

# **Slaying the last unicorn – discovery of histones in the microalga *Nanochlorum eucaryotum***

Valerie WC Soo<sup>1,2</sup> & Tobias Warnecke<sup>1,2\*</sup>

<sup>1</sup>Medical Research Council London Institute of Medical Sciences, London, United Kingdom

<sup>2</sup>Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, London, United Kingdom

\*corresponding author (tobias.warnecke@lms.mrc.ac.uk)

Running title: Histones in *Nanochlorum eucaryotum*

## ABSTRACT

Histones are the principal constituents of eukaryotic chromatin. The four core histones (H2A, H2B, H3, and H4) are conserved across sequenced eukaryotic genomes and therefore thought to be universal to eukaryotes. In the early 1980s, however, a series of biochemical investigations failed to find evidence for histones or nucleosomal structures in the microscopic green alga *Nanochlorum eucaryotum*. If true, derived histone loss in this lineage would constitute an exceptional case that might help us further understand the principles governing eukaryotic gene regulation. To substantiate these earlier reports of histone loss in *N. eucaryotum*, we sequenced, assembled and quantified its transcriptome. Following a systematic search for histone-fold domains in the assembled transcriptome, we detect orthologs to all four core histones. We also find histone mRNAs to be highly expressed, comparable to the situation in other eukaryotes. Finally, we obtain characteristic protection patterns when *N. eucaryotum* chromatin is subjected to micrococcal nuclease digestion, indicating widespread formation of nucleosomal complexes *in vivo*. We conclude that previous reports of missing histones in *N. eucaryotum* were mistaken. By all indications, *N. eucaryotum* has histone-based chromatin characteristic of most eukaryotes.

## INTRODUCTION

Nucleosomes are the fundamental repeat units of eukaryotic chromatin; four core histones (H3, H4, H2A and H2B) assembled into octameric complexes that wrap ~150 bp of DNA. All four core histones were present in the last common ancestor of eukaryotes (Makarova 2005; Talbert *et al.* 2019), have since been retained along distant eukaryotic lineages, and are considered part of the universal, indispensable molecular toolkit of eukaryotes.

This was not always the case. In the pre-genomic era, histones were at one point thought to be absent, for example, from the chromatin of several fungi including *Neurospora crassa* (Dwivedi *et al.* 1969; Leighton *et al.* 1971), *Allomyces arbusculus* (Stumm and Van Went 1968), and *Phycomyces blakesleeanus* (Leighton *et al.* 1971). The chromatin of dinoflagellates was also considered free of histones (Rizzo 1985).

Early claims of missing histones in fungi did not stand up to closer scrutiny: improved biochemical protocols soon found histones in high abundance (Hsiang and Cole 1973; Cohen and Stein 1975). The case of dinoflagellates proved more complicated. Histones do indeed seem to have lost their role as the principal packaging agent of dinoflagellate DNA. Other small basic proteins organize their permanently condensed chromatin (Gornik *et al.* 2012; Janouškovec *et al.* 2017; Irwin *et al.* 2018). Sequencing of multiple dinoflagellate genomes, however, revealed full complements of core histone genes (Marinov and Lynch 2015), which appear to be lowly expressed and whose functional roles remain unclear (Gornik *et al.* 2012).

We are aware of one other putative case of histone loss in eukaryotes. In a series of papers in the early 1980s, histones and nucleosomal structures were reported absent from the microscopic green alga *Nanochlorum eucaryotum* (Wilhelm *et al.* 1982; Zahn 1984). This claim has gone unchallenged since and been reiterated decades later (Bendich and Drlica 2000; Lodé 2012) as a potential example of deviant eukaryotic chromatin architecture.

*N. eucaryotum* belongs to the Trebouxiophyceae (Schlegel 1991; Schreiner *et al.* 1995; Yamamoto *et al.* 2001; Somogyi *et al.* 2013), which also comprise more intensively studied species of the genus *Chlorella*. Isolated from a seawater aquarium that contained material originally sampled off Rovinj, Croatia, *N. eucaryotum* was described as minute in size (~1-

2 $\mu$ m in diameter), with a typical 1:1:1 relationship between nucleus, chloroplast, and mitochondrion (Wilhelm *et al.* 1982). Its chromatin appeared decondensed throughout a closed mitosis. And, despite intensive efforts to detect histones and nucleosomal structures by biochemical means, electron microscopy and comparative hybridization, neither histones nor the classic beads-on-a-string arrangement of nucleosomes along DNA were observed (Wilhelm *et al.* 1982; Zahn 1984).

Here, we sequence the transcriptome of *N. eucaryotum* to show that histones are present, highly expressed, and closely related to histone proteins in other members of the Trebouxiophyceae. We also find evidence to support nucleosome formation, as micrococcal nuclease (MNase) digestion reveals characteristic nucleosomal ladder patterns. Based on this evidence we conclude that *N. eucaryotum* is not a unicorn.

## RESULTS & DISCUSSION

To establish whether histones are present in *N. eucaryotum* we obtained strain SAG 55.87 from the Culture Collection of Algae at the University of Göttingen (Sammlung von Algenkulturen der Universität Göttingen, SAG). This strain is identical to strain UTEX 2502 in the Culture Collection of Algae at the University of Texas at Austin (Somogyi *et al.* 2013), even though the associated species designations are not. Early comparative analyses, both phylogenetic and morphological, triggered a veritable bonanza of valid and invalid taxonomic revisions, as *Nanochlorum eucaryotum* became variously known as *Nannochloris eucaryotum* (Menzel and Wild 1989), the name still associated with UTEX 2502, and *Pseudochloris wilhelmii*, the name currently associated with SAG 55.87 (Somogyi *et al.* 2013). Below, to avoid confusion, we will use the term *Nanochlorum eucaryotum* (SAG 55.87) to refer to our results. SAG 55.87 is the strain on which the pertinent observations about histones were made and which was deposited in SAG by the original authors (C. Wilhelm, personal communication).

## *Detection of histone transcripts in Nanochlorum eucaryotum (SAG 55.87)*

No genome assembly or functional genomic data are available for SAG 55.87. To search for evidence of histones and assess their likely role in the formation of *N. eucaryotum* (SAG 55.87) chromatin, we therefore extracted RNA from a 31-day old *N. eucaryotum* (SAG 55.87) culture and assembled the algal transcriptome *de novo* (see Methods).

We then screened 6-frame translations of all assembled transcripts for histone domains using hidden Markov models from Pfam (see Methods). This search revealed histones domains in several predicted transcripts. Following manual curation, for example to only retain the longest unique polypeptides (see Methods), we used BLAST (blastp) to identify, in an unbiased manner, proteins with the highest similarity to the candidate histones in the non-redundant NCBI database. The top hits for several *N. eucaryotum* (SAG 55.87) candidate histones were annotated core histones from previously sequenced members of the Trebouxiophyceae (including *Auxenochlorella protothecoides*, *Micratinium conductrix*, and *Chlorella variabilis*, Table S1).

To place candidate hits in phylogenetic context, we added putative *N. eucaryotum* (SAG 55.87) histones to a large prior alignment of eukaryotic and archaeal histones (Stevens *et al.* 2020) and subsequently built a phylogenetic tree (see Methods) from *N. eucaryotum* (SAG 55.87) candidate histones and the histones of *C. variabilis* NC64, for which a high quality genome assembly is available. We find candidate *N. eucaryotum* (SAG 55.87) histones that have high similarity to and cluster with *C. variabilis* H2A, H2B, H3, and H4 (Figure 1). The affiliation between *C. variabilis* H4 and its likely *N. eucaryotum* (SAG 55.87) orthologs is less tight than for the other core histones (Figure 1). However, rather than indicative of unusual divergence, this likely reflects the fact that transcript reconstruction in *N. eucaryotum* (SAG 55.87) only retrieved partial rather than full length H4 candidate sequences (61 and 50 amino acids in length).

We conclude that the *N. eucaryotum* (SAG 55.87) genome encodes and expresses a full complement of core histone proteins.

## *Histone transcripts in N. eucaryotum (SAG 55.87) are abundant*

Although we detect all four core histones in the *N. eucaryotum* (SAG 55.87) transcriptome, it is conceivable that their abundance is low, precluding formation of nucleosomes genome-wide, in a manner reminiscent of dinoflagellates (Gornik *et al.* 2012). This would be consistent with earlier failed attempts to detect histones by biochemical means (Wilhelm *et al.* 1982; Zahn 1984).

To assess the abundance of histones, and therefore their likely capacity to act as global organizers of *N. eucaryotum* (SAG 55.87) chromatin, we quantified the relative abundance of reconstructed transcripts, including those that contain histone domains (see Methods). We then considered two metrics: the overall histone expression level (by summing abundance over all relevant transcripts) and the individual histone transcript with the highest expression level. On both counts, *N. eucaryotum* (SAG 55.87) histones must be considered highly expressed when compared to non-histone transcripts (Figure 2). To compare relative histone abundance in *N. eucaryotum* (SAG 55.87) to other species, we *de novo*-assembled transcriptomes of five other eukaryotes, including three algae, from publicly available RNA-Seq data (see Methods). To ensure comparability, raw reads were processed, assembled, and quantified following the same pipeline used for *N. eucaryotum* (SAG 55.87). As expected, histones are highly expressed in all eukaryotes tested. More importantly, there is no indication (Figure 2) that histones are relatively less abundant in *N. eucaryotum* (SAG 55.87).

We conclude that histone genes are likely present in sufficient abundance in *N. eucaryotum* (SAG 55.87) to mediate typical eukaryotic chromatin organization.

#### *Evidence for widespread nucleosome formation in N. eucaryotum (SAG 55.87)*

Even though histones are present and highly expressed it is still conceivable that they fail to form nucleosomal complexes. Nucleosomes in other eukaryotes provide protection from micrococcal nuclease (MNase) digestion, which produces characteristic ladder patterns when the resulting DNA fragments are separated on a gel, with individual rungs of the ladder corresponding to fragments protected by one or consecutively more nucleosomes.

To establish whether nucleosomal complexes form in *N. eucaryotum* (SAG 55.87), we therefore carried out a series of MNase digestion experiments. We find that MNase digestion of *N. eucaryotum* (SAG 55.87) chromatin produces a ladder pattern indistinguishable from

patterns observed in other eukaryotes (Figure 3). We conclude that widespread nucleosome formation occurs in *N. eucaryotum* (SAG 55.87).

In summary, we detect highly abundant mRNAs corresponding to the four core histones in the transcriptome of *N. eucaryotum* (SAG 55.87), which are similar in sequence to the histones of closely related green algae. In all aspects we have investigated, including the formation of nucleosomes inferred from MNase digestion, we find no indication that *N. eucaryotum* (SAG 55.87) is unusual when compared to other eukaryotes. While eukaryotes without histones might yet exist, they remain to be found.

## METHODS

### *Algal strain and growth conditions*

*Pseudochloris wilhelmii*/*Nanochlorum eucaryotum* strain 55.87 was obtained from the Culture Collection of Algae at the University of Göttingen (Sammlung von Algenkulturen der Universität Göttingen, SAG) and grown in liquid culture at room temperature near a window that receives 8 to 12 hours of indirect sunlight each day. Algal growth was monitored microscopically using direct cell counts derived from a haemocytometer. As recommended by SAG, SAG 55.87 was grown in Brackish Water Medium (200 mg/L KNO<sub>3</sub>, 20 mg/L K<sub>2</sub>HPO<sub>4</sub>, 20 mg/L MgSO<sub>4</sub>, 4 mg/L EDTA, 3.5 mg/L FeSO<sub>4</sub>, 0.05 mg/L H<sub>3</sub>BO<sub>3</sub>, 0.01 mg/L MnSO<sub>4</sub>, 0.005 mg/L ZnSO<sub>4</sub>, 0.005 mg/L Co(NO<sub>3</sub>)<sub>2</sub>, 0.005 mg/L Na<sub>2</sub>MoO<sub>4</sub>, 2.5 x 10<sup>-5</sup> mg/L CuSO<sub>4</sub>, 0.005 mg/L vitamin B<sub>12</sub>, 45.5% (v/v) 0.2 µm-filtered natural seawater (Sea Water NSW), and 3% (v/v) autoclaved soil extract (Canna Terra Professional Soil Mix; part number 02-075-050)). All cultures were sealed with parafilm to prevent moisture loss during incubation.

### *Total RNA extraction*

Total RNA was isolated from a 31-day-old *N. eucaryotum* (SAG 55.87) culture consisting of ~1.9 × 10<sup>7</sup> cells in total. To minimize the risk of bacterial contamination, 0.2 µm-filtered lysozyme solution was added into the culture to a final concentration of 0.5 mg/mL two days prior to cell harvest. After harvesting the culture by centrifugation (13,000 ×g, fixed-angle

rotor, 5 min, 4°C), the resulting cell pellet was resuspended in 1 mL of RNA Protection Buffer (1× NEB DNA/RNA Protection Reagent, 1% (w/v) polyvinylpyrrolidone-40, 0.3% (v/v) beta-mercaptoethanol), and was split into two 0.5-mL aliquots. An equal volume (0.5 mL) of autoclaved acid-washed glass beads (425-600 µm diameter) was added into each cell suspension aliquot, and the mixture was mechanically lysed in a Qiagen TissueLyser II at 20 Hz for five cycles of 1 min on/1 min off. Next, cellular debris was pelleted (13,000 ×g, 2 min, 4°C), and the supernatant was transferred into a new microcentrifuge tube containing an equal volume of NEB RNA Lysis Buffer. Following a thorough mixing, the mixture was applied to a NEB gDNA Removal Column to remove genomic DNA. Total RNA was subsequently column-purified as described in the NEB Monarch Total RNA Miniprep Kit, and eluted using 60 µL nuclease-free water. Two technical replicates were carried out for RNA extractions. These were multiplexed for sequencing and reads pooled prior to processing and assembly (see below).

### *RNA sequencing*

RiboMinus Plant Kit for RNA-Seq (Invitrogen) was used to deplete ribosomal RNAs (rRNAs) from the two RNA samples. The rRNA-depleted RNA samples were then used for library preparation (RNA fragmentation, cDNA synthesis, adaptor ligation and indexing) following the protocol of the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina. The final libraries were subjected to 100-bp paired-end sequencing on an Illumina HiSeq2500 v4 sequencer. Library quality and yield were assessed on a Bioanalyser 2100 and Qubit fluorometer, respectively. RNA sequencing data have been deposited in the NCBI Sequence Read Archive with accession PRJNA670301 (SRR12853753 for technical replicate 1, SRR12853754 for technical replicate 2).

### *Genomic DNA extraction*

As part of this project, we also extracted and sequenced genomic DNA from *N. eucaryotum* (SAG 55.87). We do not analyze or discuss DNA sequencing results above but document below how these data were derived so that they can be re-used by others in the future.



Genomic DNA extraction was carried out as described by (Jagielski *et al.* 2017) with minor modifications. A lysozyme-treated *N. eucaryotum* culture (corresponding to  $\sim 7.4 \times 10^7$  cells) was harvested by centrifugation (2,500  $\times g$ , swing-bucket rotor, 20 min, 4°C). The cell pellet was then washed thrice with TE buffer (10 mM Tris-HCl pH 7.5, 0.1 mM EDTA) to remove as much growth medium as possible. Next, the cell pellet was resuspended in 1.5 mL Extraction Buffer (2% (v/v) Triton X-100, 1% (v/v) SDS, 100 mM NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA) and split into two 0.75-mL aliquots. After combining each aliquot with an equal volume of autoclaved acid-washed glass beads (425-600  $\mu m$  diameter), the cell mixture was mechanically lysed in a Qiagen TissueLyser II as described above (“Total RNA extraction”). After pooling the lysate ( $\sim 1.5$  mL). To digest cellular proteins, proteinase K was added to the lysate to a final concentration of 0.15 mg/mL, and the mixture was incubated at 55°C for 1 hour to allow digestion of cellular proteins. Further added to the mixture were 0.75 mL of 5M NaCl and 0.4 mL of pre-warmed CTAB buffer (1% CTAB, 1.4M NaCl, 100 mM Tris-HCl pH 8.0, 10 mM EDTA). After a 10-min incubation at 65°C, nucleic acids were extracted using phenol:chloroform:isoamyl alcohol (25:24:1) three times, and were precipitated with 0.1 volume of 3M sodium acetate (pH 5.5) and 2 volumes of 100% ice-cold ethanol. Following 1 hour of precipitation at -20°C, nucleic acids were centrifuged at 21,000  $\times g$  for 30 min at 4°C. The resulting pellet was air-dried, solubilized in 500  $\mu L$  TE buffer containing 0.12 mg/mL RNase A, and incubated at 37°C for 2.5 hours to allow digestion of cellular RNA. Total genomic DNA was re-extracted and re-precipitated as before. The final DNA pellet was air-dried, and then re-solubilized in 100  $\mu L$  of 10 mM Tris-HCl (pH 8.0).

### *Genome sequencing*

Algal genomic DNA was sequenced on two platforms: Oxford Nanopore and Illumina Hiseq. For the former platform, 500 ng of genomic DNA was used to construct the sequencing library according to the protocol of the Ligation Sequencing Kit (Oxford Nanopore Technologies). The sequencing library was then sequenced on an Oxford Nanopore Optimised Flongle Flow Cell for a period of 24 hours using the MINKNOW software application. For the Illumina platform, 200 ng of the genomic DNA (from the same batch) was used to prepare the sequencing library (400-500 bp) following the recommended protocol of the NEBNext Ultra II FS DNA Library Prep Kit for Illumina. The final library was subjected to 100-bp paired-end sequencing on an Illumina Hiseq2500 v4 sequencer.

Genomic DNA and sequencing libraries were quantified using Qubit fluorometric assays while DNA quality was assessed on a Bioanalyser 2100. Sequencing data for genomic DNA have been deposited in the NCBI Sequence Read Archive under accession PRJNA670458.

### *In situ MNase digestion of algal chromatin*

*N. eucaryotum* (SAG 55.87) chromatin was digested with MNase as described by Potdar *et al.* (2018). Briefly, up to  $9 \times 10^7$  *N. eucaryotum* (SAG 55.87) cells were harvested by centrifugation (2,500  $\times$ g, swing-bucket rotor, 20 min, 4°C), washed once with TMC buffer (50 mM Tris-HCl pH 7.5, 5 mM MgCl<sub>2</sub>, 5 mM CaCl<sub>2</sub>), and then resuspended in 1 mL of TMC buffer. Next, the cell suspension was divided into 0.2 mL aliquots, each of which was added into a 2-mL microcentrifuge tube containing 300 mg sterile acid-washed glass beads (425-600  $\mu$ m diameter). After adding an appropriate amount of MNase (ThermoFisher Scientific; ~90 Units per reaction), the mixture was immediately vortexed for 1 min at the highest setting and was incubated at 37°C for at least 3 min. To halt enzymatic digestion, 20  $\mu$ L of 10 $\times$  STOP buffer (200 mM EDTA and 5% (v/v) SDS) was added into each mixture. The cell lysate was recovered and transferred into a new microcentrifuge tube. To recover the nucleic acid fraction, an equal volume of 2 $\times$  CTAB buffer (200 mM Tris-HCl pH 8, 40 mM EDTA, 2.8 M NaCl, 4% CTAB) was added into the cell lysate and the mixture was allowed to incubate at 65°C for 1 hour. Nucleic acid extraction was subsequently carried out using phenol:chloroform:isoamyl alcohol (25:24:1) as described above ('Genomic DNA extraction'). The resulting nucleic acid fraction was treated with 200 mg RNase A at 37°C for 2 hours, and the algal DNA was finally cleaned up by re-extraction using phenol:chloroform:isoamyl alcohol (25:24:1). To view the extent of MNase digestion of the algal chromatin, the extracted DNA was electrophoresed on 3% agarose gels at 80V for 1 hour.

### *De novo transcriptome assembly*

Reads from the two technical replicates (see above) were combined and trimmed using TrimGalore 0.6.5 (-q 30 --illumina --paired -trim1 --gzip --length 80) with cutadapt v. 2.8 to remove adapters and low quality bases.

De novo transcriptomes were assembled using Trinity (v.2.4.0) and quantified using kallisto (v. 0.46.1; parameters -b 100 for paired-end data and -b 100 -l 180 -s 20 for single-end data).

Comparative transcriptomic datasets were identified by searching NCBI GEO for Illumina (Tru-Seq)-based rRNA-depleted RNA-seq data. Reads were trimmed and transcriptomes de novo assembled and quantified as above and transcripts containing histone folds identified as detailed below. Accession numbers for these datasets, represented in Figure 2, are as follows: *Saccharomyces cerevisiae* (GEO accession: GSM1892898; SRA accession: SRR2517449); *Oryza sativa Japonica* (GEO accession: GSM1585988; SRA accession: SRR1761780); *Chlamydomonas reinhardtii* (GEO accession: GSM3069474; SRA accession: SRR6904722); *Dunaliella tertiolecta* (GEO accession: GSM1821015; SRA accession: SRR2099913); *Chromochloris zofingiensis* (GEO accession: GSM2431131; SRA accession: SRR5117383).

### *Detection of histone mRNAs*

Predicted transcripts were 6-frame translated into a polypeptide library. The standard genetic code was used for translation as we did not expect histones to be encoded by either the chloroplast or mitochondrion. Whenever a stop codon occurred, the upstream and downstream peptides were considered separate library entries. Peptides shorter than 30 amino acids were discarded. This predicted polypeptide library was then searched for hits against three Pfam domain models: PF00125 (comprising all four eukaryotic core histones), PF00538 (H1 linker histone), and PF00808 (archaeal histones/nuclear factor Y) using hmmsearch from the HMMER package (v.3.3). Hits with an E-value <0.001 were considered further and filtered for the longest unique polypeptides.

### *Phylogenetic analysis*

The remaining set of candidate histone-fold proteins were added to a pre-existing alignment of eukaryotic and archaeal histone-fold proteins (Stevens *et al.* 2020) using Mafft v7.310 (-add -reorder) along with annotated histones from *C. variabilis* NC64A obtained from Uniprot. A sub-alignment of candidate *N. eucaryotum* (SAG 55.87) and *C. variabilis* NC64A histones was extracted and used to build a maximum likelihood phylogenetic tree using

RAxML (v8.1.16 -f a -m PROTCATAUTO -N 100) with 100 rapid bootstraps (Kozlov *et al.* 2019).

## ACKNOWLEDGEMENTS

This work was supported by Medical Research Council core funding to TW. We thank members of the Molecular Systems group and Christian Wilhelm for discussions, the Sammlung von Algenkulturen der Universität Göttingen (SAG) for providing strain SAG 55.87, the LMS Genomics facility for library preparation and sequencing and John West for advice on algal culture.

## REFERENCES

- Bendich A. J., Drlica K., 2000 Prokaryotic and eukaryotic chromosomes: what's the difference? *BioEssays* **22**: 481–486.
- Cohen R. J., Stein G. S., 1975 Chromosomal proteins of *Phycomyces blakesleeana*. *Experimental Cell Research* **96**: 247–254.
- Dwivedi R. S., Dutta S. K., Bloch D. P., 1969 Isolation and characterization of chromatin from *Neurospora crassa*. *The Journal of Cell Biology* **43**: 51–58.
- Gornik S. G., Ford K. L., Mulhern T. D., Bacic A., McFadden G. I., Waller R. F., 2012 Loss of Nucleosomal DNA Condensation Coincides with Appearance of a Novel Nuclear Protein in Dinoflagellates. *Current Biology* **22**: 2303–2312.
- Hsiang M. W., Cole R. D., 1973 The isolation of histone from *Neurospora crassa*. *The Journal of Biological Chemistry* **248**: 2007–2013.
- Irwin N. A. T., Martin B. J. E., Young B. P., Browne M. J. G., Flaus A., Loewen C. J. R., Keeling P. J., Howe L. J., 2018 Viral proteins as a potential driver of histone depletion in dinoflagellates. *Nature Communications* **9**: 1535.
- Jagielski T., Gawor J., Bakuła Z., Zuchniewicz K., Żak I., Gromadka R., 2017 An optimized method for high quality DNA extraction from microalga *Prototheca wickerhamii* for genome sequencing. *Plant Methods* **13**: 77.
- Janouškovec J., Gavelis G. S., Burki F., Dinh D., Bachvaroff T. R., Gornik S. G., Bright K. J., Imanian B., Strom S. L., Delwiche C. F., Waller R. F., Fensome R. A., Leander B. S., Rohwer F. L., Saldarriaga J. F., 2017 Major transitions in dinoflagellate evolution

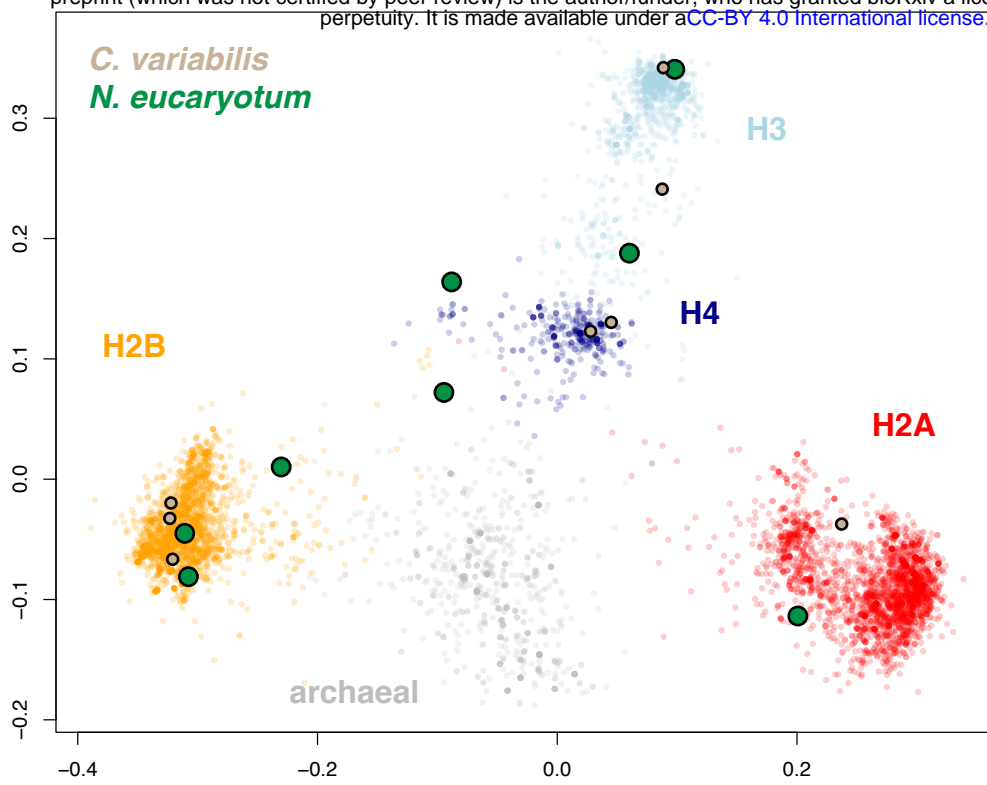
- unveiled by phylotranscriptomics. *Proceedings of the National Academy of Sciences of the United States of America* **114**: E171–E180.
- Kozlov A. M., Darriba D., Flouri T., Morel B., Stamatakis A., 2019 RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**: 4453–4455.
- Leighton T. J., Dill B. C., Stock J. J., Phillips C., 1971 Absence of Histones from the Chromosomal Proteins of Fungi. *Proceedings of the National Academy of Sciences of the United States of America* **68**: 677–680.
- Lodé T., 2012 For Quite a Few Chromosomes More: The Origin of Eukaryotes.... *Journal of Molecular Biology* **423**: 135–142.
- Makarova K. S., 2005 Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Research* **33**: 4626–4638.
- Marinov G. K., Lynch M., 2015 Diversity and Divergence of Dinoflagellate Histone Proteins. *G3: Genes|Genomes|Genetics* **6**: 397–422.
- Menzel K., Wild A., 1989 A Comparative Ultrastructural Investigation of Some Nannochloris Species (Chlorococcales) with Particular Reference to the Systematic Position of Nanochlorum eucaryotum. *Botanica Acta* **102**: 152–158.
- Potdar P., Pinto P., D'Souza N., Joshi P., Malwade A., Sen S., 2018 A Novel Method to Generate MNase Ladders Reveal Rapid Chromatin Remodeling upon Gametogenesis and Mating in Chlamydomonas. *Protist* **169**: 632–644.
- Rizzo P. J., 1985 Histones in protistan evolution. *BioSystems* **18**: 249–262.
- Schlegel M., 1991 Protist evolution and phylogeny as discerned from small subunit ribosomal RNA sequence comparisons. *European Journal of Protistology* **27**: 207–219.
- Schreiner M., Geisert M., Oed M., Arendes J., Güngerich U., Breter H. J., Stüber K., Weinblume D., 1995 Phylogenetic relationship of the green alga Nanochlorum eucaryotum deduced from its chloroplast rRNA sequences. *J Mol Evol* **40**: 428–442.
- Somogyi B., Felföldi T., Solymosi K., Flieger K., Márialigeti K., Böddi B., Vörös L., 2013 One step closer to eliminating the nomenclatural problems of minute coccoid green algae: *Pseudochloris wilhelmii*, gen. et sp. nov. (Trebouxiophyceae, Chlorophyta). *European Journal of Phycology* **48**: 427–436.
- Stevens K. M., Swadling J. B., Hoher A., Bang C., Gribaldo S., Schmitz R. A., Warnecke T., 2020 Histone variants in archaea and the evolution of combinatorial chromatin complexity. *bioRxiv* **2**: 2020.04.13.037952.
- Stumm C., Van Went J. L., 1968 Histone bei dem Phycomyceten Allomyces arbuscula (Butl.). *Experientia* **24**: 1112–1113.
- Talbert P. B., Meers M. P., Henikoff S., 2019 Old cogs, new tricks: the evolution of gene expression in a chromatin context. *Nat. Rev. Genet.* **20**: 283–297.

Wilhelm C., Eisenbeis G., Wild A., Zahn R., 1982 *Nanochlorum eucaryotum*: a Very Reduced Coccoid Species of Marine Chlorophyceae. *Zeitschrift für Naturforschung C* **37**: 107–114.

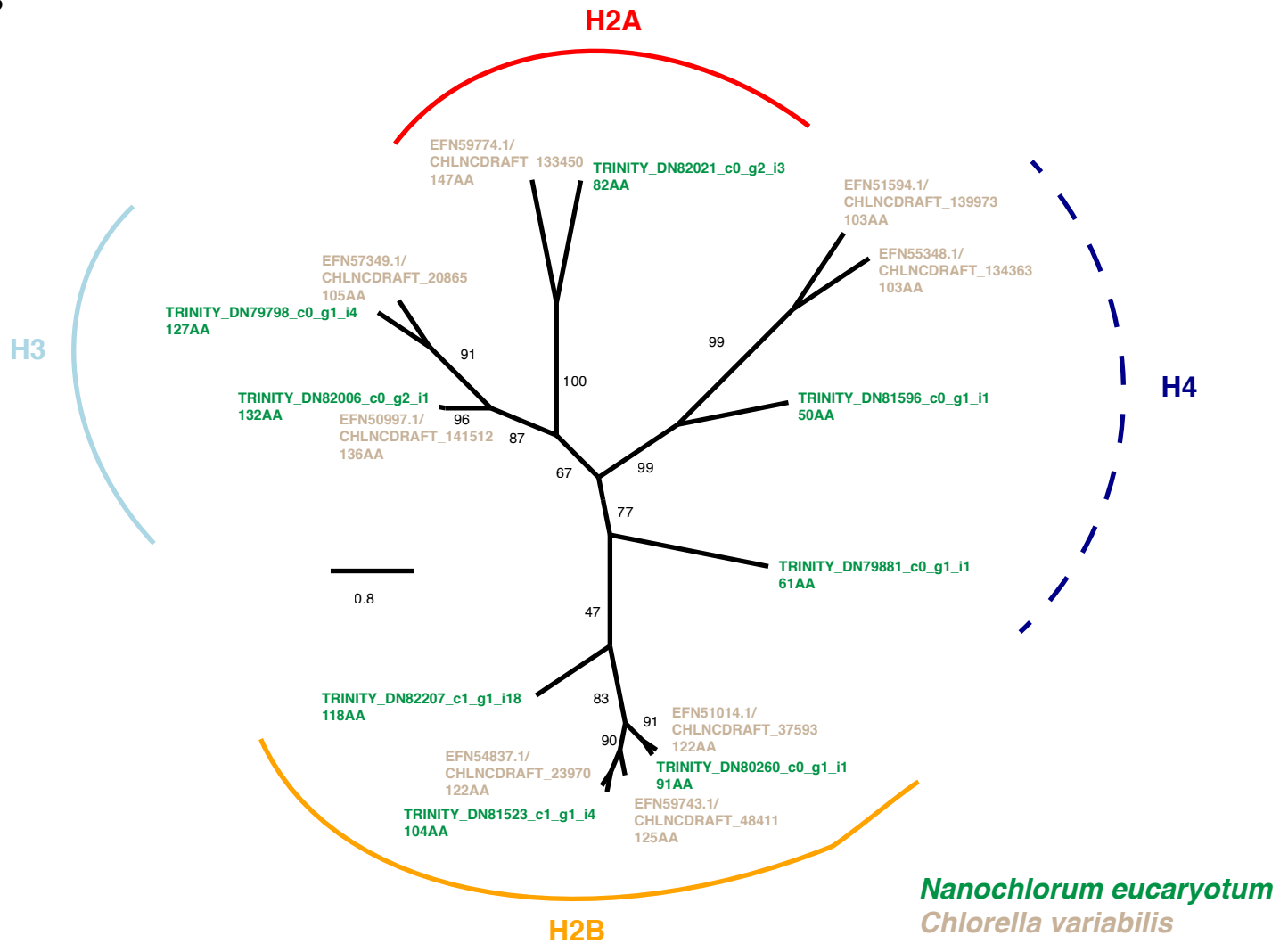
Yamamoto M., Nozaki H., Kawano S., 2001 Evolutionary relationships among multiple modes of cell division in the genus *Nannochloris* (Chlorophyta) revealed by genome size, actin gene multiplicity, and phylogeny. *Journal of Phycology* **37**: 106–120.

Zahn R. K., 1984 A green alga with minimal eukaryotic features: *Nanochlorum eucaryotum*. *Orig Life* **13**: 289–303.

**A**

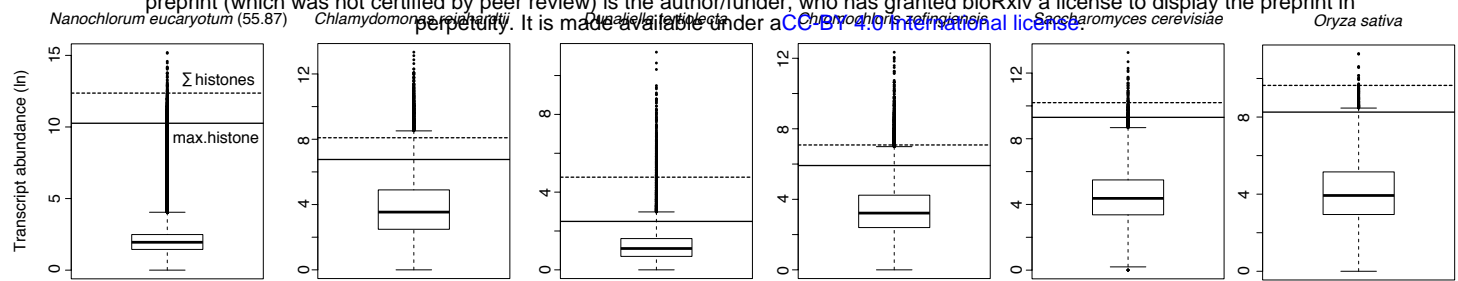


**B**

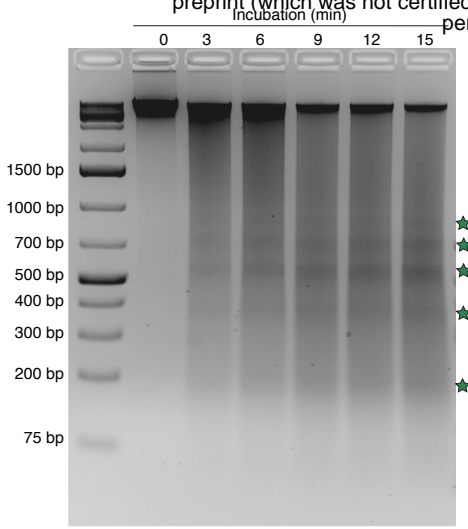


**Figure 1.** *Nanochlorum eucaryotum* (SAG 55.87) candidate histones in phylogenetic context. **(A)** Similarity of eukaryotic and archaeal histones based on classic multidimensional scaling. Histones annotated as H2A, H2B, H3, H4 or of archaeal origin are represented by distinct colours and fall into reasonably distinct clusters. Histones from *C. variabilis* and histone candidates from *Nanochlorum eucaryotum* (SAG 55.87) associate with each of the eukaryotic clusters, suggesting that orthologs for all four core histones are present in these species. The alignment underlying this analysis is provided in Stevens *et al.* (2020). **(B)** Maximum likelihood phylogenetic tree of *C. variabilis* and *Nanochlorum eucaryotum* (SAG 55.87) histones. The length of the relevant polypeptides is provided in amino acids (AA) for each sequence. *Nanochlorum eucaryotum* (SAG 55.87) candidate histones are represented by IDs from the *de novo* transcript assembly (see Table S1).





**Figure 2.** Abundance of histone transcripts compared to the remainder of de novo-assembled transcripts in *Nanochlorum eucaryotum* (SAG 55.87) and five other eukaryotes, including three algae (*C. reinhardtii*, *D. tertiolecta*, *C. zofingiensis*). Abundance of the most highly expressed histone transcript (solid lines) and the summed abundance of all histone transcripts (dashed lines) is indicated for each species. See Methods for details on the underlying RNA-Seq data, assembly and quantification process.



**Figure 3.** MNase digestion of *N. eucaryotum* (SAG 55.87) chromatin. Enriched DNA fragment sizes (green stars) are indicative of nucleosomal protection from MNase digestion. Each reaction shown in this agarose gel was carried out with 90 units of MNase and  $\sim 10^7$  algal cells.