Genome Biology

RESEARCH

The chromatin organization of a chlorarachniophyte nucleomorph genome

Georgi K. Marinov[1]* , Xinyi Chen[2], Tong Wu[3], Chuan He[3,4,5], Arthur R. Grossman[6], Anshul Kundaje[1,7] and William James Greenleaf[1,8,9,10]*

*Correspondence:
GKM359@gmail.com;
wjg@stanford.edu
[1] Department of Genetics,
Stanford University, Stanford,
CA 94305, USA
[10] Chan Zuckerberg Biohub,
San Francisco, CA, USA
Full list of author information
is available at the end of the
article

Abstract

Background: Nucleomorphs are remnants of secondary endosymbiotic events between two eukaryote cells wherein the endosymbiont has retained its eukaryotic nucleus. Nucleomorphs 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 miniaturization among all known eukaryotic genomes. Previous computational studies have suggested that nucleomorph chromatin likely exhibits a number of divergent features.

Results: In this work, we provide the first maps of open chromatin, active transcription, and three-dimensional organization for the nucleomorph genome 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 over its entire length, with few signs of polymerase pausing at transcription start sites (TSSs). At the same time, most nucleomorph TSSs show very strong nucleosome positioning. Chromosome conformation (Hi-C) maps reveal that nucleomorph chromosomes interact with one other at their telomeric regions and show the relative contact frequencies between the multiple genomic compartments of distinct origin that *B. natans* cells contain.

Conclusions: We provide the first study of a nucleomorph genome using modern functional genomic tools, and derive numerous novel insights into the physical and functional organization of these unique genomes.

Background

Endosymbiosis, especially between a eukaryotic host and a prokaryote, is a common event in the evolution of eukaryotes, and subsequent changes in the host and endosymbiont genomes often follow similar general trends. One such trend is the reduction of the endosymbiont's genome due to gene loss and endosymbiotic gene transfer [1, 2] (EGT) into the host's nucleus, the classic example of which are the extremely reduced genomes of plastids and mitochondria that evolved as the bacterial progenitors of these organelles underwent organellogenesis. This trend 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 retention of the plastid of the photosynthetic eukaryotic endosymbiont (as a secondary plastid) while the nucleus of the endosymbiont is lost entirely. However, several notable exceptions to this general rule do exist. One is the dinotoms, the result of an 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 best known from the chlorarachniophytes and the cryptophytes (but may in fact have arisen in other groups too, such as some dinoflagellates [6, 7]). Nucleomorphs retain a vestigial nucleus with a highly reduced but still functional remnant of the endosymbiont's genome [8, 9].

A remarkable feature of chlorarachniophyte and cryptophyte nucleomorphs is that they have evolved independently, from a green and a red alga, respectively, yet their genomes exhibit surprisingly 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 (∼380 kbp for the chlorarachniophyte *B. natans*). All sequenced nucleomorph genomes are organized into three highly AT-rich chromosomes, in which arrays of ribosomal RNA genes form the subtelomeric regions. These genomes are also extremely compressed, exhibiting very little intergenic space between genes, with genes even overlapping on occasions [12].The genes themselves are also often shortened [13–19].

A number of important questions about the biology associated with the extremely reduced nucleomorph genome remain unanswered, including the extent of conservation and divergence of chromatin organization and transcriptional mechanisms of these extremely reduced nuclei relative to that of a convention eukaryotic genome. Previous computational analysis of nucleomorph genome sequences [20] 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 nearly all branches of the eukaryotic tree [21], with the notable exception of dinoflagellates [22]. This is likely because these PTMs are deposited in a highly regulated manner on specific residues of histones, and are then read out by various effector proteins, thus playing crucial 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 (in what is often referred to as "histone code" [23]).

Nucleomorphs appear to be one of the few [20, 22] exceptions 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 encoded by the host nuclear genome and imported from the host's cytoplasm [24]. Sequence analyses of the H3 and H4 proteins show remarkable divergence from the typical amino acid sequence in eukaryotes; specifically, the chlorarachniophyte histones have lost nearly all key histone code residues [20]. Furthermore, the heptad repeats in the C-terminal domain (CTD) tail of the Rpb1 subunit of RNA polymerase II, which are highly

conserved in eukaryotes [25] and key to the their transcriptional cycle and mRNA processing [26], have also been lost [19, 20, 27].

These observations suggest that the nucleomorph chromatin and chromatin-based regulatory mechanisms may be unconventional compared to those of other eukaryotes. For example, nucleomorphs may organize and protect DNA differently than other eukaryotes, nucleomorph promoters may display atypical signatures of nucleosome depletion and positioning, histone modifications, etc., and relation of these marks to transcriptional activity, or they may exhibit unique 3D genomic organization. However, none of these features associated with nucleomorph chromatin or gene expression regulation has been directly studied.
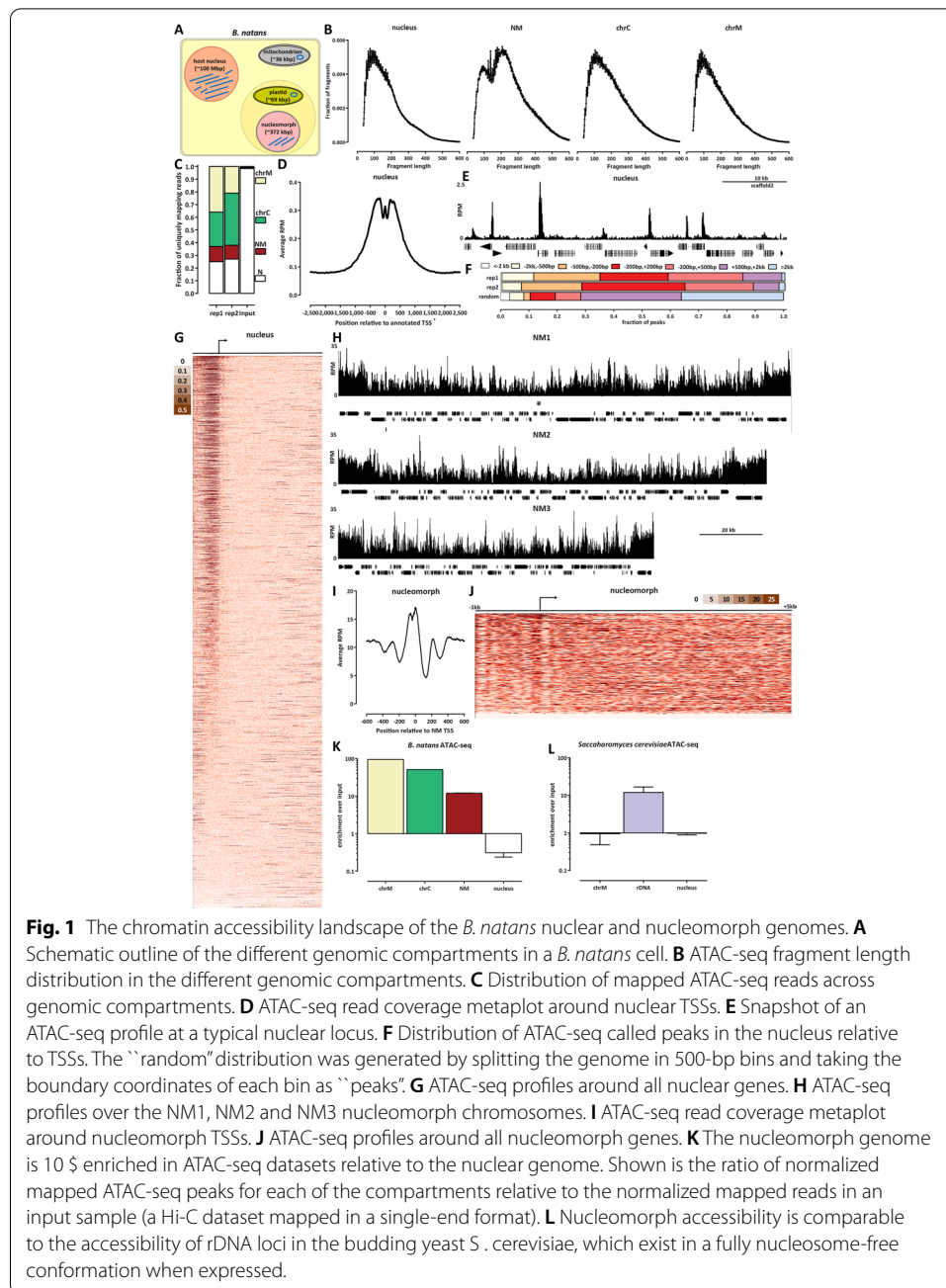
In this work, we map chromatin accessibility, active transcription, and three-dimensional (3D) genome organization in the chlorarachniophyte *Bigelowiella natans* to address these gaps in our knowledge of nucleomorph biology. We find that nucleomorph chromosomes exist in a highly accessible state, reminiscent of what is observed for ribosomal DNA (rDNA) in other eukaryotes, such as budding yeast, where rDNA is thought to be fully nucleosome-free when actively transcribed [28–30]. However, nucleomorph promoters are associated with strongly positioned nucleosomes, and they exhibit a distinct nucleosome-free region upstream of the transcription start site (TSS). Active transcription is nearly uniformly distributed across nucleomorph genomes, with the exception of elevated transcription and chromatin accessibility at the subtelomeric rDNA genes. 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 preferentially interact with each other. Curiously, the genome of the *B. natans* mitochondrion, which derives from the host, exhibits an elevated Hi-C *trans* contact frequency with the genomes of the endosymbiont compartments (the plastid and the nucleomorph) than it does with the host genome. These results provide novel insights into chlorarachniophyte nucleomorph chromatin structure and a framework for future mechanistic studies of transcriptional and regulatory biology in nucleomorphs.

Results

Chromatin accessibility in nucleomorphs

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 (Fig. 1A)—nucleus, nucleomorph, mitochondrion, and plastid—we first examined the fragment length distribution in each (Fig. 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 a more detailed discussion). The mitochondrion and the plastid fragment length distributions are unimodal, consistent with the open DNA structure expected from these compartments which do not contain nucleosomes.

We then examined the distribution of reads across the compartments (Fig. 1C). As expected from the lack of nucleosomal protection over mitochondrial and plastid DNA,

Fig. 1 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.

B. natans ATAC libraries are dominated by reads mapping to those compartments. However, curiously, nucleomorph-mapping reads also represented 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 note that previous reports have identified the nuclear genome of *B. natans* as haploid and the nucleomorph genome as diploid [31]. We observe ratios of reads in our input samples that match these proportions (Additional file 1: Fig. S1).
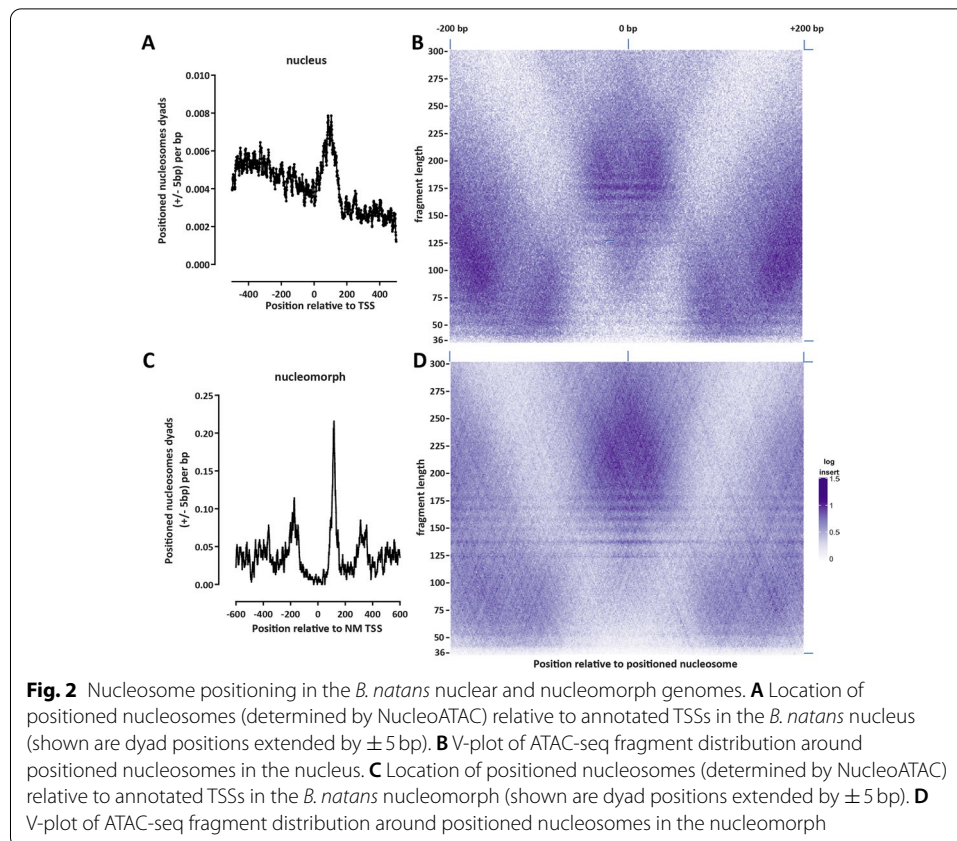
We next turned our attention to ATAC-seq profiles in the nucleus, both to characterize accessibility in the *B. natans* host genome and to verify the quality of the ATAC-seq libraries. Figure 1D shows the average ATAC-seq signal over annotated *B. natans* TSSs; it is enriched over promoters, as expected 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 actual TSSs, but only the sites of translation initiation, with most 5′UTR missing). Examination of browser tracks confirmed the enrichment over TSSs (Fig. 1E) and did not reveal obvious open chromatin sites outside promoters. We carried out peak calling using MACS2 [32], 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 a 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 standard *B. natans* culture conditions, the majority of promoters exhibit an open chromatin configuration (Fig. 1G).

Genome browser examination of ATAC-seq profiles over the nucleomorph genome (Fig. 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 (Fig. 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 (Fig. 1J).

We then quantified the extent of increased accessibility over organellar genomes by calculating the enrichment of ATAC-seq signal relative to the total DNA mass as measured by an input sample. We find that the nucleomorph is $\sim 10\times$ enriched in ATAC-seq libraries, compared to $\sim 100\times$ and $\sim 50\times$ for the mitochondrion and plastid genomes, respectively (Fig. 1K). Notably, this enrichment is comparable to what is observed for rDNA genes in the budding yeast *S. cerevisiae* (Fig. 1L), which are known to exist in an almost fully nucleosome-free configuration when actively transcribed, which is thought to be $\sim 50\%$ of the time [28–30, 33].

Thus, the nucleomorph apparently exists in a highly accessible 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 $\sim 2\times$ and they occupy a minor ($\sim 11\%$) portion of it.

However, 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 [34] over the whole nucleomorph genome and over the 1-kb regions centered on annotated 5′ gene ends in the nucleus. We identified 7251 and 1440 positioned nucleosomes in the nucleus and in the nucleomorph, respectively. The distribution of the nuclear nucleosomes
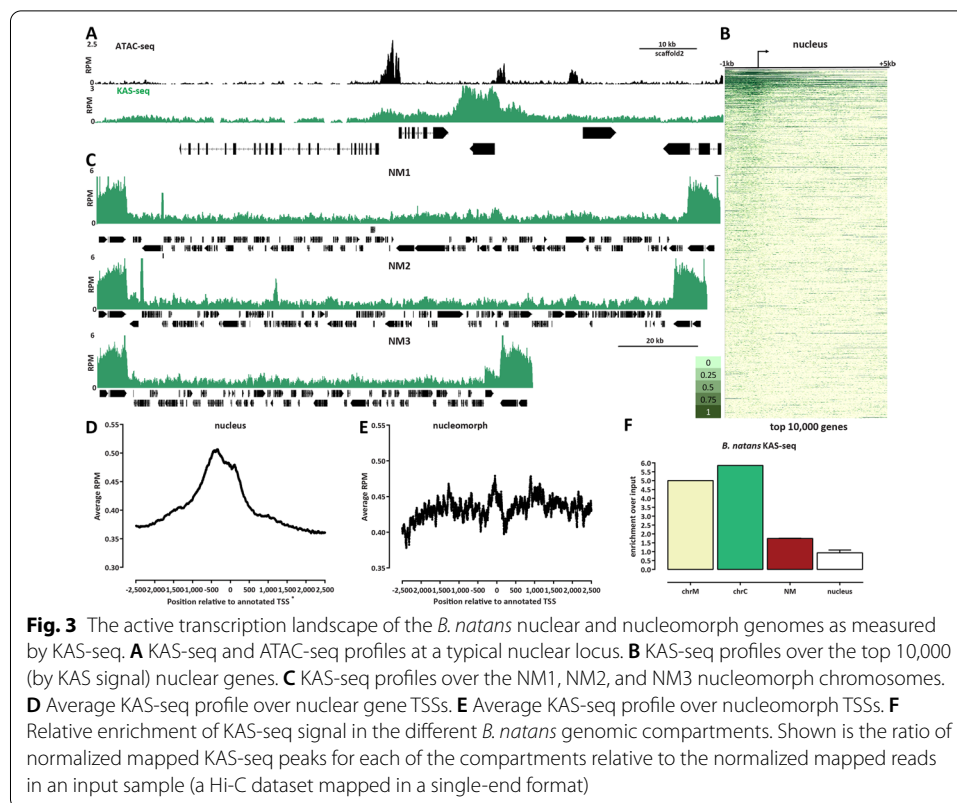
Fig. 2 Nucleosome positioning in the *B. natans* nuclear and nucleomorph genomes. **A** Location of positioned nucleosomes (determined by NucleoATAC) relative to annotated TSSs in the *B. natans* nucleus (shown are dyad positions extended by ± 5 bp). **B** V-plot of ATAC-seq fragment distribution around 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

peaked shortly downstream of TSSs (Fig. 2A), suggesting that nuclear TSSs are also associated with a positioned $+1$ nucleosome. A V-plot [35] analysis showed that the ATAC-seq fragment lengths associated with these nucleosomes are in the 175–200 bp range and that subnucleosomal fragments are located in the immediate vicinity (Fig. 2A). In contrast, in the nucleomorph, we observe three nucleosomes positioned in the vicinity of the TSS ($+1$, $+2$, and -1; Fig. 2C), but ATAC-seq fragment lengths associated with these nucleosomes are larger, in the 200–225 bp range (Fig. 2D).

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 [36], 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 [36], thus KAS-seq is a good proxy for active transcription.
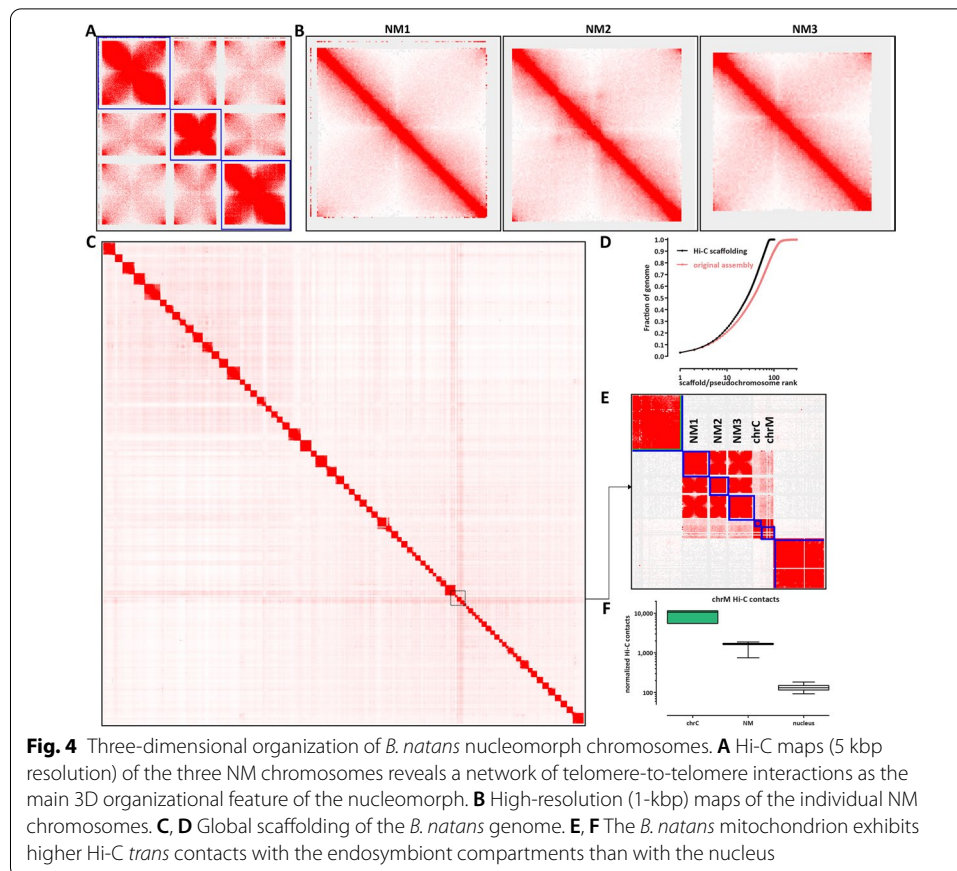
In the *B. natans* nucleus, KAS-seq shows enrichment over promoters and over actively transcribed genes (Fig. 3A, B), as expected based on patterns observed in other eukaryotes [36], indicative of RNA polymerase spending more time near the TSS. However, while we find general concordance between KAS-seq signal and

Fig. 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)

RNA-seq levels (Additional file 1: Fig. S2), we observe only very weak correlation between promoter accessibility and active transcription (Additional file 1: Fig. S3), 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 several previously reported to be highly expressed genes [37–41] (Fig. 3C–E; Additional file 1: Fig. S2). The increased transcription of the rDNA genes is consistent with their higher accessibility observed in ATAC-seq data. We 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 (Fig. 3F) relative to an input sample.

These observations, based on measuring actual active transcription, corroborate previous reports, based on transcriptomic analysis, of high and pervasive transcriptional activity over most of the nucleomorph genome [37–41]. However, rDNA genes were removed in some of these analyses [38] while we identify them as a transcriptional unit existing in a distinct state from the rest of the nucleomorph genome (in the analysis presented here, multimapping reads were retained and normalized, allowing us to measure accessibility and transcription levels over the rDNA genes; see the Methods section for more details).

Fig. 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**, **D** Global scaffolding of the *B. natans* genome. **E**, **F** The *B. natans* mitochondrion exhibits higher Hi-C *trans* contacts with the endosymbiont compartments than with the nucleus

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 [42]). We employed a modified protocol for the highly AT-rich nucleomorph genomes (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 (Fig. 4A, B). In addition, the subtelomeric regions of all nucleomorph chromosomes show high levels of Hi-C contacts with each other, implying a telomeric network of interactions (Fig. 4A). In many eukaryotes, a centromeric interaction network is also observed [43], but enriched interchromosomal interactions in nucleomorphs appear to be 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 [44] of the existing assembly of the *B. natans* nuclear genome [45], which originally consisted of 302 nuclear contigs. Our chromosome-level assembly identifies 79 pseudochromosomes; the

smallest is ~350 kbp, and the largest is ~3 Mbp. This assembly retains only 18 smaller unplaced contigs, the largest being 8753 bp (Fig. 4D).

We made one surprising observation when manually finalizing the chromosome-level assembly—although the mitochondrion is topologically derived from the host (Fig. 1A) and is separated from the nucleomorph and plastid genomes in the endosymbiont by several membranes, it exhibits elevated Hi-C *trans* contacts with both plastid and nucleomorph chromosomes. This preferential enrichment can be visually seen in the Hi-C maps themselves (Fig. 4E) and was also confirmed by a systematic analysis of chrM *trans* contacts with all other chromosomes (Fig. 4F). We also note that we obtain the same result with all available methods for normalizing Hi-C data (Additional file 1: Fig. S4). While both the plastid and the mitochondrial genomes exist in high copy numbers, the nucleomorph genome has only double the copy number of the nuclear genome (as shown by our input samples); thus, the increased likelihood of observing Hi-C contacts between the nucleomorph genome and the mitochondrial genome may represent frequent physical proximity, rather than direct contact, in the cell, which then leads to ligation events with nucleomorph chromosomes during the in situ Hi-C procedure. We note that existing electron microscopy images of the ultrastructural organization of the *B. natans* cells also support the possibility of physical proximity between the nucleomorph and mitochondria [46, 47].

Discussion

This study presents 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.

While it was previously suspected that nucleomorphs are very highly transcriptionally active, we demonstrate that this activity is also reflected at the level of chromatin structure, as nucleomorph chromosomes are much more highly accessible than those in the nucleus. Previous transcriptomic analyses also suggested pervasive largely uniform transcription levels that also do not change much between conditions [38, 39, 41], 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). Taken together, these results suggest the possibility of limited transcriptional regulation in the nucleomorph (which may also be related to the strong divergence of nucleomorph histones H3 and H4 and the absence in them of most key residues involved in regulatory functions). However, nucleomorph promoters exhibit a very prominent upstream nucleosome depleted region and strong degree of nucleosome positioning. How this promoter architecture is generated by sequence elements associated with each promoter is at present not known. It also remains opaque whether these elements merely indicate the location of transcription initiation or if sequence elements with regulatory activity can influence the levels of transcription. To dissect the function of these elements, methods for the direct genetic manipulation of nucleomorphs will be needed. Somewhat

surprisingly, this strong nucleosome positioning at TSSs is not associated with promoter pausing by the polymerase; elucidating the mechanistic details of transcription initiation and initial nucleosome clearance will likely resolve this apparent contradiction.

The presence of strongly positioned promoter-proximal nucleosomes also suggests that nucleosomes in different locations in the nucleomorph may in fact exist in distinct chromatin states, but what these might be given the lack of the classical histone post-translational modifications in the nucleomorph histones is a mystery. There exist only limited studies of the nucleomorph proteome [46], and the posttranslational modifications of nucleomorph histones are yet to be studied. The difference in nucleosome protection fragment lengths between the nuclear and the nucleomorph compartment suggests that the nucleomorph may also contain a distinct linker histone(s); these issues remain to be clarified in the future.

The mechanistic origins of the preferential association between mitochondria and endosymbiont compartments in Hi-C maps are not currently clear. The mitochondrial genome is enclosed by two membranes, while the endosymbiont is enveloped by two membranes, and the plastid inside it by another two [11]. Thus, it is six membranes that separate mitochondrial genome from the plastid genome, and four membranes plus a nuclear membrane exist between it and the nucleomorph chromosomes. More frequent physical proximity between mitochondria and the endosymbiont in the cell is the most likely candidate explanation, as permeabilization of membranes during fixation could allow for crosslinking between chromatin in different compartments. High-resolution imaging approaches [48, 49] should be able to test this hypothesis.

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, and thus determining if these organisms also exhibit elevated accessibility, strong nucleosome positioning, and lack promoter polymerase pausing will be illuminating.

Conclusions

In summary, we investigated the chromatin structure of the highly miniaturized chlorarachniophyte nucleomorph genome (as well as that of the host's nuclear genome). Our results reveal for the first time the unique properties of nucleomorph chromatin, including its elevated physical accessibility and active transcription levels, as well as its three-dimensional genome organization. The experimental approaches we have established here will also be highly useful when applied to other nucleomorph- and endosymbiont-bearing eukaryote lineages.

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 to 12-h-dark cycle.

ATAC-seq experiments

ATAC-seq experiments were performed following the omniATAC protocol [50].

Briefly, ~1 M *B. natans* cells were centrifuged at $1000\,g$, then resuspended in $500\,\mu L$ $1\times$ PBS and centrifuged again. Cells were then resuspended in $50\,\mu L$ 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 $3\,min$. Subsequently $1\,mL$ ATACRSB-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 $25\,\mu L$ $2\times$ TD buffer ($20\,mM$ Tris-HCl pH 7.6, $10\,mM$ $MgCl_2$, 20% dimethyl formamide), $2.5\,\mu L$ transposase (custom produced), and $22.5\,\mu L$ 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 [50]. 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 [36] with some modifications.

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

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

DNA was purified using AMPure XP beads ($50\,\mu L$ for a $100\,\mu L$ reaction or $100\,\mu L$ for a $200\,\mu L$ reaction), beads were washed on a magnetic stand twice with 80% EtOH, and eluted in $130\,\mu L$ $25\,mM$ 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, $10\,\mu L$ of $10\,mg/mL$ Dynabeads MyOne Streptavidin T1 beads (Life Technologies, 65602) were separated on a magnetic stand, and then washed with $300\,\mu L$ of $1\times$ 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 $300\,\mu L$ of $2\times$ 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 $300\,\mu L$ if necessary), and the beads were incubated for $\geq15\,min$ at room temperature on a rotator. After separation on a magnetic stand, the beads were washed with $300\,\mu L$ of $1\times$ TWB, and heated at $55\,°C$ in

a Thermomixer with shaking for 2 min. 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 min (in a Thermomixer, with shaking at 1000 rpm) and then at 65 °C for 30 min.

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 min, adding 3 μL USER enzyme, and incubating at 37 °C for 15 min (in a Thermomixer, with shaking at 1000 rpm) .

Beads were then separated on a magnetic stand, and washed with 300 μL TWB for 2 min 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 min.

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 s and 12 cycles of 98 °C for 10 s, 65 °C for 30 s, and 72 °C for 1 min, followed by incubation at 72 °C for 5 min.

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 [51] as follows:

B. natans cells were first crosslinked using 37% formaldehyde (Sigma) at a final concentration of 1% for 15 min 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 2000 g for 5 min, 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 min, followed by centrifugation at 2500 g for 5 min, a wash with 500 μL of cold Hi-C Lysis Buffer, and centrifugation at 2500 g for 5 min. The pellet was the resuspended in 50 μL of 0.5% SDS and incubated at 62 °C for 10 min. 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 min.

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 h 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 min 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 5 U/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 min.

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 $\geq$ 4 h with rotation.

Nuclei were then pelleted by centrifugation at 2000 g for 5 min; 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 Purification 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, and 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 $\geq$ 15 min 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 min. 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 min and then at 65 °C for 30 min.

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 min, adding 3 μL USER enzyme, and incubating at 37 °C for 15 min.

Beads were then separated on a magnetic stand and washed with 600 μL TWB for 2 min 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 min.

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 s and 12 cycles of 98 °C for 10 s, 65 °C for 30 s, and 72 °C for 1 min, followed by incubation at 72 °C for 5 min.

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

Demultiplexed FASTQ files were mapped to the v1.0 assembly for *Bigelowiella natans* CCMP2755 (with the nucleomorph sequence added) as 2 × 36mers using Bowtie [52] with the following settings: -v 2 -k 2 -m 1 --best --strata -X 1000. Duplicate reads were

Marinov *et al. Genome Biology*   (2022) 23:65

Page 14 of 18

removed using picard-tools (version 1.99). Reads mapping to the plastid, mitochondrion and the nucleomorph 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 [53].

ATAC-seq peak calling

Peak calling was carried out using version 2.1.0 of MACS2 [32] 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 [34] as follows. We used the low resolution nucleosome calling program nucleoatac occ with default parameters that requires ATAC-seq data and genomic windows of interest and returns a list of nucleosome positions based on the distribution of ATAC-seq fragment lengths centered at these positions. To cover the whole nucleomorph genome, sliding windows of 1 kbp in steps of 500 bp were taken as inputs, and redundant nucleosome positions were eventually discarded. For nuclear TSSs, 1-kbp windows centered at the TSSs were used as inputs. V plots were made by aggregating unique-mapping ATAC-seq reads centered around the positioned nucleosomes and mapping the density of fragment sizes versus fragment center locations relative to the positioned nucleosomes as previously described [34, 35].

KAS-seq data processing

Demultiplexed FASTQ files were mapped to the v1.0 assembly for *Bigelowiella natans* CCMP2755 (with the nucleomorph sequence added) as 2×36mers using Bowtie [52] 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 [53].

Mappabiltiy tracks generation

In order to estimate unique mappability, genomes from all compartments were tiled with reads of varying sizes (1×25mers, 1×36mers, 1×50mers, 1×75mers, 1×100mers) at every position. The reads were then aligned using Bowtie against the complete index containing all genomes. Mappability at each position was estimated as R/L where R is the

number of mapped reads covering it and *L* is the read length. Mappability normalization for each RPKM (Reads Per Kilobase per Million mapped reads) calculation was applied by multiplying the RPKM values by the reciprocal of its average mappability score.

RNA-seq data processing

RNA-seq datasets for *B. natans* were obtained from GEO accession GSE115762. Reads were mapped using STAR (version 2.5.3a) with the following settings: with the following settings: --limitSjdbInsertNsj10000000--outFilterMultimapNmax

   50--outFilterMismatchNmax999

   --outFilterMismatchNoverReadLmax0.04

   --alignIntronMin10--alignIntronMax1000000--alignMatesGapMax1000000--align-SJoverhangMin8--alignSJDBoverhang   1--sjdbScore1--twopassModeBasic--twopass-1readsN-1. Quantification was carried out using Stringtie [54] (version 2.0).

Hi-C data processing and assembly scaffolding

As an initial step, Hi-C sequencing reads were processed against the previously published *B. natans* assembly [45] using the Juicer pipeline [55] 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 [44] for automated scaffolding with the following parameters: --editor-coarse-resolution5000--editor-coarse-region5000--polisher-input-size 100000--polisher-coarse-resolution1000

   --polisher-coarse-region300000

   --splitter-input-size100000

   --splitter-coarse-resolution5000

   --splitter-coarse-region300000--sort-output--build-gapped-map-r 10-i 5000.

   Manual correction of obvious assembly and scaffolding errors was then carried out using Juicebox [56].

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

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-022-02639-5.

Additional file 1: Fig. S1. Relative genomic copy numbers of the different *B. natans* genomic compartments. Shown are mappability-corrected RPKM values for an input sample for each chromosome/contig. **Fig. S2.** Relationship between transcript levels (based on RNA-seq) and KAS-seq in the nucleomorph. Shown are RNA-seq TPM values (using RNA-seq data from GEO accession GSE115762) and KAS-seq RPM values. Several highly expressed genes such as *Hsp70*, *Hsp90*, *DnaK* and ribosomal RNAs are highlighted. **Fig. S3.** 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. **Fig. S4.** Impact of different Hi-C normalization methods on the quantification of Hi-C *trans* contacts between different compartments. (A) KR normalization (B) No normalization (C) Coverage normalization (VC) (D) Coverage normalization (VC SQRT).

Additional file 2. Review history.

Review history
The review history is available as Additional file 2.

Declarations

Author details

[1]Department of Genetics, Stanford University, Stanford, CA 94305, USA. [2]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA. [3]Department of Chemistry and Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL 60637, USA. [4]Department of Biochemistry and Molecular Biology and Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL 60637, USA. [5]Howard Hughes Medical Institute, The University of Chicago, Chicago, IL 60637, USA. [6]Department of Plant Biology, Carnegie Institution for Science, Stanford, CA 94305, USA. [7]Department of Computer Science, Stanford University, Stanford, CA 94305, USA. [8]Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA 94305, USA. [9]Department of Applied Physics, Stanford University, Stanford, CA 94305, USA. [10]Chan Zuckerberg Biohub, San Francisco, CA, USA.

References

1. Blanchard JL, Lynch M. Organellar genes: why do they end up in the nucleus? Trends Genet. 2000;16(7):315–20.
2. Moran NA, Bennett GM. The tiniest tiny genomes. Annu Rev Microbiol. 2014;68:195–215.
3. Keeling PJ. The endosymbiotic origin, diversification and fate of plastids. Philos Trans R Soc Lond Ser B Biol Sci. 2010;365(1541):729–48.
4. Dodge JD. A dinoflagellate with both a mesokaryotic and a eukaryotic nucleus: Part 1 fine structure of the nuclei. Protoplasma. 1971;73:145–57.
5. Figueroa RI, Bravo I, Fraga S, Garc´es E, Llaveria G. The life history and cell cycle of *Kryptoperidinium foliaceum*, a dinoflagellate with two eukaryotic nuclei. Protist. 2009;160(2):285–300.
6. Nakayama T, Takahashi K, Kamikawa R, Iwataki M, Inagaki Y, Tanifuji G. Putative genome features of relic green alga-derived nuclei in dinoflagellates and future perspectives as model organisms. Commun Integr Biol. 2020;13(1):84–8.
7. Sarai C, Tanifuji G, Nakayama T, Kamikawa R, Takahashi K, Yazaki E, et al. Dinoflagellates with relic endosymbiont nuclei as models for elucidating organellogenesis. Proc Natl Acad Sci U S A. 2020;117(10):5364–75.
8. Greenwood AD. The Cryptophyta in relation to phylogeny and photosynthesis. In: Sanders JV, Goodchild DJ, editors. Electron microscopy 1974. Australian Academy of Sciences: Canberra; 1974. p. 566–7.
9. Greenwood AD, Griffiths HB, Santore UJ. Chloroplasts and cell compartments in Cryptophyceae. Brit Phycol J. 1977;12:119.
10. Archibald JM. Nucleomorph genomes: structure, function, origin and evolution. Bioessays. 2007;29(4):392–402.
11. Archibald JM, Lane CE. Going, going, not quite gone: nucleomorphs as a case study in nuclear genome reduction. J Hered. 2009;100(5):582–90.
12. Williams BA, Slamovits CH, Patron NJ, Fast NM, Keeling PJ. A high frequency of overlapping gene expression in compacted eukaryotic genomes. Proc Natl Acad Sci U S A. 2005;102(31):10936–41.
13. Zauner S, Fraunholz M, Wastl J, Penny S, Beaton M, Cavalier-Smith T, et al. 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. 2000;97(1):200–5.

14. Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, Deng LT, et al. The highly reduced genome of an enslaved algal nucleus. Nature. 2001;410(6832):1091–6.
15. Tanifuji G, Onodera NT, Wheeler TJ, Dlutek M, Donaher N, Archibald JM. Complete nucleomorph genome sequence of the nonphotosynthetic alga Cryptomonas paramecium reveals a core nucleomorph gene set. Genome Biol Evol. 2011;3:44–54.
16. Moore CE, Curtis B, Mills T, Tanifuji G, Archibald JM. Nucleomorph genome sequence of the cryptophyte alga Chroomonas mesostigmatica CCMP1168 reveals lineage-specific gene loss and genome complexity. Genome Biol Evol. 2012;4(11):1162–75.
17. Tanifuji G, Onodera NT, Brown MW, Curtis BA, Roger AJ, Ka-Shu Wong G, et al. Nucleomorph and plastid genome sequences of the chlorarachniophyte Lotharella oceanica: convergent reductive evolution and frequent recombination in nucleomorph-bearing algae. BMC Genomics. 2014;15:374.
18. Gilson PR, Su V, Slamovits CH, Reith ME, Keeling PJ, McFadden GI. Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. Proc Natl Acad Sci U S A. 2006;103(25):9566–71.
19. Lane CE, van den Heuvel K, Kozera C, Curtis BA, Parsons BJ, Bowman S, et al. 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. 2007;104(50):19908–13.
20. Marinov GK, Lynch M. Conservation and divergence of the histone code in nucleomorphs. Biol Direct. 2016;11(1):18.
21. Postberg J, Forcob S, Chang WJ, Lipps HJ. The evolutionary history of histone H3 suggests a deep eukaryotic root of chromatin modifying mechanisms. BMC Evol Biol. 2010;10:259.
22. Marinov GK, Lynch M. Diversity and divergence of dinoflagellate histone proteins. G3 (Bethesda). 2015;6(2):397–422.
23. Jenuwein T, Allis CD. Translating the histone code. Science. 2001;293(5532):1074–80.
24. Hirakawa Y, Burki F, Keeling PJ. Nucleus- and nucleomorph-targeted histone proteins in a chlorarachniophyte alga. Mol Microbiol. 2011;80(6):1439–49.
25. Yang C, Stiller JW. Evolutionary diversity and taxon-specific modifications of the RNA polymerase II Cterminal domain. Proc Natl Acad Sci U S A. 2014;111(16):5920–5.
26. Eick D, Geyer M. The RNA polymerase II carboxy-terminal domain (CTD) code. Chem Rev. 2013;113(11):8456–90.
27. Stump AD, Ostrozhynska K. Selective constraint and the evolution of the RNA polymerase II C-Terminal Domain. Transcription. 2013;4(2):77–86.
28. Jones HS, Kawauchi J, Braglia P, Alen CM, Kent NA, Proudfoot NJ. RNA polymerase I in yeast transcribes dynamic nucleosomal rDNA. Nat Struct Mol Biol. 2007;14(2):123–30.
29. Merz K, Hondele M, Goetze H, Gmelch K, Stoeckl U, Griesenbeck J. 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. 2008;22(9):1190–204.
30. Conconi A, Widmer RM, Koller T, Sogo JM. Two different chromatin structures coexist in ribosomal RNA genes throughout the cell cycle. Cell. 1989;57(5):753–61.
31. Hirakawa Y, Ishida K. Polyploidy of endosymbiotically derived genomes in complex algae. Genome Biol Evol. 2014;6(4):974–80.
32. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. Nat Protoc. 2012;7(9):1728–40.
33. Shipony Z, Marinov GK, Swaffer MP, Sinnott-Armstrong NA, Skotheim JM, Kundaje A, et al. Longrange single-molecule mapping of chromatin accessibility in eukaryotes. Nat Methods. 2020;17(3):319–27.
34. Schep AN, Buenrostro JD, Denny SK, Schwartz K, Sherlock G, Greenleaf WJ. Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. Genome Res. 2015;25(11):1757–70.
35. Henikoff JG, Belsky JA, Krassovsky K, MacAlpine DM, Henikoff S. Epigenome characterization at single base-pair resolution. Proc Natl Acad Sci U S A. 2011;108:18318–23.
36. Wu T, Lyu R, You Q, He C. Kethoxal-assisted single-stranded DNA sequencing captures global transcription dynamics and enhancer activity in situ. Nat Methods. 2020;17(5):515–23.
37. Hirakawa Y, Suzuki S, Archibald JM, Keeling PJ, Ishida K. Overexpression of molecular chaperone genes in nucleomorph genomes. Mol Biol Evol. 2014;31(6):1437–43.
38. Tanifuji G, Onodera NT, Moore CE, Archibald JM. Reduced nuclear genomes maintain high gene transcription levels. Mol Biol Evol. 2014;31(3):625–35.
39. Suzuki S, Ishida K, Hirakawa Y. Diurnal transcriptional regulation of endosymbiotically derived genes in the chlorarachniophyte *Bigelowiella natans*. Genome Biol Evol. 2016;8(9):2672–82.
40. Sanit´a Lima M, Smith DR. Pervasive transcription of mitochondrial, plastid, and nucleomorph genomes across diverse plastid-bearing species. Genome Biol Evol. 2017;9(10):2650–7.
41. Rangsrikitphoti P, Durnford DG. Transcriptome profiling of *Bigelowiella natans* in response to light stress. J Eukaryot Microbiol. 2019;66(2):316–33.
42. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159(7):1665–80.
43. Hoencamp C, Dudchenko O, Elbatsh AMO, Brahmachari S, Raaijmakers JA, van Schaik T, et al. 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. Science. 2021;372(6545):984–9.
44. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science. 2017;356(6333):92–5.
45. Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, et al. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature. 2012;492(7427):59–65.
46. Hopkins JF, Spencer DF, Laboissiere S, Neilson JA, Eveleigh RJ, Durnford DG, et al. Proteomics reveals plastid- and periplastid-targeted proteins in the chlorarachniophyte alga Bigelowiella natans. Genome Biol Evol. 2012;4(12):1391–406.
47. Moestrup O, Sengco M. Ultrastructural studies on *Bigelowiella natans*, gen. et sp. nov., a chlorarachniophyte flagellate. J Phycol. 2001;37:624–46.

Marinov *et al. Genome Biology*   (2022) 23:65

Page 18 of 18

48. Payne AC, Chiang ZD, Reginato PL, Mangiameli SM, Murray EM, Yao CC, et al. In situ genome sequencing resolves DNA sequence and structure in intact biological samples. Science. 2021;371(6532):eaay3446.

49. Mateo LJ, Murphy SE, Hafner A, Cinquini IS, Walker CA, Boettiger AN. Visualizing DNA folding and RNA in embryos at single-cell resolution. Nature. 2019;568(7750):49–54.

50. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods. 2017;14(10):959–62.

51. Marinov GK, Trevino AE, Xiang T, Kundaje A, Grossman AR, Greenleaf WJ. Transcription-dependent domain-scale three-dimensional genome organization in the dinoflagellate *Breviolum minutum*. Nat Genet. 2021;53(5):613–7.

52. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.

53. Marinov GK, Wang J, Handler D, Wold BJ, Weng Z, Hannon GJ, et al. Pitfalls of mapping high-throughput sequencing data to repetitive sequences: Piwi's genomic targets still not identified. Dev Cell. 2015;32(6):765–71.

54. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33(3):290–5.

55. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.

56. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016;3(1):95–8.

57. Marinov GK, Chen X, Wu T, He C, Grossman AR, Kundaje A, Greenleaf WJ. 2022. The chromatin organization of a chlorarachniophyte nucleomorph genome. Datasets. Gene Expression Omnibus GSE181751 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE181751.

58. Marinov GK, Chen X. The chromatin organization of a chlorarachniophyte nucleomorph genome. Github. 2022. https://github.com/georgimarinov/GeorgiScripts.

59. Marinov GK, Chen X. The chromatin organization of a chlorarachniophyte nucleomorph genome. Source code. Github. 2022. https://github.com/chenxy19/nucleomorph.

60. Marinov GK, Chen X. The chromatin organization of a chlorarachniophyte nucleomorph genome. Source code. Zenodo; 2022. https://doi.org/10.5281/zenodo.5877028.

Publisher's Note