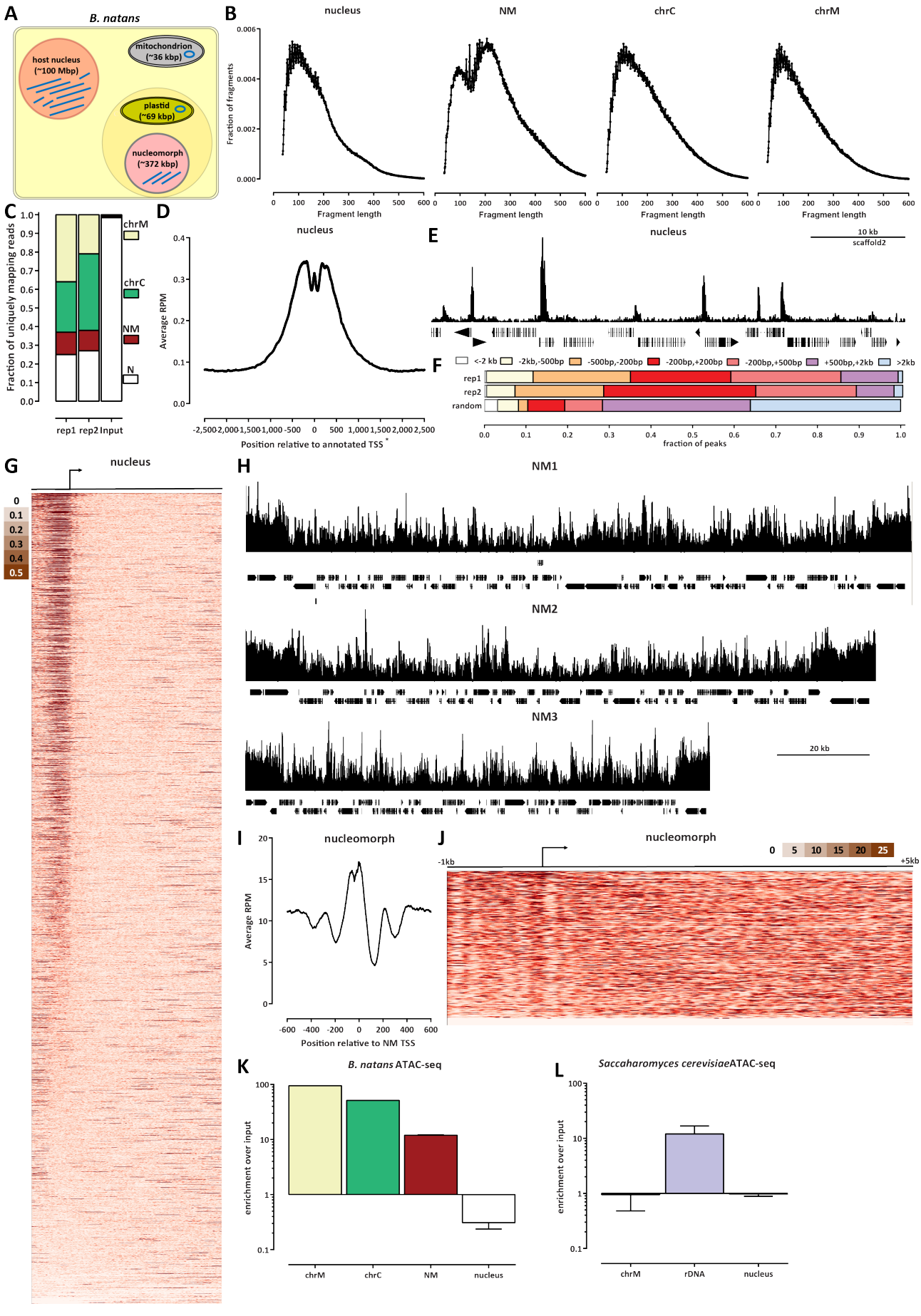
Yet in the same time nucleomorph TSSs show very strong nucleosome positioning. To more accurately analyze nucleosome positioning in both the nuclear and the nucleomorph compartments, we applied the NucleoATAC algorithm[31] over the whole nucleomorph genome and over the 1-kb regions centered on annotated 5' gene ends in the nucleus. We identified 7,251 and 1,440 positioned nucleosomes in the nucleus and in the nucleomorph, respectively. The distribution of the nuclear ones peaked shortly downstream of TSSs (Figure 3A), suggesting that nuclear TSSs are also associated with a positioned +1 nucleosome. A V-plot[32] analysis of these nucleosomes showed that the ATAC-seq fragment lengths associated with them are in the 175-200 bp range, and that subnucleosomal fragments are located in the immediate vicinity (Figure 3A). In contrast, in the nucleomorph we observed the three positioned nucleosomes (+1, +2, and -1) indicated above (Figure 3C), but the V-plots are centered at a larger fragment length, in the 200-225 bp range (Figure 3D).

Transcriptional activity in the nucleomorph genome

Next, we studied the patterns of active transcription in the nucleomorph. To this end, we deployed the KAS-seq assay[33], which maps single-stranded DNA (ssDNA) by specifically labeling unpaired guanines with N_3-kethoxal, to which biotin can then be attached using click chemistry, allowing for regions containing ssDNA to be specifically enriched. Most ssDNA in the cell is usually associated with RNA polymerase bubbles[33], thus KAS-seq is a good proxy for active transcription.

In the *B. natans* nucleus, KAS-seq shows enrichment over promoters and over actively transcribed genes (Figure 3A-B), as expected based on patterns observed in other eukaryotes[33], indicative of RNA polymerase spending more time near the TSS. Of note, we observe only very weak correlation between promoter accessibility and active transcription (Supplementary Figure 1), suggesting significant decoupling between the opening of nucleosome depleted promoter-proximal regions and the regulation of active transcription in *B. natans*.

In the nucleomorph, we see largely uniform levels of KAS-seq signal, with the exception of the rDNA genes, and three localized peaks, one on the first, and two on the second nucleomorph chromosomes (Figure 3C-E). The increased transcription of the rDNA genes is consistent with their higher accessibility observed in ATAC-seq data. We


A
B. natans
host nucleus (~100 Mbp)
mitochondrion (~36 kbp)
plastid (~69 kbp)
nucleomorph (~372 kbp)
B
nucleus
NM
chrC
chrM
Fraction of fragments
Fragment length
C
Fraction of uniquely mapping reads
chrM
chrC
NM
N
rep1 rep2 Input
D
nucleus
Average RPM
Position relative to annotated TSS
E
nucleus
10 kb
scaffold2
F
<-2 kb
-2kb,-500bp
-500bp,-200bp
-200bp,+200bp
-200bp,+500bp
+500bp,+2kb
>2kb
rep1
rep2
random
fraction of peaks
G
nucleus
H
NM1
NM2
NM3
20 kb
I
nucleomorph
Average RPM
Position relative to NM TSS
J
nucleomorph
-1kb
+5kb
K
B. natans ATAC-seq
enrichment over input
chrM
chrC
NM
nucleus
L
Saccharomyces cerevisiae ATAC-seq
enrichment over input
chrM
rDNA
nucleus

quantified the overall enrichment of active transcription in the different compartments and found that the nucleomorph is ∼2-fold enriched in KAS-seq datasets than the nucleus (Figure 3F).

These observations, based on measuring actual active transcription, corroborate previous reports, based on transcriptomic analysis, of high pervasive and largely uniform transcriptional activity over most the nucleomorph[34–37]. However, rDNA genes were removed in some of these analyses[34] while we identify them as a transcriptional unit existing in a distinct from the rest of the nucleomorph state (in the analysis presented here, multimapping reads were retained and normalized for, allowing us to measure accessibility and transcription levels over the rDNA genes; see the Methods section for more details).

Three-dimensional organization of the *B.natans* nucleomorph genome

Finally, we mapped the three-dimensional genome organization in *B.natans* using *in situ* chromosomal conformation capture (Hi-C[38]). We employed a modified for the highly AT-rich nucleomorph genomes protocol (see Methods for details) and generated high-resolution 1-kbp maps, which allow us to investigate the fine features of the small nucleomorph chromosomes.

Hi-C maps reveal that the nucleomorph chromosomes often exist in a folded conformation, in which the two chromosome ends contact each other (Figure 4A-B). In addition, the telomeric regions of all nucleomorph chromosomes physically associate with each other, forming a telomeric network of interactions (Figure 4A). In many eukaryotes, a centromeric interaction network is also observed[39], but enriched interchromosomal interactions in nucleomorphs are only telomeric. We do not observe much internal structure inside individual nucleomorph chromosomes, with the exception of NM2, in which one potential loop interaction is seen; its mechanistic origins are currently unclear as its singular nature prevents the identification of sequence drivers of its formation.

We also used our Hi-C data to generate a chromosome-level scaffolding[40] of the existing assembly of the *B. natans* nuclear genome[41], which originally consisted of 302 nuclear contigs. Our chromosome-level assembly identifies 79 pseudochromosomes, the smallest one of which is ∼350 kbp, with the largest being ∼3 Mbp, and retains only 18 smaller unplaced contigs, the largest of them being only 8,753 bp (Figure 4D).

We made one curious observation when manually finalizing the chromosome-level assembly – the mitochondrion, although it is topologically derived from the host (Figure 1A), exhibits frequent Hi-C interactions with the plastid and also with the nucleomorph chromosomes. This was seen in the maps themselves (Figure 4E) and was confirmed by a systematic analysis of chrM interactions with all other chromosomes (Figure 4F). While both the plastid and the mitochondrial genomes exist in high copy numbers, the nucleomorph genome has the same copy number as the nuclear one (as shown by our input samples), thus we conclude that these preferential interactions likely indeed represent frequent physical proximity in the cell, which then leads to ligation events with nucleomorph chromosomes during the *in situ* Hi-C procedure.

Discussion

This study present the first analysis of physical chromatin organization in a nucleomorph genome, in the chlorarachniophyte *B. natans*, using a combination of ATAC-seq, Hi-C and KAS-seq measurements. We also provide a near-complete chromosome-level scaffolding of the nuclear genome by taking advantage of the physical proximity information provided by Hi-C data, and assess the extent of physical interactions between the different genomic compartments.

The novel insights about nucleomorph chromatin that emerge from our work will guide future mechanistic studies of its inner workings.

While it was previously suspected that nucleomorphs are very highly transcriptionally active, we demonstrate that this is also reflected at the level of chromatin structure, as nucleomorph chromosomes are much more highly accessible than those in the nucleus. Previous transcriptomic

Figure 1 *(preceding page)*: The chromatin accessibility landscape of the *B. natans* nuclear and nucleomorph genomes. (A) Schematic outline of the different genomic compartments in a *B. natans* cell. (B) ATAC-seq fragment length distribution in the different genomic compartments. (C) Distribution of mapped ATAC-seq reads across genomic compartments. (D) ATAC-seq read coverage metaplot around nuclear TSSs. (E) Snapshot of an ATAC-seq profile at a typical nuclear locus. (F) Distribution of ATAC-seq called peaks in the nucleus relative to TSSs. The "random" distribution was generated by splitting the genome in 500-bp bins and taking the boundary coordinates of each bin as "peaks". (G) ATAC-seq profiles around all nuclear genes. (H) ATAC-seq profiles over the NM1, NM2 and NM3 nucleomorph chromosomes. (I) ATAC-seq read coverage metaplot around nucleomorph TSSs. (J) ATAC-seq profiles around all nucleomorph genes. (K) The nucleomorph genome is ∼10× enriched in ATAC-seq datasets relative to the nuclear genome. Shown is the ratio of normalized mapped ATAC-seq peaks for each of the compartments relative to the normalized mapped reads in an input sample (a Hi-C dataset mapped in a single-end format). (L) Nucleomorph accessibility is comparable to the accessibility of rDNA loci in the budding yeast *S. cerevisiae*, which exist in a fully nucleosome-free conformation when expressed.

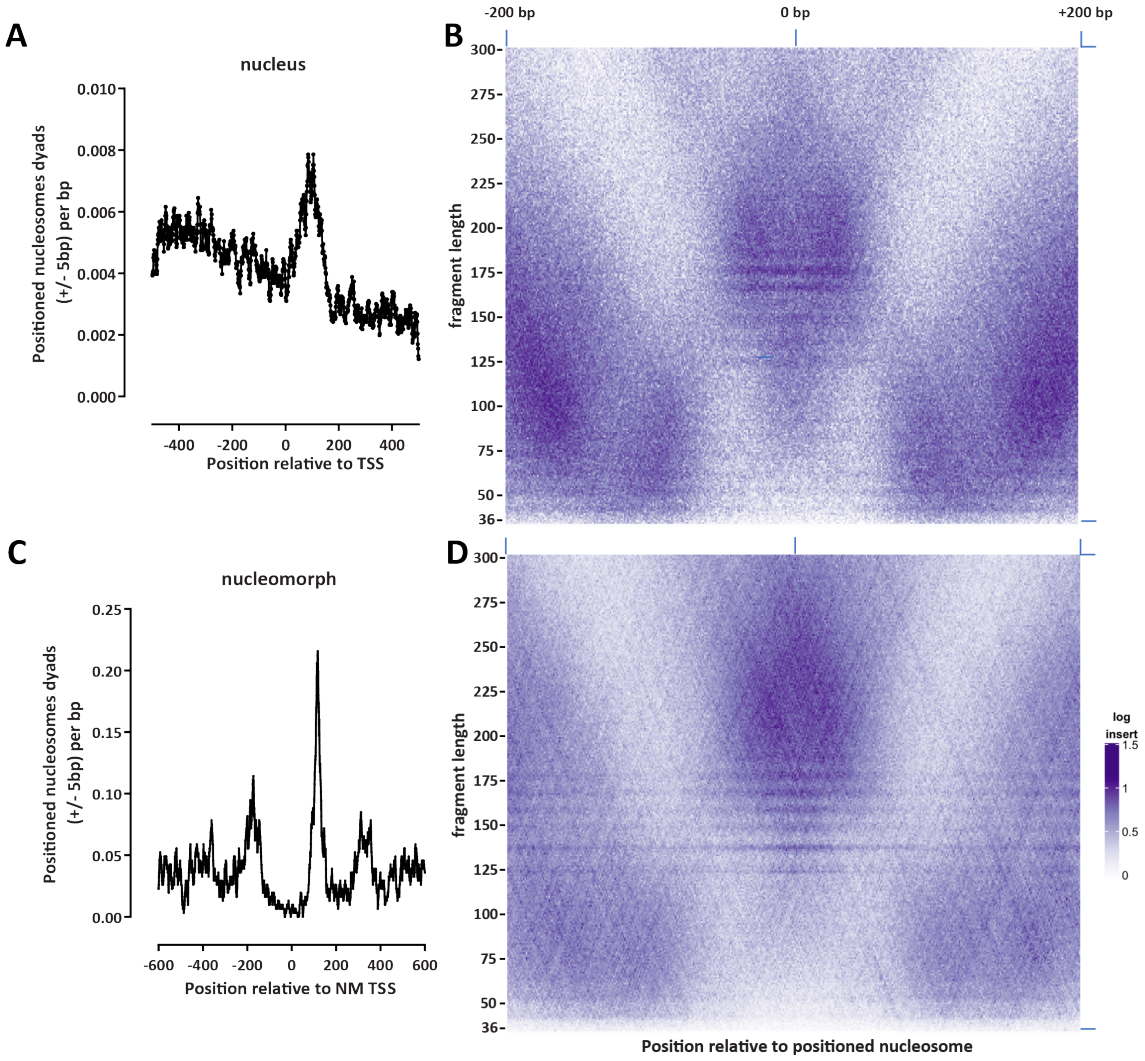


Figure 2: Nucleosome positioning in the *B. natans* nuclear and nucleomorph genomes. (A) Location of positioned nucleosomes (determined by NucleoATAC) relative to annotated TSSs in the *B. natans* nucleus (shown are dyad positions extended by ±5 bp) (B) V-plot of ATAC-seq fragment distribution around positioned nucleosomes in the nucleus. (C) Location of positioned nucleosomes (determined by NucleoATAC) relative to annotated TSSs in the *B. natans* nucleomorph (shown are dyad positions extended by ±5 bp) (D) V-plot of ATAC-seq fragment distribution around positioned nucleosomes in the nucleomorph.

analyses also suggested pervasive largely uniform transcription levels that also do not change much between conditions[34,35,37], and this is also what is seen at the level of the measurements of active transcription by KAS-seq, with the notable exception of the rDNA genes, which are much more strongly transcribed than the rest of the nucleomorph (and also exhibit elevated accessibility). This suggests the possibility of limited transcriptional regulation in the nucleomorph. However, in the same time nucleomorph promoters exhibit a very prominent upstream nucleosome depleted region and strong degree of nucleosome positioning. How this structure is generated by sequence elements associated with each promoter is at present not known, nor is it known whether these elements merely indicate the location of transcription initiation of there are in fact also elements with regulatory activity that can influence the levels of transcription and this has simply not been observed yet. For detailed dissection of these elements methods for the direct genetic manipulation of nucleomorphs will be needed while chromatin analysis of multiple related species, which might reveal conservation of key motifs, will be helpful. It is also curious that the strong nucleosome positioning is not associated with promoter pausing by the polymerase; the mechanistic details of transcription initiation and initial nucleosome clearance will be fascinating to elucidate.

The presence of strongly positioned promoter-proximal nucleosomes also suggests that nucleosomes in different locations in the nucleomorph may in fact exist in distinct

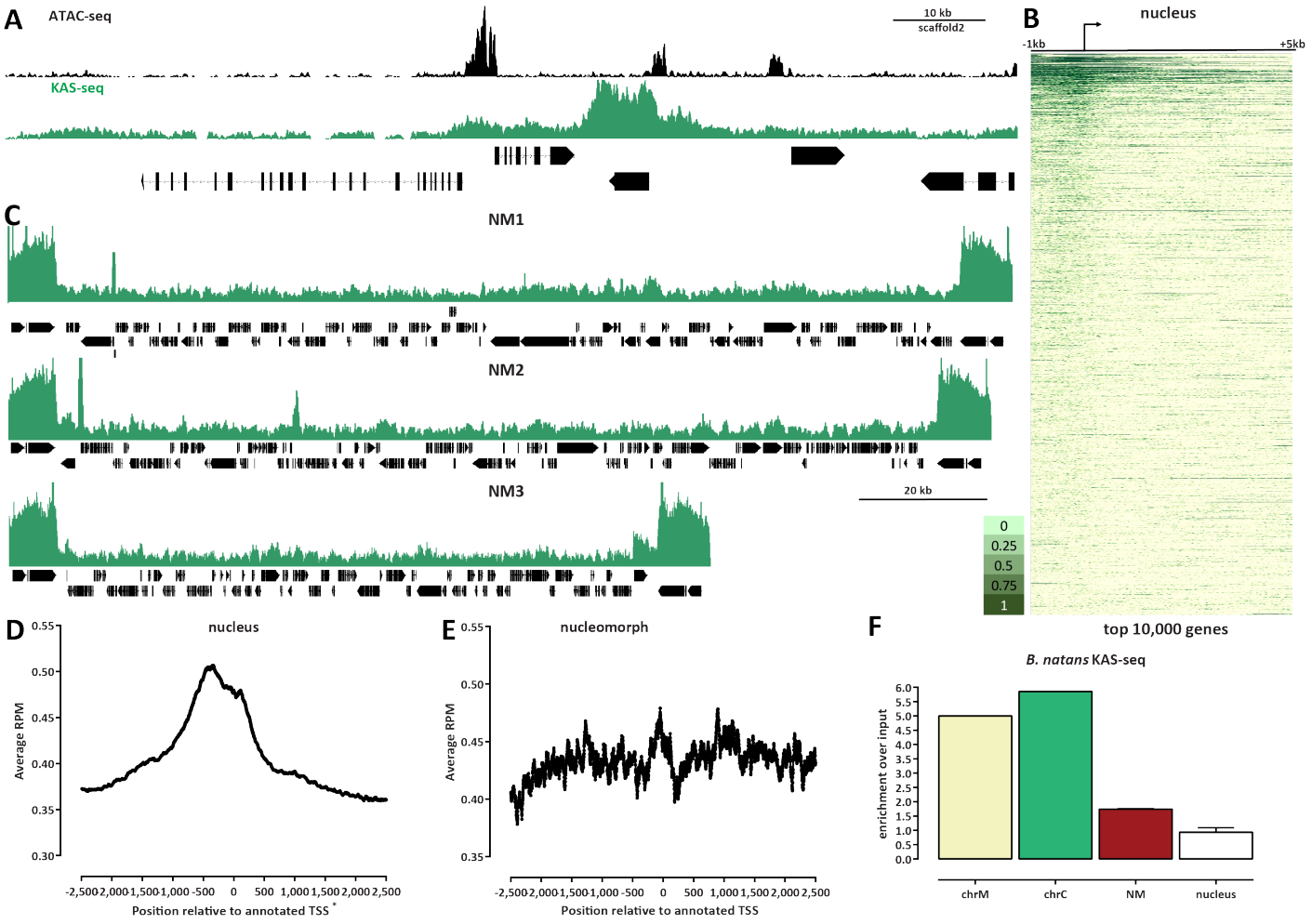


Figure 3: The active transcription landscape of the *B. natans* nuclear and nucleomorph genomes as measured by KAS-seq. (A) KAS-seq and ATAC-seq profiles at a typical nuclear locus. (B) KAS-seq profiles over the top 10,000 (by KAS signal) nuclear genes. (C) KAS-seq profiles over the NM1, NM2 and NM3 nucleomorph chromosomes. (D) Average KAS-seq profile over nuclear gene TSSs (E) Average KAS-seq profile over nucleomorph TSSs (F) Relative enrichment of KAS-seq signal in the different *B. natans* genomic compartments. Shown is the ratio of normalized mapped KAS-seq peaks for each of the compartments relative to the normalized mapped reads in an input sample (a Hi-C dataset mapped in a single-end format)

chromatin states, but what these might be given the lack of the classical histone code amino-acid residues in the nucleomorph histones is a mystery. There have been only limited studies of the nucleomorph proteome in that past [42] and the posttranslation modifications of nucleomorph histones are yet to be studied. The difference in nucleosome protection fragment lengths between the nuclear and the nucleomorph compartment point the possibility that the nucleomorph may also contain a distinct linker histone(s); this issues is to be clarified in the future.

Finally, it will be instructive to compare chromatin organization across the different nucleomorph-bearing groups. Nucleomorph histones in cryptophytes are considerably closer to the conventional state of most eukaryotes; it will be illuminative to know whether those exhibit elevated accessibility, strong nucleosome positioning, and lack or presence of promoter polymerase pausing.

Methods

B. natans cell culture

Bigelowiella natans strain CCMP2755 starting cultures were obtained from NCMA (National Center for Marine Algae and Microbiota) and cultured in L1-Si media on a 12-h-light:12-h-dark cycle.

ATAC-seq experiments

ATAC-seq experiments were performed following the omni-ATAC protocol [43].

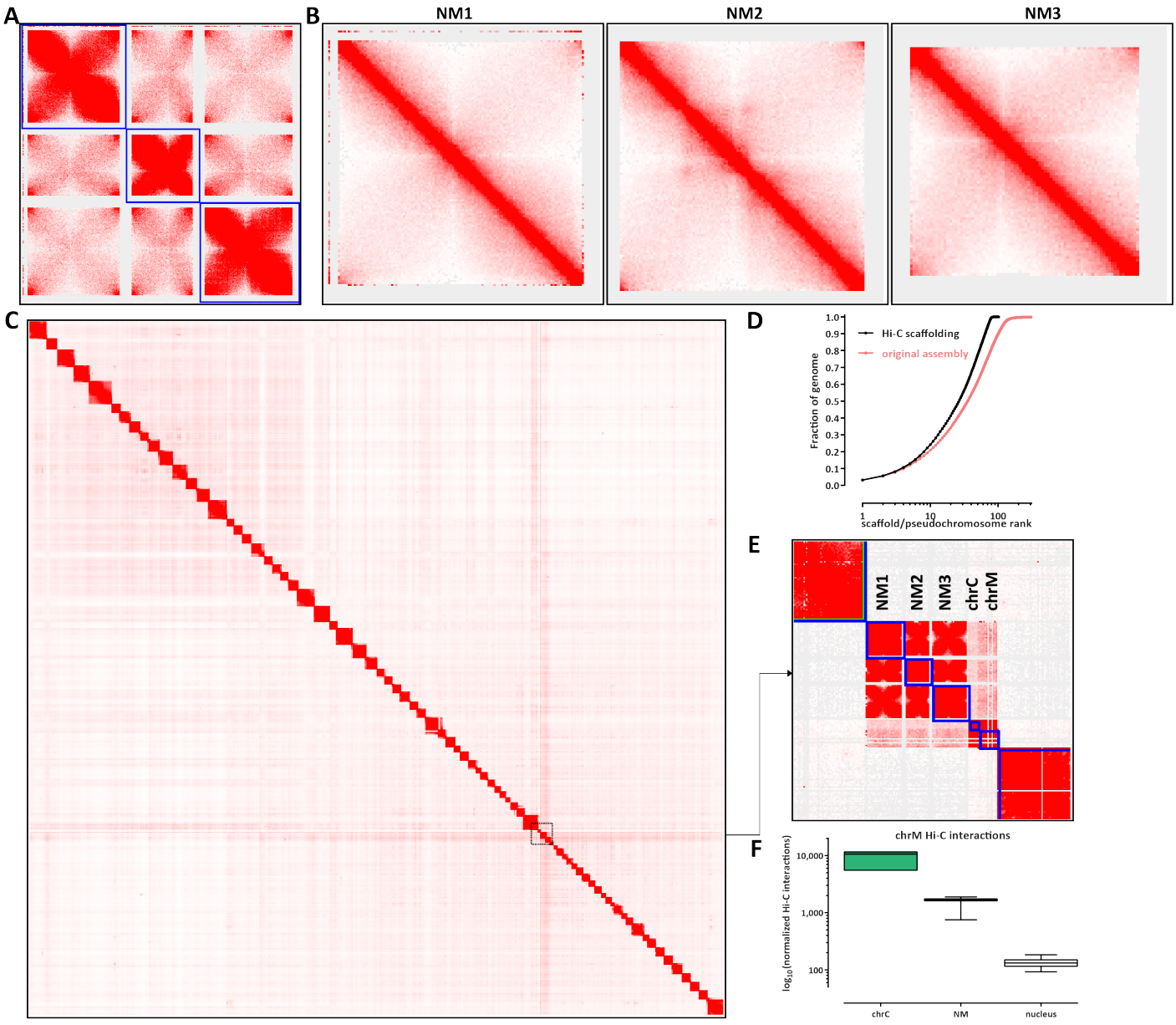


Figure 4: Three-dimensional organization of *B. natans* nucleomorph chromosomes. (A) Hi-C maps (5-kbp resolution) of the three NM chromosomes reveals a network of telomere-to-telomere interactions as the main 3D organizational feature of the nucleomorph. (B) High-resolution (1-kbp) maps of the individual NM chromosomes. (C) and (D) Global scaffolding of the *B. natans* genome. (E) and (F) The *B. natans* mitochondrion interacts physically more often with the endosymbiont compartments than with the nucleus.

Briefly, ~1M *B. natans* cells were centrifuged at 1,000 *g*, then resuspended in 500 μL 1× PBS and centrifuged again. Cells were then resuspended in 50 μL ATAC-RSB-Lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% IGEPAL CA-630, 0.1% Tween-20, 0.01% Digitonin) and incubated on ice for 3 minutes. Subsequently 1 mL ATAC-RSB-Wash buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, 0.01% Digitonin) were added, the tubes were inverted several times, and nuclei were centrifuged at 500 *g* for 5 min at 4 °C.

Transposition was carried out by resuspending nuclei in a mix of 25 μL 2× TD buffer (20 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 20% Dimethyl Formamide), 2.5 μL transposase (custom produced) and 22.5 μL nuclease-free H_2O, and incubating at 37 °C for 30 min in a Thermomixer at 1000 RPM.

Transposed DNA was isolated using the MinElute PCR Purification Kit (Qiagen Cat# 28004/28006), and PCR amplified as previously before[43]. Libraries were purified using the MinElute kit, then sequenced on a Illumina NextSeq

550 instrument as 2×36mers or as 2×75mers.

KAS-seq experiments

KAS-seq experiments were performed as previously described[33] with some modifications.

B. natans cells were pelleted by centrifugation at 1000 *g* for 5 minutes at room temperature, then resuspended in 500 μL of media supplemented with 5 M N_3-kethoxal (final concentration). Cells were incubated at room temperature for 10 minutes, then centrifuged at 1000 *g* for 5 minutes at room temperature to remove the media with the kethoxal, and resuspended in 100 μL cold 1× PBS. Genomic DNA was then extracted using the Monarch gDNA Purification Kit (NEB T3010S) following the standard protocol but with elution using 85 μL 25 mM K_3BO_3 at pH 7.0.

The click reaction was carried out by combining 87.5 μL purified and sheared DNA, 2.5 μL 20 mM DBCO-PEG4-biotin (DMSO solution, Sigma 760749), and 10 μL 10× PBS in a final volume of 100 μL. The reaction was incubated at 37 °C for 90 minutes.

DNA was purified using AMPure XP beads (50 μL for a 100 μL reaction or 100 μL for a 200 μL reaction), beads were washed on a magnetic stand twice with 80% EtOH, and eluted in 130 μL 25mM K_3BO_3.

Purified DNA was then sheared on a Covaris E220 instrument down to ∼150-400 bp size.

For streptavidin pulldown of biotin-labeled DNA, 10 μL of 10 mg/mL Dynabeads MyOne Streptavidin T1 beads (Life Technologies, 65602) were separated on a magnetic stand, then washed with 300 μL of 1× TWB (Tween Washing Buffer; 5 mM Tris-HCl pH 7.5; 0.5 mM EDTA; 1 M NaCl; 0.05% Tween 20). The beads were resuspended in 300 μL of 2× Binding Buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA; 2 M NaCl), the sonicated DNA was added (diluted to a final volume of 300 μL if necessary), and the beads were incubated for ≥15 minutes at room temperature on a rotator. After separation on a magnetic stand, the beads were washed with 300 μL of 1× TWB, and heated at 55 °C in a Thermomixer with shaking for 2 minutes. After removal of the supernatant on a magnetic stand, the TWB wash and 55 °C incubation were repeated.

Final libraries were prepared on beads using the NEBNext Ultra II DNA Library Prep Kit (NEB, #E7645) as follows. End repair was carried out by resuspending beads in 50 μL 1× EB buffer, and adding 3 μL NEB Ultra End Repair Enzyme and 7 μL NEB Ultra End Repair Enzyme, followed by incubation at 20 °C for 30 minutes (in a Thermomixer, with shaking at 1,000 rpm) and then at 65 °C for 30 minutes.

Adapters were ligated to DNA fragments by adding 30 μL Blunt Ligation mix, 1 μL Ligation Enhancer and 2.5 μL NEB Adapter, incubating at 20 °C for 20 minutes, adding 3 μL USER enzyme, and incubating at 37 °C for 15 minutes (in a Thermomixer, with shaking at 1,000 rpm) .

Beads were then separated on a magnetic stand, and washed with 300 μL TWB for 2 minutes at 55 °C, 1000 rpm in a Thermomixer. After separation on a magnetic stand, beads were washed in 100 μL 0.1 × TE buffer, then resuspended in 15 μL 0.1 × TE buffer, and heated at 98 °C for 10 minutes.

For PCR, 5 μL of each of the i5 and i7 NEB Next sequencing adapters were added together with 25 μL 2× NEB Ultra PCR Mater Mix. PCR was carried out with a 98 °C incubation for 30 seconds and 12 cycles of 98 °C for 10 seconds, 65 °C for 30 seconds, and 72 °C for 1 minute, followed by incubation at 72 °C for 5 minutes.

Beads were separated on a magnetic stand, and the supernatant was cleaned up using 1.8× AMPure XP beads.

Libraries were sequenced in a paired-end format on a Illumina NextSeq instrument using NextSeq 500/550 high output kits (2×36 cycles).

Hi-C experiments

Hi-C was carried out using the previously described *in situ* procedure[44] as follows:

B. natans cells were first crosslinked using 37% formaldehyde (Sigma) at a final concentration of 1% for 15 minutes at room temperature. Formaldehyde was then quenched using 2.5 M Glycine at a final concentration of 0.25 M. Cells were subsequently centrifuged at 2,000 *g* for 5 minutes, washed once in 1× PBS, and stored at -80 °C.

Cell lysis was initiated by incubation with 250 μL of cold Hi-C Lysis Buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% Igepal CA630) on ice for 15 minutes, followed by centrifugation at 2,500 *g* for 5 minutes, a wash with 500 μL of cold Hi-C Lysis Buffer, and centrifugation at 2,500 *g* for 5 minutes. The pellet was the resuspended in 50 μL of 0.5% SDS and incubated at 62 °C for 10 minutes. SDS was quenched by adding 145 μL of H_2O and 25 μL of 10% Triton X-100 and incubating at 37 °C for 15 minutes.

Restriction digestion was carried out by adding 25 μL of 10× NEBuffer 2 and 100 U of the MluCI restriction enzyme (NEB, R0538) and incubating for ≥2 hours at 37 °C in a Thermomixer at 900 rpm. The MluCI restriction enzyme was chosen as more suitable for the highly AT-rich nucleomorph genome. The reaction was then incubated at 62 °C for 20 minutes in order to inactivate the restriction enzyme.

Fragment ends were filled in by adding 37.5 μL of 0.4 mM biotin-14-dATP (ThermoFisher Scientific, # 19524-016), 1.5 μL each of 10 mM dCTP, dGTP and dTTP, and 8 μL of 5U/μL DNA Polymerase I Large (Klenow) Fragment (NEB M0210). The reaction was the incubated at 37 °C in a Thermomixer at 900 rpm for 45 minutes.

Fragment end ligation was carried out by adding 663 μL H_2O, 120 μL 10× NEB T4 DNA ligase buffer (NEB B0202), 100 μL of 10% Triton X-100, 12 μL of 10 mg/mL Bovine Serum Albumin (100× BSA, NEB), 5 μL of 400 U/μL T4 DNA Ligase (NEB M0202), and incubating at room temperature for ≥4 hours with rotation.

Nuclei were then pelleted by centrifugation at 2,000 *g* for 5 minutes; the pellet was resuspended in 200 μL ChIP Elution Buffer (1% SDS, 0.1 M $NaHCO_3$), Proteinase K

was added, and incubated at 65 °C overnight to reverse crosslinks.

After addition of 600 μL 1×TE buffer, DNA was sheared using a Covaris E220 instrument. DNA was then purified using the MinElute PCR Purificaiton Kit (Qiagen #28006), with elution in a total volume of 300 μL 1× EB buffer.

For streptavidin pulldown of biotin-labeled DNA, 150 μL of 10 mg/mL Dynabeads MyOne Streptavidin T1 beads (Life Technologies, 65602) were separated on a magnetic stand, then washed with 400 μL of 1× TWB (Tween Washing Buffer; 5 mM Tris-HCl pH 7.5; 0.5 mM EDTA; 1 M NaCl; 0.05% Tween 20). The beads were resuspended in 300 μL of 2× Binding Buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA; 2 M NaCl), the sonicated DNA was added, and the beads were incubated for ≥15 minutes at room temperature on a rotator. After separation on a magnetic stand, the beads were washed with 600 μL of 1× TWB, and heated at 55 °C in a Thermomixer with shaking for 2 minutes. After removal of the supernatant on a magnetic stand, the TWB wash and 55 °C incubation were repeated.

Final libraries were prepared on beads using the NEBNext Ultra II DNA Library Prep Kit (NEB, #E7645) as follows. End repair was carried out by resuspending beads in 50 μL 1× EB buffer, and adding 3 μL NEB Ultra End Repair Enzyme and 7 μL NEB Ultra End Repair Enzyme, followed by incubation at 20 °C for 30 minutes and then at 65 °C for 30 minutes.

Adapters were ligated to DNA fragments by adding 30 μL Blunt Ligation mix, 1 μL Ligation Enhancer and 2.5 μL NEB Adapter, incubating at 20 °C for 20 minutes, adding 3 μL USER enzyme, and incubating at 37 °C for 15 minutes.

Beads were then separated on a magnetic stand, and washed with 600 μL TWB for 2 minutes at 55 °C, 1000 rpm in a Thermomixer. After separation on a magnetic stand, beads were washed in 100 μL 0.1 × TE buffer, then resuspended in 16 μL 0.1 × TE buffer, and heated at 98 °C for 10 minutes.

For PCR, 5 μL of each of the i5 and i7 NEB Next sequencing adapters were added together with 25 μL 2× NEB Ultra PCR Mater Mix. PCR was carried out with a 98 °C incubation for 30 seconds and 12 cycles of 98 °C for 10 seconds, 65 °C for 30 seconds, and 72 °C for 1 minute, followed by incubation at 72 °C for 5 minutes.

Beads were separated on a magnetic stand, and the supernatant was cleaned up using 1.8× AMPure XP beads.

Libraries were sequenced in a paired-end format on a Illumina NextSeq instrument using NextSeq 500/550 high output kits (either 2×75 or 2×36 cycles).

ATAC-seq data processing

Demultipexed FASTQ files were mapped to the `v1.0` assembly for *Bigelowiella natans* CCMP2755 (with the nucleomorph sequence added) as 2×36mers using Bowtie[45] with the following settings: `-v 2 -k 2 -m 1 --best --strata -X 1000`. Duplicate reads were removed using `picard-tools` (version 1.99). Reads mapping to the plastid, mitochondrion and the nucleomoprh were filtered out for the analysis of accessibility in the nuclear genome.

Browser tracks generation, fragment length estimation, TSS enrichment calculations, and other analyses were carried out using custom-written Python scripts (`https://github.com/georgimarinov/GeorgiScripts`).

For the purpose of the analysis of rDNA arrays in nucleomorphs, alignments were carried out with unlimited multimappers with the following settings: `-v 2 -a --best --strata -X 1000`. Normalization of multimappers was performed as previously described[46].

ATAC-seq peak calling

Peak calling was carried out using version 2.1.0 of MACS2[29] with default settings.

Analysis of positioned nucleosomes

Positioned nucleosomes along the whole nucleomorph genome and in the ±500 bp regions around annotated TSSs in the nucleus were identified using NucleoATAC[31] as follows. XXX DETAILS FROM XYNYI XXX

KAS-seq data processing

Demultipexed FASTQ files were mapped to the `v1.0` assembly for *Bigelowiella natans* CCMP2755 (with the nucleomorph sequence added) as 2×36mers using Bowtie[45] with the following settings: `-v 2 -k 2 -m 1 --best --strata -X 1000`. Duplicate reads were removed using `picard-tools` (version 1.99).

Browser tracks generation, fragment length estimation, TSS enrichment calculations, and other analyses were carried out using custom-written Python scripts (`https://github.com/georgimarinov/GeorgiScripts`).

For the analysis of rDNA arrays in nucleomorphs, alignments were carried out with unlimited multimappers with the following settings: `-v 2 -a --best --strata -X 1000`. Normalization of multimappers was performed as previously described[46].

Hi-C data processing and assembly scaffolding

As an initial step, Hi-C sequencing reads were processed against the previously published *B. natans* assembly[41] using the Juicer pipeline[47] for analyzing Hi-C datasets (version 1.8.9 of Juicer Tools).

The resulting Hi-C matrices were then used as input to the 3D DNA pipeline[40] for automated scaffolding with the following parameters: `--editor-coarse-resolution 5000 --editor-coarse-region 5000 --polisher-input-size 100000 --polisher-coarse-resolution 1000 --polisher-coarse-region 300000 --splitter-input-size 100000 --splitter-coarse-resolution 5000`

`--splitter-coarse-region 300000 --sort-output --build-gapped-map -r 10 -i 5000`.

Manual correction of obvious assembly and scaffolding errors was then carried out using Juicebox[47].

After finalizing the scaffolding, Hi-C reads were reprocessed against the new assembly using the Juicer pipeline.

Author contributions

G.K.M. conceptualized the study, performed cell culture, ATAC-seq, KAS-seq and Hi-C experiments and analyzed data. X.C. carried out nucleosome positioning analysis. T.W. and C.H. provided key reagents. W.J.G., A.K., and A.R.G. supervised the study. G.K.M. wrote the manuscript with input from all authors.

Acknowledgements

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

The authors would like to thank Alexandro E. Trevino and members of the Greenleaf, Kundaje, Grossman and Pringle laboratories for helpful discussion and suggestions regarding this work.

Data Availability

Data associated with this manuscript have been submitted to GEO under accession number XXXX

Code Availability

Custom code used to process the data is available at `https://github.com/georgimarinov/GeorgiScripts` and XXXX.

Competing Interests

The authors declare no competing interests.

References

1. Blanchard JL, Lynch M. 2000. Organellar genes: why do they end up in the nucleus? *Trends Genet* **16**(7):315–320.
2. Moran NA, Bennett GM. 2014. The tiniest tiny genomes. *Annu Rev Microbiol* **68**:195–215.
3. Keeling PJ. 2010. The endosymbiotic origin, diversification and fate of plastids. *Philos Trans R Soc Lond B Biol Sci* **365**(1541):729–748.
4. Dodge JD. 1971. A dinoflagellate with both a mesokaryotic and a eukaryotic nucleus: Part 1 fine structure of the nuclei. *Protoplasma* **73**:145–157.
5. Figueroa RI, Bravo I, Fraga S, Garcés E, Llaveria G. 2009. The life history and cell cycle of *Kryptoperidinium foliaceum*, a dinoflagellate with two eukaryotic nuclei. *Protist* **160**(2):285–300.
6. Nakayama T, Takahashi K, Kamikawa R, Iwataki M, Inagaki Y, Tanifuji G. 2020. Putative genome features of relic green alga-derived nuclei in dinoflagellates and future perspectives as model organisms. *Commun Integr Biol* **13**(1):84–88.
7. Sarai C, Tanifuji G, Nakayama T, Kamikawa R, Takahashi K, Yazaki E, Matsuo E, Miyashita H, Ishida KI, Iwataki M, Inagaki Y. 2020. Dinoflagellates with relic endosymbiont nuclei as models for elucidating organellogenesis. *Proc Natl Acad Sci U S A* **117**(10):5364–5375.
8. Greenwood AD. 1974. The Cryptophyta in relation to phylogeny and photosynthesis. *In: Sanders JV, Goodchild DJ, (Eds.). Electron microscopy 1974*. Canberra: Australian Academy of Sciences. p.566–567.
9. Greenwood AD, Griffiths HB, Santore UJ. 1977. Chloroplasts and cell compartments in Cryptophyceae. *Brit Phycol J* **12**:119.
10. Archibald JM. 2007. Nucleomorph genomes: structure, function, origin and evolution. *Bioessays* **29**(4):392–402.
11. Archibald JM, Lane CE. 2009. Going, going, not quite gone: nucleomorphs as a case study in nuclear genome reduction. *J Hered* **100**(5):582–590.
12. Zauner S, Fraunholz M, Wastl J, Penny S, Beaton M, Cavalier-Smith T, Maier UG, Douglas S. 2000. Chloroplast protein and centrosomal genes, a tRNA intron, and odd telomeres in an unusually compact eukaryotic genome, the cryptomonad nucleomorph. *Proc Natl Acad Sci U S A* **97**(1):200–205.
13. Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, Deng LT, Wu X, Reith M, Cavalier-Smith T, Maier UG. 2001. The highly reduced genome of an enslaved algal nucleus. *Nature* **410**(6832):1091–1096.
14. Tanifuji G, Onodera NT, Wheeler TJ, Dlutek M, Donaher N, Archibald JM. 2011. Complete nucleomorph genome sequence of the nonphotosynthetic alga *Cryptomonas paramecium* reveals a core nucleomorph gene set. *Genome Biol Evol* **3**:44–54.

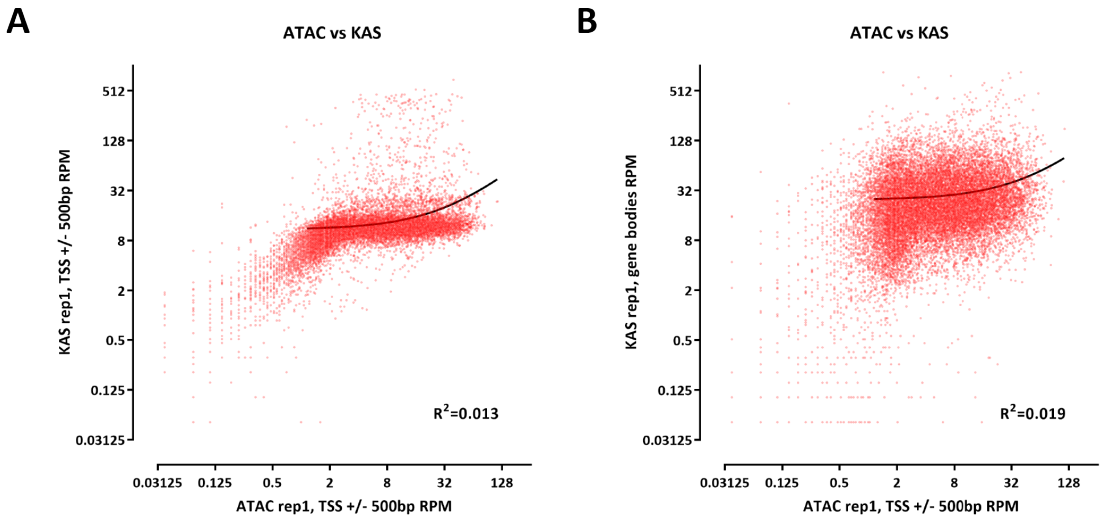
15. Moore CE, Curtis B, Mills T, Tanifuji G, Archibald JM. 2012. Nucleomorph genome sequence of the cryptophyte alga *Chroomonas mesostigmatica* CCMP1168 reveals lineage-specific gene loss and genome complexity. *Genome Biol Evol* **4**(11):1162–1175.
16. Tanifuji G, Onodera NT, Brown MW, Curtis BA, Roger AJ, Ka-Shu Wong G, Melkonian M, Archibald JM. 2014. Nucleomorph and plastid genome sequences of the chlorarachniophyte *Lotharella oceanica*: convergent reductive evolution and frequent recombination in nucleomorph-bearing algae. *BMC Genomics* **15**:374.
17. Gilson PR, Su V, Slamovits CH, Reith ME, Keeling PJ, McFadden GI. 2006. Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. *Proc Natl Acad Sci U S A* **103**(25):9566–9571.
18. Lane CE, van den Heuvel K, Kozera C, Curtis BA, Parsons BJ, Bowman S, Archibald JM. 2007. Nucleomorph genome of *Hemiselmis andersenii* reveals complete intron loss and compaction as a driver of protein structure and function. *Proc Natl Acad Sci U S A* **104**(50):19908–19913.
19. Marinov GK, Lynch M. 2016. Conservation and divergence of the histone code in nucleomorphs. *Biol Direct* **11**(1):18.
20. Postberg J, Forcob S, Chang WJ, Lipps HJ. 2010. The evolutionary history of histone H3 suggests a deep eukaryotic root of chromatin modifying mechanisms. *BMC Evol Biol* **10**:259.
21. Marinov GK, Lynch M. 2015. Diversity and Divergence of Dinoflagellate Histone Proteins. *G3 (Bethesda)* **6**(2):397–422.
22. Jenuwein T, Allis CD. 2001. Translating the histone code. *Science* **293**(5532):1074–1080.
23. Hirakawa Y, Burki F, Keeling PJ. 2011. Nucleus- and nucleomorph-targeted histone proteins in a chlorarachniophyte alga. *Mol Microbiol* **80**(6):1439–1449
24. Eick D, Geyer M. 2013. The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem Rev* **113**(11):8456–8490.
25. Yang C, Stiller JW. 2014. Evolutionary diversity and taxon-specific modifications of the RNA polymerase II C-terminal domain. *Proc Natl Acad Sci U S A* **111**(16):5920–5925.
26. Jones HS, Kawauchi J, Braglia P, Alen CM, Kent NA, Proudfoot NJ. 2007. RNA polymerase I in yeast transcribes dynamic nucleosomal rDNA. *Nat Struct Mol Biol* **14**(2):123–130.
27. Merz K, Hondele M, Goetze H, Gmelch K, Stoeckl U, Griesenbeck J. 2008. Actively transcribed rRNA genes in *S. cerevisiae* are organized in a specialized chromatin associated with the high-mobility group protein Hmo1 and are largely devoid of histone molecules. *Genes Dev* **22**(9):1190–1204.
28. Conconi A, Widmer RM, Koller T, Sogo JM. 1989. Two different chromatin structures coexist in ribosomal RNA genes throughout the cell cycle. *Cell* **57**(5):753–761.
29. Feng J, Liu T, Qin B, Zhang Y, Liu XS. 2012. Identifying ChIP-seq enrichment using MACS. *Nat Protoc* **7**(9):1728–1740.
30. Shipony Z, Marinov GK, Swaffer MP, Sinnott-Armstrong NA, Skotheim JM, Kundaje A, Greenleaf WJ. 2020. Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *Nat Methods* **17**(3):319–327.
31. Schep AN, Buenrostro JD, Denny SK, Schwartz K, Sherlock G, Greenleaf WJ. 2015. Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Res* **25**(11):1757–1770.
32. Henikoff JG, Belsky JA, Krassovsky K, MacAlpine DM, Henikoff S. 2011. Epigenome characterization at single base-pair resolution. *Proc Natl Acad Sci U S A* **108**:18318–18323
33. Wu T, Lyu R, You Q, He C. 2020. Kethoxal-assisted single-stranded DNA sequencing captures global transcription dynamics and enhancer activity *in situ*. *Nat Methods* **17**(5):515–523.
34. Tanifuji G, Onodera NT, Moore CE, Archibald JM. 2014. Reduced nuclear genomes maintain high gene transcription levels. *Mol Biol Evol* **31**(3):625–635.
35. Suzuki S, Ishida K, Hirakawa Y. 2016. Diurnal Transcriptional Regulation of Endosymbiotically Derived Genes in the Chlorarachniophyte *Bigelowiella natans*. *Genome Biol Evol* **8**(9):2672–2682.
36. Sanitá Lima M, Smith DR. 2017. Pervasive Transcription of Mitochondrial, Plastid, and Nucleomorph Genomes across Diverse Plastid-Bearing Species. *Genome Biol Evol* **9**(10):2650–2657.
37. Rangsrikitphoti P, Durnford DG. 2019. Transcriptome Profiling of *Bigelowiella natans* in Response to Light Stress. *J Eukaryot Microbiol* **66**(2):316–333.
38. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**(7):1665–1680.
39. Hoencamp C, Dudchenko O, Elbatsh AMO, Brahmachari S, Raaijmakers JA, van Schaik T, Sedeño Cacciatore Á, Contessoto VG, van Heesbeen RGHP, van den Broek B, Mhaskar AN, Teunissen H, St Hilaire BG, Weisz D, Omer AD, Pham M, Colaric Z, Yang Z, Rao SSP, Mitra N, Lui C, Yao W, Khan R, Moroz LL, Kohn A, St Leger J, Mena A, Holcroft K, Gambetta MC, Lim F, Farley E, Stein N, Haddad A, Chauss D, Mutlu AS, Wang MC, Young ND, Hildebrandt E, Cheng HH, Knight CJ, Burnham TLU, Hovel KA, Beel AJ, Mattei PJ, Kornberg RD, Warren WC, Cary G, Gómez-Skarmeta JL, Hinman V, Lindblad-Toh K, Di Palma F, Maeshima

K, Multani AS, Pathak S, Nel-Themaat L, Behringer RR, Kaur P, Medema RH, van Steensel B, de Wit E, Onuchic JN, Di Pierro M, Lieberman Aiden E, Rowland BD. 2021. 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. *Science* **372**(6545):984–989.

40. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**(6333):92–95.
41. Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, Arias MC, Ball SG, Gile GH, Hirakawa Y, Hopkins JF, Kuo A, Rensing SA, Schmutz J, Symeonidi A, Elias M, Eveleigh RJ, Herman EK, Klute MJ, Nakayama T, Oborník M, Reyes-Prieto A, Armbrust EV, Aves SJ, Beiko RG, Coutinho P, Dacks JB, Durnford DG, Fast NM, Green BR, Grisdale CJ, Hempel F, Henrissat B, Höppner MP, Ishida K, Kim E, Kořený L, Kroth PG, Liu Y, Malik SB, Maier UG, McRose D, Mock T, Neilson JA, Onodera NT, Poole AM, Pritham EJ, Richards TA, Rocap G, Roy SW, Sarai C, Schaack S, Shirato S, Slamovits CH, Spencer DF, Suzuki S, Worden AZ, Zauner S, Barry K, Bell C, Bharti AK, Crow JA, Grimwood J, Kramer R, Lindquist E, Lucas S, Salamov A, McFadden GI, Lane CE, Keeling PJ, Gray MW, Grigoriev IV, Archibald JM. 2012. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* **492**(7427):59–65.
42. Hopkins JF, Spencer DF, Laboissiere S, Neilson JA, Eveleigh RJ, Durnford DG, Gray MW, Archibald JM. 2012. Proteomics reveals plastid- and periplastid-targeted proteins in the chlorarachniophyte alga Bigelowiella natans. *Genome Biol Evol* **4**(12):1391–1406.
43. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, Kathiria A, Cho SW, Mumbach MR, Carter AC, Kasowski M, Orloff LA, Risca VI, Kundaje A, Khavari PA, Montine TJ, Greenleaf WJ, Chang HY. 2017. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**(10):959–962.
44. Marinov GK, Trevino AE, Xiang T, Kundaje A, Grossman AR, Greenleaf WJ. 2021. Transcription-dependent domain-scale three-dimensional genome organization in the dinoflagellate *Breviolum minutum*. *Nat Genet* **53**(5):613–617.
45. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**(3):R25.
46. Marinov GK, Wang J, Handler D, Wold BJ, Weng Z, Hannon GJ, Aravin AA, Zamore PD, Brennecke J, Toth KF. 2015. Pitfalls of mapping high-throughput sequencing data to repetitive sequences: Piwi's genomic targets still not identified. *Dev Cell* **32**(6):765–771
47. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**(1):95–98.

Supplementary Materials

Supplementary Figures



Supplementary Figure 1: Relationship between chromatin accessibility and active transcription as measured by KAS-seq in the *B. natans* nuclear genome. (A) Correlation between ATAC-seq signal over promoters and KAS-seq signal over promoters. (B) Correlation between ATAC-seq signal over promoters and KAS-seq signal over gene bodies.