The functional diversity and evolution of nuclear processes

by

Nicholas A. T. Irwin
BSc. Hons, University of British Columbia, 2017

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies
(Botany)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

September 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

The functional diversity and evolution of nuclear processes

submitted by Nicholas A. T. Irwin in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Botany

Examining Committee:

Dr. Patrick J. Keeling, Professor, Department of Botany, UBC

Supervisor

Dr. LeAnn J. Howe, Professor, Department of Biochemistry and Molecular Biology, UBC

Supervisory Committee Member

Dr. Brian S. Leander, Professor, Department of Zoology and Botany, UBC

Supervisory Committee Member

Dr. Ivan Sadowski, Professor, Department of Biochemistry and Molecular Biology, UBC

University Examiner

Dr. James D. Berger, Professor Emeritus, Department of Zoology, UBC

University Examiner

Dr. Joel B. Dacks, Professor, Department of Medicine, University of Alberta

External Examiner

Abstract

The nucleus is a defining characteristic of eukaryotic cells which not only houses the genome but a myriad of processes that function synergistically to regulate cellular activity. Nuclear proteins are key in facilitating core eukaryotic processes such as genome compaction, nucleocytoplasmic exchange, and DNA replication, but the interconnectedness of these processes makes them challenging to dissect mechanistically. Moreover, the antiquity of the nucleus complicates evolutionary analyses, limiting our view of nuclear evolution. Despite this, a comprehensive understanding of the function and evolution of nuclear processes is essential given their central importance in disease, basic cell biology, and eukaryotic evolution. In this dissertation, I argue that insights into nuclear biology and evolution can be obtained by examining eukaryotic diversity rather than relying solely on traditional model organisms. I begin by presenting an introduction to nuclear evolution and diversity, highlighting the existence of nuclear variation across eukaryotes from a systems perspective and underscoring the potential utility of biodiversity in studying nuclear processes (Chapter 1). In the following chapters, I test my hypothesis by examining the function and evolution of different processes in a subset of divergent nuclear systems: namely, chromatin in the dinoflagellate dinokaryon, nuclear pore complexes (NPCs) in the nucleomorphs of chlorarachniophytes and cryptophytes, and DNA replication in the ciliate macronucleus. In Chapter 2, I use an experimental evolutionary approach to investigate the drivers of histone depletion in dinoflagellates, revealing the capacity for viruses to shape cellular evolution and raising questions regarding the subfunctionalization of remnant dinoflagellate histones. In Chapter 3, I reconstruct the NPCs of endosymbiotically acquired nuclei, termed nucleomorphs, in silico, and predict a highly reduced pore structure, suggesting that a complex NPC may not be required for baseline nuclear function. Lastly, in Chapter 4, I examine the diversity of motile DNA replication systems in ciliates, highlighting new models for studying DNA replication and the capacity of cytoskeletal elements to coordinate nuclear organization and processes. Ultimately this work confirms the efficacy of examining diverse nuclear systems, provides insights into the biology and evolution of nuclear processes, and encourages a re-evaluation of how we view and select model organisms.

Lay Summary

Eukaryotic organisms, such as animals, plants, and fungi, are defined by the presence of a nucleus. Within all eukaryotic cells, the nucleus houses DNA which constitutes the genetic information defining all aspects of cellular life. Besides DNA, the nucleus contains a plethora of processes that manipulate how a cell's DNA is interpreted. These interpretations are fundamental to eukaryotic biology as they dictate how organisms respond to their environments, grow, and develop. Exploring nuclear biology is not only key to illuminating cellular function and evolution but is fundamental in understanding disease mechanisms. Research investigating the nucleus mostly relies on yeast and animal experimental models but the complexity of the nucleus warrants investigations into other organisms as they can provide unique perspectives into nuclear function. Here, I examine nuclear systems in diverse eukaryotic organisms, demonstrating how they can provide insights into nuclear function and the evolution and diversity of life.

Preface

Most of the work presented in this thesis has been published or submitted to peer-reviewed journals and was conducted in collaboration with other scientists. Here I outline the contributions and publications that correspond to each of the proceeding chapters.

Chapter 1: Introduction

The introduction constitutes an extended version of a manuscript intended to be a review article. I conducted literature reviews, designed the figures, and wrote the manuscript with input from Patrick Keeling. This chapter will be modified and revised to include recent discoveries stemming from this thesis and submitted as:

Irwin, N. A. T. and Keeling, P. J. Nuclear biology: Mechanistic insights from eukaryotic diversity. *In preparation*.

Chapter 2: Viral proteins as a potential driver of histone depletion in dinoflagellates

This study was originally conceived and designed as an extension of a pilot project carried out in LeAnn Howe's lab during my bachelor's degree. I conducted all of the experiments and analyses with the exception of the Rpb3 chromatin immunoprecipitation (ChIP) and quantitative ChIP-polymerase chain reaction (PCR) experiments which were performed by Benjamin Martin, and the micrococcal nuclease digestions which were completed with the help of LeAnn Howe. Generation of expression constructs, the immunofluorescence experiment, and acquisition of synthetic genetic array (SGA) data were completed during my bachelor's degree (approximately 20% of the data). I wrote the manuscript and created all the figures and datasets with input from all of the authors. A version of this chapter has been published as:

Irwin, N. A. T., Martin, B. J. E., Young, B. P., Browne, M. J. G., Flaus, A., Loewen, C. J. R., Keeling, P. J., and Howe, L. J. (2018). Viral proteins as a potential driver of histone depletion in dinoflagellates. *Nat. Commun. 9*, 1535.

Chapter 3: Extensive reduction of the nuclear pore complex in nucleomorphs

This project was conceived and designed by Patrick Keeling and I. I developed the bioinformatic pipelines, conducted the analyses, designed the figures, and wrote the manuscript with input from Patrick Keeling. A version of this chapter has been published as:

Irwin, N. A. T. and Keeling, P. J. (2019). Extensive reduction of the nuclear pore complex in nucleomorphs. *Genome Biol. Evol. 11*, 678–687.

Chapter 4: Function and evolution of motile DNA replication systems in ciliates

Chapter four was originally initiated as a phylogenetics project by Denis Lynn and William Bourland. After Denis Lynn's passing, I reoriented the project and altered its scope to be about the function and evolution of motile DNA replication systems. Denis Lynn isolated *Licnophora macfarlandi* and generated sequencing libraries from *L. macfarlandi* and *Phacodinium metchnikoffi*. William Bourland isolated and made microscopic observations of *P. metchnikoffi* including the scanning electron microscopy, histology, and 5-ethynyl-2′-deoxyuridine labeling experiments. I conducted all other analyses and experiments, wrote the manuscript, and made the figures (excluding Figure 13, 15, 16, and C1 which were made by William Bourland), with input from all authors. The manuscript has been published as:

Irwin, N. A. T., Pittis, A. A., Mathur, V., Howe, L. J., Keeling, P. J., Lynn, D. H., and Bourland, W. A. (2020) The function and evolution of motile DNA replication systems in ciliates. *Current Biology*. In press.

Table of Contents

List of Tables

List of Figures

List of Symbols & Abbreviations

3HA	Three hemaglutinin epitopes
AMP	Adenosine monophosphate
APC	Anaphase promoting complex
BF	Brightfield
BLAST	Basic local alignment search tool
bp	Base pair
BSA	Bovine serum albumin
BUSCO	Benchmarking universal single copy orthologues
BWA	Burrows Wheeler aligner
cAMP	Cyclic-adenosine monophosphate
cDNA	Complementary deoxyribonucleic acid
CDS	Coding sequence
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation sequencing
CONThreeP	Colpodea, Oligohymenophorea, Nassophorea, Plagiopylea, Prostomatea, and Phyllopharyngea
CRuMs	Collodictyonids, Rigifilida, Mantamonadidae
C-terminus	Carboxy-terminus
Cy3	Cyanine 3
DANPOS	Dynamic analysis of nucleosome position and occupancy by sequencing
DAPI	4′,6-diamidino-2-phenylindole
DIC	Differential interference contrast
DNA	Deoxyribonucleic acid
DVNP	Dinoflagellate viral nucleoprotein
E2F	E2 factor
EDTA	Ethylenediaminetetraacetic acid

EdU	5-Ethynyl-2'-deoxyuridine
eggNOG	Evolutionary genealogy of genes: Non-supervised orthologous groups
EMBO	European Molecular Biology Organization
ER	Endoplasmic reticulum
ESCRT	Endosomal sorting complexes required for transport
ETE3	Environment for tree exploration 3
F	Empirical base frequencies
FACT	Facilitates chromatin transcription
FG	Phenylalanine glycine
FZ	Frontal zone
G4	Gamma distributed rates with four categories
GAL1pr	*GAL1* promoter
GC	Guanine cytosine
GMP	Guanosine monophosphate
GO	Gene ontology
GT	Guanine, thiamine
GTR	General time reversible
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HLP	HU-like protein
HMM	Hidden Markov model
IP	Immunoprecipitation
IPTG	Isopropyl β-d-1-thiogalactopyranoside
IQR	Interquartile range
kbp	Kilo base pair
KS	Kolmogorov-Smirnov
LCA	Last common ancestor
LECA	Last eukaryotic common ancestor

LG	Le and Gascuel
MAFFT	Multiple alignment using fast Fourier transform
Mbp	Mega base pair
MCL	Markov clustering
MCM	Minichromosome maintenance
MCMC	Markov chain Monte Carlo
ML	Maximum likelihood
MMETSP	Marine Microbial Eukaryotic Transcriptome Sequencing Project
MNase	Micrococcal nuclease
MNase-seq	Micrococcal nuclease sequencing
mRNA	Messenger ribonucleic acid
NAPS	Nucleic Acid/Protein Service
NCBI	National Center for Biotechnology Information
NCLDV	Nucleocytoplasmic large DNA viruses
NDR	Nucleosome depleted region
NEB	New England Biolabs
NLS	Nuclear localization signal
NPC	Nuclear pore complex
NSERC	Nautral Science and Engineering Research Council
N-terminus	Amino-terminus
OD_{600}	Optical density at 600 nm
ODU	Optical density unit
ORC	Origin recognition complex
OTU	Operational taxonomic unit
PBS	Phosphate buffered saline
PBS-T	Phosphate buffered saline with tween
PCD	Postciliodesmatophora

PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PIC	Pre-initiation complex
PMSF	Phenylmethanesulfonyl fluoride
PPC	Periplastidal compartment
PSI-BLAST	Position specific iterative basic local alignment search tool
PTM	Post-translational modification
qPCR	Quantitative polymerase chain reaction
R8	Free rate model with eight categories
RAxML	Randomized accelerated maximum likelihood
RB	Replication band
REVIGO	Reduced visualized gene ontology
RNA	Ribonucleic acid
RNAi	RNA interference
RNAP	RNA polymerase
rRNA	Ribosomal ribonucleic acid
RZ	Rear zone
SAL	Spirotrichea, Armophorea, Litostomatea
SBF	SCB binding factor
SCaFoS	Selection, concatenation, and fusion of sequences
SDB	Spheroplast digestion buffer
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	Standard error of the mean
SGA	Synthetic genetic array
SK	Sorbitol potassium
SMORE-seq	Simultaneous mapping of RNA ends by sequencing

S-phase	Synthesis phase
Spiro	Spirotrichea
SV40	Simian virus 40
TATA	Thiamine, adenine, thiamine, adenine
TE	Tris with ethylenediaminetetraacetic acid
TIM	Translocase of the inner membrane
TOM	Translocase of the outer membrane
TRAMP	Trf4/Air2/Mtr4p polyadenylation complex
TSS	Transcription start site
USA	United States of America
UTR	Untranslated region
v/v	Volume per volume
w/v	Weight per volume
YPD	Yeast extract, peptone, dextrose
Δ	Deletion

Acknowledgements

This thesis represents a culmination of not only my PhD but my time at UBC as a whole. My development as a scientist has been hugely defined by the remarkable people I have had the opportunity to interact with since first being introduced to research in April 2013.

I wanted to start by expressing my gratitude to those that have supervised me over the years, particularly Patrick Keeling, LeAnn Howe, and Denis Lynn. I want to thank my PhD supervisor, Patrick Keeling, for sharing his adeptness in navigating the academic world, showing me how to find unique biological problems and communicate their importance, and giving me the opportunity to explore through field work and lab exchanges. I am also grateful to Patrick for giving me the freedom to pursue my own interests regardless of their direct relevance to the goals of the lab. I also wish to thank LeAnn Howe, who served as a member of my supervisory committee, but also mentored me during the formative years of my bachelor's degree. LeAnn's rigor, biochemical expertise, and skill in experimental design has been instrumental for me in defining my own research philosophies and goals and I have deeply enjoyed our multidisciplinary collaborations. Lastly, I want to thank Denis Lynn, to whom this thesis is dedicated, who was my first supervisor, and through his enthusiasm and egalitarianism, convinced me to pursue protistology and research at a time when I was considering alternative career paths. Denis' Zen mindset in both life and work, alongside his kindness and scientific thoroughness has been a major inspiration in my life. Denis tragically passed away on our sampling trip to Calvert Island on 26 June 2018. In his third edition of the "*The Ciliated Protozoa*", Denis described his own advisers in the framework of an academic family, specifically his academic "father", "uncle", and "grandfather". I think this reflected the stages in which his mentors contributed to his career, but also the connection and gratitude he felt to them (and probably also their ages). Perhaps I could describe Patrick, LeAnn, and Denis as my academic father, aunt, and grandfather?

I also want to thank the many mentors and colleagues who have played an important role in teaching me how to do science, both theoretically and practically, and have largely influenced the direction I hope to take my own research. These include (in chronological order), Vittorio

Boscaro, Christopher Howe, Ellen Nisbet, Benjamin Martin, Elisabeth Hehenberger, Yoshihisa Hirakawa, Alexandros Pittis, William Bourland, and Thomas Richards.

I am indebted to all the members of the Keeling lab and UBC community, past and present, who have provided exceptional company, inspired scientific thought, and made my time in the lab enjoyable and memorable. I particularly wish to thank Juan Saldarriaga for introducing me to teaching as a teaching assistant; I appreciated being involved in the planning and execution of the course and enjoyed the good conversations. I also want to thank Brian Leander for his supervision as a member of my supervisory committee and for the Californian atmosphere he imparts on the Biodiversity Research Centre.

Lastly, I want to sincerely thank my family and friends, especially my parents, my brothers Daniel and Liam, and my partner, Varsha. Without their love and support through both happy and heart-wrenching periods, it would have been impossible for me to get where I am today. I will always be grateful for their endless encouragement, interest, and tolerance.

Nicholas Irwin
May 2020

P.S. During the writing of this thesis, while quarantined in my house during the COVID-19 pandemic, I saw the following birds from my window:

American Goldfinch	*Bushtit*	*Spotted Towhee*
American Robin	*Dark-eyed Junco*	*Steller's Jay*
Anna's Humming Bird	*Downy Woodpecker*	*Vaux's Swift*
Bald Eagle	*Hammond's Fly Catcher*	*Violet Green Swallow*
Bewick's Wren	*Hermit Thrush*	*Western Tanager*
Black Capped Chickadee	*North Western Crow*	*Wilson's Warbler*
Black Headed Grosbeak	*Northern Flicker*	*Yellow Rumped Warbler*

For their company during a stark time, I acknowledge them too.

To Dr. Lynn

Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference.[1]

[1] Excerpt from "The road not taken"
Frost, R. 1916. Mountain Interval. Henry Holt and Company.

1. Introduction

1.1 The nucleus as a molecular meshwork and eukaryotic hallmark

The nucleus is a eukaryotic hallmark that serves as a molecular hub, coordinating the biological processes that define complex cells. In contrast to prokaryotes, eukaryotic genomes reside within a nucleus which is surrounded by a nuclear envelope that is continuous with the endoplasmic reticulum (ER) and studded with nuclear pore complexes (NPC). Within the nucleus, eukaryotic genomes are packaged within a dynamic chromatin environment where a myriad of functions occur, such as transcription, DNA replication, DNA repair, and nucleocytoplasmic exchange. However, despite the differences between these activities, nuclear processes are deeply interconnected (Figure 1). From a broad perspective, spatial organization of the genome into topologically associated domains could be considered the foundation of eukaryotic genome regulation and ultimately cellular function (Solovei et al., 2016). But spatial organization is in part modulated by the nuclear cytoskeleton which also interacts with histone octamers and NPCs (Rowley and Corces, 2016; Solovei et al., 2016; Taddei et al., 2006). Histones not only influence chromosome condensation but regulate the recruitment and processivity of transcription and DNA replication machinery (Kurat et al., 2017; Lawrence et al., 2016; Wyrick et al., 1999). Nuclear pore complexes and their constituent subunits mediate nuclear trafficking, but also bind histones, which together further influence genome organization and manipulate gene expression and the DNA damage response (Beck and Hurt, 2017; Ptak and Wozniak, 2016). Moreover, the actions of transcription and DNA replication themselves influence histone deposition and the recruitment of chromatin remodelers, modification enzymes, transcription factors, and architectural proteins, all of which can subsequently affect one another and many aspects of nuclear functionality in general (Khorasanizadeh, 2004; Lai and Pugh, 2017) (Figure 1). This limited snap shot of the nucleus and its processes underscores the challenges associated with dissecting these mechanisms as their roles in eukaryotic biology are convoluted by feedback loops and uncertainty regarding cause and effect relationships. Regardless, disentangling nuclear

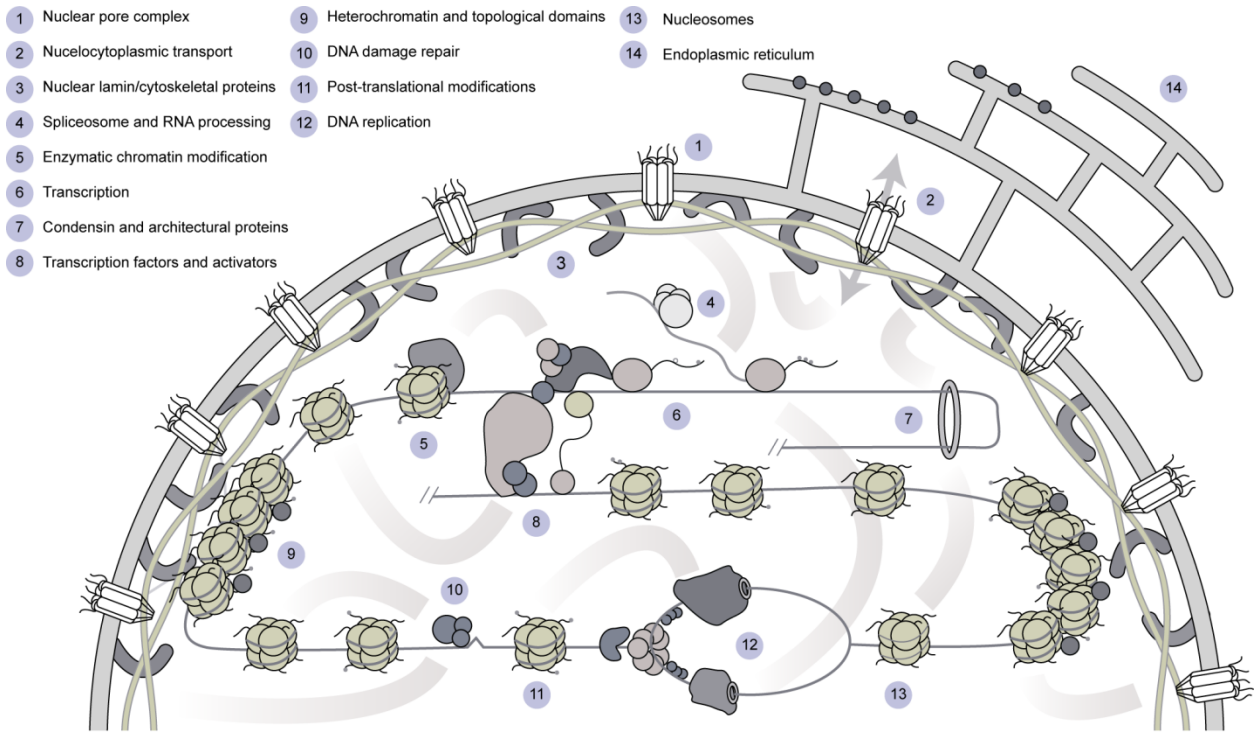


Figure 1. The nucleus is a hub of diverse, complex, and interconnected molecular processes. A simplified map depicting the structures and functions that exist within a canonical, non-dividing, eukaryotic nucleus. Proteins and complexes are not drawn to scale and shapes have been simplified. Examples of interactions, whether physical or indirect, are represented with grey bars. Not all nuclear processes or details are depicted here but those shown are mentioned below and exemplify the nature of nuclear complexity.

---

complexity and developing a comprehensive understanding of the functions, interconnections, and evolutionary origins of these systems is pertinent given their significance regarding disease, evolutionary theory, and basic cell biology.

The brunt of nuclear research has been conducted using yeast and animal models which has driven major advances in our understanding of nuclear processes. Within the fields of molecular and cell biology, these organisms are often considered to be especially distantly related and therefore good comparisons, however, they actually represent a rather limited taxonomic breadth relative to eukaryotic diversity as a whole, which in turn limits our

comparative capacity. Indeed, the presence or absence of orthologues, protein domains, or biochemical pathways in metazoans (i.e., multicellular animals) and yeast is often cited as evidence for universal importance or novelty. In some cases, this may be true, but these assumptions often fail to account for lineage- and species-specific innovation and loss driven by environmental or genetic circumstances, such as ecological interactions and genetic bottlenecks, as well as horizontal gene transfer, and perhaps most importantly, phylogenetic relatedness. Despite diverging around a billion years ago, fungi and metazoans are both part of the Opisthokonta, one of at least six eukaryotic supergroups (Burki et al., 2020; Keeling and Burki, 2019) (Figure 2). Each of these supergroups, contains a remarkable amount of diversity, not only in terms of morphology and phylogeny, but in their cellular and molecular processes. Many of these diverse organisms provide rule-bending exceptions to eukaryotic paradigms, such as the histone depleted dinoflagellate nucleus and genome rearrangements in ciliates, but also provide useful systems for exploring important mechanistic nuances that have been challenging to comprehend in traditional models. Likewise, a deeper understanding of how the nucleus and its processes originated is key to interpreting and even predicting the mechanisms and diversity we observe across eukaryotes, including in animals. Here I review both recent advances and older observations relating to the origin and evolution of the nucleus and its processes and describe nuclear diversity from a systems perspective, highlighting underutilized and unexplored model systems. I discuss unanswered questions relating to nuclear biology and hypothesize that the answers to these questions may be revealed more simply by exploring eukaryotic diversity as opposed to relying solely on traditional model organisms and future technological innovations.

1.2 The origin and evolution of the nucleus and its processes

In order to understand the structure, function, and diversity of the modern eukaryotic nucleus, an evolutionary context outlining its origin and function are essential. Evolutionary hypotheses regarding the origin of the nucleus are both numerous and varied (Martin, 2005). This likely reflects the challenges associated with studying such an ancient event, as experiments become indirect and evidence becomes reliant on often serendipitous observations. A comprehensive hypothesis for the origin of the nucleus must account for the basis of the nuclear envelope, the evolutionary drivers that would justify the compartmentalization of the genome, and the sources of nuclear novelty. To this end, nuclear origin hypotheses tend to fall within two camps, those

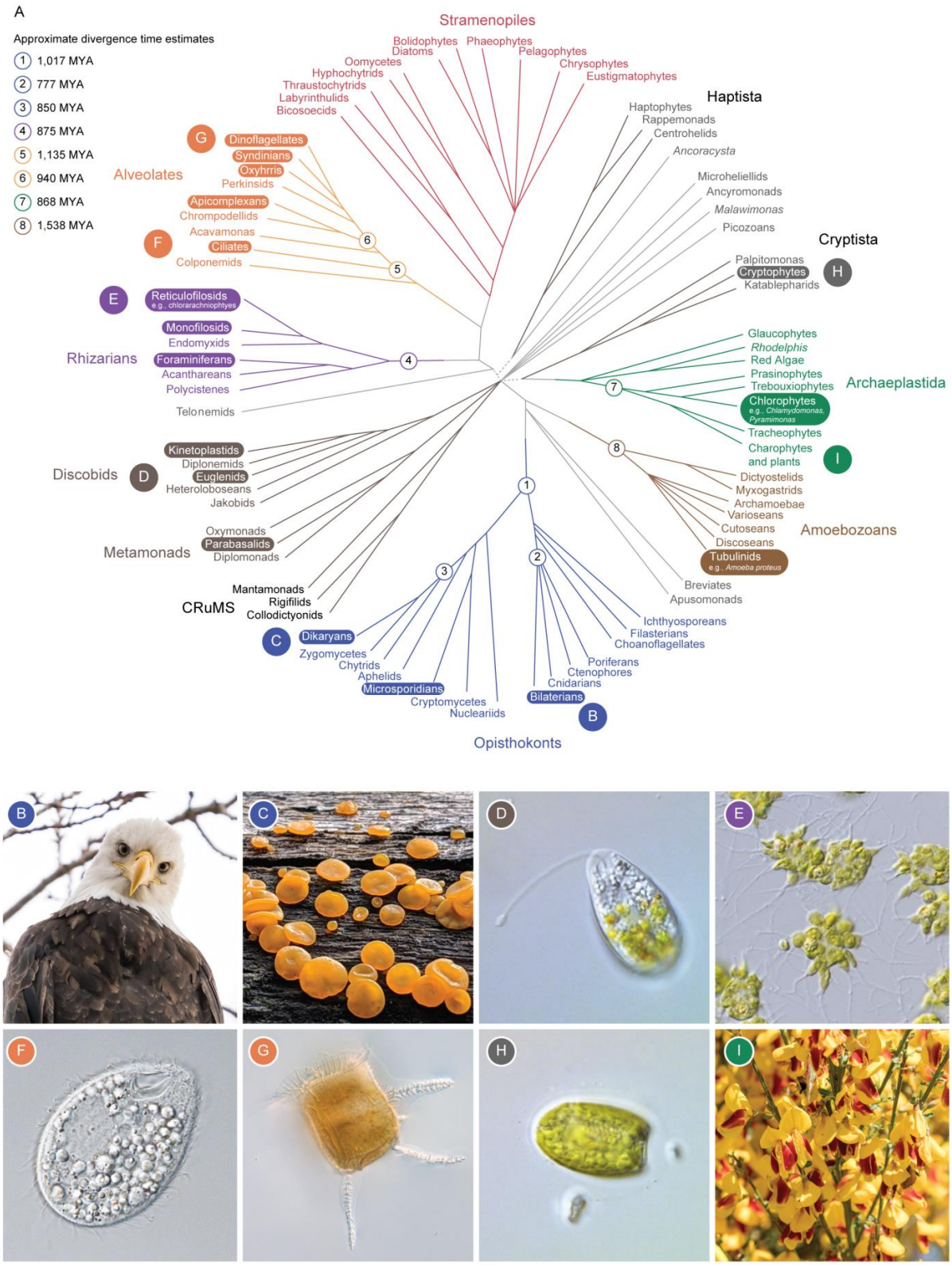


Figure 2. Nuclear diversity exists across the tree of eukaryotes. A. An unrooted phylogenetic tree representing the diversity of eukaryotes, adapted from Keeling and Burki (2019). Major

eukaryotic clades are labeled and lineages discussed below, that exhibit nuclear diversity, are highlighted. Although the topology of the tree reflects relatedness, branch lengths are not indicative of divergence time or evolutionary rates. For orientation, the approximate divergence times of select nodes have been noted using TimeTree (Kumar et al., 2017) which is based on the median divergence time across available molecular dating studies. 1, last common ancestor (LCA) between metazoans and fungi; 2, LCA of metazoans; 3, LCA between Dikarya (i.e., ascomycetes and basidiomycetes) and chytrids; 4, LCA of the Rhizaria supergroup; 5, LCA between ciliates and myzozoans (i.e., apicomplexans and dinoflagellates); 6, LCA between dinoflagellates and apicomplexans; 7, LCA between chlorophytes and land plants; 8, LCA of the Amoebozoa supergroup. Examples of an animal (*Haliaeetus leucocephalus*) (**B**), a fungus (*Dacrymyces palmatus*, courtesy of Liam Irwin) (**C**), a euglenid (*Petalomonas* sp.) (**D**), a chlorarachniophyte (*Amorphochlora amoebiformsi*, courtesy of Dr. Yoshihisa Hirakawa) (**E**), a ciliate (*Tetrahymena* sp.) (**F**), a dinoflagellate (*Ceratocorys horrida*) (**G**), a cryptophyte (unidentified cryptomonad) (**H**), and a plant (*Cytisus scoparius*) (**I**) are shown.

---

purporting an autogenous (i.e., karyogenic) and endosymbiotic (i.e., endokaryogenic) nuclear precursor (Hendrickson and Poole, 2018; Lake and Rivera, 1994). Autogenous hypotheses posit that the nucleus evolved through the elaboration and diversification of pre-existing cellular structures, particularly the endomembrane system (Cavalier-Smith, 1988; Ford Doolittle, 1980; Taylor, 1976). In contrast, endosymbiotic theories claim that the nucleus was derived from a symbiont, typically an archaeon or a DNA virus in a bacterial or archaeal host, respectively (Bell, 2019, 2001; Horiike et al., 2001; Lake and Rivera, 1994; Lake et al., 1982; Margulis et al., 2000; Takemura, 2001). Recent discoveries have provided further insights into eukaryogenesis and nuclear evolution, prompting a re-evaluation of these hypotheses.

Autogenous models suggest that the nuclear envelope arose following the invagination and functional diversification of the plasma membrane or ER around membrane-associated chromosomes (Cavalier-Smith, 1988; Dacks and Field, 2007; Gould et al., 2016; Jékely, 2008; Wilson and Dawson, 2011). In accordance with this, eukaryotic chromosomes adhere to the nuclear cytoskeleton and the nuclear envelope, which in itself represents an extension of the endomembrane system (Figure 1). Moreover, proteins central to nuclear structure, specifically

nucleoporins and their β-propeller and α-solenoid domains, are homologous to coatomer proteins involved in vesicular trafficking, providing further support for an evolutionary connection between nuclear proteins and the endomembrane (Devos et al., 2004; Field and Rout, 2019). Evidence for the directionality of this association (i.e., whether nuclear components were derived from the endomembrane system or vice versa) has come following the discovery of coatomer proteins and components of the ESCRT III vesicle biogenesis machinery in Asgard archaea, which may be the closest prokaryotic relatives of eukaryotes (Klinger et al., 2016; Williams et al., 2020; Zaremba-Niedzwiedzka et al., 2017). The single cultured Asgard archaea, *Prometheoarchaeum syntrophicum*, lacks internalized structure, but is characterized by an elaborate out-folding of its plasma-membrane (Imachi et al., 2020). Perhaps the original role of these proteins was to increase cellular surface area through projection as opposed to invagination, but regardless, the presence of these proteins suggests that coatomers and endomembrane-related components were likely present in eukaryotes immediately following their divergence from archaea and prior to the origin of the nucleus (Imachi et al., 2020). Unlike *P. syntrophicum*, planctomycete bacteria, such as *Gemmata obscuriglobus*, display complex invaginations of their internal membrane, contain coatomer proteins, and NPC-like structures that are internally localized and convergently comprised of β-propeller and α-solenoid domains (Fuerst and Sagulenko, 2012; Sagulenko et al., 2014, 2017). The exact nature of planctomycete compartmentalization is still under debate, but the apparent separation between transcription and translation in combination with the convergence in pore architecture suggests that *G. obscuriglobus* may be useful when considering nuclear evolution in eukaryotes and indicates that certain protein domains may be predisposed to pore cooption (Devos, 2014; Fuerst and Sagulenko, 2012; Hendrickson and Poole, 2018; Jogler et al., 2019; Shiratori et al., 2019).

In contrast to the autogenous model, evidence identifying the source of the nuclear envelope given endosymbiotic hypotheses is scarcer. The idea that the two lipid bilayers surrounding the nucleus reflect a bona fide double membrane, as observed in bacterial-derived organelles such as mitochondria and chloroplasts, is a misconception as the membranes are continuous and function as an extension of the ER (Figure 1) (Gould et al., 2016). Accordingly, it is unclear how prokaryotic or viral membranes could give rise to the nuclear envelope, as membrane continuity would either require the ER lumen to be homologous to the endosymbiont cytoplasm or complex membrane biogenesis events would need to occur (Gould et al., 2016;

López-García and Moreira, 2006). However, recent reports have revealed that large *Pseudomonas* phages can form nucleus-like compartments during infection (Chaikeeratisak et al., 2017a, 2017b; Mendoza et al., 2020). Although the compartmental boundaries are proteinaceous, these nucleus-like structures are notably similar to eukaryotic nuclei as they facilitate the segregation of proteins based on functionality, separating proteins involved in DNA metabolism from metabolic enzymes and ribosomes, and they employ a dynamic bipolar spindle that positions the nucleus at the cellular midpoint (Chaikeeratisak et al., 2017a). Additionally, some eukaryote-infecting nucleocytoplasmic large DNA viruses (NCDLVs) replicate in nucleus-like membrane-bound viral factories, structured using host-derived membranes (Raoult and Forterre, 2008). Although an explicit endosymbiotic source for the nuclear envelope is still unclear, this suggests that viruses at least have the capacity to induce cellular compartmentalization and form nucleus-like structures in their hosts.

The identification of nuclear similarities in viruses and prokaryotes provide comparative models for investigating the reasons underlying nuclear evolution in eukaryotes (Hendrickson and Poole, 2018). However, without further experimentation (e.g., evidence for nuclear homology and similar functional roles) it will remain unclear the extent to which these systems are comparable. Nonetheless, one of the primary outcomes of nuclear compartmentalization is the segregation of transcription from translation. In prokaryotes, mRNA is translated as it emerges from RNA polymerase, providing a rapid transition from transcription to protein production (McCarthy and Gualerzi, 1990). In contrast, eukaryotic transcripts must be transcribed, processed, and exported from the nucleus prior to translation, which not only delays gene expression but requires additional energy (Köhler and Hurt, 2007). Reasons for autogenous nuclear compartmentalization have recently been summarized (see Hendrickson and Poole, 2018) and typically center around the separation of molecular conflicts, the limitation of transposable elements and gene transfer, or the biophysical properties of genomes, all of which may or may not have an adaptive basis. For example, the isolation of translation from transcription could have initially permitted the evolution of mRNA processing which may have permitted intron expansion, blocking a reversion to an anucleate state (López-García and Moreira, 2006; Martin and Koonin, 2006). Additionally, reduced cross-talk between mRNA and non-coding RNA may be important for limiting unintentional transcriptional impairment (Hendrickson and Poole, 2018), the importance of which could be proportional to genome size

and rates of permissive transcription. Enclosure of the genome may also block the insertion of transposable elements or limit horizontal gene transfer, but little evidence supports this notion as transposable elements are often abundant in eukaryotic genomes and horizontal gene transfer still occurs, albeit at lower rates than observed in prokaryotes (Hendrickson and Poole, 2018; Husnik and McCutcheon, 2018; Keeling and Palmer, 2008; Leger et al., 2018). It is also feasible that the biophysical properties of genomes or chromosomes could drive autogenous nuclear formation. Thermodynamic calculations suggest that nucleoid formation is energetically favorable for genomes larger than 10 Mbp due to phase separation (Braun, 2008). Furthermore, the nuclear envelope may have been necessitated by the advent of cytoskeletal motor proteins which could have induced chromosomal shearing (Cavalier-Smith, 2010) or the emergence of heterochromatin which could have generated mechanical load on the cell membrane forcing its invagination (Wilson and Dawson, 2011). Endosymbiotic models are similarly speculative, especially when considering why gene flow would occur from the host genome to the endosymbiont, a pattern which is inverse to known endosymbiotic systems (Husnik and McCutcheon, 2018; Timmis et al., 2004). Viral host-shutoff strategies have been proposed as one mechanism through which gene transfer could be redirected from the host to a viral genome during a mild persistent infection (Bell, 2001). For example, the targeted degradation of host derived mRNA could create a selective pressure for gene re-localization (Bell, 2001; Gaglia et al., 2012; Svenson and Karlström, 1976). However, even given this, the circumstances that would result in the complete loss of the host genome are unclear and unparsimonious.

Regardless of how or why nuclear compartmentalization evolved, the origin of the nucleus coincided with a great expansion of functional novelty. Eukaryotes exhibit an array of nuclear machinery and processes that differentiate them from prokaryotes such as heterochromatin, the mitotic spindle, cyclin and cyclin dependent kinases, multiple RNA polymerases, and DNA replication complexes such as ORC (origin recognition complex), and the MCM (minichromosome maintenance) complex (Margolin and Bernander, 2004). *In silico* reconstructions of nuclear systems in the last eukaryotic common ancestor (LECA) using comparative genomics has revealed that these nuclear processes were largely established prior to the radiation of extant eukaryotes. For example, LECA is now presumed to have had an elaborate nuclear pore complex (Neumann et al., 2010), lamins (Koreny and Field, 2016), a complete kinetochore (van Hooff et al., 2017), basal transcription factors, histone post-

translational modifications, chromatin remodeling enzymes (Iyer et al., 2008; Koster et al., 2015), cell cycle regulators (Medina et al., 2016), and an MCM complex (Liu et al., 2009). But where did these proteins come from? A recent evolutionary analysis investigating the source of kinetochore proteins suggested that the majority of the complex was derived through gene duplication (Tromer et al., 2019), a process which seems to have been instrumental in the evolution of eukaryotic complexity generally (Lynch and Conery, 2000, 2003; Maere et al., 2005). Although some kinetochore proteins, as with NPC, replication, and transcription components, contain domains with prokaryotic homologues, many contain eukaryote specific-domains or are functionally distinct from prokaryotic relatives (Field and Dacks, 2009; Tromer et al., 2019; Vosseberg and Snel, 2017). It is possible that some of these genes evolved from prokaryotic homologues, perhaps following duplication or gene fusion/fission, and subsequently became so derived that sequence and structural homology is now undetectable. However, the fundamental nature of nuclear proteins and their strong conservation in cellular organisms makes this scenario uncertain and therefore the mechanism underlying the extensive and elaborate diversification of the eukaryotic nuclear proteome remains unclear. Endosymbiotic hypotheses provide an immediate source of nuclear innovations, particularly when considering a viral partner (Bell, 2001). Viruses not only experience high evolutionary rates due to the decreased fidelity of their DNA polymerases and recurrent genetic bottlenecks (Duffy, 2018; Sanjuán et al., 2010), but are also under pressure to manipulate the nucleic acid-based machinery of their hosts to promote their own replication. This suggests that viruses could act as nuclear tinkerers in eukaryotic evolution. Indeed, nuclear protein homologues, such as polymerases, topoisomerase, and histones, are observed within viral genomes, although whether viruses derived these genes from eukaryotes or vice versa is occasionally uncertain but generally favors the former scenario (Erives, 2017). A recent phylogenetic analysis of RNA polymerase (RNAP) genes in eukaryotes suggested that although RNAP III was inherited vertically from archaea, RNAP II and RNAP I were obtained following successive transduction events of RNAP homologues between stem eukaryotes (i.e., pre-LECA) and the NCLDV (Guglielmini et al., 2019). In some cases, viral proteins, particularly DNA and RNA polymerases, have also been shown to replace cellular machinery as in mitochondria (Shutt and Gray, 2006), chloroplasts (Hedtke et al., 1997), and the endosymbiotically-derived nuclei of chlorarachniophytes (Suzuki et al., 2016). Similarly, in fungi, a viral-derived protein, SBF, replaced E2F transcription factors which regulate the cell

cycle in metazoans (Medina et al., 2016) and in dinoflagellate algae, viral proteins replaced histones as the predominant chromosome packaging proteins (Gornik et al., 2012). Thus, viruses have and importantly, have had, the capacity to manipulate eukaryotic nuclear proteomes, presenting an intriguing source of nuclear novelty during eukaryogenesis.

Neither autogenous nor endosymbiotic theories for the origin of the nucleus provide a comprehensive or infallible view of nuclear evolution, but this may in part be reconciled by taking a hybrid stance. For example, it is feasible that the nucleus arose autogenously through the functional diversification of the endomembrane system, but perhaps this occurred with contributions from DNA viruses and other sources of horizontal gene transfer, which could have been a source of novelty and provided selective pressures and the elevated evolutionary rates required to drive nuclear evolution. Indeed, the initiation of eukaryogenesis coincided with a rapid increase in substitutions rates followed by a decrease during the radiation of extant eukaryotes, which is apparent in universal gene phylogenies (Pittis and Gabaldón, 2016). Without a detailed view of the phylogenetic origins of eukaryotic nuclear proteins and a deeper understanding of the functional implications of putative nuclear analogs in prokaryotes and viruses, the evolution of the nucleus will remain obscure. However, this is easier said than done, particularly given the evolutionary time scales that this problem encompasses. Another approach which may provide insights into not only nuclear evolution but the biology and mechanisms underlying modern nuclear processes, is examining nuclear diversity in a wider range of extant eukaryotes to understand how the nucleus has evolved and functionally diversified in the comparatively recent past.

1.3 Nuclear diversity: Insights from unexpected divergence

The nucleus and its processes are typically presumed to be so highly conserved as to be almost functionally static. This presumption is in part valid, given that some nuclear proteins, such as histones, are among the most highly conserved proteins known, and accordingly, many nuclear proteins are preserved throughout the eukaryotic tree (Koonin et al., 2004). However, nuclear diversity not only exists but covers a breadth of nuclear processes such as nuclear shape, the nuclear envelope and NPCs, chromatin biology and transcription, and DNA replication and chromosome segregation. Natural variation can provide unique perspectives into how these

systems function and evolve, similar to how genetic manipulation in the laboratory can facilitate mechanistic deduction, and may be key in understanding the important, yet subtle, nuances underpinning nuclear processes and their functional interconnections. An overview of these processes and examples of variation amongst eukaryotes follows.

1.3.1 Nuclear shape and structure

The nucleus represents a mechanosensitive organelle that responds to physical stress, often with mechanistic consequences. The shape and malleability of the nucleus is determined largely by the nucleoskeleton, namely lamins, chromatin, and the cytoskeleton (Stephens et al., 2019a). Lamins and chromatin fibers provide the rigidity and flexibility required to cope with normal mechanical stress, but the flexibility of these systems is limited and, as a result, nuclear deformation can be a restrictive property for cellular migration (Davidson et al., 2014; McGregor et al., 2016). Failure of nuclear structure, whether due to excessive force or mutation, can also result in transcriptional disregulation and DNA damage, and these outcomes can manifest as cancers and genetic disorders (Goldman et al., 2004; Stephens et al., 2019b; Xia et al., 2018). Recent advances have highlighted chromatin fibers as the basis for the nuclear mechanical response, but questions remain regarding how chromatin and lamins control nuclear shape and whether the downstream effects of altering chromatin structure convolute our mechanistic understanding of this process (Stephens et al., 2019a).

Two examples of distantly related organisms that may provide insights into the determinants of nuclear shape and malleability are ciliates and microsporidians (Figure 2). The ciliates comprise a phylum of ciliated protozoa, found prominently in freshwater and marine ecosystems, where they serve as key microbial predators (Lynn, 2008). From a nuclear perspective, ciliates are well known for exhibiting nuclear dimorphism, where each cell contains a small transcriptionally silent germline micronucleus used in sexual processes, and a larger somatic macronucleus in which gene expression takes place (Lynn, 2008). Ciliates also display dramatic variation in macronuclear shape that differs not only amongst taxonomic classes but even within individual genera (Figure 3). Whereas some ciliates have typical oval or circular macronuclei, others are elongated, lobate, or branching as demonstrated by the H-shape

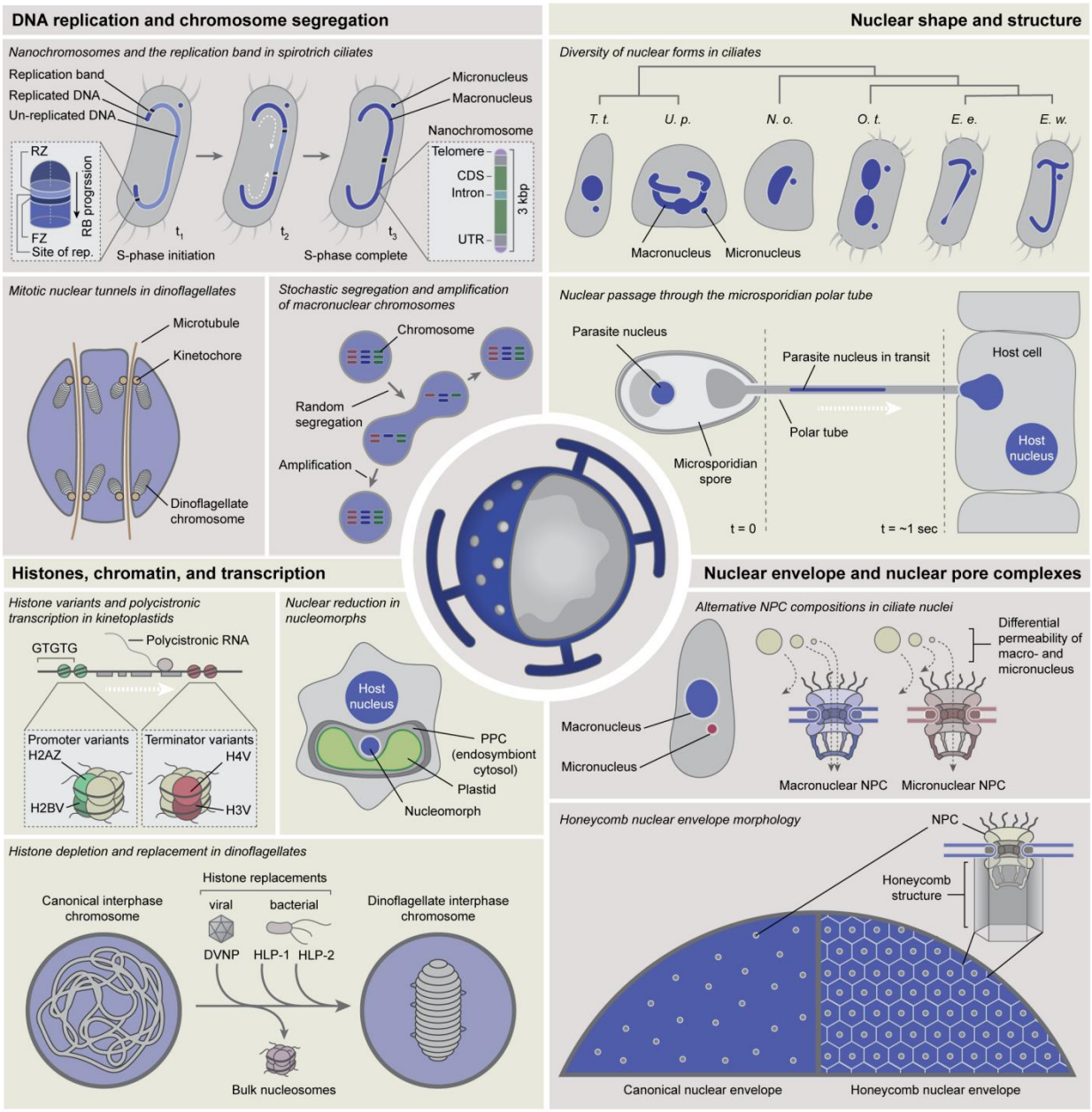


Figure 3. Nuclear variation from a systems perspective. A summary schematic depicting many of the divergent nuclear processes discussed. Acronyms defined (clockwise from top right): *T. t.*, *Tetrahymena thermophila*; *U. p.*, *Urceolaria parakorschelti*; *N. o.*, *Nyctotherus ovalis*; *O. t.*, *Oxytricha trifallax*; *E. e.*, *Euplotes enigma*; *E. w.*, *Euplotes woodruffi*; NPC, Nuclear pore complex; DVNP, dinoflagellate viral nucleoprotein; HLP, HU-like protein; PPC, periplastidal compartment; RZ, rear zone; FZ, frontal zone; CDS, coding sequence; UTR, untranslated region. In the top left, t_1, t_2, and t_3 denote early, mid, and late S-phase, respectively.

macronucleus of *Urceolaria parakorshelti* or the "question-mark" nucleus of *Euplotes enigma* (Figure 3) (Boscaro et al., 2018a; Irwin and Lynn, 2015). In contrast to the macronucleus, ciliate micronuclei tend to conform to canonical nuclear morphologies, thus providing a system where both typical and atypical nuclear shapes are present in the same cell simultaneously. Investigating the molecular mechanisms underpinning alterations to nuclear shape in ciliates may reveal important architectural determinants and could prove invaluable in parsing out the individual roles of lamins, chromatin, and cytoskeletal elements in nuclear structure.

Microsporidian nuclei push nuclear morphology in a different direction. These widespread, intracellular parasites of animals, are closely related to fungi (Figure 2) (Capella-Gutiérrez et al., 2012), and have evolved a unique and elaborate infection machinery to facilitate host invasion (Corradi et al., 2010; Keeling and Fast, 2002). Specifically, microsporidians utilize a characteristic polar tube, which acts as a hypodermic needle by exploding out of a spore, piercing the host plasma membrane, and serving as a tunnel through which the micorsporidian cytoplasm can travel and enter a host cell (Figure 3) (Keeling and Fast, 2002). This entire process occurs over a millisecond timescale and can require firing the nucleus, or at least the genomic DNA, through a 400 nm wide tube at over 250 $\mu m \cdot s^{-1}$ (Jaroenlak et al., 2020). Remarkably, this involves distorting the aspect ratio of the nuclear contents (width/height) by nearly 800% from a circular form to an elongated string, before reconstituting in the host cell (Figure 3) (Jaroenlak et al., 2020). Examining the cellular conditions that permit such extreme alterations to nuclear shape in microsporidia could identify factors defining nuclear deformability.

1.3.2 The nuclear envelope and nuclear pore complexes

The basis of nuclear shape and structure is the nuclear envelope. By compartmentalizing the nucleus, this membrane generates a spatial barrier with the rest of the cell, and this necessitates NPCs to mediate nucleocytoplasmic exchange. Besides being intrinsic to nuclear architecture and stability, the nuclear envelope and NPCs play important roles in regulating genome organization, the nuclear proteome, and cellular biology, generally. Membrane imbedded proteins such as SUN-domain proteins and nespirins, which localize to the inner and outer nuclear membranes, respectively, provide a physical connection between the nucleus and cytoskeleton allowing the nuclear envelope to link nuclear and cellular alterations (Adam, 2017;

De Magistris and Antonin, 2018). Moreover, NPCs not only facilitate nuclear trafficking but can also serve as euchromatic oases at the nuclear periphery, which is typically dominated by heterochromatin (Blobel, 1985; Raices and D'Angelo, 2017). Genomic re-localization to NPCs can be important for gene expression, particularly for inducible metabolic genes in yeast (Taddei et al., 2006). Additionally, NPC abundance in the nuclear envelope fluctuates in response to cell cycle progression (typically in anticipation of cell division) and metabolic queues, individual nucleoporins can act as transcriptional regulators in the nucleoplasm when dissociated from NPCs, and variation in NPC composition and post-translational modifications can be important for defining cell states (Beck and Hurt, 2017; Doucet et al., 2010; Dultz and Ellenberg, 2010; Gomar-Alba and Mendoza, 2020; Kim et al., 2018; Raices and D'Angelo, 2017; Sun et al., 2019). Unsurprisingly, given the multitude of functions carried out by the nuclear envelope and NPCs, a number of unanswered yet fundamental questions persist, including how protein content and membrane dynamics are regulated at the nuclear envelope, how NPCs define chromatin organization, and how NPC content and variation impact differentiation and development (Gomar-Alba and Mendoza, 2020; De Magistris and Antonin, 2018; Sun et al., 2019).

Variability in nuclear membrane architecture may provide unique perspectives into the structure and regulation of nuclear membranes. For example, the nuclear envelope of *Amoeba proteus* is studded with cylinders where the cylindrical axis faces the nuclear surface (Figure 2, 3) (Flickinger, 1970; Pappas, 1956). The top of the cylinder is associated with the inner nuclear membrane and centered around an NPC, whereas the bottom is open and continuous with the nucleoplasm (Figure 3) (Flickinger, 1970). These structures produce a striking "honeycomb" appearance, and similar morphologies have been observed in diverse eukaryotes including apicomplexans, formaniniferans, and even in metazoan nuclei such as in the nervous system of the leech, *Herudo medicinalis* (Figure 2, 3) (Beams et al., 1957; Dalhgren, 1967; Flickinger, 1970; Gray and Guillery, 1963; Pappas, 1956). The ubiquity of these structure suggests they could be broadly important in nuclear function yet the composition and role of these cylinders is unclear, although they have been speculated to function in nucleocytoplasmic transport or nuclear structure (Flickinger, 1970). Intriguingly, mutations in lamin B resulting in impaired farnesylation produce a honeycomb lamin structure (Jung et al., 2013), however a mechanistic connection between this and honeycombed nuclear envelopes remains unfounded.

Nuclear pore complex diversity also provides insights into nuclear biology more broadly. Nuclear pore complex reconstructions and structural studies across the eukaryotic tree have revealed variability in NPC composition. For example, ancestral state reconstructions across eukaryotes revealed divergence in the NPC scaffold, cytoplasmic fibrils, and transmembrane domains relative to Opisthokonts (Neumann et al., 2010). Likewise, cryoelectron-microscopy studies in the green alga, *Chlamydomonas reinhardtii*, identified an enlarged outer ring diameter and an oligomeric inner ring, unlike that observed in humans (Mosalaganti et al., 2018). Additionally, experimental investigations of NPCs in trypanosome parasites revealed novel nucleoporins, underscoring the inadequacy of relying solely on compositional comparisons to humans and yeast (Field et al., 2014; Holden et al., 2014; Obado et al., 2016). By correlating divergence in NPC structure with the genome and transcriptional regulatory mechanisms in these organisms, a clearer view of the function of the NPC beyond trafficking may be attainable. Similarly, eukaryotes with variable NPC abundance, such as in the ciliate *Tetrahymena pyriformis*, where pore abundance varies between 70 and 190 between stationary and logarithmic growth phases (Wunderlich, 1972), may provide useful models for investigating NPC regulation and assembly dynamics. Finally, recent work has suggested that compositional regulation and modification of NPCs can be important in cell fate determination and development as altered NPC structure can impact nuclear proteome content and dissociated nucleoporins can alter gene expression (Gomar-Alba and Mendoza, 2020; Kumar et al., 2018; Ori et al., 2013; Sun et al., 2019). In line with this, NPCs in the macronuclei and micronuclei of the ciliate, *Tetrahymena thermophila*, are compositionally differentiated and these differences coincide with altered permeability (Figure 3) (Iwamoto et al., 2009, 2017). This suggests that NPC compositional variation may be a common regulatory approach across eukaryotes and systems such as the dimorphic nuclei of ciliates may provide a simple model for probing and generalizing this mechanism.

1.3.3 Histones, chromatin, and transcription

Although factors such as nuclear shape and membrane dynamics impact nuclear function, molecular responses are often mediated through interactions with chromatin. The nucleosome is the fundamental repeating unit of eukaryotic chromatin, that comprises a protein octamer containing two copies of each of the four core histones, histone H2A, H2B, H3, and H4, that is

wrapped in approximately 146 bp of DNA (Kornberg, 1974; Luger et al., 1997). Nucleosomes have classically been perceived as genomic repressors as they sterically occlude transcription and replication machinery by occupying regulatory elements such as TATA boxes and replication origins as well as facilitate chromosome condensation (Grimaldi et al., 2014; Han and Grunstein, 1988; Lipford and Bell, 2001). However, through the actions of chromatin remodeling enzymes, histone chaperones, histone variants, and post-translational modifications (PTMs), nucleosome dynamics form the basis for differential genome regulation and are key in directing stress responses, DNA replication coordination, and chromosome segregation as well as defining genomic boundaries (Lai and Pugh, 2017). Additionally, nucleosomes are strongly interconnected with transcriptional mechanisms, not only because activating PTMs are important for establishing transcriptional states, but because nucleosomes enhance transcription itself (Nagai et al., 2017; Puerta et al., 1993; Ramachandran et al., 2015). This could reflect the influence of nucleosomes on the DNA topology of transcribed genes which undergo positive supercoiling in advance of RNAP II (Ma and Wang, 2016; Teves and Henikoff, 2014). In fact, similar to eukaryotes, archaeal histone homologues may also promote transcription as they define promoter regions upstream of transcription start sites which are important for the assembly of the transcription pre-initiation complex (PIC) and limit permissive transcription from non-promoter regions (Ammar et al., 2012). This implies, that the ancestral role of nucleosomes in eukaryotes and archaea could have been potentiating, opposed to repressive as has been hypothesized previously (Talbert et al., 2019). Therefore, dissecting and understanding the multifaceted functionality of nucleosomes and their impact on transcription and other nuclear processes is fundamental to interpreting eukaryotic chromatin biology and evolution.

Because nucleosomes are deeply interconnected with nearly every nuclear process, species which exhibit nuclear divergence can provide simplified systems from which nucleosome function can be deduced. Arguably the most striking example of chromatin divergence is observed in dinoflagellates where bulk nucleosomes have been replaced by viral derived proteins termed DVNPs (dinoflagellate viral nucleoproteins) (Figure 2, 3) (Gornik et al., 2012; Kato et al., 1997; Rizzo and Noodén, 1972). Acquisition of these proteins coincided with genome enlargement, the partial replacement of genomic thiamine with the rare base, 5'-hydroxymethyluracil, the loss of repressive histone modification sites, and alterations to chromosome morphology (Gornik et al., 2012; Marinov and Lynch, 2015; Okamoto et al., 2012;

Soyer-Gobillard et al., 1999; Wisecaver and Hackett, 2011). Later in dinoflagellate evolution, DVNPs were supplemented with bacterial HU-like proteins (HLPs), the acquisition of which correlates with the emergence of the unmistakable dinoflagellate nucleus, or dinokaryon, characterized by permanently condensed, birefringent, liquid crystalline chromosomes with unique architectural features including non-canonical banding and loops (Figure 3) (Chow et al., 2010; Chudnovsky et al., 2002; Golyshev et al., 2018; Janouškovec et al., 2016; Soyer-Gobillard et al., 1999; Wisecaver and Hackett, 2011). Despite the bulk replacement of histones, histone genes, albeit divergent ones, are retained and constitutively expressed in dinoflagellates (Beauchemin and Morse, 2017; Marinov and Lynch, 2015; Riaz et al., 2019; Roy and Morse, 2012). Histone activating PTM sites are also retained, and histone chaperones such as FACT (facilitates chromatin transcription), which is important in maintaining nucleosome stability in the wake of transcription, are conserved (Marinov and Lynch, 2015). This implies that dinoflagellate histones may be retained to facilitate transcription rather than repress gene expression. This could reflect a reversion to an ancestral proto-eukaryotic state, but regardless, suggests that dinoflagellates could be a unique model for investigating the non-repressive roles of nucleosomes. Experiments examining the functional impact of DVNPs and HLPs on chromatin structure will be required before the function of dinoflagellate histones can be interpreted.

Similar to dinoflagellates, other disparate eukaryotic lineages display condensed interphase chromosomes and histone divergence. Examples of species with non-mitotic condensed chromosomes include parabasalids (Lingle and Salisbury, 1995), euglenids (Haapala and Soyer, 1975), and cercozoans such as ebridians, but unfortunately little is known about how these chromosomes are condensed (Figure 2) (Hargraves, 2002; Hoppenrath and Leander, 2006). Similarly, motile gametes in diverse protozoans have condensed chromosomes, as do animals which use protamines as opposed to histones for nuclear condensation in sperm (Braun, 2001; Dacks and Kasinsky, 1999). Investigations into these systems could shed light on the mechanisms governing chromosomes compaction and will be interesting to compare to the better studied case in the dinoflagellates.

Histone divergence has also been reported in different eukaryotic groups. One example is observed in nucleomorphs which are highly reduced, endosymbiotically acquired nuclei that

were derived following secondary endosymbioses between a heterotrophic predator and a eukaryotic alga (Figure 3). Although the nuclei of endosymbionts typically undergo complete reductive evolution, this process halted independently in chlorarachniophytes and cryptophytes, which now have two distinct nuclei: a canonical host nucleus and a reduced green or red algal-derived nucleus, respectively (Figure 2) (Douglas et al., 2001; Grisdale et al., 2019; Keeling, 2013; Moore and Archibald, 2009). The extensive reduction of nucleomorphs poses a unique opportunity to study a nucleus dependent on all but the most core nuclear components (Curtis et al., 2012; Marinov and Lynch, 2016). Indeed, histone PTM sites, including those involved in transcriptional activation, mitosis, and heterochromatin, have been lost in chlorarachniophyte nucleomorphs whereas activating sites have been retained in cryptophyte nucleomorphs (Hirakawa et al., 2011; Marinov and Lynch, 2016). Moreover, the C-terminal domain of nucleomorph RNAP II was lost in both instances, despite its role in transcriptional initiation and elongation and general conservation across eukaryotes (Hsin and Manley, 2012; Liu et al., 2010; Marinov and Lynch, 2016).

Another case of histone diversification is observed in kinetoplastids, such as trypanosomes, which are common parasites responsible for diseases such as sleeping sickness and Chagas disease (Figure 2). Unlike canonical eukaryotic systems, trypanosome transcription occurs polycistronically, generating large transcripts that are on average 150 kbp long and encode 55 genes. These transcripts are fragmented by trans-splicing prior to translation (Sutton and Boothroyd, 1986). The size of the polycistronic gene clusters means that transcription need only initiate from around 200 sites throughout the genome which are denoted by histone variants and GT-rich sequence motifs as opposed to classical promoter elements (Figure 3) (Daniels et al., 2010; Kolev et al., 2010; Müller et al., 2018; Siegel et al., 2009; Wedel et al., 2017). In particular, H2AZ and H2BV along with H4K10 acetylation mark transcription start sites, whereas H3V and H4V associate with termination sites (Figure 3) (Müller et al., 2018; Siegel et al., 2009). Additionally, trypanosomes exhibit non-canonical histone modifications such as hyper-acetylation of histone H2A and a dearth of modifications in histone N-terminal tails, which are the predominant site of histone PTMs in other organisms (Janzen et al., 2006). Comparing the impact of novel histone variants and modification strategies on trypanosome transcription with mechanisms in other eukaryotes that display polycistronic transcription, such

as dinoflagellates, could provide insights into the interplay between chromatin and transcriptional dynamics (Shoguchi et al., 2013).

A final notable example of histone divergence, that is not directly related to the nucleus but serves as an interesting instance of functional diversification, are histrosomes in the green alga *Pyramimonas parkae* (Figure 2) (Yamagishi et al., 2015). Histrosomes are ejective organelles comprising core histones bound to N-acetyl-glucosamine polymers (Yamagishi et al., 2015). The function of histrosomes is unclear, although superficially similar ejectosomes in cryptophytes have been speculated to function in cellular defense (Cavalier-Smith, 1982). The presence of these structures in related green algae have yet to be revealed and the consequences that histone neofunctionalization has had, if any, on nuclear chromatin in *P. parkae* remains unexplored.

1.3.4 DNA replication and chromosome segregation

Although chromatin structure forms the core of nuclear regulation, without proper cell division, prior cellular responses are conducted in vain. In order for cell division to succeed, the faithful duplication and separation of nuclear chromosomes is required. As such, DNA replication and chromosome segregation represent conserved and highly coordinated processes. Although the action of DNA synthesis is confined to S-phase, DNA replication is an ongoing and chronological process that initiates with the licensing of replication origins in preceding cell cycle phases. Licensing involves the establishment of the pre-replication complex over cis-encoded replication origins, which comprises the ORC (origin recognition complex), Cdc6, Ctd1, and two MCM (minichromosome maintenance) replicative helicases (Aze and Maiorano, 2018; Prioleau and MacAlpine, 2016). Activation of the replication complex requires the action of cyclin-dependent and Dbf4-dependent kinases which facilitate the assembly of the active helicase (Fragkos et al., 2015). Although thousands of licensed origins can occur across the genome, selective activation generates spatial and temporal replication patterns which are in part consistent between cell cycles and over evolutionary time, but also exhibit cell-type and species-specific differences (Aze and Maiorano, 2018; Rhind and Gilbert, 2013; Ryba et al., 2010). A major correlate of these replication timing profiles is chromatin structure, as euchromatin and heterochromatin are permissive and restrictive to replication initiation, respectively (Schwaiger et al., 2010; Vogelauer et al., 2002). Regardless of variation in replication patterns, genome

duplication is achieved in anticipation of chromosome segregation. The separation of replicated chromosomes during cell division then requires a complex interplay between the centromere and its centromeric histone H3 variant, the kinetochore, and the microtubule-based mitotic spindle (Hinshaw and Harrison, 2018). However, chromosome segregation characteristics vary widely across the eukaryotic tree whether in terms of nuclear envelope dynamics (e.g., open, closed, and semi-closed mitosis) or kinetochore composition, making it a challenging system to analyze (van Hooff et al., 2017; Sazer et al., 2014). Thus, despite the central importance of these processes, certain details remain unresolved, such as the factors defining replication origins (as not all are apparent at the sequence level), the functional role of spatiotemporal replication patterns, the reasons dictating the differential importance of kinetochore components in different species, and the nature of mechanistic variation in both processes (Aze and Maiorano, 2018; Hinshaw and Harrison, 2018).

Diversity in chromosome morphology may provide insights into the nuances underpinning DNA replication and chromosome segregation. For example, in spirotrich ciliates, the macronuclear genome is encoded on tens of thousands of gene-sized endoreplicated "nanochromosomes" that typically contain a single gene flanked by short untranscribed regions and telomeres (Figure 2, 3) (Aeschlimann et al., 2014; Swart et al., 2013). The abundance and variable frequency of these chromosomes create issues for coordinating and ensuring the fidelity of replication but spirotrich ciliates may overcome these challenges through the use of a replication band (RB). The replication band is a large, localized, and motile hub that traverses the macronucleus while replicating DNA (Figure 3) (Olins and Olins, 1994). Replication bands typically form at the poles of an elongated tubular macronucleus and subsequently translocate lengthwise before merging and disappearing at the macronuclear midpoint (Figure 3) (Olins and Olins, 1994). Replication associated proteins such as telomerase and PCNA are coupled and travel with the RB which functions as a stratified structure with different proteins and modifications localized to the frontal and rear zones (Figure 3) (Fang and Cech, 1995; Olins and Olins, 1994; Olins et al., 1989). Understanding licensing dynamics, replication origin specification, propulsion and organizational mechanisms, and the impact that a sequential DNA replication pattern has on nuclear processes in spirotrichs could potentially provide novel perspectives into DNA replication regulation and function in general. Likewise, similar chromosome morphology is observed in parallel in other ciliate lineages (Maurer-Alcalá et al.,

2018; Riley and Katz, 2001), yet these groups are not known to contain RBs. Investigating the mechanisms through which these lineages cope with chromosome fragmentation could provide additional comparative models.

Despite the complexity of divergent DNA replication machinery, such as the RB, ensuring the faithful replication of the genome is only half the battle as it must be followed by accurate chromosome segregation. Ciliate macronuclear chromosomes lack centromeres and kinetochore machinery is highly reduced, alluding to an alternative chromosome segregation mechanism that appears to proceed stochastically and has been shown to be dependent on nuclear architectural proteins such as condensins (Cervantes et al., 2006; Flickinger, 1965; van Hooff et al., 2017). Polyploidy and chromosome amplification following division may permit random segregation as only a single copy of each chromosome may be required as a template to re-establish the original nucleus, but this remains obscure (a hypothetical model for how this could occur is illustrated in Figure 3). In comparison to the ciliate system, the diversification of dinoflagellate chromosomes also appears to have led to alternative mitotic mechanisms (Figure 2, 3). In most dinoflagellates, closed mitosis occurs but involves the invagination of the nuclear envelope to form trans-nuclear tunnels and in some cases an endomembrane network that features membrane-bound kinetochores to which chromosomes can attach (Figure 3) (Bhaud et al., 2000; Gavelis et al., 2019). Understanding mitotic and kinetochore diversification across eukaryotes and linking observations to chromosome and genome characteristics may facilitate the deduction of chromosome segregation mechanisms and could help link kinetochore and centromeric components to specific functions (Makarova and Oliferenko, 2016).

1.4 Thesis overview and aims

Despite preconceptions, the nucleus evidently represents a functionally diverse organelle. Nuclear variability is not isolated to obscure molecular processes but rather touches every aspect of nuclear functionality (Figure 1, 3). Likewise, alternative nuclear systems are present across the tree of eukaryotes and are represented in major eukaryotic clades such as dinoflagellates and ciliates, which comprise speciose, environmentally ubiquitous, and ecologically fundamental groups of organisms (Figure 2). Although technological advances and experimental innovations in classic model organisms have driven nuclear research in the past decades, the sheer

complexity of the nucleus has left a swath of unanswered yet pertinent questions regarding nuclear processes. In this dissertation, I hypothesize that exploiting diverse nuclear systems and employing alternative model organisms will provide insights into the functional diversity and evolution of nuclear processes. To this end, I sought to utilize a combination of experimental and computational approaches to inspect three separate nuclear processes in three diverse eukaryotic systems, namely chromatin in the dinoflagellate dinokaryon, NPCs in the nucleomorphs of chlorarachniophytes and cryptophytes, and DNA replication in the ciliate macronucleus. In particular, I had three aims which are addressed in the subsequent chapters:

1. Identify the evolutionary mechanism through which histones were replaced by DVNP in dinoflagellates and explore how they interact with each other in a canonical chromatin environment.

2. Investigate the extent of nuclear reduction in nucleomorphs by characterizing the nuclear pore complex in chlorarachniophytes and cryptophtyes.

3. Examine the function and evolution of motile DNA replication systems in spirotrich ciliates.

The ultimate goal of these aims was to demonstrate the informative potential of nuclear diversity, gain insights into the function of nuclear processes, and develop a better understanding of nuclear evolution in relation to eukaryogenesis and eukaryotic diversification.

2. Viral proteins as a potential driver of histone depletion in dinoflagellates

2.1 Introduction

The conserved organization of DNA in the eukaryotic nucleus is a paradigm in biology. Within the nucleus, DNA is bound to highly conserved protein octamers comprised of two copies of each of the four core histones: histone H2A, H2B, H3, and H4 (Luger et al., 1997). These histones, in combination with approximately 146 bp of DNA, coalesce to form nucleosomes which act as the fundamental repeating units of eukaryotic chromatin and serve to facilitate DNA condensation (Khorasanizadeh, 2004; Luger et al., 1997). Furthermore, histones are often post-translationally modified, especially on their intrinsically disordered N-terminal tails, leading to altered nucleosome dynamics and the recruitment of transcription, replication, and DNA repair factors (Downs et al., 2000; Lipford and Bell, 2001; Rando and Winston, 2012; Strahl and Allis, 2000). As a result, nucleosomes play a fundamental role in genomic regulation and consequently, histones constitute some of the most highly conserved proteins known. For example, both histones H3 and H4 share roughly 90% amino acid sequence identity between yeast and humans despite around a billion years of divergence (Douzery et al., 2004). Therefore, histones contribute heavily to the growth and development of eukaryotic organisms and, given their conservation, are often viewed as a prerequisite for complex cellular life.

The dinoflagellates, a group of ecologically important unicellular eukaryotic algae, are a striking exception to the above paradigm as they have abandoned histones as their primary DNA packaging proteins (Gornik et al., 2012; Rizzo and Noodén, 1972). Phylogenetic analyses have revealed that histone depletion coincided with dramatic changes in nuclear characteristics including massive genome enlargement, the emergence of liquid crystalline chromosomes, and the acquisition of apparently viral-derived proteins termed DVNPs (dinoflagellate viral nucleoproteins) (Gornik et al., 2012; Janouškovec et al., 2016; Kato et al., 1997; Talbert and Henikoff, 2012). In basal dinoflagellates, DVNPs represent the predominant basic nucleoprotein and localize to chromosomes, suggesting that they play a direct role in chromosome organization (Gornik et al., 2012; Kato et al., 1997). Accordingly, it has been hypothesized that these

nucleoproteins could have been transferred from viruses to dinoflagellate progenitors with canonical chromatin and eventually replaced the majority of histones as chromatin packaging proteins.

Even though the bulk of their chromatin has diverged, dinoflagellates retain a full complement of histone genes (Gornik et al., 2012; Lin et al., 2010; Marinov and Lynch, 2015; Roy and Morse, 2012). The function of these remnant histones remains unclear, yet their low expression levels, relaxed conservation, and the presence of histone chaperones may indicate some degree of subfunctionalization to certain cellular processes such as transcription (Marinov and Lynch, 2015). Thus, not only the evolutionary mechanisms that drove dinoflagellate chromatin divergence but also the exact contributions of DVNPs and histones to dinoflagellate chromatin structure and function have yet to be resolved.

The above questions have remained unanswered in large part due to the technical challenges associated with studying dinoflagellate biology. In particular, a lack of genetic transformation methods and comprehensive genomic data, resulting from the size and complexity of dinoflagellate genomes, have created experimental restrictions. One way of avoiding these issues is to utilize model organisms. *Saccharomyces cerevisiae* represents a suitable model for investigating chromatin evolution because of its well-characterized and typical chromatin biology, its genetic malleability, and its well annotated genome. Therefore, to circumvent the limitations associated with dinoflagellates and gain insights into the initial transition between histone and DVNP-based chromatin, we employed an experimental evolutionary approach utilizing *S. cerevisiae* to assess how DVNP interacts with canonical eukaryotic chromatin. To this end, we found that DVNP antagonizes chromatin by localizing to histone binding sites, displacing nucleosomes, impairing transcription, and ultimately inhibiting growth. However, DVNP toxicity can be attenuated through histone depletion and cells reduce their histones following DVNP expression. These results reveal that histone depletion is an adaptive response to DVNP and emphasize the role that horizontal gene transfer, and possibly pathogenic stresses, can play in driving cellular evolution.

2.2 Results

2.2.1 DVNP enters the nucleus and impairs growth in S. cerevisiae

In order to examine the interactions between DVNP and nucleosomal chromatin, we first codon optimized and synthesized *Hematodinium* sp. DVNP.5 and placed it under the control of the galactose-inducible and dextrose-repressible *GAL1* promoter (Appendix Figure A1). SV40 nuclear localization signals (NLS) and or three hemaglutinin (3HA) epitope tags were added to the N- or C- terminus and protein expression was confirmed by immunoblot following galactose induction (Figure 4A, B). Immunofluorescence revealed co-localization between DVNP and Hoescht stain with all constructs, suggesting that DVNP localized to the nucleus independent of the additional NLS (Figure 4C). We also noted DVNP-dense regions associated with the nucleus (Figure 4D). This may reflect partial nucleolar localization as nucleoli are depleted of DNA dyes and because the cationic N-terminus of DVNP could act as a general nucleolar targeting signal (Emmott and Hiscox, 2009; Sirri et al., 2008).

To investigate the phenotypic effects of DVNP in yeast, we performed growth assays and found that DVNP expression impaired growth (Figure 4D), consistent with a previous report in *Toxoplasma gondii*, which is a closer relative of dinoflagellates than yeast (Gornik et al., 2012). Moreover, the addition of an N-terminal tag abrogated DVNP toxicity, either as a result of impaired function or diminished expression as the sequence composition, and therefore the immunogenicity, of the N- and C-terminal 3HA tags varied (Figure 4B, D; Appendix Figure A1F). In contrast, the addition of a C-terminal NLS accentuated the growth defect suggesting that toxicity may be dependent on nuclear localization (Figure 4D).

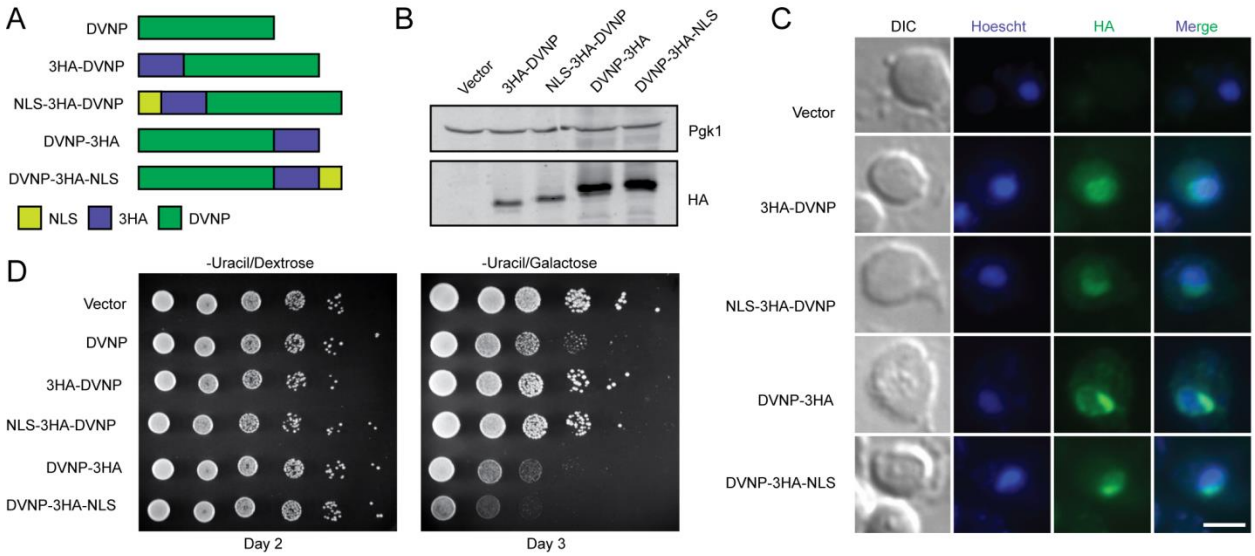


Figure 4. DVNP is localized to the nucleus and impairs growth in *S. cerevisiae*. A. Schematics of DVNP constructs. **B.** Immunoblot on total protein extracts from galactose-induced cells. **C.** Immunofluorescence micrographs showing colocalization between the HA epitopes on DVNP (green) and Hoescht stain (blue). Scale bar, 3 μm. This experiment was repeated twice with the same results. **D.** Serial dilution growth assays for cells containing indicated expression constructs. Cells were grown on selective media lacking uracil in the presence of either dextrose or galactose.

2.2.2 DVNP disrupts nucleosomal chromatin in S. cerevisiae

Given the possible dependency of DVNP toxicity on nuclear localization and the capacity of DVNP to strongly and non-specifically associate with DNA in vitro (Gornik et al., 2012), we hypothesized that DVNP was associated with the yeast genome. To assess this, we performed chromatin-immunoprecipitation (ChIP) using anti-HA antibodies and recovered a 17.6-fold increase in immunoprecipitated DNA in DVNP-3HA-NLS-expressing cells relative to the vector control (Appendix Figure A2A). To investigate the genomic localization of DVNP, we sequenced the immunoprecipitated DNA and inputs (ChIP-seq). In contrast to previous in vitro results (Gornik et al., 2012), we found that rather than associating non-specifically to areas of free DNA, such as the nucleosome depleted regions (NDRs) over promoters, DVNP was

depleted at NDRs and enriched upstream and downstream of transcription start sites (TSS) (Figure 5A). This binding profile is reminiscent of nucleosome binding, characterized by prominent -1 and +1 nucleosome peaks upstream and downstream of the TSS (Figure 5A) (Mavrich, 2008). These data therefore indicate that DVNP localizes preferentially to nucleosome bound regions of the genome.

The similarities in the binding profiles between DVNP and nucleosomes suggested that DVNP interacts with chromatinized DNA. To investigate whether this alters chromatin structure, we compared nucleosome profiles between DVNP-3HA-NLS-expressing cells and a vector control using micrococcal nuclease sequencing (MNase-seq). Nucleosomal peak height and trough depths decreased in the presence of DVNP, which is indicative of nucleosome disruption (Figure 5B) (van Bakel et al., 2013; Gossett and Lieb, 2012). To assess whether nucleosome loss was DVNP-dependent, genomic windows were binned by DVNP enrichment and changes in nucleosome occupancy were inspected in these bins (Figure 5C; Appendix Figure A2B). This revealed a negative association between DVNP enrichment and nucleosome loss that was independent of ChIP-seq inputs (Figure 5C; Appendix Figure A2C, D). We also found that displacement predominantly occurred at the -1 and +1 nucleosomes, mirroring the localization of DVNP (Figure 5D; Appendix Figure A2E). DANPOS (Dynamic Analysis of Nucleosome Position and Occupancy by Sequencing) reaffirmed that DVNP and reduced nucleosome occupancy were associated, but connections between changes in fuzziness or position and DVNP were less apparent (Appendix Figure A2F) (Chen et al., 2013). We also investigated the association between nucleosome loss and stability by binning nucleosomes by their predicted occupancy, inferred from nucleosomal sequence preference (Figure 5E; Appendix Figure A2G) (Kaplan et al., 2009). We found that weaker nucleosomes experienced significantly greater loss than more stable nucleosomes, suggesting that nucleosomal stability prevents DVNP disruption. Overall, these data suggest that DVNP preferentially associates with nucleosomal regions of the genome and induces histone displacement.

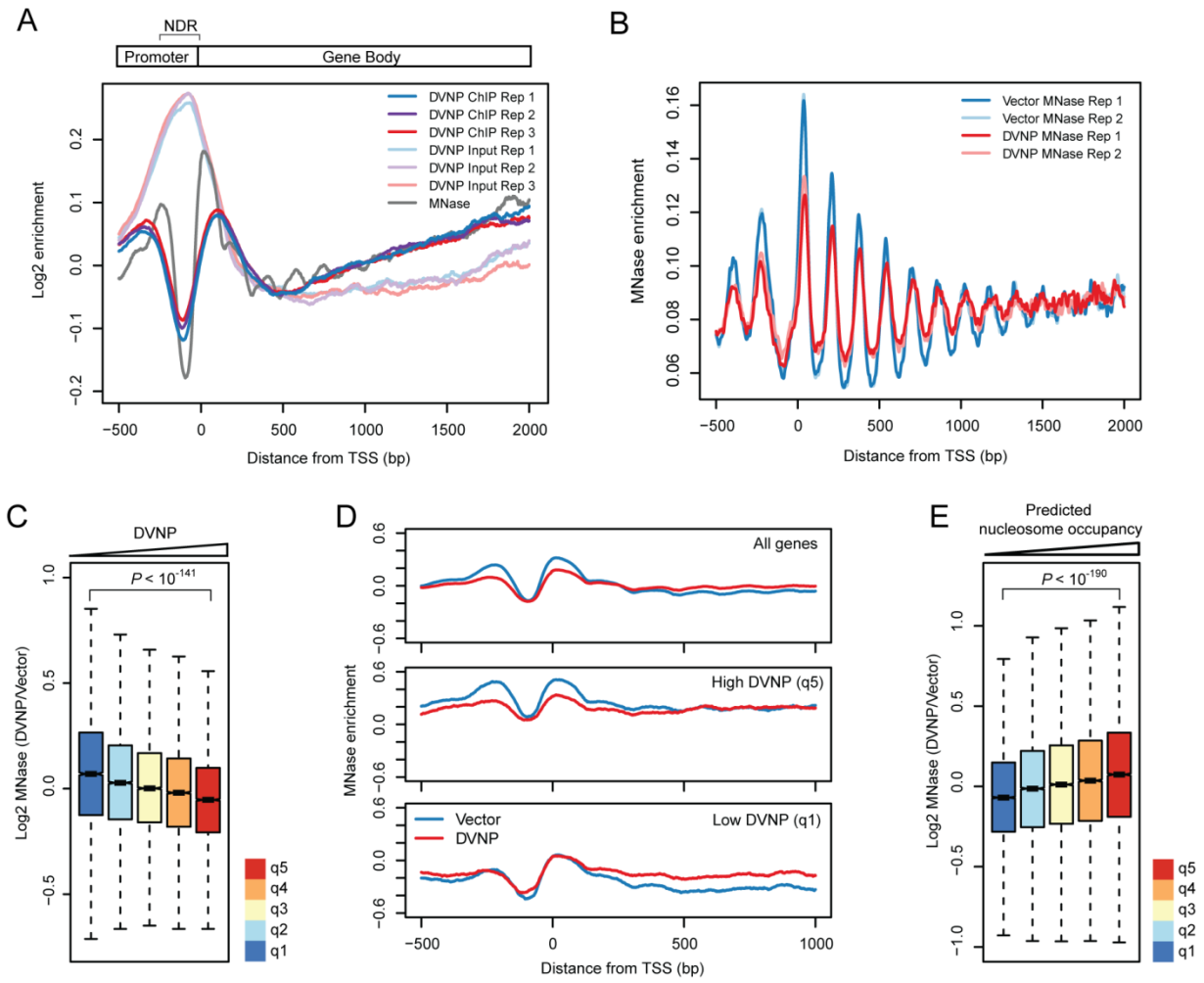


Figure 5. DVNP binds nucleosomal regions and induces histone loss in *S. cerevisiae*. A. Average gene plot showing the relative enrichment of DVNP immunoprecipitates (IPs), and ChIP inputs from DVNP-3HA-NLS expressing cells. Three biological replicates are shown for the ChIP and input. Also shown is the enrichment profile for MNase-digested control cells. A schematic of an average gene is shown above, with the nucleosome depleted region (NDR) labeled. DVNP ChIP and input replicates were combined in all subsequent analyses. **B.** Average gene plot showing MNase digested DNA from two biological replicates obtained from DVNP-3HA-NLS expressing and control cells. Enrichment was calculated from read midpoints and Gaussian smoothed with a standard deviation of 4. MNase-seq replicates were combined in all further analyses. **C.** Genome wide relationship between DVNP and nucleosome loss. Five hundred base pair genomic windows were binned into 20% DVNP quintiles (q1-5, see Appendix Figure A2D) and nucleosome loss is shown in these bins (n = 9,659). Outliers are shown in

Appendix Figure A2G. **D.** Average gene plots showing MNase profiles for DVNP-3HA-NLS expressing and control cells. Averages were calculated from all genes (top panel), or genes with the top (q5, middle panel) or bottom (q1, bottom panel) 20% DVNP enrichment over the gene body (see Appendix Figure A2E). **E.** Genome wide association between predicted nucleosome occupancy and nucleosome loss due to DVNP expression. Nucleosome bound sites were binned into 20% predicted occupancy quintiles (q1-5, see Appendix Figure A2G) and nucleosome loss is shown in these bins (n = 13,509). Outliers are shown in Appendix Figure A2H. All P-values were calculated using two-sided Welch's t-tests. Box plot notches represent an estimate of the 95% confidence interval of the median.

2.2.3 DVNP impairs transcription in S. cerevisiae

Previous work has emphasized the importance of nucleosomes in regulating the recruitment and processivity of RNA polymerase II (RNAP II) (Fitz et al., 2016; Gossett and Lieb, 2012; Kaplan et al., 2003; Nock et al., 2012; Venkatesh et al., 2012). This led us to investigate whether DVNP adversely affects transcription by performing ChIP-seq for Rpb3, the third largest subunit of RNAP II, in DVNP-3HA-NLS-expressing and control cells. Using spike-in controls for normalization, we identified a ~35% global reduction in RNAP II occupancy that was corroborated by quantitative PCR (Figure 6A, B; Appendix Figure A3A, B). We also observed a reduction of the Rpb3 peak over the TSS, consistent with the localization of DVNP (Figure 6A; Appendix Figure A3A). However, the loss of Rpb3 was not dependent on transcriptional rate or DVNP abundance and was only weakly positively associated with nucleosome loss (Appendix Figure A3B-E).

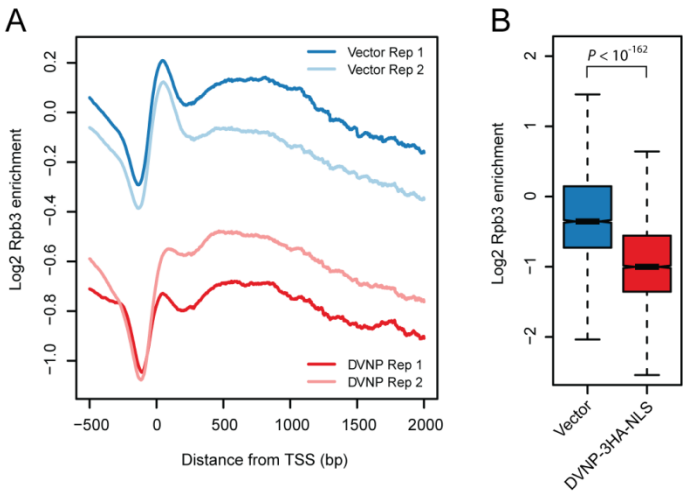


Figure 6. DVNP expression results in transcriptional impairment in *S. cerevisiae*. A. Spike-in normalized average gene plot showing the enrichment of two biological replicates of Rpb3 ChIPs from DVNP-3HA-NLS expressing and control cells. **B.** Differences in Rpb3 following DVNP expression. Values represent the mean Rpb3 occupancy calculated over gene bodies after averaging biological replicates (n = 4,793). Outliers are shown in Appendix Figure A3F. The P-value was calculated using a two-sided Welch's t-test. Box plot wedges represent an estimate of the 95% confidence interval of the median.

2.2.4 Histone reduction relieves DVNP toxicity in S. cerevisiae

Given the deleterious effects of DVNP on cell growth, we next wondered how an ancestral organism with canonical chromatin could have come to tolerate this protein. We therefore tested whether genetic changes could facilitate resistance to DVNP toxicity using a synthetic genetic array (SGA), whereby the relative growth of 5,426 non-essential yeast deletion mutants expressing DVNP-3HA-NLS was assessed. Functional classification of gene deletions causing improved growth revealed chromatin and transcription associated categories as the most significant functional hits (Figure 7A) (Mewes et al., 2002). This chromatin connection and the loss of histones in dinoflagellates led us to investigate whether histone expression altered DVNP toxicity. By analyzing 42 gene deletions previously shown to affect histone gene expression (Eriksson et al., 2012; Kurat et al., 2014a), we found that reducing and increasing histone

production relieved and exacerbated DVNP toxicity, respectively (Figure 7B). Moreover, we identified a loss of total histones H3, H4, and H2B following DVNP induction in wild type cells (Figure 7C-F). These data suggest that histone reduction is an adaptive response to DVNP toxicity and that cells cope with DVNP by maintaining a lower abundance of histones.

Of all the genes inspected, only four exceptions were noted. In particular, deletions of *SPT21*, a sequence specific histone gene activator, and *HHF2*, one of the two genes encoding histone H4, increased toxicity in the SGA (Figure 7B) (Eriksson et al., 2012; Kurat et al., 2014b). However, neither of these deletions were detrimental when manually assessed and a newly generated *spt21*Δ mutant in a different strain background relieved toxicity despite DVNP levels being unchanged (Appendix Figure A4A-E). Furthermore, deletion of components of the TRAMP (Trf4/Air2/Mtr4p polyadenylation) complex and Xrn1, which negatively regulate histone levels (Mullen and Marzluff, 2008; Reis and Campbell, 2007), improved growth (Figure 7B). However, these proteins are involved in general RNA degradation (Callahan and Butler, 2010; Pirkl et al., 2013), so their removal could promote RNA stability and relieve problems associated with transcriptional defects.

To reaffirm our results, we also assessed whether histone depletion could be a non-specific adaptive response to the over-expression of any toxic exogenous or endogenous nuclear protein. To examine this, we performed the same analyses with published SGA data for two proteins expressed from the same *GAL1* promoter, TDP-43 and Hho1. TDP-43 is a toxic mammalian DNA-binding protein whereas Hho1 is the yeast homolog of histone H1, which resembles the size and basicity of DVNP (Appendix Figure A4F, G) (Armakola et al., 2012; Lawrence et al., 2017). Although both of these proteins significantly impair growth when over-expressed (Armakola et al., 2012; Lawrence et al., 2017), no similar effect of histone levels on the TDP-43 or *HHO1*over-expression phenotype was observed, revealing that DVNP's genetic interactions are not ubiquitous.

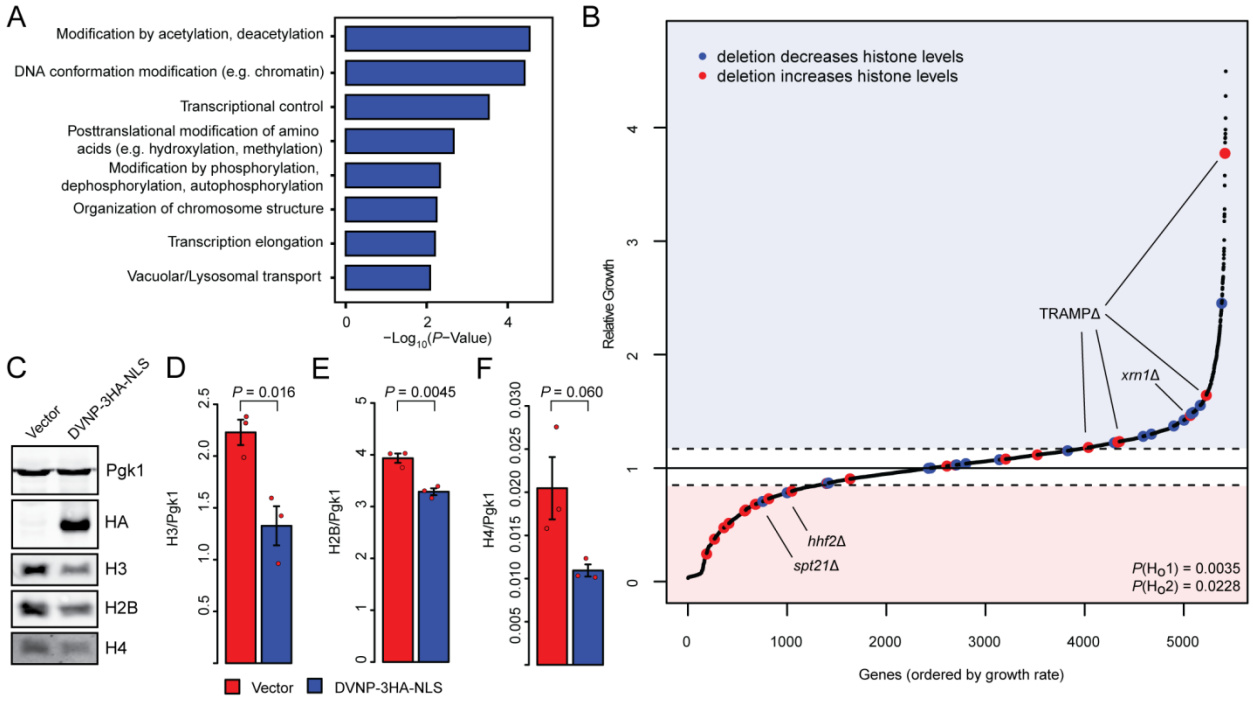


Figure 7. Histone loss is an adaptive response to DVNP toxicity in *S. cerevisiae*. A. Significant MIPS functional classifications of gene deletions that relieved DVNP toxicity in the SGA screen. **B.** Relative growth of ~5,500 non-essential deletion strains expressing DVNP-3HA-NLS from a *GAL1* promoter. Gene deletions known to increase and decrease histone expression are shown in red and blue, respectively. Two null hypotheses were rejected by χ^2 test: H_o1: gene deletions that affect histone levels are randomly distributed and, H_o2: gene deletions that increase and decrease histones are enriched below and above the growth thresholds by chance. Dashed lines denote positive and negative growth thresholds. Exceptions to the trend have been labeled. **C.** Immunoblot on total protein extracts from DVNP-3HA-NLS-expressing and control cells. **D-F.** Quantified immunoblots showing histone levels in DVNP-3HA-NLS-expressing and control cells (n = 3 biological replicates). Histone signal was normalized to the loading control, Pgk1, which remained constant between conditions. These experiments were repeated three times with similar results. *P*-values were obtained by two-sided Welch's t-test. Error bars represent the standard error of the mean (SEM).

2.3 Discussion

Here we sought to investigate how the dinoflagellate DNA-binding protein, DVNP, interacts with the canonical chromatin of yeast to gain insights into dinoflagellate chromatin divergence. To this end, we showed that DVNP interacts antagonistically with nucleosomal chromatin, causing histone displacement, transcriptional impairment, and growth inhibition, but that histone reduction partially mitigates this toxicity. It is possible that histone depletion relieves toxicity through transcriptional up-regulation, as is seen in ageing yeast with reduced histones (Hu et al., 2014), or by reducing excess displaced histones which are cytotoxic (Gunjan and Verreault, 2003). In either case, it leads to a model for the origin of dinoflagellate nuclear organization based on a stepwise increase in DVNP and corresponding depletion of histones. On one hand, forced exposure to DVNP, such as during viral infection, may have prompted histone depletion as a mechanism for limiting DVNP toxicity. Alternatively, DVNP may have been introduced during a transiently histone-depleted stage, for example following histone dilution in the wake of genome expansion. In the first instance, histone depletion would be a direct response to the most deleterious effects of DVNP, which in turn would open the door to more DVNP binding ultimately resulting in a large-scale displacement of histones by DVNP. In the second instance, it is possible that DVNP had a mild short-term benefit in an already histone depleted system, and its presence may have prevented the re-colonization of chromatin by histones over time. In either case, it would appear that something about the underlying biology of the ancestral dinoflagellate made it possible for the invasion of DVNP to lead to a progressively shifting balance between histones and DVNP, over time leading to a functional replacement by DVNP as the major genome packaging protein.

Although DVNP is likely of viral origin given the homology it shares with proteins in algae-infecting viruses (Gornik et al., 2012; Shoguchi et al., 2013), the actual source of DVNP remains to be clarified. No virus of this kind has been found in dinoflagellates, although the diversity of phycodnaviruses is not well sampled, and our model infers such an infection in the distant evolutionary past. But given the deleterious effects of DVNP expression, the most likely context would be pathogenesis, since this gives a powerful selective force for the depletion of histones, which in normal contexts would itself be deleterious. Moreover, viruses frequently

utilize chromatin effectors during pathogenesis to manipulate host processes and defenses. For example, foot-and-mouth virus protease 3C and adenovirus protein VII disrupt cellular expression and signaling by interacting with host nucleosomes (Avgousti et al., 2016; Tesar and Marquardt, 1990). Despite this, other sources of DVNP are also possible. Ancestral dinoflagellates could have been less susceptible to DVNP, facilitating passive acquisition from a virus, food, or commensal symbiont. However, DVNP is unknown in cellular genomes outside dinoflagellates, making this less likely. The activity of DVNP in modern viruses infecting their current hosts would presumably shed some light on these possibilities.

Why DVNP is not observed in other organisms is an interesting question. DVNP-related proteins have been identified in viruses infecting diverse algae such as the stramenopile *Ectocarpus siliculosus* and green alga *Micromonas*, yet these organisms have not acquired DVNP and their histone-based chromatin is unaffected (Gornik et al., 2012; Shoguchi et al., 2013).This may simply be due to the low probability of initiating such drastic change to chromatin structure. However, it may also be that other unique aspects of dinoflagellate biology 'preconditioned' the system such that this radical reaction to the introduction of DVNP was more likely. For example, the tree of dinoflagellates shows that gene expression using specialized trans-splicing mRNAs predates the rise of DVNP (Hearne and Pitula, 2011; Zhang et al., 2007, 2011). This suggests that gene expression in the dinoflagellate ancestor was already very unusual and that control of expression had shifted from transcriptional to post-transcriptional mechanisms. Genetic systems with such characteristics could react very differently to a perturbation such as the introduction of DVNP and histone displacement.

Despite the antagonism we observe between DVNP and histones, the retention of histone genes in dinoflagellates suggests that these proteins still have some role in dinoflagellate chromatin regulation. Here we find that replacement of histones with DVNP results in a net loss of RNAP II occupancy, indicating that the yeast transcriptional machinery is less equipped to deal with DVNP than with nucleosomes. Recent bioinformatic analyses have revealed that dinoflagellate histones have relaxed selection over heterochromatin-associated modification sites whereas activating sites have been conserved, suggesting that dinoflagellates may lack nucleosomal heterochromatin (Marinov and Lynch, 2015). This role is seemingly filled by DVNP and may be accomplished through its enhanced ability to repress transcription relative to

histones. We also showed that DNA with a high GC content, which is predicted to form more stable nucleosomes, was more resistant to nucleosome displacement by DVNP (Kaplan et al., 2009; Tillo and Hughes, 2009). Interestingly, dinoflagellates have a strong GC codon bias, which may promote nucleosome stability in open reading frames (Williams et al., 2017). This, together with the known conservation of activating histone modifications and histone chaperones in dinoflagellates (Marinov and Lynch, 2015), suggest that nucleosomes and DVNP may function in euchromatic and heterochromatic environments, respectively. If nucleosomes have subfunctionalized in dinoflagellates, then biochemical investigations into the activities of dinoflagellate histones may provide unique insights into the roles histones play in diverse eukaryotes by highlighting some of their specific functions, beyond bulk chromatin condensation.

2.4 Materials and Methods

2.4.1 Plasmids and yeast strains

DVNP.5 from the dinoflagellate *Hematodinium* sp. (accession number: AFY23231.1) was codon optimized for expression in *S. cerevisiae* and synthesized by GenScript into a pUC57 vector (Gornik et al., 2012). DVNP was then amplified with the addition of 5' SpeI and NdeI restriction sites and a 3' XmaI site using high fidelity Kapa TaqReadyMix (Kapa Biosystems) polymerase chain reaction (PCR). The DVNP amplicon was then cloned into a pRS416 expression vector containing a *GAL1* promoter (*GAL1pr*) using SpeI and XmaI restriction enzymes (New England BioLabs (NEB)) (Mumberg et al., 1994; Sikorski and Hieter, 1989) (Appendix Figure A1A-E). To add a 3HA epitope tag to the C-terminus of the protein, DVNP was cloned into a pRS416 vector containing a C-terminal 3HA tag (Lawrence et al., 2017). In contrast, a 3HA tag was added to the N-terminus of DVNP by performing gap repair on an NdeI (NEB) digested DVNP plasmid using a 3HA gene block synthesized by Integrated DNA Technologies (Fairhead et al., 1996). The discrepancies between these two methods led to sequence variation between the N- and C-terminal 3HA tags (see Appendix Figure A1F). Lastly, SV40 NLS were added to the N and C termini of DVNP by PCR (Goldfarb et al., 1986). All plasmids were confirmed by sequencing which was conducted by the Nucleic Acid/Protein Service (NAPS) Unit at the University of British Columbia.

Plasmids were transformed into yeast using a lithium acetate-based protocol (Gietz and Woods, 2006). All yeast strains used in this study are listed in Table A1. *SPT21* was deleted from the wild type strain (FY602 (Sterner et al., 1999)) by targeted homologous integration using an amplicon containing the *HIS3* marker gene flanked by the ends of the *SPT21* gene (Bähler et al., 1998). Deletion was confirmed by PCR using primers directed within the marker gene and upstream of *SPT21*.

2.4.2 Total protein extraction and immunoblotting

Yeast strains containing expression vectors were grown to saturation at 30 °C in synthetic dropout media lacking uracil and supplemented with 2% dextrose. To induce DVNP expression, cells were collected by centrifugation at 3000 xg for 3 minutes and then washed twice and ultimately resuspended in the same dropout media containing 2% galactose. Cultures were grown in galactose for 16 hours at 30 °C to an optical density (OD_{600}) of 0.8 ± 0.1 prior to collection. Cell numbers were normalized by OD_{600} and total protein was isolated using a mild 0.2M NaOH alkali treatment (Kushnirov, 2000).

Protein samples were heated for 5 minutes at 95 °C and separated using 15% SDS-PAGE (sodium dodecyl sulfate - polyacrylamide gel electrophoresis). Following electrophoresis, gels were equilibrated in an SDS buffer (62.5 mM Tris pH 6.8, 2.3% SDS) for 30 minutes prior to being transferred to a nitrocellulose membrane in an ethanolamine transfer solution (0.15% ethanolamine, 0.017 mM glycine, 20% methanol). Transfer efficiency and equal protein loading were confirmed by ponceau staining prior to blocking in 2% powdered milk in PBS-T (0.68 M NaCl, 13.4 mM KCl, 50 mM Na_2HPO_4, 8.8 mM KH_2PO_4 pH 7.4, 1% Tween-20) for two hours at room temperature. Membranes were incubated with the following primary antibodies: HA (Roche, High affinity 3F10 clone, 1:2500, 16 hours, 4 °C), Pgk1 (Novex, 459250, 1:10000, 1 hour, 20 °C), H3 (Genscript, rabbit polyclonal raised to antigen CKDIKLARRLRGERS, 1:5000, 16 hours, 4 °C), H4 (Abcam, ab31830, 1:2000, 16 hours, 4 °C), or H2B (Active Motif, 39237, 1:2000, 16 hours, 4 °C). Following primary antibody incubation, membranes were washed three times in PBS-T and incubated with anti-rat (LiCOR, 926–32219), anti-mouse (LiCOR, 926–32221), and or anti-rabbit (LiCOR, 926–32210) secondary antibodies at 1:15000 dilutions for one hour at room temperature. Membranes were washed in PBS-T for 25

minutes and imaged using a LiCOR Odyssey imaging system. Protein quantification was performed using LiCOR Odyssey Infrared Imaging software v3.0.

2.4.3 Immunofluorescence microscopy

Immunofluorescence was conducted using a previously developed protocol (Pringle et al., 1989), with some modifications. Cells constitutively expressing DVNP from pRS416 vectors containing *HHT2* promoters (Lawrence et al., 2017) were grown in synthetic dropout media lacking uracil to an OD_{600} of 0.4 before being harvested and fixed in 3.7% formaldehyde for one hour at 25 °C. Fixed cells were pelleted at 9000 xg for 30 seconds before being washed twice in SK buffer (1 M sorbitol, 50 mM KPO_4, pH 7.5) and stored at 4 °C for 48 hours. Fixed cells were then applied to poly-L lysine (Sigma) coated slides and allowed to settle for 5 minutes. The cell solution was then aspirated and the slide was washed twice with SK buffer prior to being submerged in a -20 °C methanol bath and -20 °C acetone bath for 6 and 3 minutes, respectively. Non-specific sites were blocked with 3% bovine serum albumin (BSA) in PBS for 20 minutes. The slides were then incubated with 1:100 HA antibody (Roche, High affinity 3F10 clone) diluted in 3% BSA PBS for one hour at 37 °C in a humidified chamber. Following this, the slides were washed with PBS and incubated as above with 1:2000 fluorescein conjugated anti-rat antibody for 45 minutes. Slides were finally washed again and mounted with fluoromount aqueous mounting media (Sigma) containing 2.5 μg/mL Hoescht stain. Micrographs were acquired on a Zeiss Axio Observer inverted microscope equipped with a Zeiss CoilibriLED illuminator and a ZeissAxiocam ultrahigh-resolution monochrome digital camera Rev 3.0. Immunofluorescent images were analyzed using Zeiss ZEN software v2.1 and ImageJ.

2.4.4 Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) experiments were performed based on previously outlined protocols (Maltby et al., 2012). Cells were grown as described above (see Total protein extraction and immunoblotting) before being fixed in 1% formaldehyde for 30 minutes at room temperature. Excess formaldehyde was quenched with 125mM glycine for 15 minutes and then cells were pelleted at 3000 xg for 3 minutes at 4 °C and washed with cold PBS. After three washes, cells were normalized to 40 OD units (ODU) before being frozen at -80 °C.

Following thawing, cells were resuspended in lysis buffer (50 mM HEPES pH 7.5, 140 mM NaCl, 0.5 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 1 mM phenylmethanesulfonyl fluoride (PMSF), and 1X Protease inhibitor cocktail (Roche)) and lysed by bead beating. Cell lysates were pelleted at 15,000 xg for 30 minutes at 4 °C, washed and resuspended in lysis buffer, and sonicated for 30 cycles of 30 seconds on/30 seconds off at high power at 4 °C using a Biorupter sonicator (Diagenode). Sonicated lysates were then pre-cleared with protein G conjugated magnetic beads (Dynabeads, Thermo Fisher) for 1 hour at 4 °C. After clearing, 6% of the lysate was collected as 'input' and 1:400 anti-HA (Roche, High affinity 3F10 clone) antibody or 1:1250 anti-Rpb3 (Abcam, ab81859, monoclonal clone 1y26[1y27]) antibody was added prior to 16 hours of rotation at 4 °C. Antibodies were extracted using protein G conjugated magnetic beads (Dynabeads, Thermo Fisher) for 4 hours at 4 °C and the beads were subsequently washed twice with lysis buffer, twice with lysis buffer supplemented with 500 mM NaCl, twice with lithium buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.6% NP-40, 0.5% Na-deoxycholate, 1 mM EDTA pH 8), and once with TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). Immunoprecipitates were eluted with elution buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8, 1% SDS, 150 mM NaCl, 5 mM DTT) at 65 °C and then treated with 80 µg/mL proteinase K at 65 °C for 16 hours and 300 µg/mL RNase A at 37 °C for 2 hours. For the Rpb3 ChIP, prior to DNA purification, spike-in DNA (10:1 ($2x10^{-4}$: $2x10^{-5}$ ng/µL) spike-in 1:2) was added to a 1:400 and 1:3.33 dilution in the ChIPs and inputs, respectively. DNA purification was performed using a Qiagen Minelute PCR purification kit or by phenol:chloroform:isoamyl extraction. DNA fragmentation and concentration were assessed using a 1% agarose gel containing syto60 dye (Invitrogen) and a high sensitivity Qubit fluorometer (Thermo Fisher), respectively.

2.4.5 Micrococcal nuclease digestion

Micrococcal nuclease (MNase) digestions were performed as described previously (Lawrence et al., 2017). Cells were grown as for ChIP (see Chromatin immunoprecipitation) before being normalized to 25 ODUs. Cells were resuspended in 1M sorbitol, 5mM β-mercaptoethanol and 10 mg/mL zymolyase prior to being incubated at 37 °C for 10 minutes. Spheroplasts were washed in 1M sorbitol, twice in spheroplast digestion buffer (SDB: 1 M sorbitol, 50 mM NaCl, 10 mM Tris pH 8, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM β -mercaptoethanol, 0.5 mM spermidine, 0.075%

NP40) and resuspended in SDB before being digested with MNase for 2 minutes. Digestions were stopped with 5mM EDTA and 1% SDS and crosslinks were reversed by overnight incubation at 65 °C. Proteinase and RNase treatment as well as DNA fragmentation assessments and concentration were done as above (see Chromatin immunoprecipitation).

2.4.6 ChIP-quantitative PCR

ChIP-quantitative PCR was performed using previously developed protocols (Martin et al., 2017). In particular, Rpb3 ChIP eluates (see chromatin immunoprecipitation) were diluted by a factor of 20 and quantified by quantitative polymerase chain reaction (qPCR). qPCR reactions were performed in technical triplicate, using SYBR green for detection in an Applied Biosystems StepOnePlus Real-Time PCR System, and quantified against a standard curve of genomic DNA. Primers used for qPCR are listed in Table A2.

2.4.7 Sequencing and bioinformatic analysis

Sequencing libraries were constructed using 2 ng of DNA using a low-input protocol (Brind'Amour et al., 2015). Briefly, samples were end repaired (1X T4 DNA ligase buffer (NEB), 0.4 mM dNTP mix, 2.25 U T4 DNA polymerase (NEB), 0.75 U Klenow DNA polymerase (NEB), and 7.5 U of T4 polynucleotide kinase (NEB), incubated at room temperature for 30 minutes), A-tailed (1X NEB buffer 2, 0.4 mM dATP, and 3.75 U of Klenow (exo-) (NEB) incubated at 37 °C for 30 minutes), ligated to adaptors (1X Quick DNA ligase buffer (NEB), 1 mM Illumina PE adapters, and 1600 U Quick DNA-ligase (NEB), incubated at room temperature for 1 hour) and PCR amplified (1X NEBNext master mix (NEB) and 0.4 μM indexed primers (Illumina)) using 12 PCR cycles with a 65 °C annealing temperature and a 30 second extension time. DNA was purified between each step using two volumes of NucleoMag NGS DNA purification beads (Macherey-Nagel) except after adaptor ligation and PCR amplification where 0.8 volumes were used to facilitate size selection. Library yield and size distribution was assessed using a high sensitivity Qubit fluorometer (Thermo Fisher) and an Agilent Tape Station, respectively.

Libraries were pooled and size selected on a 2% agarose gel to between 100 and 1000 bp. Pooled libraries were then sequenced using either 80 bp paired end reads on an Illumina MiSeq

using a v3 reagent kit (DVNP ChIPs, inputs, and MNase-seq) or on an Illumina HiSeq with 100 bp paired end reads using a HiSeq SBS v4 reagent kit (Rpb3 ChIPs and inputs). FASTQ files were initially assessed using FastQC v0.11.4 prior to being aligned to saccer3, the most recent build of the yeast genome (released February 3, 2011; downloaded from http://www.yeastgenome.org), using the Burrows Wheeler aligner (BWA) algorithm v0.7.13 (Andrews, 2010; Li and Durbin, 2009). Samtools v0.1.19 was then used to filter out mapped reads with mapping quality scores less than 10 (Li et al., 2009). Sequence fragment sizes were filtered to exclude excessively large and small fragments as inferred from fragment size distributions. Subsequent analyses and statistics were performed using the Java Genomics toolkit (downloaded from http://palpant.us/javagenomics-toolkit/), DANPOS v2 (Chen et al., 2013), and R v3.4.0.

Average gene profiles were obtained by averaging the sequencing coverage, which was normalized to the average genomic coverage, at each base, 500 bp upstream and 2500 bp downstream of the transcription start site (as defined by simultaneous mapping of RNA ends by sequencing (SMORE-Seq) (Parky et al., 2014) of 4793 genes. Genes were included in these calculations until their polyadenylation sites were reached (as defined by SMORE-Seq (Parky et al., 2014). For the Rpb3 ChIP calculations, genes were included until 300 bp from their polyadenylation sites due to large peaks at the 3' ends of many genes that skewed quantification. Genome wide analyses were performed by either dividing the genome into 500 bp windows with 250 bp steps or by calculating occupancy over nucleosomal sites (Brogaard et al., 2012). To avoid DNA accessibility bias, all correlations were observed in and out of input-controlled bins. With regards to box plots, boxes span from the first to third quartile with whiskers extending 1.5 times the interquartile range (IQR). Black bars represent the median and notches represent an approximation of the 95% confidence interval and extend ±1/58 IQR/sqrt(*n*).

2.4.8 Synthetic genetic array

Synthetic genetic array (SGA) analysis was carried out using a ROTOR colony manipulation robot (Singer Instruments) in combination with the non-essential yeast deletion array as previously outlined (Tong and Boone, 2006; Tong et al., 2001; Winzeler et al., 1999). The SGA starting strain, Y7093, was transformed with the pRS416-*GAL1pr*-DVNP-3HA-NLS plasmid and mated with the deletion array. Diploids were selected using YPD (1% yeast extract, 2%

peptone, 2% dextrose) supplemented with 0.25 mg/mL G418 and 0.1 mg/mL nourseothricin and sporulated on depleted media (1% KOAc, 0.5% yeast extract, 0.5% dextrose, 0.001% sporulation amino acid mix, 2% agar, 0.25 mg/mL G418) for 11 days at 30 °C. Double mutant haploids were then selected and cultured on germination media (0.7% yeast nitrogenous base without ammonium sulphate, 0.2% complete supplement mixture lacking arginine, lysine, histidine, and uracil, 2% dextrose, 2% agar, 0.05 mg/mL thialysine, 0.05 mg/mL canavanine, 0.25 mg/mL G418 and 0.1 mg/mL nourseothricin). The resulting strains were then plated onto either germination media containing 2% galactose and 2% raffinose (experimental plates) or germination media containing 2% galactose, 2% raffinose, and 2 mg/mL 5-fluoroorotic acid (5-FOA), a drug which selects for loss of *URA3*-based plasmids (control plates). Plates were imaged on a flat-bed scanner and colony size and relative growth were quantified and analyzed using Balony v1.2.1 using the default settings (Young and Loewen, 2013). Default thresholds for growth and lethality were set and resulting mutants above the rescue threshold were inspected using FunSpec (Robinson et al., 2002). Functional categories were assigned using Munich Information Centre for Protein Sequences (MIPS) functional classifications (Mewes et al., 2002).

2.4.9 Data availability

The ChIP-seq and MNase-seq data sets have been deposited in the Gene Expression Omnibus under accession number GSE102280. The SGA data is available in Supplementary Data 1 of the published manuscript (https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-018-03993-4/MediaObjects/41467_2018_3993_MOESM1_ESM.pdf).

2.5 Chapter-specific acknowledgements

I acknowledge Savrina Manhas, Vivien Measday, Michelle Moksa, Martin Hirst, James Kwan, Corey Nislow, and Dolph Schluter for providing assistance and advice on the manuscript and analyses. This work was supported by a Natural Sciences and Engineering Research Council (NSERC) Discovery Grant awarded to LeAnn J. Howe. Nicholas Irwin and Benjamin Martin were supported by graduate student fellowships from NSERC.

3. Extensive reduction of the nuclear pore complex in nucleomorphs

3.1 Introduction

The nuclear pore complex (NPC) is a large macromolecular assembly that structures the membranous pores of the nuclear envelope and serves as a gate into the nucleus (Beck and Hurt, 2017; Strambio-De-Castillia et al., 2010). The NPC consists of over thirty protein subunits, termed nucleoporins, that are typically arranged with eight-fold symmetry around the central channel of the pore. These subunits are organized into individual subcomplexes: the cytoplasmic complex, which regulates nuclear import and export; the outer, inner, and transmembrane rings, which form structural scaffolds; the central channel, which mediates passage through the pore; and the nuclear basket, which interacts with nuclear factors (Beck and Hurt, 2017). These subcomplexes act in concert to regulate nucleo-cytoplasmic exchange through two mechanisms (Stewart, 2007). Firstly, the NPC generates a size exclusion barrier that bars the passive diffusion of molecules greater than 40 kDa (Paine et al., 1975). Secondly, the NPC mediates the passage of materials exceeding this threshold through an active transport system dependent on nuclear localization signals, transport receptors, and import proteins (Stewart, 2007). Through these mechanisms, the NPC facilitates the bidirectional exchange of proteins into and out of the nucleus. But nucleoporins can also function beyond trafficking. Indeed, nucleoporins facilitate dynamic chromatin regulation and gene expression through trafficking-independent mechanisms by manipulating genome architecture and activating transcription (Bermejo et al., 2012; Capelson et al., 2010; Liang and Hetzer, 2011). Therefore, the NPC not only represents a trafficking hub, but a highly integrated nuclear component that is ubiquitous and highly conserved amongst eukaryotes (Neumann et al., 2010).

One potential exception to the conservation of the NPC are nucleomorphs. Nucleomorphs are highly reduced, relict nuclei that were derived by secondary endosymbiosis (Gilson et al., 2006; Greenwood, 1974; Ludwig and Gibbs, 1989; McFadden and Gilson, 1995). In contrast to primary endosymbiosis, where a bacterium is incorporated into a eukaryotic cell, secondary

endosymbiosis involves the uptake of one eukaryote by another, or more specifically, the uptake of a primary alga such as a green or red alga by another eukaryote (Keeling, 2017). Secondary endosymbiosis typically results in the complete reduction of the endosymbiont, leaving only the plastid; however, in two instances, the nucleus of the endosymbiont has remained (Keeling, 2013). These nuclei, termed nucleomorphs, were retained in two lineages, the chlorarachniophytes and the cryptophyceans (plastid containing cryptista) (Greenwood, 1974; Ludwig and Gibbs, 1989). The nucleomorphs of chlorarachniophytes and cryptophyceans were derived independently from green and red alga, respectively. However, these structures exhibit strong convergent reductive evolution both in terms of genome organization and nuclear functionality. For example, both chlorarachniophyte and cryptophycean nucleomorph genomes house three small linear chromosomes and typically conserved nuclear features such as histone post-translational modifications and the C-terminal domain of RNA polymerase II have been highly reduced or lost in both instances (Douglas et al., 2001; Gilson et al., 2006; Marinov and Lynch, 2016). Consequently, nucleomorphs represent a unique, highly reduced nuclear system, but whether this reduction has affected the structure and function of other conserved features, including nuclear pores, remains unclear.

A previous investigation into whether NPC genes are encoded in the nucleomorph genome itself failed to identify any homologues in both chlorarachniophyte and cryptophycean nucleomorphs (Neumann et al., 2006). This suggests at least three non-exclusive possibilities: 1) that nucleomorph NPC genes have been transferred to the host nucleus; 2) that host NPC proteins are dual-targeted to both the host nucleus and nucleomorph; and 3) that the nucleomorph has lost its NPC (Neumann et al., 2006). To discern between these possibilities, we used nuclear genomic and transcriptomic data from chlorarachniophytes and cryptophyceans to identify and phylogenetically classify all NPC proteins in nucleomorph-containing algal lineages, as well as in red and green algal representatives, and non-photosynthetic lineages closely-related to the host. If nucleomorph NPC genes were transferred to the host nucleus, we would expect to identify two sets of nucleoporins: one of which would be of algal origin, and the other of host origin. If a dual-targeting mechanism had been employed, we would observe the expression of multiple isoforms of host homologues, some of which would contain the signal and transit peptides required for nucleomorph targeting (Gould et al., 2006; Hirakawa et al., 2009, 2010). Finally, if the nucleomorph lacks an NPC, then only host-derived genes should be found,

comparable to outgroup taxa. Overall, this search supports the latter explanation: chlorarachniophytes and cryptophyceans mostly encode single, host-related NPC genes lacking signal peptides. Two exceptions were Nup98 and Rae1 which had clear nucleomorph-derived homologues, but these two proteins are insufficient to structure a canonical NPC alone and are known to have other, non-NPC functions in the nucleus. Ultimately, these data suggest that nucleomorphs represent eukaryotic nuclei that function without a canonical NPC.

3.2 Results and Discussion

3.2.1 Identification of nucleoporins in nucleomorph-containing algae and their relatives

To characterize and identify the origin of nucleomorph nucleoporins, we used profile hidden-Markov models (HMMs), generated from alignments of previously-identified NPC proteins from diverse eukaryotes (Neumann et al., 2010), to search for NPC proteins in the transcriptomes and genomes of nucleomorph-containing chlorarachniophytes and cryptophyceans (eight chlorarachniophyte transcriptomes, eight cryptophycean transcriptomes, one chlorarachniophyte genome, and two cryptophycean genomes). We also examined three genomes from green algae and two from red algae, representing close relatives of the endosymbiotic partners of the chlorarachniophytes and cryptophyceans, respectively. Additionally, we examined two transcriptomes and a genome from nucleomorph-lacking relatives of chlorarachniophytes and three transcriptomes from nucleomorph-lacking cryptistans.

Our HMM search succeeded in identifying NPC proteins in all the examined lineages. Of the 32 nucleoporins, 27 were identified in chlorarachniophytes and their relatives (Figure 8: 27 in outgroups and 24 in chlorarachniophytes) and 18 were found in cryptistans (Figure 9: 16 in outgroups and 16 in cryptophyceans). Furthermore, 25 NPC proteins were found in green algae and 19 in red algae. Animal and fungal-specific proteins, Nup37, Nup358, Pom34, and Pom121, were not identified in our datasets, highlighting the specificity of the search (Field et al. 2014). However, Pom152, traditionally considered to be a fungal-specific protein, was observed in 10 of the 13 cryptistan species. This protein was previously reported from the cryptophycean genome of *Guillardia theta* and was speculated to have been spread through horizontal gene transfer (Field et al., 2014). We sought to investigate this further by generating a phylogeny with

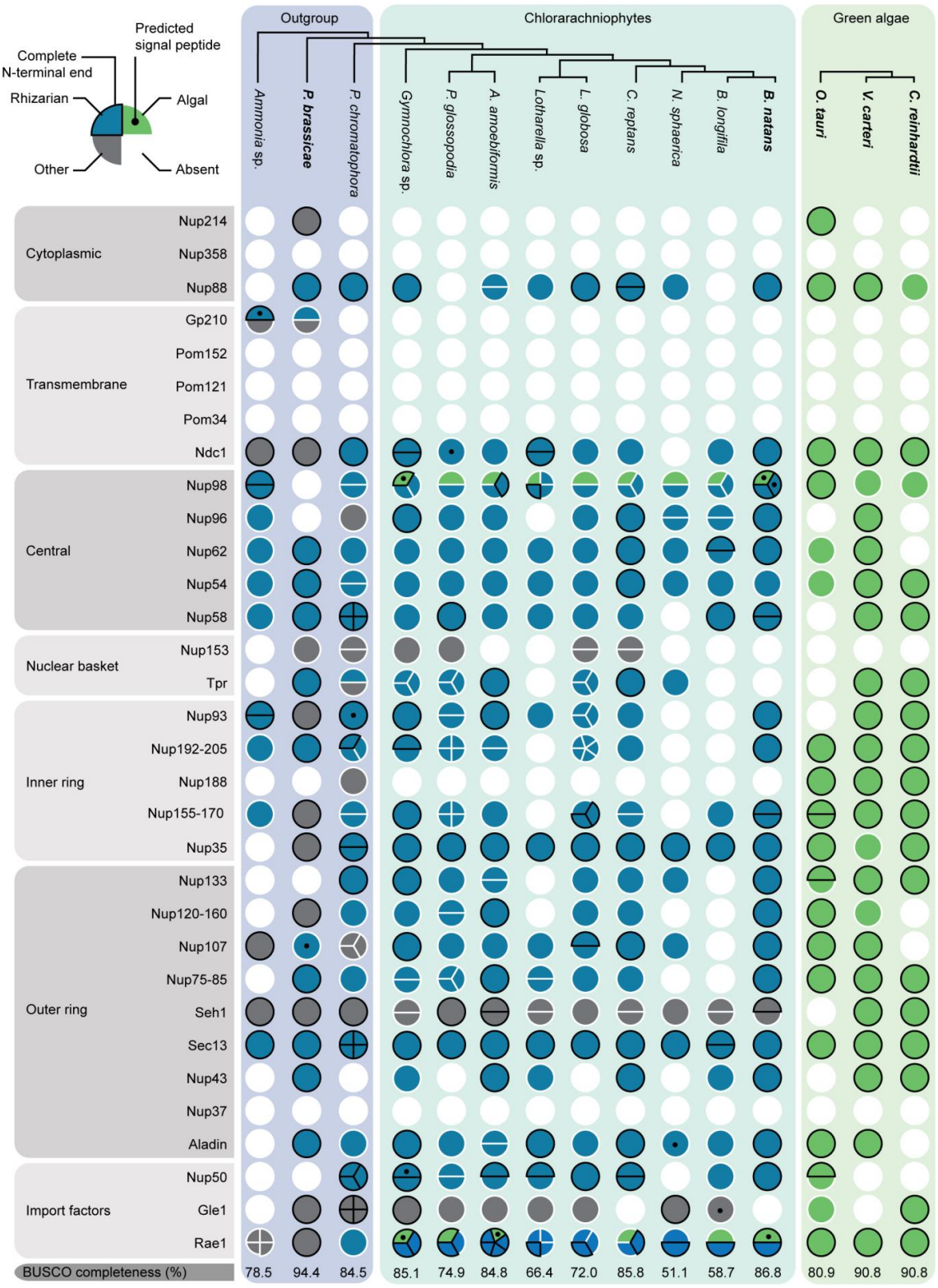


Figure 8. Nucleoporins identified in chlorarachniophytes, host-related lineages, and green algae. Circles represent proteins with each row corresponding to a different nucleoporin and

each column representing a taxon. The different proteins are shown on the left and grouped into subcomplexes (Beck and Hurt, 2017; Strambio-De-Castillia et al., 2010). Note that some proteins, such as Nup98, can be found in different parts of the NPC. A schematic phylogenetic tree is shown at the top and based on phylogenomic analyses (Irwin et al., 2019). Coloured circles represent found proteins whereas white circles represent absent proteins. Circles were subdivided to represent the number of protein copies that were identified and were coloured based on the phylogenetic affinity of the protein. In particular, circles were coloured blue, green, or grey depending on whether they clustered with rhizarian outgroups, algae or plants, or other taxa. Circles or wedges were outlined in black if they were predicted to have complete N-termini and black spots are present when a signal peptide was predicted using SignalP v4.1. Taxa with genome data have their names written in bold. The BUSCO completeness for each dataset is provided in the last row.

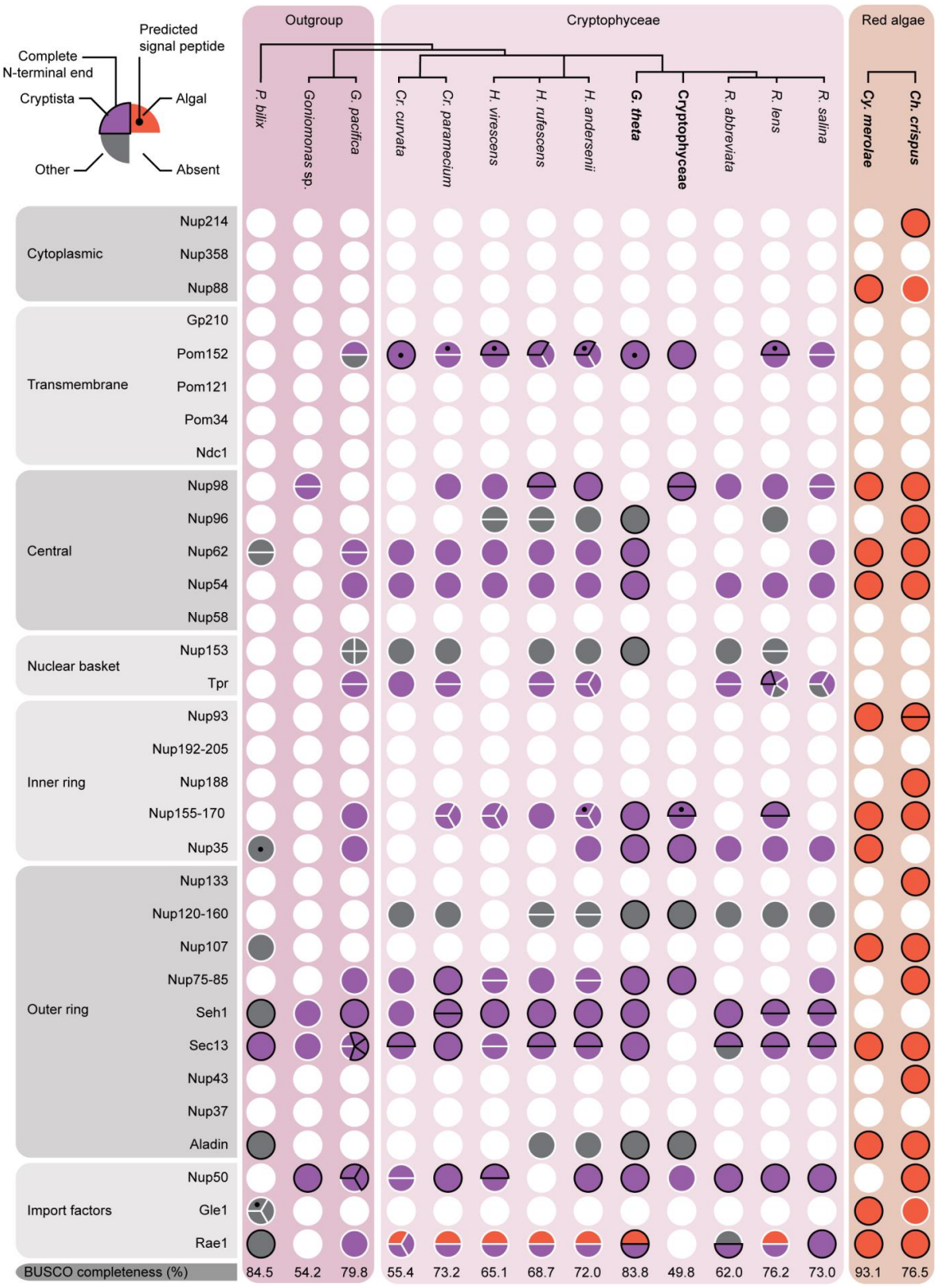


Figure 9. Nucleoporins from cryptistans and red algae. For a complete description of this figure, see the figure legend for Figure 8. A schematic phylogenetic tree is shown at the top and

based on phylogenomic analyses (Burki et al., 2016). Circles were coloured based on the phylogenetic affinity of the protein. In particular, circles were coloured purple, red, or grey depending on whether they clustered with the cryptistan outgroups, algae or plants, or other taxa.

---

cryptistan and fungal Pom152 proteins as well as prokaryotic homologues identified using the Pom152 HMM profile (Appendix Figure B1). However, the resulting phylogeny was largely inconclusive, revealing a three-clade tree lacking obvious directionality.

Mapping the identified proteins to known NPC subcomplexes showed that the search results were comprehensive for both lineages. Within rhizarians (i.e., chlorarachniophytes and their relatives) and green algae, the identified nucleoporins corresponded to each of the NPC subcomplexes, whereas proteins structuring the cytoplasmic complex and nuclear basket were conspicuously absent from cryptistans and red algae, respectively. BUSCO (Benchmarking Universal Single Copy Orthologs) analysis revealed that transcriptomic dataset completeness was variable but generally comparable to the genomic datasets, indicating that the data coverage was representative (Figure 8, 9) (Simão et al., 2015). Furthermore, consistent absences in multiple taxa, independent of dataset completeness, support the absence of these proteins (Figure 8, 9). However, failure to identify homologues of subcomplex components does not necessarily mean the substructure itself is absent. Indeed, previous investigations into the structure of trypanosome NPCs identified novel nucleoporins and divergence in peripheral structures (i.e., the cytoplasmic complex and nuclear basket) (Holden et al., 2014; Obado et al., 2016). This was suggested to be a result of cytoplasmic and nuclear functional divergence which also could have occurred in some of the groups examined here (Obado et al., 2016).

Overall, the strongest trend in the data was the consistency of both copy number and phylogenetic affinity of the nucleoporins, suggesting that single isoforms of host origin predominate in both chlorarachniophytes and cryptophyceans (Figure 8, 9). Multiple copies of some genes were identified in a few species, but they were exceptions and their narrow distribution suggests they are recent gene duplications, alternative transcripts, or mis-assemblies. Moreover, phylogenetic analysis showed that nearly all isoforms were derived from the host lineage, and were not related to the algal endosymbiont (Figure 8, 9). Consistent with this,

almost no host-related proteins were predicted to contain signal peptides, with the exception of Pom152, a transmembrane nucleoporin. These data are therefore inconsistent with both the dual-targeting of host proteins, and the large-scale transfer of NPC genes from the nucleomorph to the host genome.

The few nucleoporins that broke from this pattern were interesting exceptions, but none that suggested a nucleomorph NPC. First, several proteins were found to have an unresolved phylogeny. Nup96 was identified in five cryptophycean lineages as well as a red alga but was absent from cryptistan outgroups. Still, these cryptophycean sequences did not phylogenetically cluster with the endosymbiont lineage (Figure 9). Moreover, Nup153, Seh1 and Gle1 in the chlorarachniophytes, along with Nup153, Nup120-160, and Aladin in the cryptophyceans had uncertain origins due to phylogenetic ambiguities, or a lack of sampling of outgroups and/or algal proteins (Figure 8, 9). Previous phylogenetic analyses have revealed that host-encoded nucleomorph genes do not necessarily cluster phylogenetically with the expected algal homologues, potentially because of artifacts such as long branch attraction (Gile and Keeling, 2008; Hirakawa et al., 2011; Onuma et al., 2017). Hence, we cannot rule out that these nucleoporins are of nucleomorph origin. However, none of these proteins were predicted to contain signal peptides, which is expected for any protein targeted to the nucleomorph (Hirakawa et al., 2011), and some of those proteins, such as Seh1and Gle1 in chlorarachniophytes, even branched with the outgroup lineages but with poor support (< 70 ultrafast bootstrap), making the case for their nucleomorph localization and origin weak. Lastly, multiple isoforms of some nucleoporins (Nup192-205, Seh1, and Nup50 in chlorarachniophytes and Tpr and Sec13 in cryptophyceans) were observed in the majority of taxa. Although these proteins were not predicted to have signal peptides, many lacked complete N-terminal ends making the targeting of these proteins inconclusive. Despite this, the general lack of multiple isoforms and the absence of signal peptides strongly indicates that the widespread presence of host-derived nucleoporins in the nucleomorph is unlikely. Similarly, there is no evidence for the dual-targeting of host nucleoporins based on alternative spicing or alternative transcription start sites. These genes would be very similar throughout the mature protein, so in principal could be easily overlooked. However, the improbability of overlooking the presence of alternative transcripts for every protein from every taxon analyzed makes this a weak explanation as well.

3.2.2 Nup98 and Rae1 are the last remaining nucleomorph-related NPC proteins

In contrast to the ambiguous cases described above, two NPC proteins were clearly nucleomorph-related, Nup98 in chlorarachniophytes, and Rae1 in both chlorarachniophytes and cryptophyceans. Chlorarachniophytes consistently encoded multiple copies of both Nup98 and Rae1 (Figure 8, 9), with one copy branching with the algal endosymbiont lineage and the other with outgroup taxa (Figure 10). Moreover, signal peptides were predicted in the algal-related Nup98 isoforms that had complete N-termini, and in three of the five algal-related Rae1 proteins (Figure 8, 10). In cryptophyceans, an algal-related homolog of Rae1 was also found (Figure 9). Cryptophycean Rae1 genes were not predicted to encode signal peptides (Figure 9), which may reflect a relict nucleomorph protein that no longer undergoes trafficking, but is more likely a mis-prediction since most of the algal-related homologues are also incomplete.

That both Nup98 and Rae1 would stand out is notable because both are also uniquely functionally diverse nucleoporins. Unlike most NPC proteins, Nup98 associates with either face of the pore and functions both at the nuclear envelope and within the nucleoplasm (Griffis et al., 2002, 2003). Similarly, Rae1 associates with the NPC but is also transient, facilitating mRNA export through interactions with microtubules and importins (Blower et al., 2005; Murphy et al., 1996). Nup98 is also known to play additional roles in diverse nuclear systems. For example, distinct Nup98 isoforms are used to distinguish the macronuclei and micronuclei of dikaryotic ciliates (Iwamoto et al., 2009), and in metazoans, Nup98 is important for NPC disassembly during mitosis (Laurell et al., 2011). Nup98 and Rae1 are also both known to function in RNA trafficking through mechanisms that are, at least in part, dependent on physical interactions between one another (Blevins et al., 2003; Blower et al., 2005; Ren et al., 2010). Furthermore, Nup98 and Rae1 form a cell cycle regulating complex capable of activating the anaphase promoting complex (APC), which facilitates cell cycle progression through the ubiquitin-dependent degradation of securins (Babu et al., 2003; Jeganathan et al., 2005). Given the functional diversity of these proteins, and in particular their functions outside the NPC, their role in the nucleomorph cannot be concluded to be NPC-specific. However, Nup98 also associates with the central channel and the nucleomorph-related homologues still retain FG-repeat domains,

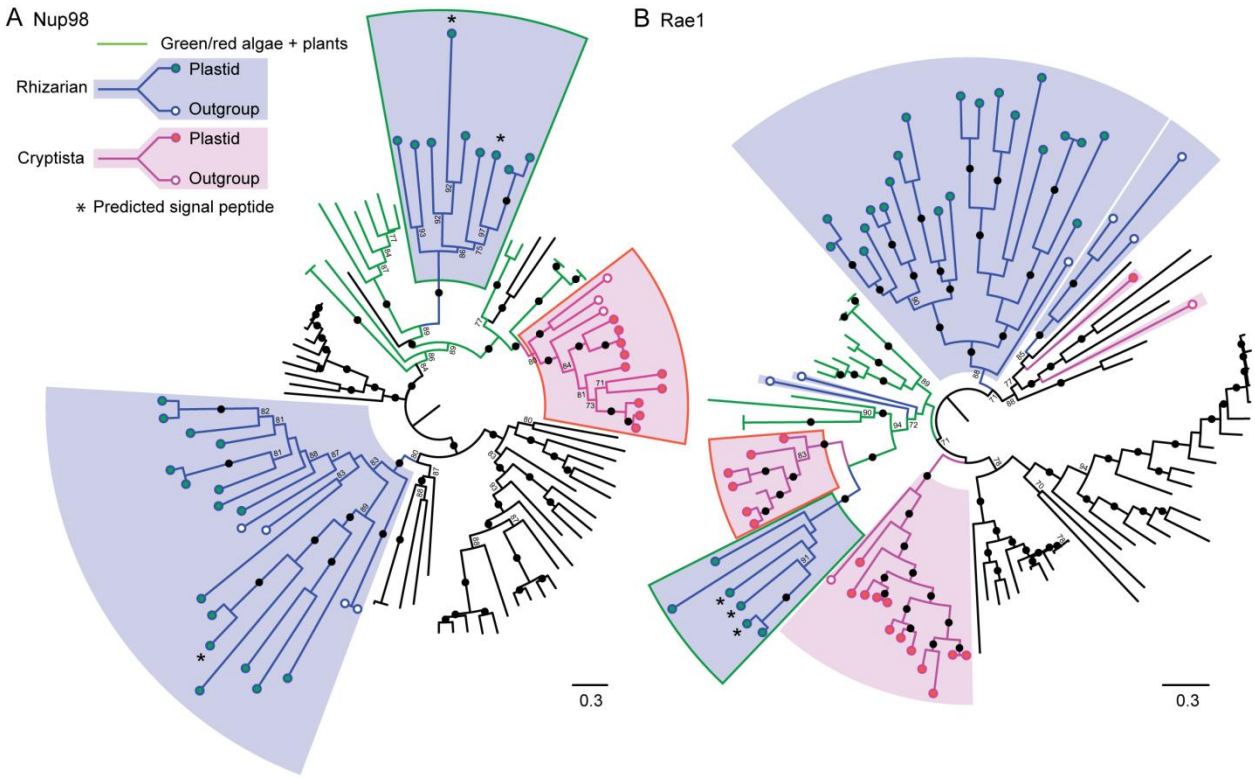


Figure 10. Phylogenetic analysis of Nup98 and Rae1. Maximum likelihood phylogenies for Nup98 (**A**) and Rae1 (**B**) were generated in IQ-Tree using the VT+F+R5 model (Nup98) and the LG+F+R5 model (Rae1) as selected by ModelFinder. Statistical support was assessed using 1000 ultrafast bootstraps and the resulting values are shown at the nodes. Values below 70 are not shown and black circles represent values over 95. Rhizarians, cryptistans, and algae and plants are labeled with blue, pink, and green, respectively. Rhizarian and cryptistan taxa containing nucleomorphs are denoted with green and red filled circles, whereas nucleomorph-lacking taxa are labeled with empty circles. Other eukaryotic taxa are shown in black and the trees were rooted at the midpoint. Host and endosymbiont-derived clades are outlined in white and green or red, respectively. The fully annotated trees are available in Appendix Figure B2.

---

which are required for NPC function (Beck and Hurt, 2017). Therefore, it is possible that Nup98 could be structuring a highly simplified NPC in chlorarachniophyte nucleomorphs.

A number of host-encoded nucleomorph-targeted proteins are implicated in cell cycle progression, and have been predicted to provide the host with control over the division of its

endosymbiont. These proteins include a DNA polymerase in chlorarachniophytes, as well as histones in both chlorarachniophytes and cryptophyceans (Hirakawa et al., 2011; Onuma et al., 2017; Suzuki et al., 2016). This is also possible for Nup98 and Rae1, however neither nucleomorph-related Nup98 nor Rae1 is differentially expressed over the cell cycle in the chlorarachniophyte, *Bigelowiella natans* (Suzuki et al., 2016), suggesting that if they are involved in the cell cycle, downstream regulators must be present. Furthermore, APC proteins have not been identified in nucleomorph genomes and those annotated in the nuclear genome of *B. natans* are not predicted to be nucleomorph targeted (Curtis et al., 2012; Suzuki et al., 2016). Hence, the role of Nup98 and Rae1 in the chlorarachniophyte nucleomorph remains uncertain. The nucleomorph-related Rae1 in cryptophyceans is even more functionally ambiguous.

3.2.3 A nucleus without a nuclear pore complex?

The retention of only one or two clearly nucleomorph-derived NPC genes in both the chlorarachniophytes and cryptophyceans, and the lack of any evidence for the targeting of host-derived nucleoporins, reveals that nucleomorphs have convergently lost the molecular machinery required to structure a canonical NPC (Figure 11). Nucleomorphs are accordingly the only known eukaryotic nuclei that appear to function without a typical NPC, which is otherwise highly conserved (Neumann et al., 2010). However, the nucleus must maintain contact with the cytoplasm, and transmission electron microscopy has revealed pore-like structures in the nucleomorph envelope (Ludwig and Gibbs, 1989). Perhaps one clue to resolving this is in the observation that fluorescent fusion proteins targeted to the periplastidal compartment (PPC, or the endosymbiont-derived cytoplasm) of the chlorarachniophyte, *Amorphochlora amoebiformis*, localize to both the PPC and the nucleomorph, despite exceeding the typical NPC size exclusion threshold of 40 kDa (Hirakawa et al., 2009; Paine et al., 1975). This suggests that there is a pore, but a functionally different one, possibly only facilitating passive transport and lacking the canonical regulatory capacity of the NPC. This would indicate that the nucleoplasm and cytoplasm were much more similar in these reduced systems than in their more complex ancestors.

In the absence of a canonical NPC (i.e., the specific and conserved set of proteins associated with pore), how a nucleomorph nuclear pore could be structured, particularly in cryptophyceans which lack Nup98, is unclear. Previous studies have revealed homology between

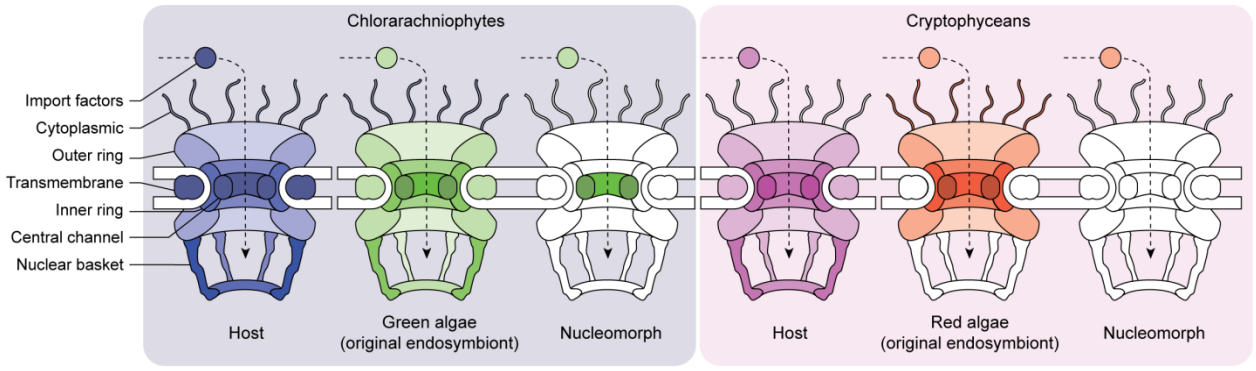


Figure 11. Nuclear pore complex in chlorarachniophytes, cryptistans, their nucleomorphs, and their algal endosymbionts. A summary schematic depicting the different subcomplexes of the NPC. If a single protein was found corresponding to a given subcomplex, the structure was coloured in. If no proteins were found, the subcomplex was left white.

---

nucleoporins and endomembrane components such as coatomers, which are a family of membrane-bending proteins, including those found on COPI and COPII-coated vesicles (Devos et al., 2004). It is thought that the NPC evolved by making these generic functions progressively more elaborate (Devos et al., 2004; Wilson and Dawson, 2011). One hypothesis is that nucleomorph pores could have reverted to something akin to such an ancestral state, perhaps utilizing coatomer proteins for a much simpler structure involved only in membrane bending. It is also possible that some nucleomorph nucleoporins have been retained but in a highly derived state that cannot be detected with current bioinformatic methods. Moreover, nucleomorphs may utilize novel nucleoporins as is the case with trypanosomes, although even trypanosomes still retain the majority of the canonical NPC proteins (Holden et al., 2014; Neumann et al., 2010; Obado et al., 2016). Ultimately, biochemical analyses will be required to fully understand how nucleomorph pores are structured and function in such a reduced state.

While the loss of the NPC is unique, the reduction in the complexity of protein trafficking systems may be a common occurrence during the reductive evolution of organelles. Reduced mitochondria, including hydrogenosomes and mitosomes, often lack seemingly essential mitochondrial import machinery such as the TIM (Translocase of the Inner Membrane) and TOM (Translocase of the Outer Membrane) complexes (Burri and Keeling, 2007; Dolezal et al.,

2005; Pyrihová et al., 2018). It is possible that trafficking proteins involved in nuclear transport could be similarly dispensable for baseline function, and thus the simplification of trafficking systems may represent a common theme in reductive evolution.

3.3 Conclusions

Here we show the nucleomorphs of chlorarachniophytes and cryptophyceans lack a canonical NPC as only two nucleomorph-derived nucleoporins, Nup98 and Rae1, could be identified in these lineages. These proteins are also involved in other cellular processes and are alone insufficient to structure a canonical NPC, revealing nucleomorphs to be the only known eukaryotic nuclei to lack the complex. Despite the likely retention of some kind of pore in the nucleomorph envelope, its exact structure, function, and role (if any) in regulation remain unknown. The lack of a canonical NPC suggests that investigations into the physical structure of nucleomorph pores should reveal unique insights into the function and evolution of the NPC as well as the reduction of endosymbionts and organelles.

3.4 Materials and Methods

3.4.1 Data acquisition and completeness analysis

Transcriptomic data was obtained for eight chlorarachniophytes (*Bigelowiella longifila* MMETSP1359; *Norisiella sphaerica*, MMETSP0113; *Chlorarachnion reptans*, MMETSP0109; *Amorphochlora amoebiformis* MMETSP0042, *Gymnochlora* sp. MMETSP0110; *Partenskyella glossopodia* MMETSP1318; *Lotharella globosa* MMETSP0111; *Lotharella* sp. MMETSP0040) one rhizarian outgroup (*Ammonia* sp. MMETSP1384), and 11 cryptistans (*Palpitomonas bilix* MMETSP0780; *Goniomonas* sp. MMETSP0114; *Goniomonas pacifica* MMETSP0108; *Cryptomonas curvata* MMETSP1050; *Cryptomonas paramecium* MMETSP0038; *Hemiselmis virescens* MMETSP1356; *Hemiselmis rufescens* MMETSP1357; *Hemiselmis andersenii* MMETSP0043; *Rhodomonas abbreviata* MMETSP1101; *Rhodomonas lens* MMETSP0484; *Rhodomonas salina* MMETSP1047) from the reassembled set of transcriptomes generated during the Marine Microbial Eukaryotic Transcriptome Sequencing Project (Johnson et al., 2019; Keeling et al., 2014). The previously generated transcriptome of *Paulinella chromatophora* (SRR3221671) was also utilized (Nowack et al., 2016). The genomes of *Bigelowiella natans*,

Plasmodiophora brassicae, *Guillardia theta*, Cryptophyceae sp. CCMP2293, *Chlamydomonas reinhardtii*, *Volvox carteri*, and *Ostreococcus tauri* were downloaded from the Joint Genome Institute (JGI) database (Curtis et al., 2012; Merchant et al., 2007; Palenik et al., 2007; Prochnik et al., 2010; Schwelm et al., 2015). The red algal genomes of *Cyanidioschyzon merolae* and *Chondrus crispus*, downloaded from EnsemblPlants, were also utilized (Collen et al., 2013; Nozaki et al., 2007). Proteins were predicted from the transcriptomic data using TransDecoder v5.1.0 (Haas et al., 2013). In order to assess the completeness of each dataset, the presence of universal single copy orthologues was determined using BUSCO v3.0.2 and the eukaryotic BUSCO database (Simão et al., 2015).

3.4.2 Nucleoporin identification

Although nuclear pore complexes are conserved across the tree of eukaryotes, sequence divergence makes BLAST-dependent identification of nuclear pore proteins problematic (Neumann et al., 2010). To circumvent this, we used profile hidden Markov models (HMMs) to identify nuclear pore proteins based on domain structure as well as sequence composition, as has been successfully done in previous analyses (Neumann et al., 2010). Previously curated sets of nucleoporins from diverse eukaryotic taxa were obtained (Neumann et al., 2010) and realigned using the high accuracy L-INS-i algorithm of MAFFT v7.222 (Katoh and Standley, 2013). Profile HMMs were generated using these alignments and HMM searches were conducted on all transcriptomes and genomes using HMMER v3.1 and an E-value threshold of 10^{-5} (Finn et al., 2011). The best 200 hits, which included all the hits (with the exception of Aladin, Nup37, Nup43, Rae1, Sec13, Seh1, and Tpr), were then extracted and incorporated into the original alignments and realigned as before. The resultant alignments were then used to generate phylogenies in IQ-Tree v.1.5.4 using the LG+G4 substitution model and statistical support was assessed using 1000 ultrafast bootstrap pseudoreplicates (Hoang et al., 2018; Nguyen et al., 2015). Concurrently, all the proteins identified in the HMM searches were used as queries in position specific-iterative BLAST (PSI-BLAST) searches against the SWISS-PROT database (Altschul et al., 1990; Boeckmann et al., 2003). The phylogenies were then visualized in FigTree v1.4.2 and the OTUs (operational taxonomic units) were annotated with their best blast hit (E-value < 10^{-5}) to facilitate the removal of paralogous clades and the identification of true orthologues (Rambaut, 2012). The trees were cleaned over three iterative cycles that involved

removing obvious paralogues, realigning the sequences, and remaking the trees. Hits were considered true positives if they clustered with proteins from related species and/or if their best blast hit was the protein of interest. However, they were discarded if they clearly grouped with paralogues. Once the nucleoporins had been identified, long branching reference taxa were removed (typically parasitic species), and the remaining sequences were re-aligned and used to generate maximum likelihood phylogenies in IQ-Tree v.1.5.4 (Nguyen et al., 2015). Phylogenetic models were selected for each tree based on Bayesian Information Criteria using ModelFinder as implemented in IQ-Tree, and statistical support was assessed using 1000 ultrafast bootstrap pseudoreplicates (Hoang et al., 2018; Kalyaanamoorthy et al., 2017). The finalized phylogenetic trees were visually inspected in FigTree v1.4.2. To ensure the reproducibility of the search and to check that nucleoporins had not been missed, the search was conducted twice using modified thresholds and identification criteria. In both cases the same datasets were recovered. The phylogenetic affinity of each of the nucleoporins was then assessed using the finalized trees. Nucleoporins were considered host- or algal-related if they branched with outgroups or algae with an ultrafast bootstrap support over 70.

Once the NPC proteins had been identified, each was individually assessed for the presence of a signal peptide, a motif required for nucleomorph targeting (Hirakawa et al., 2009). Signal peptide predictions were performed using SignalP v.4.1 using the sensitive settings as well as SignalP v.3.0 using both the HMM and neural network based approaches (Bendtsen et al., 2004; Petersen et al., 2011). In all cases the same conclusions were drawn. Signal peptides are typically found at the N-termini of proteins and therefore, in order to facilitate the identification of false-negatives due to incomplete protein predictions, we also sought to determine whether or not each identified protein had a complete N-terminus. To this end, we used completeness predictions from TransDecoder and also manually inspected alignments to assess the presence of aligned N-terminal methionines (Haas et al., 2013). Besides false-negatives, SignalP also occasionally predicted signal peptides in proteins with incomplete N-terminal ends, suggesting that some predictions were false positives.

All the data obtained, including HMM profiles, identified nucleoporins, signal peptide prediction scores, alignments, phylogenies and their associated models are available from Dryad (https://doi.org/10.5061/dryad.b0hs8gr).

3.4.3 Pom152 analysis

In order to more closely investigate the evolutionary history of Pom152, the alignment and tree were supplemented with prokaryotic homologues. Bacterial and archaeal reference proteomes were downloaded from UniProt (downloaded November 2016) and searched using the Pom152 HMM with an E-value cutoff of 10^{-5}. The resulting hits were added to the Pom152 alignment and re-aligned using MAFFT L-INS-i (Katoh and Standley, 2013). The phylogeny was generated as above using IQ-Tree and support was inferred from 1000 ultrafast bootstrap pseudoreplicates.

3.5 Chapter-specific acknowledgements

I thank Filip Husnik and Alexandros Pittis for bioinformatic assistance, and Vittorio Boscaro for comments on the manuscript. This work was supported by a Natural Sciences and Engineering Research Council (NSERC) Discovery grant awarded to Patrick Keeling. (grant number 227301). Nicholas Irwin was supported by an NSERC Canadian Graduate Scholarship.

4. The function and evolution of motile DNA replication systems in ciliates

4.1 Introduction

DNA replication is the process by which genomic DNA is duplicated in anticipation of cell division. To ensure that the genome is replicated accurately and only once, a variety of mechanisms are employed that coordinate the assembly and initiation of replication complexes (Bleichert et al., 2017). In eukaryotes, hundreds to thousands of dispersed DNA-encoded replication origins serve as docking sites upon which the pre-replication machinery can assemble (Fragkos et al., 2015; Prioleau and MacAlpine, 2016). However, the varying accessibility of eukaryotic genomes constrains cis regulators since replication origin firing is strongly influenced by local chromatin environments (Bleichert et al., 2017; Devbhandari et al., 2017; Kurat et al., 2017; Vogelauer et al., 2002). Whereas open, transcriptionally active chromatin is permissive to replication, silenced heterochromatin restricts, or at least delays initiation (Fragkos et al., 2015; McNairn and Gilbert, 2003; Schwaiger et al., 2010). In turn, accessibility to replication origins generates temporal patterns in DNA replication as cells establish replication timing profiles in which certain sites are consistently replicated before others (Hiratani et al., 2008; Rhind and Gilbert, 2013). This has been proposed to have functional implications for the establishment of transcriptional and chromatin states following cell division (Lande-Diner et al., 2009; Rhind and Gilbert, 2013). Indeed, replication patterns have been demonstrated in yeast and animals, and replication timing profiles can be evolutionarily conserved, as is observed in mammals (Rhind and Gilbert, 2013; Ryba et al., 2010). In addition to temporal control, chromatin-dependent DNA replication results in spatial coordination, as replication origins localized to the interior of the nucleus fire earlier than those found at the nuclear periphery, a pattern which correlates with the distribution of euchromatin and heterochromatin and is independent of the position of specific chromosomes (Fox et al., 1991; Fragkos et al., 2015). However, detailed investigations of DNA replication patterns and mechanisms have only been carried out in a small number of eukaryotic model systems, despite the degree of nuclear diversity observed across the tree of eukaryotes. These diverse and underexplored systems can hold important clues to understanding nuclear

mechanisms, such as whether replication timing profiles and spatial organization are functionally significant or, instead, are mostly byproducts of chromatin structure and transcriptional activity.

One system that can provide insights into the function and mechanisms governing the regulation of DNA replication is the macronuclear replication band (RB) in spirotrich ciliates. The spirotrichs are a large and diverse group of ciliated protozoa which includes model organisms such as *Oxytricha* and *Euplotes* (Lynn, 2008). Like all ciliates, spirotrichs have both a transcriptionally quiescent germline micronucleus and a transcriptionally active somatic macronucleus, but they have also evolved a number of additional unusual nuclear characteristics. For instance, spirotrich macronuclear genomes are encoded on tens of thousands of endoreplicated gene-sized chromosomes that typically contain single genes flanked by short untranscribed regions and telomeres (Riley and Katz, 2001; Swart et al., 2013). The abundance, small size, and distribution of these chromosomes within the macronucleus likely pose challenges for coordinating DNA replication, which appear to be circumvented, at least in part, by the physical and temporal coordination of DNA replication in a replication band (RB). Spirotrich replication bands are disk-like structures that typically form at each pole of a long, ellipsoidal macronucleus during the initiation of S-phase, and then migrate towards one another generating a wave of chromatin modifications and replicated DNA before merging at the macronuclear midpoint (Figure 12) (Gall, 1959; Olins and Olins, 1994). The structure of the RB is stratified into multiple zones defined by differential chromatin states (Arikawa et al., 2000; Olins et al., 1981, 1991). The frontal zone (FZ) is characterized by abundant protein phosphorylation and chromatin packaged in 40 nm fibers, whereas in the rear zone (RZ), histones become acetylated and fiber size is reduced to 10 nm (Figure 12B) (Fauré-Fremiet et al., 1957; Olins and Olins, 1994). At the interface between the FZ and RZ, replication proteins are localized and DNA replication occurs at hundreds of foci that initially fire from the nuclear envelope before proceeding towards the nuclear interior (Figure 12B) (Fang and Cech, 1995; Olins et al., 1989; Postberg et al., 2005). Thus far, the molecular machinery organizing the RB and controlling its orderly translocation are unknown. The altered chromatin states in the FZ and RZ suggest that chromatin modifiers play a role, however, the mechanisms coordinating the motility of the band and its capacity for initiating replication are more enigmatic, perhaps involving signaling cascades (Jaffe, 1999; Postberg et al., 2005). Given the vectorial nature of DNA replication in this system, the RB represents a unique, but as of yet, largely unexplored

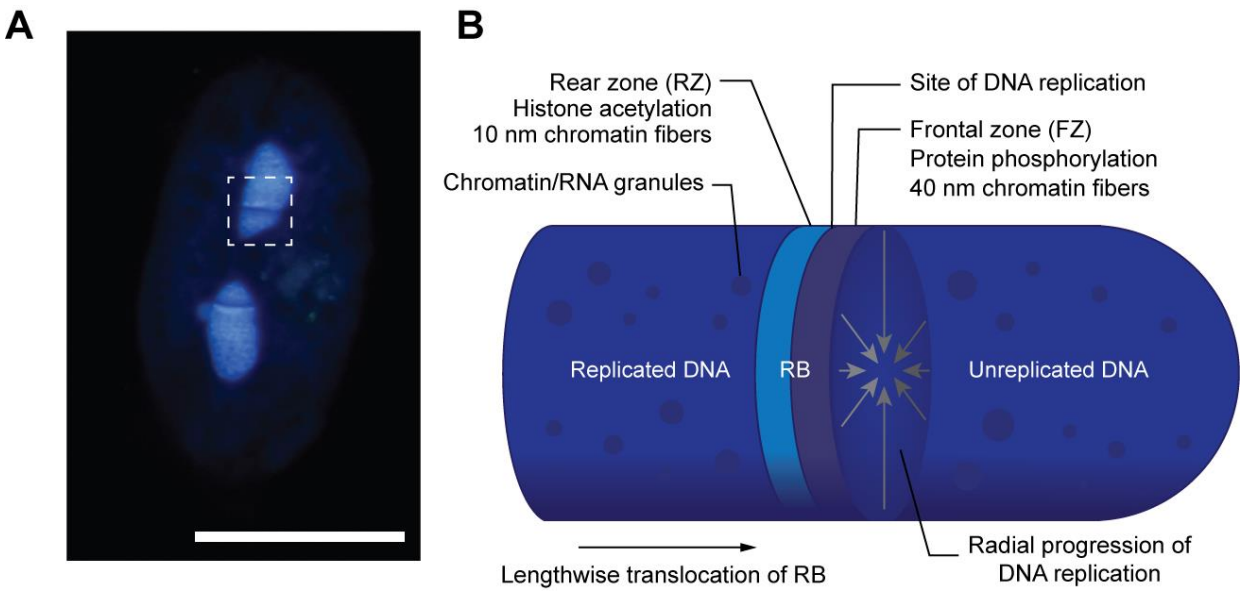


Figure 12. Representation of the replication band in *Oxytricha trifallax*.
A. A fluorescent micrograph depicting the replication band (RB) in *O. trifallax* following DAPI staining. Note the two lobes of the macronucleus and the stratified appearance of the RB. The bands will travel lengthwise along the linear axis of the elongated macronuclear lobes, before merging and then disappearing at the nuclear midpoint. Scale bar, 50 µm. **B.** A diagrammatic representation of the spirotrich replication band with key features labeled. As the RB progresses, DNA replication advances radially from the nuclear periphery towards the interior.

---

model for investigating the underlying chromatin dynamics, nuclear changes, and spatiotemporal patterns associated with DNA replication.

To better understand the function and evolution of the RB, and how it relates to spatial control over DNA replication in nuclear genomes, we aimed to reconstruct the evolutionary history of the RB and to identify processes related to RB function using phylogenomics and comparative genomics. Using transcriptomic data from two phylogenetically significant ciliate lineages related to the spirotrichs, *Licnophora macfarlandi* and *Phacodinium metchnikoffi*, we show that the RB has a paraphyletic distribution due to the phylogenetic position of *P. metchnikoffi*. Rather than possessing a typical RB, *P. metchnikoffi* uses a unique system, here termed a replication envelope, where DNA replication initiates simultaneously at the entire

nuclear periphery before progressing into the nuclear interior in a concerted wave. Based on the phylogenetic distribution of RBs, we identified genes correlated with the canonical RB, many of which were involved in intracellular transport and some of which were required for DNA replication and progression through the cell cycle in *Oxytricha trifallax*, a model spirotrich exhibiting typical RBs. These data support a model suggesting a role for calcium signaling and cytoskeletal elements in coordinating DNA replication and nuclear processes in general.

4.2 Results

4.2.1 Phylogenomics shows the paraphyly of the replication band

To reconstruct the evolutionary history of the RB, we generated transcriptomic data from two ciliates thought to be divergent relatives of the well-studied spirotrichs, *Licnophora macfarlandi* and *Phacodinium metchnikoffi*. *Licnophora macfarlandi* is a sessile marine ciliate that is known to exhibit RBs and inhabits the respiratory tree of the California sea cucumber (*Parastichopus californicus*) (Figure 13A-D) (Balamuth, 1941; da Silva-Neto et al., 2012). In contrast, *P. metchnikoffi* is a free-living ciliate inhabiting fresh water pools that appears to lack an RB (Figure 13E-S) (Fernández-Galiano and Calvo, 1992). Current phylogenetic placement of *P. metchnikoffi* is ambiguous, with different molecular phylogenies placing *P. metchnikoffi* either within or sister to the spirotrichs, indicative of a paraphyletic or monophyletic RB, respectively (Boscaro et al., 2018b; Gao et al., 2016; Lynn and Strüder-Kypke, 2002; Shin et al., 2000). In order to reconcile these topologies, we sought to construct a phylogenomic dataset incorporating *L. macfarlandi* and *P. metchnikoffi*. We combined transcriptomic data from *L. macfarlandi* and *P. metchnikoffi* with 52 previously published genomic and transcriptomic datasets from other ciliates (n = 40) and outgroup taxa (n = 12), to construct a matrix encompassing 231 concatenated genes and 73,533 amino acid sites (Table C1). Maximum likelihood analyses of the concatenated alignment produced a robustly supported phylogeny that was consistent when using either free rate heterogeneity or empirical profile mixture substitution models (Figure 14A). Similarly, Bayesian analyses recovered a nearly identical phylogeny with all four chains converging on similar, equally well supported topologies (chain bipartition discrepancies: max diff. = 0.175857, mean diff. = 8.37×10^{-4}) (Figure 14A).

The resulting trees fully resolved the deep phylogeny of the ciliate phylum and reconciled

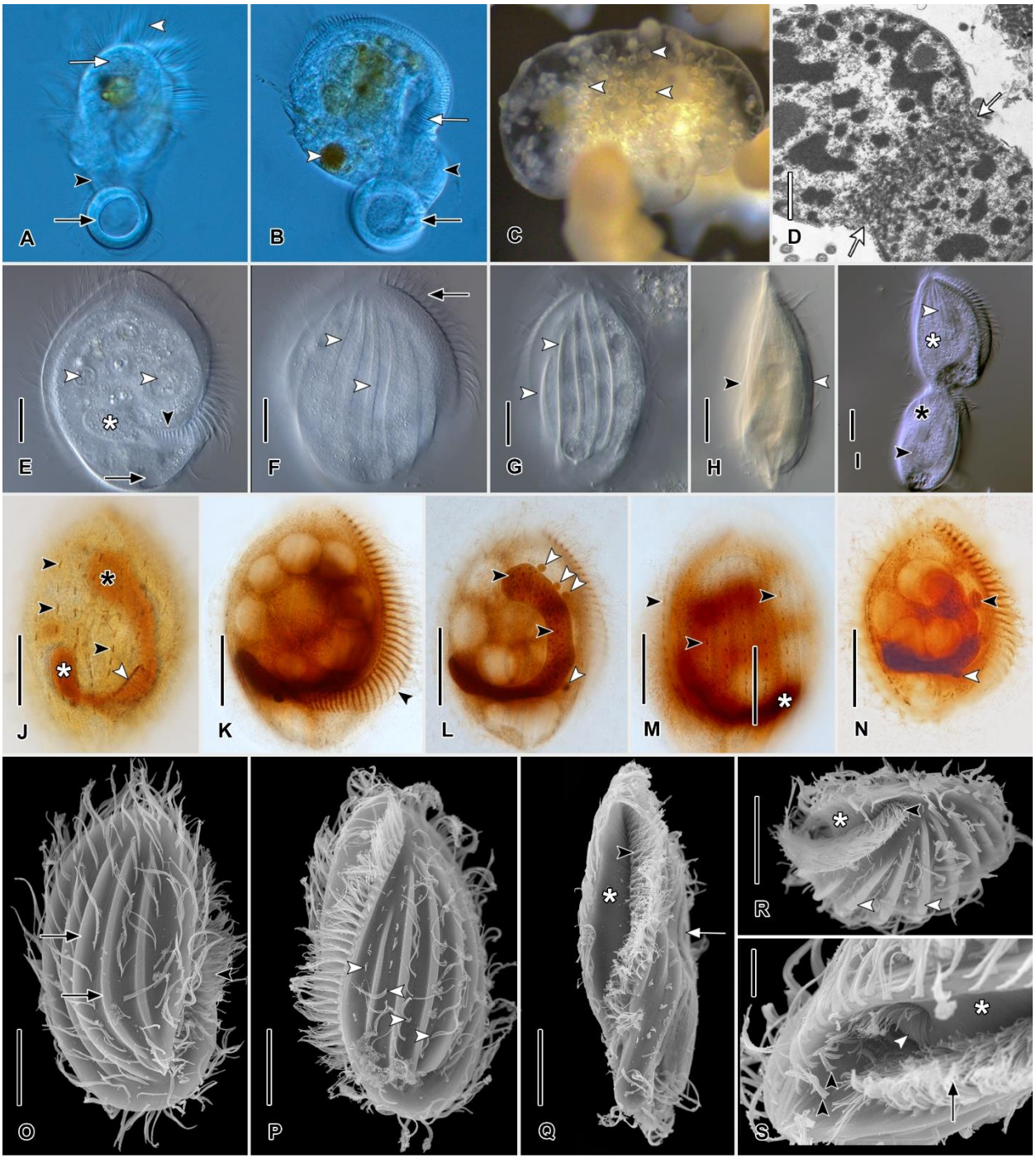


Figure 13. Morphology and identification of *Licnophora* spp. and *Phacodinium metchnikoffi*.

Micrographs were obtained *in vivo* using differential interference contrast illumination (**A-C, E-I**), modified from Silva-Neto et al. (2012) with permission (**D**), after protargol impregnation (**J-N**), or using a scanning electron microscope (**O-S**). **A**. *Licnophora* sp. (likely *L. macfarlandi*)

from the respiratory tree of the California sea cucumber (*Parastichopus californicus*) *in vivo* showing the distinctive hourglass shape with oral (white arrow) and adhesive (black arrow) discs united by the neck region (black arrowhead). Cilia of the adoral membranelles (white arrowhead) wreathe the oral disc. **B.** *Licnophora* sp. from same host as (A) showing the adoral zone of membranelles around the periphery of the oral disc (white arrow), the neck region (black arrowhead), the adhesive disc (black arrow), and a food vacuole (white arrowhead). **C.** Low magnification view of *Licnophora* sp. cells (white arrowheads) in a respiratory tree ampulla of the sea cucumber. **D.** Transmission electron micrograph of *Licnophora chattoni* macronuclear nodule showing a replication band (white arrows). **E.** *Phacodinium metchnikoffi.* Optical section showing food vacuoles (white arrowheads), the posterior lobe of the macronucleus (asterisk), the contractile vacuole (black arrow) and the posterior end of the adoral zone of membranelles (black arrowhead). **F.** Ventral side of cell showing cortical ridges (white arrowheads). **G.** Dorsal side cortical ridges (white arrowheads). **H.** Lateral view, optical section showing dorsoventral flattening (cf. Figure 13Q) with the relatively flat ventral side (black arrowhead) and the convex dorsal side (white arrowhead). **I.** Late divider showing anterior (white asterisk) and posterior (black asterisk) daughter cells (proter and opisthe, respectively). The anterior lobe of the macronucleus (white arrowhead) of the proter and the posterior lobe of the macronucleus in the opisthe (black arrowhead) are visible. **J.** Ventral view showing the anterior (black asterisk) and posterior (white asterisk) lobes of the macronucleus, somatic kineties composed of linear polykinetids or "palettes" (black arrowheads), and part of the paroral membrane (white arrowhead). **K.** Ventral view showing adoral zone of membranelles (black arrowhead). **L.** Ventral view of interphase cell showing the macronucleus with a myriad of small nucleoli (black arrowheads) and several micronuclei (white arrowheads). **M.** Dorsal view showing somatic kineties (black arrowheads). **N.** Optical section showing typical micronucleus (white arrowhead) and a micronucleus in anaphase (black arrowhead). Details of the macronucleus are obscured. **O.** Left ventrolateral view showing cortical ridges (black arrows) and the cilia of the adoral membranelles (black arrowhead). **P.** Dorsal view showing somatic kineties (white arrowheads) composed of discontinuous linear polykinetids ("palettes") between the cortical ridges. **Q.** Left lateral view showing the roof of the buccal cavity (asterisk), adoral membranelles (black arrowhead), and the dorsal surface (white arrow). **R.** Anterior apical view showing the anterior end of the adoral zone of membranelles (black arrowhead), the roof of the buccal cavity

(asterisk), and the dorsal cortical ridges (white arrowheads). **S.** Detail view of the posterior portion of the buccal cavity showing short linear somatic polykinetids (black arrowheads), the roof of the buccal cavity (asterisk), the adoral zone of membranelles (black arrow), and the cilia of the paroral membrane (white arrowhead). Scale bars: 25 μm (E–R), 10 μm (S), 1 μm (D). No scale was available for (A–C). The size of a typical individual of *L. macfarlandi* is approximately 100 μm × 55 μm (Balamuth, 1941). Nonlinear adjustments applied to (E–N). Background subtracted from (O–S).

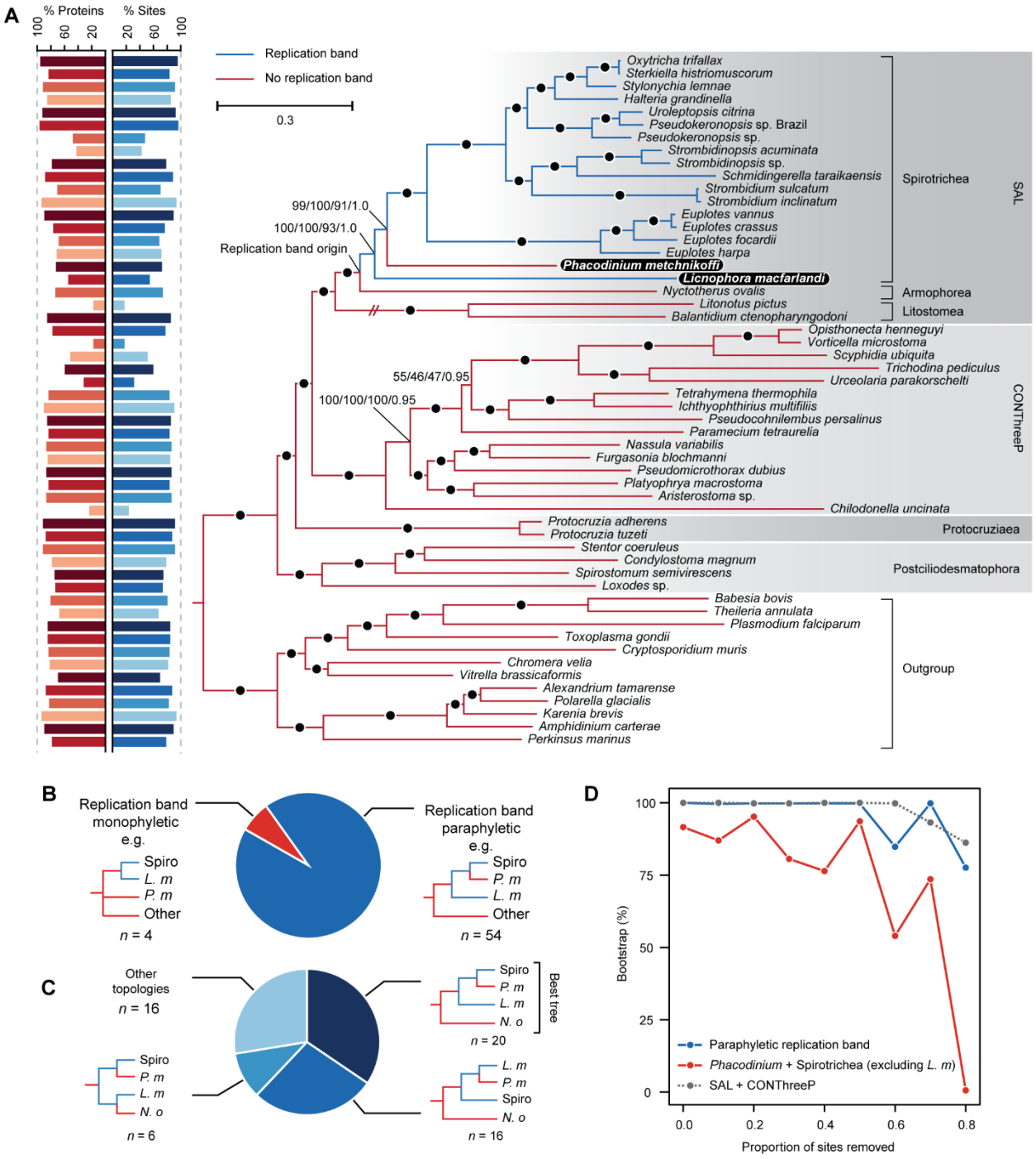


Figure 14. Phylogenomics resolve the paraphyly of the replication band and the deep phylogeny of the ciliates.

A. Maximum likelihood phylogeny based on 231 concatenated proteins comprising 73,533 amino acid sites generated using the LG+C40+F+G4 substitution model as implemented in IQ-Tree. Statistical support is displayed at the nodes and was assessed using 1,000 ultrafast

bootstraps with the LG+C40+F+G4 model, 1,000 ultrafast and non-parametric bootstraps using the LG+F+R8 model, and posterior probabilities determined using the GTR+CAT+G4 model in PhyloBayes. Black circles represent full statistical support. The percent of proteins and sites present for each taxon are shown on the left and the presence and absence of the replication band is denoted with blue and red branches, respectively. The length of the hashed branch leading to the Litostomea was reduced by half for simplicity. *Phacodinium metchnikoffi* and *Licnophora macfarlandi* have been bolded and highlighted. **B, C.** The proportion of phylogenomic jackknives (12 gene subsamples, $n = 58$) that recovered a paraphyletic replication band (**B**) and individual tree topologies (**C**). **D.** Non-parametric bootstrap support ($n = 500$) for given topologies following the removal of the fastest evolving sites as inferred using the LG+F+R8 model in IQ-Tree. Spiro, Spirotrichia; *P. m*, *Phacodinium metchnikoffi*; *L. m*, *Licnophora macfarlandi*; *N. o*, *Nyctotherus ovalis*.

---

the placement of *L. macfarlandi* and *P. metchnikoffi*. All analyses separated the ciliates into four well-supported and previously defined clades: SAL (Spirotrichea, Armophorea, Litostomatea), CONThreeP (Colpodea, Oligohymenophorea, Nassophorea, Plagiopylea, Prostomatea, and Phyllopharyngea), Protocruziaea, and the subphylum, Postciliodesmatophora (Figure 14A) (Gentekaki et al., 2014). Inclusion of the karyorelictid ciliate, *Loxodes* sp., reaffirmed the independent branching of *Protocruzia*, as previous investigations had either placed it sister to the hypotrichs, albeit with poor support, or independently branching in the absence of karyorelictid representation (Chen et al., 2015; Gao and Katz, 2014; Gentekaki et al., 2014). Furthermore, despite previous ambiguity in the topology of SAL (Gao and Katz, 2014; Gao et al., 2016; Gentekaki et al., 2014), our analyses recovered the Spirotrichea being sister to Armophorea, with Litostomatea branching at the base of the clade with full statistical support (Figure 14A), consistent with recent work (Rotterová et al., 2020). Within SAL, we found that *L. mafarlandi* and *P. metchnikoffi* both branched with spirotrichs, where they formed the two deepest branches: *L. macfarlandi* branching as sister to all other spirotrichs and *P. metchnikoffi* branching next as sister to all remaining spirotrichs (Figure 14A).

Based on this topology, the RB would be paraphyletic due to its probable absence in *P. metchnikoffi* (Figure 14A). To assess whether this topology could be a result of gene sampling

bias, we sub-sampled the 231 gene set into 5% (12 gene) concatenations by jackknifing (i.e., random sampling without replacement; $n = 58$) and used the resulting alignments for phylogenetic analysis. The majority of the resulting phylogenies showed a paraphyletic RB (93%) and many recovered the best tree topology (34%), suggesting that these phylogenies are not strongly influenced by gene selection (Figure 14B, C). Similarly, the sequential removal of fast-evolving sites did not influence RB paraphyly or the sister relationship between SAL and CONThreeP, suggesting that these topologies are not biased by long branch attraction (Figure 14D). However, the exact relationship between *P. metchnikoffi* and spirotrichs aside from *L. macfarlandi* was dependent on gene sampling and site removal, making this placement tentative (Figure 14B-D), but not in any way consistent with a monophyletic RB.

4.2.2 Characterization of the replication envelope in Phacodinium metchnikoffi

Phylogenomic analyses indicated that the RB had either been lost in *P. metchnikoffi* or that previous studies had failed to observe it. To discern between these possibilities, we investigated nuclear morphology in *P. metchnikoffi* using a variety of cytological techniques used to observe the RB in other species. In agreement with previous work, staining with methyl green-rhodamine B (Figure 15A), Feulgen (Figure 15B), and protargol (Figure 15C), did not reveal an RB in *P. metchnikoffi*. We next sought to visualize replication directly using EdU (5-Ethynyl-2'-deoxyuridine) incorporation followed by Cy3-azide labeling. As a positive control, Cy3-EdU labeling and Hoechst staining was applied to the spirotrich, *Euplotes* sp., which revealed a fluorescent pattern characteristic of the RB, wherein the nuclear regions through which the band had passed, exhibited Cy3 fluorescence consistent with replicated DNA (Figure 15D, E). Cy3-EdU labeling in *P. metchnikoffi* also distinguished replicated DNA, but showed an alternative pattern where fluorescence appeared initially throughout the nuclear periphery, before progressing inwards towards the nuclear interior in a single concerted wave (Figure 15F-I, Figure 16A-C, Figure C1). A re-evaluation of the protargol staining preparations revealed cells with a comparable nuclear structure, suggesting that these observations are independent of staining procedures (Figure 15C). Unlike the RB, the replication envelope lacked defined or stratified structures (i. e., the FZ and RZ were not observed). Similarly, replication appeared to initiate globally at the nuclear envelope contrasting with the focality observed with the RB. To distinguish this from a canonical RB, we term this a replication envelope.

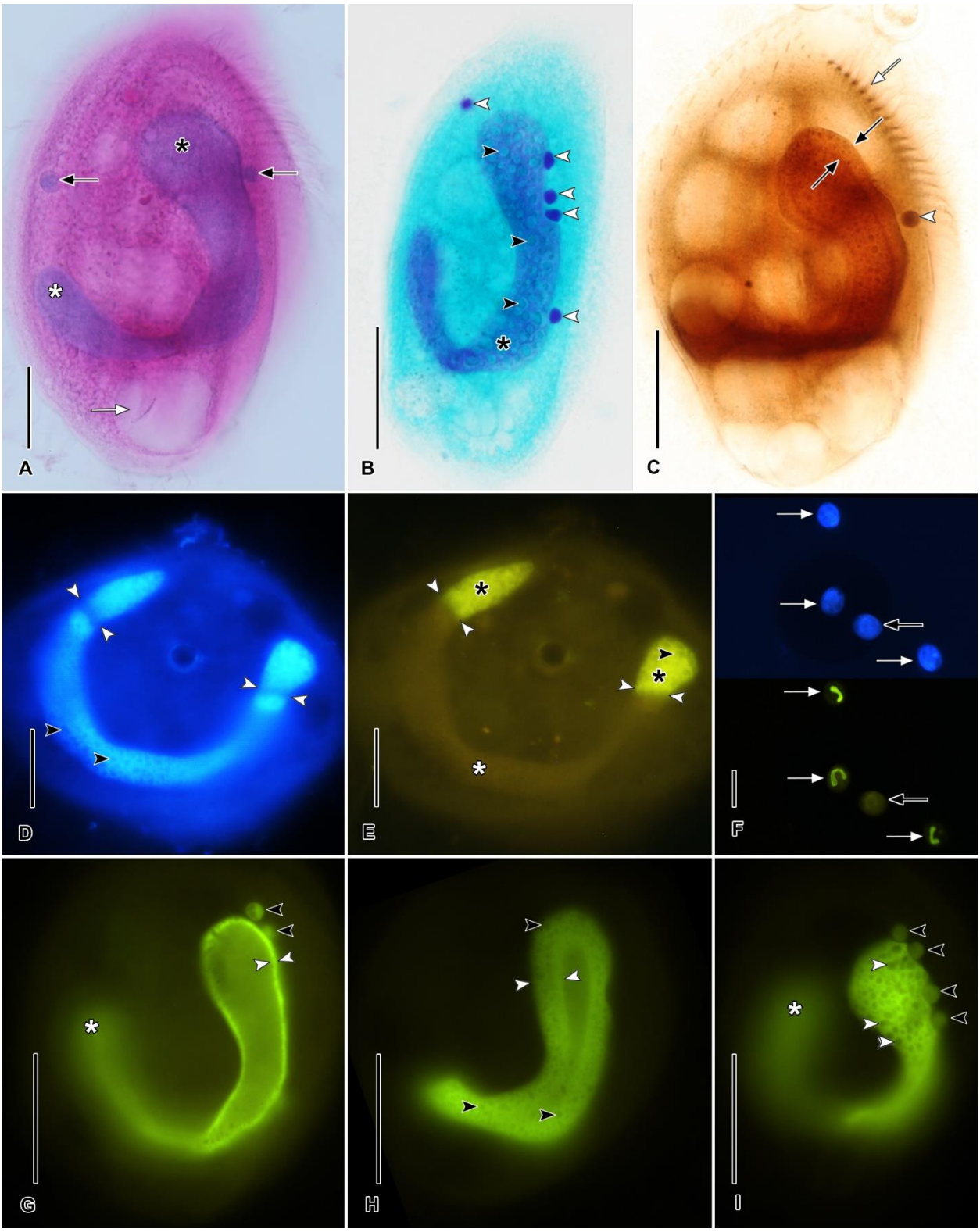


Figure 15. *Phacodinium metchnikoffi* exhibits a unique DNA replication morphology.

A. Methyl green-rhodamine B stain of *P. metchnikoffi* showing the anterior and posterior lobes

of the macronucleus (black and white asterisks, respectively), several micronuclei (black arrows) and the contractile vacuole (white arrow). **B.** *Phacodinium metchnikoffi* after Feulgen stain for DNA showing the macronucleus (asterisk) with many small nucleoli (black arrowheads) and multiple micronuclei (white arrowheads). **C.** *Phacodinium metchnikoffi* (ventral view, probable early S-phase, cf. Figure 15H) showing a putative replication envelope (black arrows), one micronucleus (white arrowhead), and the adoral zone of ciliary membranelles (white arrow). **D.** *Euplotes* sp. in early S-phase stained with Hoechst 33342 showing replication bands (white arrowheads) at each end of the macronucleus. Nucleoli (black arrowheads) appear as small lucencies. **E.** Same cell as Figure 15D after a 13 hr pulse with EdU and Cy3 azide labeling. Replication bands (white arrowheads) advance in front of regions with EdU-labeled replicated DNA (black asterisks). The as -yet unreplicated DNA is unlabeled (white asterisk). **F.** Low magnification view of *P. metchnikoffi* cells labeled with Hoechst 33342 (blue, top) and after an 18 hr pulse with EdU (green, bottom) showing three cells in various stages of S-phase (white arrows) and one non S-phase cell (black arrow). **G.** *Phacodinium metchnikoffi* macronucleus (larger anterior lobe) in early S-phase showing the very thin peripheral envelope of replicating DNA (white arrowheads) surrounding the relatively lucent central volume of unreplicated DNA. The posterior lobe (asterisk) is out of the focal plane. Micronuclei (black arrowheads) are also beginning to replicate DNA. **H.** Progression of the replication envelope (white arrowheads) toward the center of the macronucleus as the volume of unreplicated DNA diminishes. Note that nucleoli (black arrowheads) are visible only within regions of replicated DNA. **I.** Completion of macronuclear DNA replication, the replication envelope has disappeared and nucleoli (white arrowheads) are now distributed throughout the nucleoplasm. Micronuclei (black arrowheads) appear to be completing S-phase. Scale bars: 100 µm (**F**), 25 µm (**A–C**, **G–I**), 10 µm (**D**, **E**). Figure 15**G–I** show three cells from the same slide, all at 300 ms exposure time.

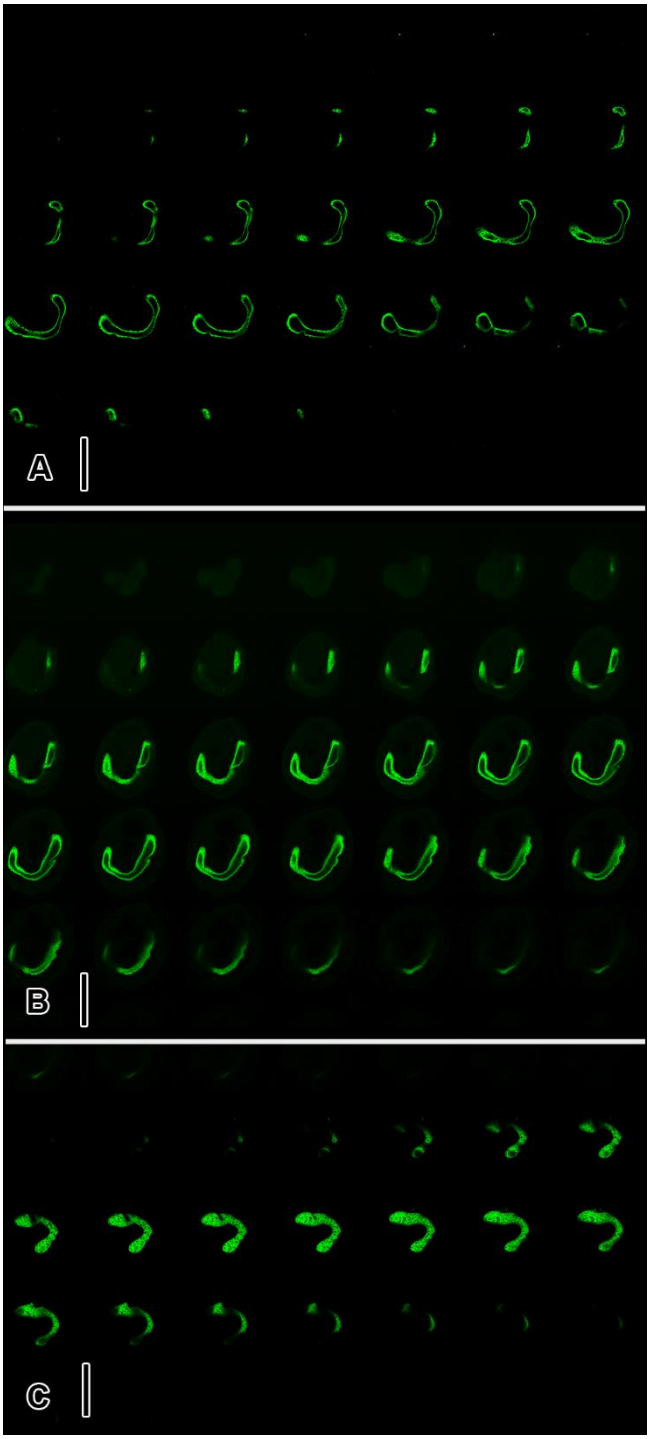


Figure 16. Confocal fluorescence imaging of the replication envelope in *Phacodinium metchnikoffi*.

Confocal imaging montages of three separate cells (**A–C**). **A.** Early S-phase. **B.** Slightly later S-phase than (**A**). **C.** Completed S-phase. Scale bars: 50 µm (**A–C**). Step size: 0.5 µm.

4.2.3 Transport and cytoskeletal-related genes are associated with the function of the replication band and DNA replication

Given the divergence between the canonical spirotrich RB and the replication envelope of *P. metchnikoffi*, we next sought to leverage RB paraphyly in order to identify genes that could be associated with its structure and function. In particular, we hypothesized that the differential presence/absence and expansion/duplication rates of protein families between related lineages with and without the RB may suggest functional relatedness. However, it is important to note that this analysis would only identify RB-associated genes as opposed to motile replication system genes in general. To this end, we clustered proteins from diverse ciliate genomic and transcriptomic datasets into protein families (orthologous groups) and compared the frequency of proteins within each family (Table C1). Of the 37,318 protein families examined, 406 and 23 were found to be enriched or depleted in RB-containing clades, respectively ($P < 0.05$, permutation test, $n = 10{,}000$) (Figure 17A, Table C2). Of those, 186 enriched families were found in both *L. macfarlandi* and the model spirotrich, *Oxytricha trifallax*, which had on average 1.8 and 2.7 times more homologues per family compared to *P. metchnikoffi*. Consistent with this, principal component analysis based on differentially present families produced discernible RB-containing and RB-lacking species clusters (Figure 17B), and the RB-association analysis produced distinct results compared to a spirotrich-association analysis (Figure C2).

Functional annotation and gene ontology (GO) analysis of protein families enriched in RB-containing lineages revealed an association with functions such as ion transport and regulation of the cytoskeleton (Figure 17C, Table C3). For example, biological processes such as cellular ion homeostasis (GO:0006873, $P < 1x10^{-6}$), ion transport (GO:0006811, $P = 1x10^{-3.19}$), and regulation of transport (GO:0051049, $P = 1x10^{-3.10}$) were significantly enriched owing to the prevalence of calcium transporters (e.g., polycystin-2, sodium/calcium exchangers, and calcium channels) (Figure 17C, D, Figure C3A). Moreover, genes involved in cytoskeletal organization (e.g., GO:0007010, $P = 1x10^{-2.10}$) including nuclear distribution protein homologues, which regulate nuclear microtubules (Vergnolle and Taylor, 2007), and gelsolin repeat proteins, which are calcium dependent actin regulators (Sun et al., 1999), were also correlated with the RB (Figure 17C, D, Figure C3B, Table C2). Other functions including responses to chemicals

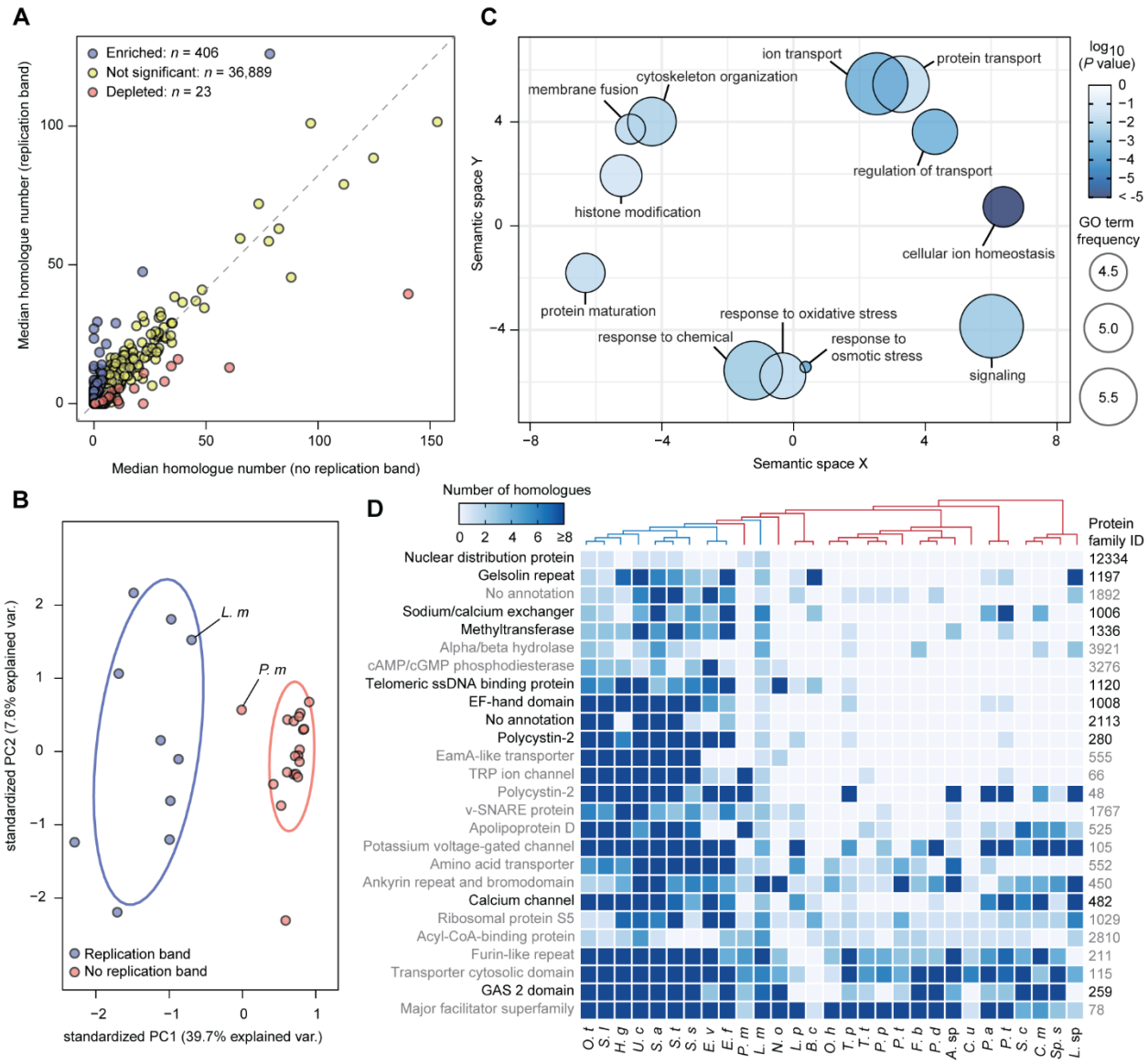


Figure 17. Comparative genomics identifies replication band-associated protein families. **A.** Scatter plot depicting the median number of homologues per protein family for clades containing and lacking the replication band (RB). Significantly enriched and depleted families are denoted in blue and red, respectively ($P < 0.05$, permutation test, $n = 10{,}000$). **B.** Principal component analysis based on the number of homologues in enriched and depleted protein families for taxa included in the analysis. *P.m*, *Phacodinium metchnikoffi*; *L. m*, *Licnophora macfarlandi*. **C.** Semantic similarity-based scatter plot generated using REVIGO that displays significant gene ontology (GO) terms for replication band-associated protein families following summarization. Point colour and size are dependent on the *P*-value (permutation test, *n* =

100,000) and the frequency of the GO-term within UniProt. **D.** Heatmap displaying the frequency of the 26 most highly enriched RB-associated protein f0amilies ($P < 1x10^{-5}$, permutation test, $n = 10,000$) across examined taxa that had representation from both *L. macfarlandi* and *Oxytricha trifallax*. The phylogeny at the top is based on Figure 14 and abbreviated species names are found at the bottom. Protein identifiers and functional annotations are listed on the right and left, respectively. Protein families highlighted in bold were further examined with RNA interference (see Figure 18).

(GO:0042221, $P = 1x10^{-2.39}$) and histone modifications (GO:0016570, $P = 1x10^{-1.30}$) were also enriched due to the presence of proteins containing, for example, cyclic-AMP/GMP phosphodiesterase domains and acetylated lysine binding bromodomains, respectively (Figure 17C, D). In agreement with the biological process terms, molecular functions including transmembrane transporter activity (GO:0022857, $P = 1x10^{-3.60}$), enzyme binding (GO:0019899, $P = 1x10^{-2.62}$), cytoskeletal protein binding (GO:0008092, $P =1x10^{-1.47}$), and phosphatase activity (GO:0016791, $P = 1x10^{-1.49}$), were also connected to RB-related proteins (Table C3).

Given the correlative nature of the preceding analysis, the list of RB-associated protein families naturally contains some families that are enriched in RB-containing species but not functionally related to the replication band. To investigate which of these RB-associated proteins are required for RB function, we used RNA interference (RNAi)-induced gene knockdowns and assessed cell cycle states as well as RB progression and morphology in the model spirotrich, *Oxytricha trifallax*. We selected ten of the most highly enriched protein families ($P < 1x10^{-5}$) (Figure 17D, Table C2) with functions related to significant gene ontology terms, and generated RNAi constructs capable of knocking down the expression of all RB-associated paralogues for each group (for a total of 64 proteins knocked down). Compared to the non-specific RNAi, knockdown of a positive control, the essential S-phase initiating protein, CDC6, significantly decreased the proportion of cells with RBs (Figure 18A), and when bands were observed, CDC6-RNAi resulted in impaired RB progression (measured as the distance of the RB from its origin divided by the length of the nuclear lobe) and an altered RB morphology where the FZ and RZ were difficult to discern (Figure 18B,C). This confirmed that at least serious cell cycle and DNA replication defects can be detected by the experimental system. Knockdown of protein families

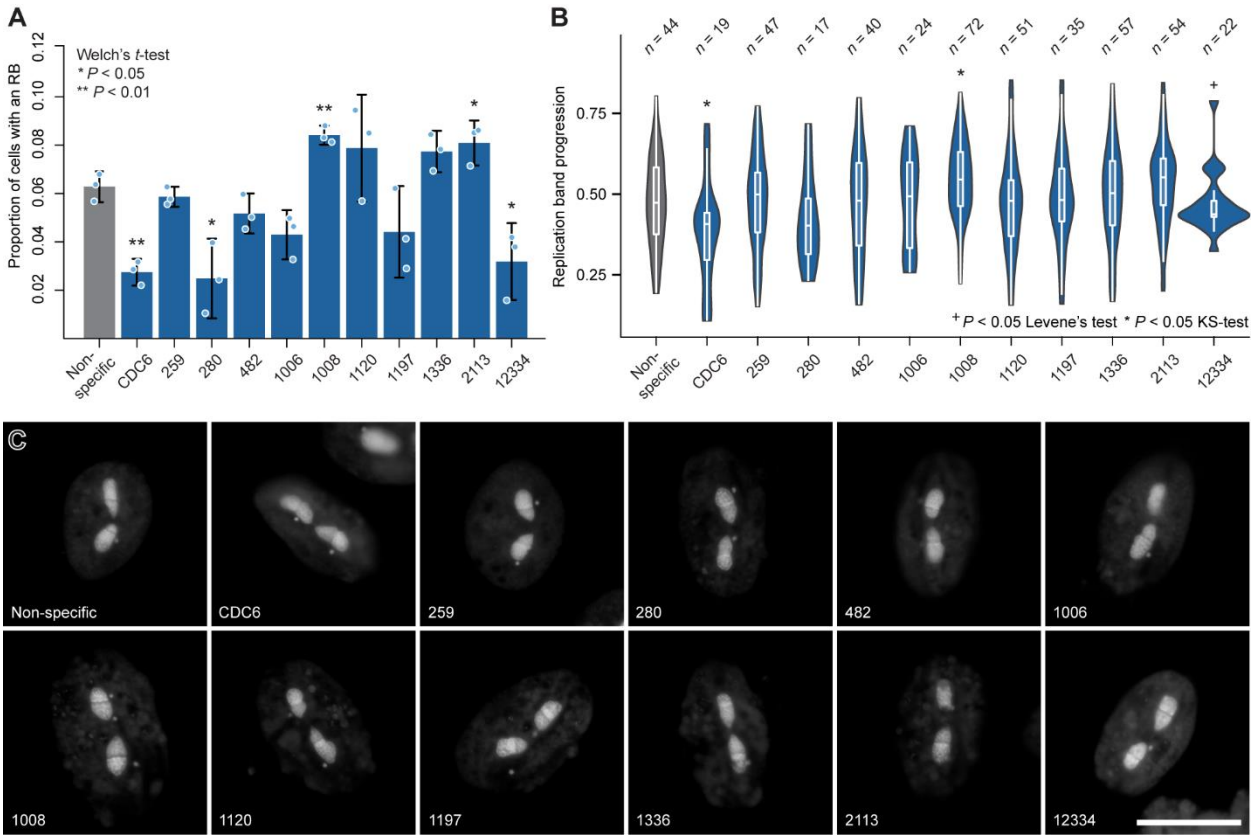


Figure 18. Replication band-associated protein families are required for DNA replication and cell cycle progression.

A. The proportion of *Oxytricha trifallax* cells with replication bands (RB) following RNAi knockdown of specified protein families (n = 3 biological replicates). For each replicate, an average of 238.7 cells were observed (151-325, standard deviation = 37.4). Error bars represent 95% confidence intervals and significance was calculated using Welch's *t*-tests. **B.** Replication band progression in cells following RNAi knockdown of specified protein families. Progression was calculated as the distance of the RB from the nuclear pole divided by the total length of the nuclear lobe. For each RB containing cell, the progression within a single lobe was measured. Significance was assessed using Levene's test and Kolmogorov-Smirnov (KS) tests for variance and distribution equality, respectively. **C.** Fluorescent micrographs taken following DAPI staining showing representative RB morphology in cells treated with RNAi. Scale bar, 50 µm. The targeted protein families are labeled on the bottom of each plot along with the non-specific control RNAi to which samples were compared. The corresponding protein functions are listed in Figure 17D.

identified as correlating with RBs confirmed that several are required for normal DNA replication and cell cycle progression: protein families 280 (polycystin calcium transporter), 1006 (sodium/calcium exchanger), and 12334 (nuclear distribution protein) all demonstrated a decrease in the proportion of cells exhibiting RBs (Figure 18A). Knockdown of 280 also resulted in an altered RB morphology comparable to knockdown of CDC6, and the depletion of 12334 appeared to impair band progression (Figure 18B, C). Knockdown of families 1008 (EF-hand calcium binding protein) and 2113 (no annotation) increased the proportion of cells exhibiting RBs (Figure 18A). Depletion of 1008 also coincided with altered band progression and morphological discrepancies; specifically, the FZ was often thicker and more clearly defined than the wild-type, whereas knockdown of 2113 had no apparent effect on band movement or appearance (Figure 18B, C). Depletion of protein families 259, 482, 1120, 1197, and 1336 did not result in significant phenotypic changes, which might indicate they are not involved in the RB, but could equally reflect technical limitations of the RNAi screen. Overall, these results confirm the prediction that RB-associated proteins, particularly those comprising families 280, 1008, 2113, and 12334, are important for DNA replication, RB structure and function, or both.

4.3 Discussion

Despite the central importance of DNA replication, nuclear genomes exhibit a variety of mechanisms that regulate how the genome is duplicated. However, regardless of the mechanism, the function of the system is conserved, and divergent cases can offer unique perspectives into how this genome duplication is achieved. In the case of spirotrich ciliates, the fragmentation of the genome into hundreds of thousands of small, single-gene chromosomes alleviates the common problem of quickly replicating large chromosomes (which generally require multiple replication origins), but elevates the challenge of 'bookkeeping'. The spirotrich macronucleus contains so many chromosomes that ensuring they are all replicated once and only once likely requires a novel degree of control over the spatial organization of DNA replication, manifested in the replication band.

The *P. metchnikoffi* replication envelope shows that motile DNA replication systems are evolutionarily malleable, but it probably evolved from a more canonical replication band. This is likely as phylogenomic analyses indicate that the last common ancestor of all spirotrichs,

including *P. metchnikoffi*, contained a replication band. Moreover, both motile DNA replication systems adhere to the core function of moving the DNA replication zone through the nucleus in a spatially and temporally coordinated fashion so that the replication status of any chromosome is evident simply by its physical location in the nucleus. At a finer scale, other similarities between the replication band and envelope are evident. In the replication band, the DNA replication hub progresses lengthwise along the axis of the macronucleus as replication proceeds radially from the nuclear envelope towards the interior (Figure 1) (Postberg et al., 2005). Accordingly, the replication envelope may reflect a loss of this lengthwise movement along the axis of the macronucleus. Given the relative simplicity of the replication envelope, it is also tempting to speculate that this system could reflect a reversion to the ancestral state of the RB. However, Cy3-EdU labeling in spirotrich relatives, including the armophoreans and litostomateans, which similarly lack RBs but contain nanochromosomes, has been inconclusive, limiting ancestral character reconstruction.

Regardless, the identification of the replication envelope in *P. metchnikoffi* provides an opportunity to exploit natural diversity that may facilitate the mechanistic dissection of motile DNA replication systems. Based on this, we identified a number of protein families that were significantly enriched in RB-containing species and hypothesize that many of these proteins may be involved in the lengthwise progression of the RB. RNAi knockdowns corroborated this for many proteins by demonstrating their requirement for normal S-phase progression, RB movement, or morphology. The over-representation within these protein families of calcium transporter paralogues and cytoskeletal regulators such as calcium-influenced gelsolin repeat proteins and nuclear distribution protein homologues allude to an overall mechanism whereby calcium signaling waves trigger the translocation of the band along a nuclear cytoskeleton. A role for calcium waves in band progression has indeed been hypothesized previously (Jaffe, 1999), and the spirotrich nucleus has also previously been shown to possess calcium-dependent contractility (Arikawa et al., 2003), and RB sensitivity to cAMP-phosphodiesterase and calmodulin inhibitors (Olins and Olins, 1993). Additionally, transmission electron microscopy studies have identified the presence of 10 nm non-chromatin filaments that run through the FZ in *Euplotes eurystomus* (Olins and Olins, 1990). A comparable mechanism was recently described in human fibroblasts, where G-protein-coupled receptor signaling and calcium release triggers linear F-actin polymerization radially from the inner nuclear membrane towards the nuclear

interior (Wang et al., 2019). Furthermore, nuclear actin dynamics have been implicated in DNA replication control and the regulation of other nuclear processes, partly through enzyme relocalization and clustering (Hurst et al., 2019; Parisis et al., 2017; Percipalle and Visa, 2006; Wei et al., 2020). This not only argues for the feasibility of calcium-dependent motility driving the spatial and temporal coordination of a DNA replication system but also suggests that the replication band could be a derivation of a broadly conserved eukaryotic nuclear trait. Congruent with this, the protein family enrichment analysis indicated that RB innovation is largely a product of gene duplication.

Ultimately, the spirotrich ciliates provide a unique case from which functional, regulatory, and evolutionary insights into the nature of DNA replication systems can be derived. Although motile DNA replication systems are currently perceived to be a ciliate-specific trait, the flexibility of these systems, as observed in *P. metchnikoffi*, and the mechanistic similarities to processes observed in human cells (Wang et al., 2019), suggest that related, perhaps morphologically subtler, systems may exist in other eukaryotes. This work also highlights the potential for the replication band to serve as a model to not only dissect the consequences of replication timing, but the mechanisms underpinning DNA replication and the nuances defining nuclear organization and regulation.

4.4 Materials and Methods

4.4.1 Phacodinium metchnikoffi

Phacodinium metchnikoffi was isolated from soil percolates from a grass lawn at Julia Davis Municipal Park in Boise, Idaho, USA (43°36'22.7"N 116°11'43.2"W, elev. 816 m) on 22 November 2017. Enriched non-clonal cultures were established in a medium composed of 10% v/v soil extract (1% w/v sterile commercial garden soil boiled in distilled water for 10 minutes, filtered with Grade 102 filter paper after cooling, and then autoclaved and buffered through the addition of 0.5 g K_2HPO_4/L [2.9 mM]) in 0.22 µm filtered site water with a rice grain added to support growth of bacteria as food for the ciliates. Cultures were maintained at room temperature (about 19.4 °C) out of direct sunlight and sub-culturing was performed once weekly. The population was verified as *P. metchnikoffi* according to morphological characteristics (da Silva

Neto, 1993) and a partial 18S rRNA gene sequence that was >99 % identical to *P. metchnikoffi* (Genbank accession number AJ277877) (Shin et al., 2000).

4.4.2 Licnophora macfarlandi

Licnophora macfarlandi was collected from the respiratory trees of the California sea cucumber (*Parastichopus californicus*) as described previously (Lynn and Strüder-Kypke, 2002). The body wall of the sea cucumber was opened and the respiratory trees were transferred to a petri dish and inspected for the presence of *L. macfarlandi*. Cells were flushed from the host tissue using filtered sea water and isolated using a glass micropipette. The identity of *L. macfarlandi* was confirmed based on morphology, the host species, and a partial 18S rRNA gene which was >99% identical to *L. macfarlandi* (Genbank accession number AF527758) (Lynn and Strüder-Kypke, 2002).

4.4.3 Oxytricha trifallax

Oxytricha trifallax (strain JRB310) was cultured in Pringsheim media (0.11 mM Na_2HPO_4, 0.08 mM $MgSO_4$, 0.85 mM $Ca(NO_3)_2$, 0.35 mM KCl, pH 7.0). at 20 ºC on a 12:12 hour light:dark cycle in disposable Petri dishes (Bracht et al., 2012; Pringsheim, 1964). Cultures were fed daily with variable amounts of *Chlamydomonas reinhardtii* (typically around 0.05 OD_{600}/mL) that had been washed with Pringsheim media. Every second day, *Klebsiella* sp., that had been cultivated overnight, was added to the cultures (1:1,000 of the culture volume) and the media was transferred to a fresh dish. *Chlamydomonas reinhardtii* (strain CC-125) was cultured in rich media (12 mM $NaCH_3COOH$, 0.1% (w/t) beef extract, 0.2% tryptone, 0.2% yeast extract, and 0.07 mM $CaCl_2 \cdot 2H_2O$), sub-cultured every second day, and fed to *O. trifallax* three days post-inoculation.

4.4.4 RNA-sequencing and transcriptome assembly

Phacodinium metchnikoffi and *L. macfarlandi* cells were isolated manually using drawn glass micropipettes, washed three successive times in 0.22 µm-filtered water, and placed in 2 µL of cell lysis buffer (0.2% Triton X-100 and RNAse inhibitor) (Picelli et al., 2014). RNA was extracted and cDNA was synthesized from pools of 10 cells according to the Smart-Seq2

protocol (Picelli et al., 2014). cDNA quantification was achieved using a Qubit 2.0 fluorometer and library assembly was conducted using a Nextera XT DNA library preparation kit. Libraries generated from *P. metchnikoffi* and *L. macfarlandi* cDNA were sequenced on an Illumina MiSeq using either 250 or 300 bp paired-end reads.

Raw paired-end Illumina reads were merged using PEAR v0.9.6 and read quality was assessed with FastQC (Andrews, 2010; Zhang et al., 2014). Adapters and low quality regions were removed using Trimmomatic v0.36 prior to *de novo* transcriptome assembly using Trinity v2.4.0 (Bolger et al., 2014; Grabherr et al., 2011). Bacterial contamination was identified and removed by comparing transcriptome contigs to UniProt reference proteomes using Diamond BLASTx (e-value < $1x10^{-25}$) (Buchfink et al., 2014; Uniprot Consortium, 2015). Proteins were then predicted with TransDecoder v5.1.0 following BLASTp searches against the SWISS-PROT database (Altschul et al., 1990; Boeckmann et al., 2003; Haas et al., 2013). The genetic codes selected for translation were based on information provided in the NCBI (National Center for Biotechnology Information) Taxonomy database (https://www.ncbi.nlm.nih.gov/Taxonomy).

4.4.5 Phylogenomic dataset assembly

In order to generate a phylogenomic dataset encompassing ciliate diversity, ciliate transcriptomes and genomes (Table C1) were searched for a set of 263 homologues that have been used previously to reconstruct the phylogeny of alveolates and eukaryotes in general (Burki et al., 2016; Mathur et al., 2019). Previously curated protein sequences, derived from a diversity of eukaryotes, were used as queries for BLASTp searches (e-value < $1x10^{-20}$, query coverage > 50%) against databases of transcriptomic and genomic protein predictions (Burki et al., 2016). The resulting hits were combined with the original protein sequences and single gene phylogenies were generated in RAxML v8.2.11 using the PROTGAMMALG model and 100 rapid bootstraps following alignment with MAFFT L-INS-i v7.222 and subsequent trimming with trimAl v1.2 using a gap threshold of 80% (Capella-Gutiérrez et al., 2009; Katoh and Standley, 2013; Stamatakis, 2014). Paralogues and contaminants were identified and removed manually after visually inspecting the phylogenies in FigTree v1.4.2 (Rambaut, 2012). The curated alignments were then filtered to remove those with less than 60% of species present and

the remaining 231 alignments were concatenated in SCaFoS v1.2.5 (Roure et al., 2007). The final concatenation comprised 231 proteins spanning 73,533 amino acid sites from 52 taxa.

4.4.6 Phylogenomic tree building and analysis

Maximum likelihood (ML) phylogenies were generated from the concatenated alignment in IQ-Tree v1.5.5 using either the free-rate heterogeneity model, LG+F+R8, as selected based on Bayesian information criteria in ModelFinder, or the empirical profile mixture model, LG+C60+F+G4 (Kalyaanamoorthy et al., 2017; Nguyen et al., 2015; Quang et al., 2008). Statistical support was assessed using both non-parametric bootstrap (n = 1,000, LG+F+R8) and ultrafast bootstrap pseudoreplicates (n = 1,000, LG+F+R8 and LG+C60+F+G4) (Hoang et al., 2018; Nguyen et al., 2015). Bayesian analyses were conducted using PhyloBayes MPI v1.4 using the GTR matrix with the CAT infinite mixture model and four gamma rate categories after the removal of constant sites (Lartillot et al., 2009). Four MCMC chains were run for at least 10,000 generations with every second tree saved. Chain convergence was assessed in PhyloBayes using a burn-in of 20%.

Topology robustness was assessed using alignment jackknifing and fast site removal to test for gene sampling bias and long branch attraction, respectively. Proteins comprising the concatenated alignment were randomly sampled without replacement to generate 12 gene concatenations (5% of the 231 genes, n = 58) from which ML phylogenies were generated using IQ-Tree v1.5.5 and the LG+F+R8 substitution model (Nguyen et al., 2015). Topologies were inspected in FigTree v1.4 and the frequencies of recovering a paraphyletic replication band, the best tree, and other topological variations were assessed (Rambaut, 2012). Moreover, site-specific rates were calculated across the 231 gene alignment using IQ-Tree v1.5.5 and the fastest sites were removed in 10% increments. Following fast site removal, ML phylogenies and non-parametric bootstraps (n = 500) were calculated in IQ-Tree v1.5.5 (Nguyen et al., 2015). Bootstrap support for the paraphyly of the replication band and the sister relationship between SAL and CONThreeP as well as *P. metchnikoffi* and previously sampled spirotrichs was assessed using ETE3 (Huerta-Cepas et al., 2016a).

4.4.7 Comparative genomics and gene ontology

To identify RB-associated homologues, we first compared all proteins predicted from ciliate transcriptomes and genomes (Table C1) to one another using BLASTp (e-value < $1x10^{-5}$, query coverage ≥ 50%) (Altschul et al., 1990). With the resulting comparison, we clustered proteins into orthologous groups (protein families) using the Markov clustering (MCL) algorithm with an inflation value of 1.4 (Enright et al., 2002). Clusters only containing proteins from a single species were discarded. To identify protein families that were significantly associated with the RB, we calculated the median number of homologues within each protein family for clades with (Spirotrichs, *L. macfarlandi*) and without the RB (*P. metchnikoffi*, *Nyctotherus ovalis*, Litostomatids, CONThreeP, Protocruziaea, and Postciliodesmatophora), and then in RB containing and lacking groups, generally. For each protein family, the difference in the median count between groups with and without the RB was then examined, and significance was assessed using two-sided permutation tests with a test distribution generated by randomizing the families 10,000 times by randomly sampling ciliate proteins without replacement. Proteins within enriched families ($P < 0.05$), that had representation from *O. trifallax* and *L. macfarlandi*, were subsequently annotated using InterproScan v5.36, eggNOG-Mapper v2, and BLASTp (e-value < $1x10^{-5}$, maximum target sequences = 1) in combination with Pfam, EggNOG, and SWISS-PROT databases, respectively (Altschul et al., 1990; Boeckmann et al., 2003; El-Gebali et al., 2019; Huerta-Cepas et al., 2016b, 2017; Quevillon et al., 2005). Protein annotation was achieved by assigning the predominant annotation from the constituent proteins to the group as a whole. Principal component analysis was conducted in R using differentially present protein families that had representation in *O. trifallax* and *L. macfarlandi*. Importantly, these RB-associated protein families are only correlated with the RB and must be experimentally validated to demonstrate a functional connection with the RB itself (see the RNAi analyses detailed below).

To explore functional enrichments amongst RB-associated proteins, gene ontology (GO) terms were assigned to individual proteins using eggNOG-Mapper v2 (Gene Ontology Consortium, 2004; Huerta-Cepas et al., 2017). GO annotations connected to individual proteins were then related to the protein's respective family before being simplified by mapping the terms to the *Saccharomyces cerevisiae* GO-slim subset using Map2Slim (see https://github.com/owlcollab/owltools/wiki/Map2Slim). GO enrichment was assessed by comparing the frequency of the terms in RB-associated families (that contained sequences from

O.trifallax and *L. macfarlandi*) to their frequency across all ciliate protein families. *P*-values were calculated using permutation tests wherein the test distributions were produced by generating 100,000 randomized, equally sized, test sets through random selection of families without replacement. Significantly enriched GO terms ($P < 0.05$) were summarized and visualized using REVIGO (Supek et al., 2011).

4.4.8 RNAi construct generation and knockdowns

RNAi plasmids were assembled using L4440 vectors and synthesized RNAi constructs (Timmons and Fire, 1998). RNAi inserts comprised variable numbers of tandem 100-150 bp fragments with homology to RB-associated paralogues in *O. trifallax*. Fragment specificity was confirmed using BLASTn against *O.trifallax* coding sequences (no identical matches over 17 bp) and inserts were synthesized by Integrated DNA Technologies. RNAi constructs capable of knocking down protein families CDC6 (2 paralogues), 259 (12), 280 (11), 482 (13), 1006 (3), 1008 (8), 1120 (6), 1197 (2), 1336 (3), 2113 (9), and 12334 (1) were synthesized (Table C2). CDC6 was chosen as a positive control as it is an essential regulator of DNA replication whereas other genes were selected based on technical feasibility (i.e., the number of paralogues), functional relatedness to RB-associated gene ontology terms (three cytoskeletal regulators and three genes involved in ion homeostasis), as well as four genes of unrelated function (Table C2). These families were also selected as they had association *P*-values $< 1x10^{-5}$ (Table C2). The RNAi constructs were amplified by polymerase chain reaction (PCR) using a high fidelity Phusion polymerase (New England BioLabs, NEB) and SacII and XbaI restriction sites were integrated at the 5' and 3' ends, respectively. RNAi constructs were inserted into the L4440 multiple cloning site following digestion with SacII and XbaI and ligation using T4 DNA ligase (NEB). The sequence composition of the resulting RNAi plasmids was confirmed by Sanger sequencing conducted by GeneWiz.

RNAi-expressing bacteria were produced as previously described (Paschka et al., 2003). Assembled plasmids were initially transformed into HT115 *Escherichia coli* and transformants were inoculated into lysogeny broth (LB) supplemented with ampicilin (100 µg/mL) and tetracycline (10 µg/mL), and grown over night at 37 ºC. Inoculants were then diluted to 0.02 OD_{600} in 2xYP media with antibiotics (Timmons and Fire, 1998) and cultures were incubated at

37 ºC until an OD_{600} of 0.4 was reached. RNAi-expression was then induced through the addition of 0.4 mM IPTG (Isopropyl β- d-1-thiogalactopyranoside) and induction was allowed to proceed over four hours. Cultures were harvested by centrifugation (3000xg for 10 minutes) and washed twice with Pringsheim media before being resuspended to an OD_{600} of 4.0 in Pringsheim media containing 2% glycerol. Prior to storage at -20 ºC, *E. coli* were heat killed following a five minute incubation at 65 ºC.

Feeding-based RNAi knockdowns were performed on *O. trifallax* by diluting stationary cultures to 2,000 cells/mL before growing them as described above with a reduced amount of *C. reinhardtii* (~95% reduction) and the addition of RNAi expressing *E. coli*. Cultures were slowly acclimatized to the presence of *E. coli* over three days by sequentially increasing the amount of *E. coli* in the cultures daily from an OD_{600} of 0.02, to 0.035, and finally to 0.05. After three days of pre-treatment, the cultures were again diluted to 2,000 cells/mL and each subsequent day, RNAi-expressing *E. coli* was added to 0.05 OD_{600}. After 36 hours, during log-phase growth, *O. trifallax* cultures were collected for RB analysis. Cultures were concentrated by centrifugation at 200xg for 2 minutes prior to the addition of 0.2 µg/mL DAPI (4′,6-diamidino-2-phenylindole). After a 10 minute incubation, the cells were fixed with 10% 3:1 methanol:acetic acid and were stored at room temperature in the dark. Replication band presence and morphology were inspected using a Zeiss Axioplan 2 microscope fitted with an X-Cite120LED fluorescence illuminator and micrographs were acquired using a Sony A7RIII digital camera attached with an LMScope digital SLR universal adapter. Replication band progression was measured in ImageJ and was calculated as the distance between the nuclear pole and the RB divided by the total length of the nuclear lobe. Band progression was measured in one nuclear lobe from each RB-containing cell.

4.4.9 Light microscopy and protargol staining

The morphology of living and protargol-impregnated cells of *Phacodinium metchnikoffi* (see Figure 13E–N) were examined under an Olympus BX51 microscope equipped with a Canon EOS 6 digital camera and cell measurements were made using an ocular micrometer. Differential interference contrast (DIC) and brightfield illumination (BF) were used to observe living and protargol-impregnated cells, respectively. For protargol staining, the following protocol was

employed: Cells were selected with micropipettes under the dissecting microscope, fixed in 10% v/v formalin (final concentration about 4 –5%) for 15 min, washed with tap water, placed in a few drops of 1% Mayer's albumin (1:1 egg albumin:glycerol), dropped on glass slides, and air-dried for at least 2 h. Dried slides were submerged in 95% isopropanol, 70% isopropanol, and tap water in staining jars for 5 min each. Slides were then placed in 0.2% potassium permanganate for 2 min, washed in tap water for 30 s, followed by 3 min in 2.5% oxalic acid then transferred to tap water for two 3 min soaks followed by 3 min in distilled water. Washed slides were then placed in pre-warmed 0.4% protargol solution (Polysciences) at 60 °C in a warming oven for 20 min (Note: Since 2013, protargol effective in the impregnation of ciliates is no longer commercially available. Commercial vendors may still offer "Strong Protargol" or "Protargol-S" which are ineffective for the demonstration of the ciliature. If stocks of an effective older product are unavailable, a protocol for synthesis of protargol effective for ciliate impregnation has been published (Pan et al., 2013).). Dieckmann's acetone developer was added to several drops of warm protargol on the slides to differentiate the ciliature and development was controlled under a compound microscope. When sufficiently developed, the slides were promptly washed in tap water for 15 s and then placed in 2.5% $Na_2S_2O_3$ for 3 s before being washed for 5 min in tap water, dehydrated for 10 min in both 70% and 100% isopropanol, and finally mounted in euparal (Hempstead Halide Inc.).

4.4.10 Scanning electron microscopy

For scanning electron microscopy, cells were selected under the dissecting microscope, fixed for 30–60 min (1% osmium tetroxide and 2.5% glutaraldehyde), and thoroughly washed at least three times in tap water under the dissecting microscope to remove debris. Cells were then placed in a small drop on 10 µm plankton net filters in custom-made brass chambers (Foissner, 2014). Cell containing chambers were subsequently transferred through a graded ethanol series (30%, 50%, 70%, 95%, 100%, 100% for 5 min each) for dehydration, and then processed in an EMS 850 critical point dryer. Dried cells were scattered onto adhesive carbon tabs on aluminum and sputtered with gold (argon pressure 0.15 Pa, energy 6–8 mA, ten 90 s sputtering exposures between 60 s rests) in a manual sputter coater (Agar Scientific). Processed samples (see Figure 13O–S) were examined using an SU3500 scanning electron microscope (Hitachi, USA) at an accelerating voltage of 10 kV in the secondary electron imaging mode.

4.4.11 Fluorescence microscopy and EdU labeling

To label cells with EdU (Invitrogen), cells were grown in 3 ml of medium to log phase in sterile 10 × 35 mm polystyrene culture dishes (Corning), 1 µL/mL of EdU (20mM) was added to the dish (final EdU conc. 20µM) and dishes were left for the desired interval (2–48 hr) before selecting cells under the stereo microscope for labeling with sulfo-Cy3-azide (Lumioprobe). Cells were collected in the wells of glass depression-slides before being washed three times in 0.22 µm-filtered EdU-free medium. Washed cells were then fixed for 15 min in 10% formalin, washed once in 100 mM Tris buffer, and then placed in fresh Tris for 5 min. Tris was removed and replaced with 0.01% TritonX-100 for 10 min to permeabalize the cells. The cells were then washed three times in tap water and the Cy3-labeling mixture was prepared (See notes below). Water was removed leaving cells in as small a volume as possible, before the cells were incubated in the dark for 30 min after the addition of 200 µL of labeling solution. Following the incubation, cells were washed three times and then stained with 0.5 µg/mL Hoechst 33342 (Sigma Aldrich) for 5 min. After staining, the cells were washed three times in tap water, and several drops of 1% Mayer's albumin were added before the cells were dropped onto clean glass slides and air-dried for at least 2 hr in the dark. Lastly, slides were dehydrated and mounted as in the protargol staining protocol. Slides were examined with a Zeiss Aksioskop 2 plus epifluorescence microscope with a 100x oil immersion objective (N.A. 1.3) and Zeiss filter sets 20 HE and 01 for Cy3 and Hoechst 3342, respectively. Images were taken with a Flex CCD camera (Diagnostic Instruments). Confocal fluorescent microscopy was done using a Zeiss LSM 510 Meta confocal imaging system using a rhodamine laser line (excitation 543 nm/emission 560 nm) and a Plan-Apochoromat 63X (N.A. 1.4) oil immersion objective (X-Y voxel size 0.8 µm, Z voxel size 0.5 µm).

4.4.12 Quantification and statistical analysis

Statistical support for phylogenies was obtained using 1,000 non-parametric bootstraps (using the LG+F+R8 substitution model), 1,000 ultrafast bootstraps (using the LG+C60+F+G4 and LG+F+R8 models), and Bayesian posterior probabilities calculated after at least 10,000 generations (GTR+CAT model, four chains, chain bipartition discrepancies: max diff. = 0.175857, mean diff. = 8.37×10^{-4}) (Figure 14A). Non-parametric bootstrap support for clades

following fast site removal was calculated from 500 non parametric bootstraps using the LG+F+R8 substitution model (Figure 14D). Statistical support for the correlation between protein families and defined phenotypes was assessed using permutation tests and 10,000 permutations (Figure 17A, Figure C2A) whereas gene ontology enrichment was assessed using 100,000 permutations (Figure 17C, Figure C2C, D). Principal component analyses were conducted in R (Figure 17B, Figure C2B). Comparisons between replication band prevalence following RNA interference were made using Welch's *t*-tests (n = 3, where each biological replicate represents the proportion of cells with replication bands assessed from an average of 238.7 cells (151-325, standard deviation = 37.4)) (Figure 18A). Replication band progression (Figure 18B) was compared between RNA interference treatments using both Levene's tests and Kolmogorov-Smirnov (KS) tests for variance and distribution equality, respectively (sample size is denoted in Figure 18B, where each replicate represents a single measured cell. Cell measurements were pooled from three biological replicates (see Figure 18A)).

4.5 Chapter-specific acknowledgements

We gratefully acknowledge Laura Landweber, Sheela George, and Talya Yerlici for providing strains of *Oxytricha trifallax* and detailed culturing instructions, Catharine Rankin and Alvaro Luna for supplying HT115 *Escherichia coli* strains and L4440 vectors, Jae-Hyeok Lee and Moyan Jia for providing *Chlamydomonas* strains, and Thomas Richards and David Milner for sharing their non-specific RNAi control vector. We also thank Elizabeth Cooney and Emma George for culturing assistance and Lumiprobe Corp. for the gift of Cy3-azide. This work was supported by grants from the Natural Sciences and Engineering Research Council (NSERC) to Denis Lynn and Patrick Keeling (RGPIN-2014-03994). Nicholas Irwin was supported by an NSERC Canadian Graduate Scholarship and Alexandros Pittis was supported by a European Molecular Biology Organization (EMBO) Long-Term Fellowship.

5. Conclusions and perspectives

5.1 Chapter summaries and future directions

In this dissertation, I hypothesized that investigating the functional diversification and evolution of nuclear processes in non-model eukaryotes with alternative nuclear systems would provide insights into the function and evolution of the nucleus and its processes in general.

In Chapter 2, I explored the evolutionary mechanisms underlying the transition between canonical and DVNP-based chromatin in dinoflagellates by heterologously expressing DVNP in the budding yeast, *S. cerevisiae*. I showed that DVNP impairs growth and antagonizes chromatin by localizing to histone binding sites, displacing nucleosomes, and impairing transcription. Furthermore, I demonstrated that DVNP toxicity could be relieved through histone depletion and that cells diminish their histones in response to DVNP expression, suggesting that histone reduction could have been an adaptive response to these viral proteins (Irwin et al., 2018).

This work is significant in that it demonstrates the capacity of viruses to drive chromatin evolution and provides functional insights into the role of DVNP and remnant histones in the dinoflagellate nucleus. DVNPs likely represent a viral innovation that was co-opted by dinoflagellates following viral infection. This exemplifies the role of viruses as nuclear tinkerers, as they can invent nuclear proteins to facilitate their reproduction which host organisms can subsequently adapt to. As hosts develop tolerance to these proteins, the likelihood of their integration into host systems increases, which can culminate in dramatic cellular alterations, as is observed in dinoflagellates. This raises questions regarding the origins of nuclear proteins more broadly, the role of viral-eukaryotic gene transfer in eukaryogenesis, and whether other transfers have occurred under adaptive or neutral evolutionary pretenses. Systematically investigating the phylogenetic history of nuclear proteins, while exploiting the recently expanded availability of viral genomic data from metagenomic surveys, may provide answers to these questions. Moreover, by demonstrating the repressive activity of DVNP and the conflict between DVNP and nucleosomes, this work supports the possibility that nucleosomes still serve an activating role in the dinokaryon which is an important possible function of dinoflagellate histones. Could the relaxed negative selection on these histones reflect that histone conservation is linked

predominantly to its silencing role? How have chromatin modifying enzymes and chaperones diverged alongside histones? Further investigations into the chromatin biology of the dinoflagellate nucleus will provide unique insights into the functionality of nucleosomes and assist in understanding the interconnectedness between histones and other nuclear components.

In Chapter 3, I investigated the composition of the NPC in nucleomorphs by using genomic and transcriptomic data to identify and phylogenetically classify NPC proteins in nucleomorph-containing algae. Although I found NPC proteins in all examined lineages, nearly all of those found in chlorarachniophytes and cryptophytes were single copy, host-related proteins that lacked signal peptides. Two exceptions were Nup98 and Rae1, which had clear nucleomorph-derived homologues. However, these proteins alone are likely insufficient to structure a canonical NPC and previous reports have revealed that Nup98 and Rae1 have other nuclear functions. These data indicate that nucleomorphs represent eukaryotic nuclei without a canonical NPC, raising fundamental questions about their structure and function and the importance and composition of the NPC during eukaryotic evolution (Irwin and Keeling, 2019).

The extensive and convergent reduction of NPCs in nucleomorphs raises the possibility that other nuclear processes and structures have been similarly altered. Reconstructing nuclear components such as cell cycle pathways, the kinetochore complex, and DNA replication machinery in nucleomorphs will help reveal baseline requirements for nucleomorph function and may shed light on host-endosymbiont control strategies. This could highlight proteins and mechanisms that are indispensable for maintaining core nuclear functionality. Understanding and exploiting the consequences of these losses will also provide information on how these processes interact with different aspects of the nucleus. For example, nucleomorphs provide a model to study the function of Nup98 exclusive of its interactions with other nucleoporins. Similarly, developing a mechanistic view of nucleomorphs may influence how we imagine a primitive nucleus functioning during eukaryogenesis. It seems improbable that a complex NPC was required during the initial stages of nuclear evolution, so this intricate system likely arose by an amalgamation of additional proteins to a simpler initial system where one or two proteins were capable of sufficiently mediating nucleocytoplasmic exchange. Recent advances in genetically modifying chlorarachniophytes have established these organisms as tractable models, which will

drive nucleomorph research in the future and permit more detailed mechanistic investigations of these systems (Fukuda et al., 2020).

Lastly, in Chapter 4, I examined the function and evolutionary history of motile DNA replication systems, particularly the replication band, in spirotrich ciliates. I showed that the replication band can take unique physical forms in different species, ranging from polar bands to a "replication envelope", where replication initiates at the nuclear periphery and advances centripetally inwards. Furthermore, I identified genes involved in cellular transport, including calcium transporters and cytoskeletal regulators, that are associated with the RB and may be involved in its function and translocation. These findings revealed the complex evolution and diversity of motile DNA replication systems and raise new possibilities regarding the regulation of nuclear organization and processes.

Further biochemical characterization of the RB and replication envelope may provide new perspectives from which general DNA replication mechanisms can be investigated. Detailed functional investigation of individual replication band-associated proteins, such as those linking calcium signaling to cytoskeletal regulation (e.g., gelsolin homologues), will be essential for understanding ciliate nuclear function and may be applicable to other eukaryotes. The large dataset of replication band-associated proteins generated here, will hopefully serve as a useful resource for guiding future studies. Additionally, perseverance regarding the description of replication patterns in non-spirotrich ciliates, which contain fragmented chromosomes, including armophoreans, litostomateans, and phylopharyngeans, will be important for further understanding the functional and evolutionary malleability of DNA replication systems. By characterizing these systems thoroughly, we will attain the basic knowledge required to interpret and use these organisms as experimental models for studying chromosome biology, DNA replication, and nuclear organization more generally. The small size of nanochromosomes permits their synthesis and modification by simple PCR methods and the ability to microinject them directly into cells makes them a remarkable model for studying chromosome structure and nucleosome dynamics. Indeed, recent studies have highlighted the connection between DNA methylation and nucleosome occupancy using nanochromosomes in *O. trifallax* (Beh et al., 2019) and in the past, the discovery of telomerase occurred following experiments in *T. thermophila* (Greider and Blackburn, 1985). Similarly, the vectorial nature of DNA replication in

these systems, the tight localization of replication machinery to molecular hubs, and the extravagant displays of nuclear organization, make these organisms promising models for future research.

These results ultimately confirm the utility of studying divergent nuclear systems, provide insights into functional and evolutionary nuclear mechanisms, and prompt a reassessment of how we view eukaryotic diversity and select model organisms.

5.2 Utilizing biodiversity to drive cellular research and the nature of model organisms

Eukaryotic biodiversity offers abundant opportunities to explore cellular and molecular processes from alternative perspectives. The studies presented in this dissertation advance our basic understanding of some of these diverse systems which is essential before more detailed investigations can take place. Examining divergent cellular systems is not only interesting because of their dependency on alternative mechanisms, but because in some instances they present a simplified view of complex processes. This is largely a result of gene loss and subfunctionalization which reduces the complexity of certain processes by simplifying nuclear interconnectedness. This is apparent in dinoflagellates and nucleomorphs where histone divergence and NPC reduction present a simplified view of nucleosome biology and nucleoporin interactions, respectively. Evidently, these eukaryotes are promising model organisms for studying the nucleus and its processes, but why are they underexploited for nuclear research? This may in part reflect a changing view of what we consider to be a model organism. In the past, these organisms were studied at the cellular level with a greater frequency, particularly in the 1970s and 80s, when they were recognized as having potential to illuminate the function of eukaryotic processes. However, as research has reoriented to focus on applied, as opposed to basic objectives, it appears that the concept of a model organisms has shifted as well. Rather than a species that presents a simplified model for studying a molecular process, model organisms now seem to be increasingly regarded as species from which insights into animal or crop biology can specifically be attained. In accordance with this, technological advances, such as the development of CRISPR-Cas9 genome editing, has improved the experimental tractability of human cell lines, questioning the relevance of model organisms in this respect. This is a valid,

but limited view of model organisms that results in their under-utilization. I believe that re-evaluating our concepts of model organisms and utilizing eukaryotic diversity to its full potential to select experimental systems that present simplified cellular processes, including those beyond the nucleus, will be key in driving basic cellular research in the years to come.

5.3 Evolutionary cell biology as an emerging field - how do we proceed?

The concept of studying diverse cellular systems to infer functional and evolutionary insights has come to be described as evolutionary cell biology (Lynch et al., 2014). Regardless of the terminology, the expansion of modern cell biology research into diverse organisms has been primed by recent methodological and technological advances. Individual and consortium efforts have led to the development and improvement of genetic methods and genomic resources for non-model eukaryotes (Faktorová et al., 2020; Keeling et al., 2014). In addition to this, protocols for growing and isolating cells have progressed, and the field is poised to incorporate more experimental research. The development of RNAi methods, vector and homologous recombination-based genetic modification systems, and other techniques, such as ChIP-seq, organellar proteomics, and proximity labeling, now facilitate the detailed experimental analysis of cellular systems in non-model organisms. In order for these to be used to their full potential, protocols, but perhaps more importantly cell cultures, need to be shared openly and freely. Culture collections have failed in some regards as they often lack the expertise to culture certain organisms, availability is often variable, and pricing can be exorbitant. This is not the fault of these institutions per se, but likely reflects the difficulty of trying to maintain a large diversity of species under one roof and a lack of appreciation from funding bodies for these resources. Indeed, many cultures and strains, including those from which genome-sequenced isolates were obtained, are grown and maintained solely in individual laboratories. Finding ways of sharing these organisms, advertising their availability, and compiling methods for their effective cultivation will be essential in alleviating the daunting task of establishing new species in the lab and encouraging the use of alternative model organisms. Finally, in order for the relevance of discoveries in diverse eukaryotes to be appreciated and integrated into our textbook view of the eukaryotic cell, I believe it will be important to emphasize findings from a systems perspective (e.g., we are studying DNA replication in ciliates, not ciliate DNA replication) and reconciling

the disparity between classical cell biologists, biochemists, protistologists, and evolutionary biologists will also be key in driving the field forward.

Bibliography

Adam, S.A. (2017). The nucleoskeleton. Cold Spring Harb Perspect Biol *9:a023556*.

Aeschlimann, S.H., Jönsson, F., Postberg, J., Stover, N.A., Petera, R.L., Lipps, H.-J., Nowacki, M., and Swart, E.C. (2014). The draft assembly of the radically organized *Stylonychia lemnae* macronuclear genome. Genome Biol. Evol. *6*, 1707–1723.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Ammar, R., Torti, D., Tsui, K., Gebbia, M., Durbic, T., Bader, G.D., Giaever, G., and Nislow, C. (2012). Chromatin is an ancient innovation conserved between Archaea and Eukarya. Elife *1*, e00078.

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Ref. Source.

Arikawa, M., Watanabe, A., Watanabe, K., and Suzaki, T. (2000). High-resolution scanning electron microscopy of chromatin bodies and replication bands of isolated macronuclei in the hypotrichous ciliate *Euplotes aediculatus*. Eur. J. Protistol. *36*, 40–45.

Arikawa, M., Momokawa, N., Saito, A., Omura, G., Khan, S.M.M.K., Suetomo, Y., Kakuta, S., and Suzaki, T. (2003). Ca^{2+}-dependent contractility of isolated and demembranated macronuclei in the hypotrichous ciliate *Euplotes aediculatus*. Cell Calcium *33*, 113–117.

Armakola, M., Higgins, M.J., Figley, M.D., Barmada, S.J., Scarborough, E.A., Diaz, Z., Fang, X., Shorter, J., Krogan, N.J., Finkbeiner, S., et al. (2012). Inhibition of RNA lariat debranching enzyme suppresses TDP-43 toxicity in ALS disease models. Nat. Genet. *44*, 1302–1309.

Avgousti, D.C., Herrmann, C., Kulej, K., Pancholi, N.J., Sekulic, N., Petrescu, J., Molden, R.C., Blumenthal, D., Paris, A.J., Reyes, E.D., et al. (2016). A core viral protein binds host nucleosomes to sequester immune danger signals. Nature *535*, 173–177.

Aze, A., and Maiorano, D. (2018). Recent advances in understanding DNA replication: Cell type–specific adaptation of the DNA replication program. F1000Research *7*, 1351.

Babu, J.R., Jeganathan, K.B., Baker, D.J., Wu, X., Kang-Decker, N., and Van Deursen, J.M. (2003). Rae1 is an essential mitotic checkpoint regulator that cooperates with Bub3 to prevent chromosome missegregation. J. Cell Biol. *160*, 341–353.

Bähler, J., Wu, J.Q., Longtine, M.S., Shah, N.G., McKenzie, A., Steever, A.B., Wach, A., Philippsen, P., and Pringle, J.R. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. Yeast *14*, 943–951.

van Bakel, H., Tsui, K., Gebbia, M., Mnaimneh, S., Hughes, T.R., and Nislow, C. (2013). A compendium of nucleosome and transcript profiles reveals determinants of chromatin architecture and transcription. PLoS Genet. *9*, e1003479.

Balamuth, W. (1941). Microscopic anatomy of *Licnophora macfarlandi*. J. Morphol. *68*, 241–277.

Beams, H.W., Tahmisian, T.N., Devine, R., and Anderson, E. (1957). Ultrastructure of the nuclear membrane of a gregarine parasitic in grasshoppers. Exp. Cell Res. *13*, 200–204.

Beauchemin, M., and Morse, D. (2017). A proteomic portrait of dinoflagellate chromatin reveals abundant RNA-binding proteins. Chromosoma *127*, 29–43.

Beck, M., and Hurt, E. (2017). The nuclear pore complex: Understanding its function through structural insight. Nat. Rev. Mol. Cell Biol. *18*, 73–89.

Beh, L.Y., Debelouchina, G.T., Clay, D.M., Thompson, R.E., Lindblad, K.A., Hutton, E.R., Bracht, J.R., Sebra, R.P., Muir, T.W., and Landweber, L.F. (2019). Identification of a DNA N6-adenine methyltransferase complex and its impact on chromatin organization. Cell *177*, 1781-1796.e25.

Bell, P.J. (2019). Evidence supporting a viral origin of the eukaryotic nucleus. BioRxiv https://doi.org/10.1101/679175.

Bell, P.J.L. (2001). Viral eukaryogenesis: Was the ancestor of the nucleus a complex DNA

virus? J. Mol. Evol. *53*, 251–256.

Bendtsen, J.D., Nielsen, H., Von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. *340*, 783–795.

Bermejo, R., Kumar, A., and Foiani, M. (2012). Preserving the genome by regulating chromatin association with the nuclear envelope. Trends Cell Biol. *22*, 465–473.

Bhaud, Y., Guillebault, D., Lennon, J.F., Defacque, H., Soyer-Gobillard, M.O., and Moreau, H. (2000). Morphology and behaviour of dinoflagellate chromosomes during the cell cycle and mitosis. J. Cell Sci. *113*, 1231–1239.

Bleichert, F., Botchan, M.R., and Berger, J.M. (2017). Mechanisms for initiating cellular DNA replication. Science *355*, eaah6317.

Blevins, M.B., Smith, A.M., Phillips, E.M., and Powers, M.A. (2003). Complex formation among the RNA export proteins Nup98, Rae1/Gle2, and TAP. J. Biol. Chem. *278*, 20979–20988.

Blobel, G. (1985). Gene gating: A hypothesis. Proc. Natl. Acad. Sci. U. S. A. *82*, 8527–8529.

Blower, M.D., Nachury, M., Heald, R., and Weis, K. (2005). A Rae1-containing ribonucleoprotein complex is required for mitotic spindle assembly. Cell *121*, 223–234.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., et al. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. *31*, 365–370.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Boscaro, V., Syberg-Olsen, M.J., Irwin, N.A.T., del Campo, J., and Keeling, P.J. (2018a). What can environmental sequences tell us about the distribution of low-rank taxa? The case of *Euplotes* (Ciliophora, Spirotrichea), including a description of *Euplotes enigma* sp. nov. J. Eukaryot. Microbiol. *66*, 281–293.

Boscaro, V., Santoferrara, L.F., Zhang, Q., Gentekaki, E., Syberg-Olsen, M.J., del Campo, J., and Keeling, P.J. (2018b). EukRef-Ciliophora: A manually curated, phylogeny-based database of

small subunit rRNA gene sequences of ciliates. Environ. Microbiol. *20*, 2218–2230.

Bracht, J.R., Perlman, D.H., and Landweber, L.F. (2012). Cytosine methylation and hydroxymethylation mark DNA for elimination in *Oxytricha trifallax*. Genome Biol. *13*, R99.

Braun, F.N. (2008). Thermodynamics of the prokaryote nuclear zone. Int. J. Astrobiol. *7*, 183–185.

Braun, R.E. (2001). Packaging paternal chromosomes with protamine. Nat. Genet. *28*, 10–12.

Brind'Amour, J., Liu, S., Hudson, M., Chen, C., Karimi, M.M., and Lorincz, M.C. (2015). An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. Nat. Commun. *6*, 6033.

Brogaard, K., Xi, L., Wang, J.-P., and Widom, J. (2012). A map of nucleosome positions in yeast at base-pair resolution. Nature *486*, 496–501.

Buchfink, B., Xie, C., and Huson, D.H. (2014). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60.

Burki, F., Kaplan, M., Tikhonenkov, D. V., Zlatogursky, V., Minh, B.Q., Radaykina, L. V., Smirnov, A., Mylnikov, A.P., and Keeling, P.J. (2016). Untangling the early diversification of eukaryotes: A phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. Proc. R. Soc. B Biol. Sci. *283*, 20152802.

Burki, F., Roger, A.J., Brown, M.W., and Simpson, A.G.B. (2020). The new tree of eukaryotes. Trends Ecol. Evol. *35*, 43–55.

Burri, L., and Keeling, P.J. (2007). Protein targeting in parasites with cryptic mitochondria. Int. J. Parasitol. *37*, 265–272.

Callahan, K.P., and Butler, J.S. (2010). TRAMP complex enhances RNA degradation by the nuclear exosome component Rrp6. J. Biol. Chem. *285*, 3540–3547.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.

Capella-Gutiérrez, S., Marcet-Houben, M., and Gabaldón, T. (2012). Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. BMC Biol. *10*, 47.

Capelson, M., Liang, Y., Schulte, R., Mair, W., Wagner, U., and Hetzer, M.W. (2010). Chromatin-bound nuclear pore components regulate gene expression in higher eukaryotes. Cell *140*, 372–383.

Cavalier-Smith, T. (1982). The evolutionary origin and phylogeny of eukaryote flagella. Symp. Soc. Exp. Biol. *35*, 465–493.

Cavalier-Smith, T. (1988). Origin of the cell nucleus. BioEssays *9*, 72–78.

Cavalier-Smith, T. (2010). Origin of the cell nucleus, mitosis and sex: Roles of intracellular coevolution. Biol. Direct *5*, 7.

Cervantes, M.D., Coyne, R.S., Xi, X., and Yao, M.-C. (2006). The condensin complex is essential for amitotic segregation of bulk chromosomes, but not nucleoli, in the ciliate *Tetrahymena thermophila*. Mol. Cell. Biol. *26*, 4690–4700.

Chaikeeratisak, V., Nguyen, K., Khanna, K., Brilot, A.F., Erb, M.L., Coker, J.K.C., Vavilina, A., Newton, G.L., Buschauer, R., Pogliano, K., et al. (2017a). Assembly of a nucleus-like structure during viral replication in bacteria. Science *355*, 194–197.

Chaikeeratisak, V., Nguyen, K., Egan, M.E., Erb, M.L., Vavilina, A., and Pogliano, J. (2017b). The phage nucleus and tubulin spindle are conserved among large *Pseudomonas* phages. Cell Rep. *20*, 1563–1571.

Chen, K., Xi, Y., Pan, X., Li, Z., Kaestner, K., Tyler, J., Dent, S., He, X., and Li, W. (2013). DANPOS: Dynamic analysis of nucleosome position and occupancy by sequencing. Genome Res. *23*, 341–351.

Chen, X., Zhao, X., Liu, X., Warren, A., Zhao, F., and Miao, M. (2015). Phylogenomics of non-model ciliates based on transcriptomic analyses. Protein Cell *6*, 373–385.

Chow, M.H., Yan, K.T.H., Bennett, M.J., and Wong, J.T.Y. (2010). Birefringence and DNA condensation of liquid crystalline chromosomes. Eukaryot. Cell *9*, 1577–1587.

Chudnovsky, Y., Li, J.F., Rizzo, P.J., Hastings, J.W., and Fagan, T.F. (2002). Cloning, expression, and characterization of a histone-like protein from the marine dinoflagellate *Lingulodinium polydrum* (Dinophyceae). J. Phycol. *38*, 543–550.

Collen, J., Porcel, B., Carre, W., Ball, S.G., Chaparro, C., Tonon, T., Barbeyron, T., Michel, G., Noel, B., Valentin, K., et al. (2013). Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. Proc. Natl. Acad. Sci. U. S. A. *110*, 5247–5252.

Corradi, N., Pombert, J.-F., Farinelli, L., Didier, E.S., and Keeling, P.J. (2010). The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. Nat. Commun. *1*, 77.

Curtis, B.A., Tanifuji, G., Burki, F., Gruber, A., Irimia, M., Maruyama, S., Arias, M.C., Ball, S.G., Gile, G.H., Hirakawa, Y., et al. (2012). Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature *492*, 59–65.

Dacks, J.B., and Field, M.C. (2007). Evolution of the eukaryotic membrane-trafficking system: Origins, tempo and mode. J. Cell Sci. *120*, 2977–2985.

Dacks, J.B., and Kasinsky, H.E. (1999). Nuclear condensation in protozoan gametes and the evolution of anisogamy. Comp. Biochem. Physiol. *124*, 287–295.

Dalhgren, L. (1967). On the nuclear distribution of RNA and DNA and on the ultrastructure of nuclei and adjacent cytoplasm of the foraminifers *Ovammina opaca* Dahlgren, *Hippocrepinella alba* Heron-Allen and Earland, and *Globobulimina turgida* (Bailey). Zool. Bridrag. Uppsala *37*, 113–138.

Daniels, J.-P., Gull, K., and Wickstead, B. (2010). Cell biology of the trypanosome genome. Microbiol. Mol. Biol. Rev. *74*, 552–569.

Davidson, P.M., Denais, C., Bakshi, M.C., and Lammerding, J. (2014). Nuclear deformability constitutes a rate-limiting step during cell migration in 3-D environments. Cell. Mol. Bioeng. *7*, 293–306.

Devbhandari, S., Jiang, J., Kumar, C., Whitehouse, I., and Remus, D. (2017). Chromatin

constrains the initiation and elongation of DNA replication. Mol. Cell *65*, 131–141.

Devos, D.P. (2014). PVC bacteria: Variation of, but not exception to, the Gram-negative cell plan. Trends Microbiol. *22*, 14–20.

Devos, D., Dokudovskaya, S., Alber, F., Williams, R., Chait, B.T., Sali, A., and Rout, M.P. (2004). Components of coated vesicles and nuclear pore complexes share a common molecular architecture. PLoS Biol. *2*, e380.

Dolezal, P., Smíd, O., Rada, P., Zubácová, Z., Bursać, D., Suták, R., Nebesárová, J., Lithgow, T., and Tachezy, J. (2005). *Giardia* mitosomes and trichomonad hydrogenosomes share a common mode of protein targeting. Proc. Natl. Acad. Sci. U. S. A. *102*, 10924–10929.

Doucet, C.M., Talamas, J.A., and Hetzer, M.W. (2010). Cell cycle-dependent differences in nuclear pore complex assembly in metazoa. Cell *141*, 1030–1041.

Douglas, S., Zauner, S., Fraunholz, M., Beaton, M., Penny, S., Deng, L.-T., Wu, X., Reith, M., Cavalier-Smith, T., and Maier, U.G. (2001). The highly reduced genome of an enslaved algal nucleus. Nature *410*, 1091–1096.

Douzery, E.J.P., Snell, E.A., Bapteste, E., Delsuc, F., and Philippe, H. (2004). The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? Proc. Natl. Acad. Sci. U. S. A. *101*, 15386–15391.

Downs, J.A., Lowndes, N.F., and Jackson, S.P. (2000). A role for *Saccharomyces cerevisiae* histone H2A in DNA repair. Nature *408*, 1001–1004.

Duffy, S. (2018). Why are RNA virus mutation rates so damn high? PLoS Biol. *16*, e3000003.

Dultz, E., and Ellenberg, J. (2010). Live imaging of single nuclear pores reveals unique assembly kinetics and mechanism in interphase. J. Cell Biol. *191*, 15–22.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., et al. (2019). The Pfam protein families database in 2019. Nucleic Acids Res. *47*, D427–D432.

Emmott, E., and Hiscox, J.A. (2009). Nucleolar targeting: the hub of the matter. EMBO Rep. *10*,

231–238.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. *30*, 1575–1584.

Eriksson, P.R., Ganguli, D., Nagarajavel, V., and Clark, D.J. (2012). Regulation of histone gene expression in budding yeast. Genetics *191*, 7–20.

Erives, A.J. (2017). Phylogenetic analysis of the core histone doublet and DNA topo II genes of Marseilleviridae: Evidence of proto-eukaryotic provenance. Epigenetics and Chromatin *10*, 55.

Fairhead, C., Llorente, B., Denis, F., Soler, M., and Dujon, B. (1996). New vectors for combinatorial deletions in yeast chromosomes and for gap-repair cloning using "split-marker" recombination. Yeast *12*, 1439–1457.

Faktorová, D., Nisbet, R.E.R., Fernández Robledo, J.A., Casacuberta, E., Sudek, L., Allen, A.E., Ares, M., Aresté, C., Balestreri, C., Barbrook, A.C., et al. (2020). Genetic tool development in marine protists: Emerging model organisms for experimental cell biology. Nat. Methods *17*, 481–494.

Fang, G., and Cech, T.R. (1995). Telomerase RNA localized in the replication band and spherical subnuclear organelles in hypotrichous ciliates. J. Cell Biol. *130*, 243–253.

Fauré-Fremiet, E., Rouller, C., and Gauchery, M. (1957). La réorganisation macronucléaire chez les *Euplotes*. Exp. Cell Res. 135–144.

Fernández-Galiano, D., and Calvo, P. (1992). Redescription of *Phacodinium metchnikoffi* (Ciliophora, Hypotrichida): General morphology and taxonomic position. J. Protozool. *39*, 443–448.

Field, M.C., and Dacks, J.B. (2009). First and last ancestors: Reconstructing evolution of the endomembrane system with ESCRTs, vesicle coat proteins, and nuclear pore complexes. Curr. Opin. Cell Biol. *21*, 4–13.

Field, M.C., and Rout, M.P. (2019). Pore timing: the evolutionary origins of the nucleus and nuclear pore complex. F1000Research *8*, 369.

Field, M.C., Koreny, L., and Rout, M.P. (2014). Enriching the pore: Splendid complexity from humble origins. Traffic *15*, 141–156.

Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: Interactive sequence similarity searching. Nucleic Acids Res. *39*, 29–37.

Fitz, V., Shin, J., Ehrlich, C., Farnung, L., Cramer, P., Zaburdaev, V., and Grill, S.W. (2016). Nucleosomal arrangement affects single-molecule transcription dynamics. Proc. Natl. Acad. Sci. U. S. A. *113*, 12733–12738.

Flickinger, C.J. (1965). The fine structure of the nuclei of *Tetrahymena pyriformis* throughout the cell cycle. J. Cell Biol. *27*, 519–529.

Flickinger, C.J. (1970). The fine structure of the nuclear envelope in amebae: Alterations following nuclear transplantation. Exp. Cell Res. *60*, 225–236.

Foissner, W. (2014). An update of "basic light and scanning electron microscopic methods for taxonomic studies of ciliated protozoa." Int. J. Syst. Evol. Microbiol. *64*, 271–292.

Ford Doolittle, W. (1980). Revolutionary concepts in evolutionary cell biology. Trends Biochem. Sci. *5*, 146–149.

Fox, M.H., Arndt-Jovin, D.J., Jovin, T.M., Baumann, P.H., and Robert-Nicoud, M. (1991). Spatial and temporal distribution of DNA replication sites localized by immunofluorescence and confocal microscopy in mouse fibroblasts. J. Cell Sci. *99*, 247–253.

Fragkos, M., Ganier, O., Coulombe, P., and Méchali, M. (2015). DNA replication origin activation in space and time. Nat. Rev. Mol. Cell Biol. *16*, 360–374.

Fuerst, J.A., and Sagulenko, E. (2012). Keys to eukaryality: Planctomycetes and ancestral evolution of cellular complexity. Front. Microbiol. *3*, 167.

Fukuda, K., Cooney, E.C., Irwin, N.A.T., Keeling, P.J., and Hirakawa, Y. (2020). High-efficiency transformation of the chlorarachniophyte *Amorphochlora amoebiformis* by electroporation. Algal Res. *48*, 101903.

Gaglia, M.M., Covarrubias, S., Wong, W., and Glaunsinger, B.A. (2012). A common strategy for

host RNA degradation by divergent viruses. J. Virol. *86*, 9527–9530.

Gall, J.G. (1959). Macronuclear duplication in the ciliated protozoan *Euplotes*. J. Biophys. Biochem. Cytol. *5*, 295–308.

Gao, F., and Katz, L.A. (2014). Phylogenomic analyses support the bifurcation of ciliates into two major clades that differ in properties of nuclear division. Mol. Phylogenet. Evol. *70*, 240–243.

Gao, F., Warren, A., Zhang, Q., Gong, J., Miao, M., Sun, P., Xu, D., Huang, J., Yi, Z., and Song, W. (2016). The all-data-based evolutionary hypothesis of ciliated protists with a revised classification of the phylum Ciliophora (Eukaryota, Alveolata). Sci. Rep. *6*, 24874.

Gavelis, G.S., Herranz, M., Wakeman, K.C., Ripken, C., Mitarai, S., Gile, G.H., Keeling, P.J., and Leander, B.S. (2019). Dinoflagellate nucleus contains an extensive endomembrane network, the nuclear net. Sci. Rep. *9*, 839.

Gene Ontology Consortium (2004). The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. *32*, D258–D261.

Gentekaki, E., Kolisko, M., Boscaro, V., Bright, K.J., Dini, F., Di Giuseppe, G., Gong, Y., Miceli, C., Modeo, L., Molestina, R.E., et al. (2014). Large-scale phylogenomic analysis reveals the phylogenetic position of the problematic taxon *Protocruzia* and unravels the deep phylogenetic affinities of the ciliate lineages. Mol. Phylogenet. Evol. *78*, 36–42.

Gietz, R.D., and Woods, R.A. (2006). Yeast transformation by the LiAc/SS Carrier DNA/PEG method. Yeast Protoc. 107–120.

Gile, G.H., and Keeling, P.J. (2008). Nucleus-encoded periplastid-targeted EFL in chlorarachniophytes. Mol. Biol. Evol. *25*, 1967–1977.

Gilson, P.R., Su, V., Slamovits, C.H., Reith, M.E., Keeling, P.J., and McFadden, G.I. (2006). Complete nucleotide sequence of the chlorarachniophyte nucleomorph: Nature's smallest nucleus. Proc. Natl. Acad. Sci. U. S. A. *103*, 9566–9571.

Goldfarb, D.S., Gariépy, J., Schoolnik, G., and Kornberg, R.D. (1986). Synthetic peptides as

nuclear loclaization signals. Nature *322*, 641–644.

Goldman, R.D., Shumaker, D.K., Erdos, M.R., Eriksson, M., Goldman, A.E., Gordon, L.B., Gruenbaum, Y., Khuon, S., Mendez, M., Varga, R., et al. (2004). Accumulation of mutant lamin A causes progressive changes in nuclear architecture in Hutchinson-Gilford progeria syndrome. Proc. Natl. Acad. Sci. U. S. A. *101*, 8963–8968.

Golyshev, S., Berdieva, M., Musinova, Y., Sheval, E., and Skarlato, S. (2018). Ultrastructural organization of the chromatin elements in chromosomes of the dinoflagellate *Prorocentrum minimum*. Protistology *12*, 163–172.

Gomar-Alba, M., and Mendoza, M. (2020). Modulation of cell identity by modification of nuclear pore complexes. Front. Genet. *10*, 1301.

Gornik, S.G., Ford, K.L., Mulhern, T.D., Bacic, A., McFadden, G.I., and Waller, R.F. (2012). Loss of nucleosomal DNA condensation coincides with appearance of a novel nuclear protein in dinoflagellates. Curr. Biol. *22*, 2303–2312.

Gossett, A.J., and Lieb, J.D. (2012). In vivo effects of histone H3 depletion on nucleosome occupancy and position in *Saccharomyces cerevisiae*. PLoS Genet. *8*, e1002771.

Gould, S.B., Sommer, M.S., Kroth, P.G., Gile, G.H., Keeling, P.J., and Maier, U.G. (2006). Nucleus-to-nucleus gene transfer and protein retargeting into a remnant cytoplasm of cryptophytes and diatoms. Mol. Biol. Evol. *23*, 2413–2422.

Gould, S.B., Garg, S.G., and Martin, W.F. (2016). Bacterial vesicle secretion and the evolutionary origin of the eukaryotic endomembrane system. Trends Microbiol. *24*, 525–534.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

Gray, E.G., and Guillery, R.W. (1963). On nuclear structure in the ventral nerve cord of the leech *Hirudo medicinalis*. Zeitschrift Für Zellforsch. Und Mikroskopische Anat. *59*, 738–745.

Greenwood, A. (1974). The Cryptophyta in relation to phylogeny and photosynthesis. In 8th

International Congress of Electron Microscopy, J. Sanders, and D. Goodchild, eds. (Canberra: Australian Academy of Sciences), pp. 566–567.

Greider, C.W., and Blackburn, E.H. (1985). Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. Cell *43*, 406–413.

Griffis, E.R., Altan, N., Lippincott-Schwartz, J., and Powers, M.A. (2002). Nup98 Is a mobile nucleoporin with transcription-dependent dynamics. Mol. Biol. Cell *13*, 1282–1297.

Griffis, E.R., Xu, S., and Powers, M.A. (2003). Nup98 localizes to both nuclear and cytoplasmic sides of the nuclear pore and binds to two distinct nucleoporin subcomplexes. Mol. Biol. Cell *14*, 600–610.

Grimaldi, Y., Ferrari, P., and Strubin, M. (2014). Independent RNA polymerase II preinitiation complex dynamics and nucleosome turnover at promoter sites in vivo. Genome Res. *24*, 117–124.

Grisdale, C.J., Smith, D.R., and Archibald, J.M. (2019). Relative mutation rates in nucleomorph-bearing algae. Genome Biol. Evol. *11*, 1045–1053.

Guglielmini, J., Woo, A.C., Krupovic, M., Forterre, P., and Gaia, M. (2019). Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. Proc. Natl. Acad. Sci. U. S. A. *116*, 19585–19592.

Gunjan, A., and Verreault, A. (2003). A Rad53 kinase-dependent surveillance mechanism that regulates histone protein levels in *S. cerevisiae*. Cell *115*, 537–549.

Haapala, O.K., and Soyer, M.-O. (1975). Organization of chromosome fibrils in *Euglena gracilis*. Hereditas *80*, 185–194.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. *8*, 1494–1512.

Han, M., and Grunstein, M. (1988). Nucleosome loss activates yeast downstream promoters in

vivo. Cell *55*, 1137–1145.

Hargraves, P.E. (2002). The ebridian flagellates *Ebria* and *Hermesinum*. Plankt. Biol. Ecol. *49*, 9–16.

Hearne, J.L., and Pitula, J.S. (2011). Identification of two spliced leader RNA transcripts from *Perkinsus marinus*. J. Eukaryot. Microbiol. *58*, 266–268.

Hedtke, B., Börner, T., and Weihe, A. (1997). Mitochondrial and chloroplast phage-type RNA polymerases in *Arabidopsis*. Science *277*, 809–811.

Hendrickson, H.L., and Poole, A.M. (2018). Manifold routes to a nucleus. Front. Microbiol. *9*, 2604.

Hinshaw, S.M., and Harrison, S.C. (2018). Kinetochore function from the bottom up. Trends Cell Biol. *28*, 22–33.

Hirakawa, Y., Nagamune, K., and Ishida, K. (2009). Protein targeting into secondary plastids of chlorarachniophytes. Proc. Natl. Acad. Sci. U. S. A. *106*, 12820–12825.

Hirakawa, Y., Gile, G.H., Ota, S., Keeling, P.J., and Ishida, K.I. (2010). Characterization of periplastidal compartment-targeting signals in chlorarachniophytes. Mol. Biol. Evol. *27*, 1538–1545.

Hirakawa, Y., Burki, F., and Keeling, P.J. (2011). Nucleus- and nucleomorph-targeted histone proteins in a chlorarachniophyte alga. Mol. Microbiol. *80*, 1439–1449.

Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.W., Lyou, Y., Townes, T.M., Schübeler, D., and Gilbert, D.M. (2008). Global reorganization of replication domains during embryonic stem cell differentiation. PLoS Biol. *6*, e245.

Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. *35*, 518–522.

Holden, J.M., Koreny, L., Obado, S., Ratushny, A. V., Chen, W.-M., Chiang, J.-H., Kelly, S., Chait, B.T., Aitchison, J.D., Rout, M.P., et al. (2014). Nuclear pore complex evolution: a trypanosome Mlp analogue functions in chromosomal segregation but lacks transcriptional

barrier activity. Mol. Biol. Cell *25*, 1421–1436.

van Hooff, J.J., Tromer, E., van Wijk, L.M., Snel, B., and Kops, G.J. (2017). Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. EMBO Rep. *18*, e201744102.

Hoppenrath, M., and Leander, B.S. (2006). Ebriid phylogeny and the expansion of the Cercozoa. Protist *157*, 279–290.

Horiike, T., Hamada, K., Kanaya, S., and Shinozawa, T. (2001). Origin of eukaryotic cell nuclei by symbiosis of Archea in Bacteria revealed by homology-hit analysis. Nat. Cell Biol. *3*, 210–214.

Hsin, J.P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev. *26*, 2119–2137.

Hu, Z., Chen, K., Xia, Z., Chavez, M., Pal, S., Seol, J.H., Chen, C.C., Li, W., and Tyler, J.K. (2014). Nucleosome loss leads to global transcriptional up-regulation and genomic instability during yeast aging. Genes Dev. *28*, 396–408.

Huerta-Cepas, J., Serra, F., and Bork, P. (2016a). ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. Mol. Biol. Evol. *33*, 1635–1638.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M., et al. (2016b). eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. *44*, D286–D293.

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., and Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. Mol. Biol. Evol. *34*, 2115–2122.

Hurst, V., Shimada, K., and Gasser, S.M. (2019). Nuclear actin and actin-binding proteins in DNA repair. Trends Cell Biol. *29*, 462–476.

Husnik, F., and McCutcheon, J.P. (2018). Functional horizontal gene transfer from bacteria to

eukaryotes. Nat. Rev. Microbiol. *16*, 67–79.

Imachi, H., Nobu, M.K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., Takano, Y., Uematsu, K., Ikuta, T., Ito, M., et al. (2020). Isolation of an archaeon at the prokaryote–eukaryote interface. Nature *577*, 519–525.

Irwin, N.A.T., and Keeling, P.J. (2019). Extensive reduction of the nuclear pore complex in nucleomorphs. Genome Biol. Evol. *11*, 678–687.

Irwin, N.A.T., and Lynn, D.H. (2015). Molecular phylogeny of mobilid and sessilid ciliates symbiotic in eastern Pacific limpets (Mollusca: Patellogastropoda). J. Eukaryot. Microbiol. *62*, 543–552.

Irwin, N.A.T., Martin, B.J.E., Young, B.P., Browne, M.J.G., Flaus, A., Loewen, C.J.R., Keeling, P.J., and Howe, L.J. (2018). Viral proteins as a potential driver of histone depletion in dinoflagellates. Nat. Commun. *9*, 1535.

Irwin, N.A.T., Tikhonenkov, D. V., Hehenberger, E., Mylnikov, A.P., Burki, F., and Keeling, P.J. (2019). Phylogenomics supports the monophyly of the Cercozoa. Mol. Phylogenet. Evol. *130*, 416–423.

Iwamoto, M., Mori, C., Kojidani, T., Bunai, F., Hori, T., Fukagawa, T., Hiraoka, Y., and Haraguchi, T. (2009). Two distinct repeat sequences of Nup98 nucleoporins characterize dual nuclei in the binucleated ciliate *Tetrahymena*. Curr. Biol. *19*, 843–847.

Iwamoto, M., Osakada, H., Mori, C., Fukuda, Y., Nagao, K., Obuse, C., Hiraoka, Y., and Haraguchi, T. (2017). Compositionally distinct nuclear pore complexes of functionally distinct dimorphic nuclei in the ciliate *Tetrahymena*. J. Cell Sci. *130*, 1822–1834.

Iyer, L.M., Anantharaman, V., Wolf, M.Y., and Aravind, L. (2008). Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. Int. J. Parasitol. *38*, 1–31.

Jaffe, L.F. (1999). Organization of early development by calcium patterns. BioEssays *21*, 657–667.

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., et al. (2016). Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. Proc. Natl. Acad. Sci. U. S. A. *114*, E171–E180.

Janzen, C.J., Fernandez, J.P., Deng, H., Diaz, R., Hake, S.B., and Cross, G.A.M. (2006). Unusual histone modifications in *Trypanosoma brucei*. FEBS Lett. *580*, 2306–2310.

Jaroenlak, P., Cammer, M., Davydov, A., Sall, J., Usmani, M., Liang, F.-X., Ekiert, D.C., and Bhabha, G. (2020). 3-dimensional organization and dynamics of the microsporidian polar tube invasion machinery. BioRxiv https://doi.org/10.1101/2020.04.03.024240.

Jeganathan, K.B., Malureanu, L., and Van Deursen, J.M. (2005). The Rae1-Nup98 complex prevents aneuploidy by inhibiting securin degradation. Nature *438*, 1036–1039.

Jékely, G. (2008). Origin of the nucleus and Ran-dependent transport to safeguard ribosome biogenesis in a chimeric cell. Biol. Direct *3*, 31.

Jogler, C., Wiegand, S., and Devos, D.P. (2019). Commentary: Manifold routes to a nucleus. Front. Microbiol. *10*, 1198.

Johnson, L.K., Alexander, H., and Brown, C.T. (2019). Re-assembly, quality evaluation, and annotation of 678 microbial eukaryotic reference transcriptomes. Gigascience *8*, giy158.

Jung, H.J., Nobumori, C., Goulbourne, C.N., Tu, Y., Lee, J.M., Tatar, A., Wu, D., Yoshinaga, Y., De Jong, P.J., Coffinier, C., et al. (2013). Farnesylation of lamin B1 is important for retention of nuclear chromatin during neuronal migration. Proc. Natl. Acad. Sci. U. S. A. *110*, 1923–1932.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589.

Kaplan, C.D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription inititaion from cryptic sites. Science *239*, 17–18.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., Leproust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., et al. (2009). The DNA-encoded nucleosome

organization of a eukaryotic genome. Nature *458*, 362–366.

Kato, K.H., Moriyama, A., Huitorel, P., Cosson, J., Cachon, M., and Sato, H. (1997). Isolation of the major basic nuclear protein and its localization on chromosomes of the dinoflagellate, *Oxyrrhis marina*. Biol. Cell *89*, 43–52.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

Keeling, P.J. (2013). The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. Annu. Rev. Plant Biol. *64*, 583–607.

Keeling, P.J. (2017). Chlorarachniophytes. In Handbook of the Protists, A.G.B. Simpson, C.H. Slamovits, and J.M. Archibald, eds. (Springer), pp. 765–781.

Keeling, P.J., and Burki, F. (2019). Progress towards the tree of eukaryotes. Curr. Biol. *29*, R808–R817.

Keeling, P.J., and Fast, N.M. (2002). Microsporidia: Biology and evolution of highly reduced intracellular parasites. Annu. Rev. Microbiol. *56*, 93–116.

Keeling, P.J., and Palmer, J.D. (2008). Horizontal gene transfer in eukaryotic evolution. Nat. Rev. Genet. *9*, 605–618.

Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., et al. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol. *12*, e1001889.

Khorasanizadeh, S. (2004). The nucleosome: From genomic organization to genomic regulation. Cell *116*, 259–272.

Kim, S.J., Fernandez-Martinez, J., Nudelman, I., Shi, Y., Zhang, W., Raveh, B., Herricks, T., Slaughter, B.D., Hogan, J.A., Upla, P., et al. (2018). Integrative structure and functional anatomy of a nuclear pore complex. Nature *555*, 475–482.

Klinger, C.M., Spang, A., Dacks, J.B., and Ettema, T.J.G. (2016). Tracing the archaeal origins of

eukaryotic membrane-trafficking system building blocks. Mol. Biol. Evol. *33*, 1528–1541.

Köhler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. Nat. Rev. Mol. Cell Biol. *8*, 761–773.

Kolev, N.G., Franklin, J.B., Carmi, S., Shi, H., Michaeli, S., and Tschudi, C. (2010). The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. PLoS Pathog. *6*, e1001090.

Koonin, E. V., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Krylov, D.M., Makarova, K.S., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., et al. (2004). A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol. *5*, R7.

Koreny, L., and Field, M.C. (2016). Ancient eukaryotic origin and evolutionary plasticity of nuclear lamina. Genome Biol. Evol. *8*, 2663–2671.

Kornberg, R.D. (1974). Chromatin structure: A repeating unit of histones and DNA. Science *184*, 868–871.

Koster, M.J.E., Snel, B., and Timmers, H.T.M. (2015). Genesis of chromatin and transcription dynamics in the origin of species. Cell *161*, 724–736.

Kumar, A., Sharma, P., Gomar-Alba, M., Shcheprova, Z., Daulny, A., Sanmartín, T., Matucci, I., Funaya, C., Beato, M., and Mendoza, M. (2018). Daughter-cell-specific modulation of nuclear pore complexes controls cell cycle entry during asymmetric division. Nat. Cell Biol. *20*, 432–442.

Kumar, S., Stecher, G., Suleski, M., and Hedges, S.B. (2017). TimeTree: A resource for timelines, timetrees, and divergence times. Mol. Biol. Evol. *34*, 1812–1819.

Kurat, C.F., Recht, J., Radovani, E., Durbic, T., Andrews, B., and Fillingham, J. (2014a). Regulation of histone gene transcription in yeast. Cell. Mol. Life Sci. *71*, 599–613.

Kurat, C.F., Lambert, J.-P., Petschnigg, J., Friesen, H., Pawson, T., Rosebrock, A., Gingras, A.-C., Fillingham, J., and Andrews, B. (2014b). Cell cycle-regulated oscillator coordinates core

histone gene transcription through histone acetylation. Proc. Natl. Acad. Sci. U. S. A. *111*, 14124–14129.

Kurat, C.F., Yeeles, J.T.P., Patel, H., Early, A., and Diffley, J.F.X. (2017). Chromatin controls DNA replication origin selection, lagging-strand synthesis, and replication fork rates. Mol. Cell *65*, 117–130.

Kushnirov, V. V. (2000). Rapid and reliable protein extraction from yeast. Yeast *16*, 857–860.

Lai, W.K.M., and Pugh, B.F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. Nat. Rev. Mol. Cell Biol. *18*, 548–562.

Lake, J.A., and Rivera, M.C. (1994). Was the nucleus the first endosymbiont? Proc. Natl. Acad. Sci. U. S. A *91*, 2880–2881.

Lake, J.A., Henderson, E., Clark, M.W., and Matheson, A.T. (1982). Mapping evolution with ribosome structure: Intralineage constancy and interlineage variation. Proc. Natl. Acad. Sci. U. S. A. *79*, 5948–5952.

Lande-Diner, L., Zhang, J., and Cedar, H. (2009). Shifts in replication timing actively affect histone acetylation during nucleosome reassembly. Mol. Cell *34*, 767–774.

Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics *25*, 2286–2288.

Laurell, E., Beck, K., Krupina, K., Theerthagiri, G., Bodenmiller, B., Horvath, P., Aebersold, R., Antonin, W., and Kutay, U. (2011). Phosphorylation of Nup98 by multiple kinases is crucial for NPC disassembly during mitotic entry. Cell *144*, 539–550.

Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral thinking: How histone modifications regulate gene expression. Trends Genet. *32*, 42–56.

Lawrence, M.B.D., Coutin, N., Choi, J.K., Martin, B.J.E., Irwin, N.A.T., Young, B., Loewen, C., and Howe, L.J. (2017). Histone acetylation, not stoichiometry, regulates linker histone binding in *Saccharomyces cerevisiae*. Genetics *207*, 347–355.

Leger, M.M., Eme, L., Stairs, C.W., and Roger, A.J. (2018). Demystifying eukaryote lateral gene

transfer. BioEssays 1700242.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Liang, Y., and Hetzer, M.W. (2011). Functional interactions between nucleoporins and chromatin. Curr. Opin. Cell Biol. *23*, 65–70.

Lin, S., Zhang, H., Zhuang, Y., Tran, B., and Gill, J. (2010). Spliced leader-based metatranscriptomic analyses lead to recognition of hidden genomic features in dinoflagellates. Proc. Natl. Acad. Sci. U. S. A. *107*, 20033–20038.

Lingle, W.L., and Salisbury, J.L. (1995). Ultrastructure of the parabasalid protist *Holomastigotoides*. J. Eukaryot. Microbiol. *42*, 490–505.

Lipford, J.R., and Bell, S.P. (2001). Nucleosomes positioned by ORC facilitate the initiation of DNA replication. Mol. Cell *7*, 21–30.

Liu, P., Kenney, J.M., Stiller, J.W., and Greenleaf, A.L. (2010). Genetic organization, length conservation, and evolution of RNA polymerase II carboxyl-terminal domain. Mol. Biol. Evol. *27*, 2628–2641.

Liu, Y., Richards, T.A., and Aves, S.J. (2009). Ancient diversification of eukaryotic MCM DNA replication proteins. BMC Evol. Biol. *9*, 60.

López-García, P., and Moreira, D. (2006). Selective forces for the origin of the eukaryotic nucleus. BioEssays *28*, 525–533.

Ludwig, M., and Gibbs, S.P. (1989). Evidence that the nucleomorphs of *Chlorarachnion reptans* (Chlorarachniophyceae) are vestigial nuclei: Morphology, division and DNA-DAPI fluorescence. J. Phycol. *25*, 385–394.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal

structure of the nucleosome core particle at 2.8 Å resolution. Nature *389*, 251–260.

Lynch, M., and Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. Science *290*, 1151–1155.

Lynch, M., and Conery, J.S. (2003). The origins of genome complexity. Science *302*, 1401–1404.

Lynch, M., Field, M.C., Goodson, H. V., Malik, H.S., Pereira-Leal, J.B., Roos, D.S., Turkewitz, A.P., and Sazer, S. (2014). Evolutionary cell biology: Two origins, one objective. Proc. Natl. Acad. Sci. U. S. A. *111*, 16990–16994.

Lynn, D.H. (2008). The ciliated protozoa: Characterization, classification, and guide to the literature: Third edition (New York: Springer Science & Business Media).

Lynn, D.H., and Strüder-Kypke, M. (2002). Phylogenetic position of *Licnophora*, *Lechriopyla*, and *Schizocaryum*, three unusual ciliates (phylum Ciliophora) endosymbiotic in echinoderms (phylum Echinodermata). J. Eukaryot. Microbiol. *49*, 460–468.

Ma, J., and Wang, M.D. (2016). DNA supercoiling during transcription. Biophys. Rev. *8*, 75–87.

Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M., and Van De Peer, Y. (2005). Modeling gene and genome duplications in eukaryotes. Proc. Natl. Acad. Sci. U. S. A. *102*, 5454–5459.

De Magistris, P., and Antonin, W. (2018). The dynamic nature of the nuclear envelope. Curr. Biol. *28*, R487–R497.

Makarova, M., and Oliferenko, S. (2016). Mixing and matching nuclear envelope remodeling and spindle assembly strategies in the evolution of mitosis. Curr. Opin. Cell Biol. *41*, 43–50.

Maltby, V.E., Martin, B.J.E., Brind'Amour, J., Chruscicki, A.T., McBurney, K.L., Schulze, J.M., Johnson, I.J., Hills, M., Hentrich, T., Kobor, M.S., et al. (2012). Histone H3K4 demethylation is negatively regulated by histone H3 acetylation in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. U. S. A. *109*, 18505–18510.

Margolin, W., and Bernander, R. (2004). How do prokaryotic cells cycle? Curr. Biol. *14*, 768–

770.

Margulis, L., Dolan, M.F., and Guerrero, R. (2000). The chimeric eukaryote: Origin of the nucleus from the karyomastigont in amitochondriate protists. Proc. Natl. Acad. Sci. U. S. A. *97*, 6954–6959.

Marinov, G.K., and Lynch, M. (2015). Diversity and divergence of dinoflagellate histone proteins. G3 *6*, 397–422.

Marinov, G.K., and Lynch, M. (2016). Conservation and divergence of the histone code in nucleomorphs. Biol. Direct *11*, 18.

Martin, W. (2005). Archaebacteria (Archaea) and the origin of the eukaryotic nucleus. Curr. Opin. Microbiol. *8*, 630–637.

Martin, W., and Koonin, E. V. (2006). Introns and the origin of nucleus-cytosol compartmentalization. Nature *440*, 41–45.

Martin, B.J.E., Mcburney, K.L., Maltby, V.E., Jensen, K.N., Brind'Amour, J., and Howe, L.J. (2017). Histone H3K4 and H3K36 methylation independently recruit the NuA3 histone acetyltransferase in *Saccharomyces cerevisiae*. Genetics *205*, 1113–1123.

Mathur, V., Kolísko, M., Hehenberger, E., Irwin, N.A.T., Leander, B.S., Kristmundsson, Á., Freeman, M.A., and Keeling, P.J. (2019). Multiple independent origins of apicomplexan-like parasites. Curr. Biol. *29*, 2936–2941.

Maurer-Alcalá, X.X., Yan, Y., Pilling, O.A., Knight, R., and Katz, L.A. (2018). Twisted tales: insights into genome diversity of ciliates using single-cell 'omics. Genome Biol. Evol. *10*, 1927–1939.

Mavrich, T.N. (2008). A barrier nucleosome model for statistical positioning of nucleosome throughout the yeast genome. Genome Res. *18*, 1073–1083.

McCarthy, J.E.G., and Gualerzi, C. (1990). Translation control of prokatyotic gene expression. Trends Genet. *6*, 78–85.

McFadden, G., and Gilson, P. (1995). Something borrowed, something green: Lateral transfer of

chloroplasts by secondary endosymbiosis. Trends Ecol. Evol. *10*, 12–17.

McGregor, A.L., Hsia, C.R., and Lammerding, J. (2016). Squish and squeeze - the nucleus as a physical barrier during migration in confined environments. Curr. Opin. Cell Biol. *40*, 32–40.

McNairn, A.J., and Gilbert, D.M. (2003). Epigenomic replication: Linking epigenetics to DNA replication. BioEssays *25*, 647–656.

Medina, E.M., Turner, J.J., Gordân, R., Skotheim, J.M., and Buchler, N.E. (2016). Punctuated evolution and transitional hybrid network in an ancestral cell cycle of fungi. Elife *5*, e09492.

Mendoza, S.D., Nieweglowska, E.S., Govindarajan, S., Leon, L.M., Berry, J.D., Tiwari, A., Chaikeeratisak, V., Pogliano, J., Agard, D.A., and Bondy-Denomy, J. (2020). A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases. Nature *577*, 244–248.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, J., Witman, G.B., Terry, A., Salamov, A., Fritz-laylin, L.K., Maréchal-drouard, L., et al. (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science *318*, 245–250.

Mewes, H.-W., Frishman, D., Güldener, U., Mannhaupt, G., Mayer, K., Mokrejs, M., Morgenstern, B., Münsterkötter, M., Rudd, S., and Weil, B. (2002). MIPS: A database for genomes and protein sequences. Nucleic Acid Res. *30*, 31–34.

Moore, C.E., and Archibald, J.M. (2009). Nucleomorph genomes. Annu. Rev. Genet. *43*, 251–264.

Mosalaganti, S., Kosinski, J., Albert, S., Schaffer, M., Strenkert, D., Salomé, P.A., Merchant, S.S., Plitzko, J.M., Baumeister, W., Engel, B.D., et al. (2018). In situ architecture of the algal nuclear pore complex. Nat. Commun. *9*, 2361.

Mullen, T.E., and Marzluff, W.F. (2008). Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5’ to 3’ and 3’ to 5’. Genes Dev. *22*, 50–65.

Müller, L.S.M., Cosentino, R.O., Förstner, K.U., Guizetti, J., Wedel, C., Kaplan, N., Janzen, C.J., Arampatzi, P., Vogel, J., Steinbiss, S., et al. (2018). Genome organization and DNA

accessibility control antigenic variation in trypanosomes. Nature *563*, 121–125.

Mumberg, D., Mulier, R., and Funk, M. (1994). Regulatable promoters of *Saccharomyces cerevisiae*: Comparison of transcriptional activity and their use for heterologous expression. Nucleic Acids Res. *22*, 5767–5768.

Murphy, R., Watkins, J.L., and Wente, S.R. (1996). GLE2, a *Saccharomyces cerevisiae* homologue of the *Schizosaccharomyces pombe* export factor RAE1, is required for nuclear pore complex structure and function. Mol. Biol. Cell *7*, 1921–1937.

Nagai, S., Davis, R.E., Mattei, P.J., Eagen, K.P., and Kornberg, R.D. (2017). Chromatin potentiates transcription. Proc. Natl. Acad. Sci. U. S. A. *114*, 1536–1541.

Neumann, N., Jeffares, D.C., and Poole, A.M. (2006). Outsourcing the nucleus: nuclear pore complex genes are no longer encoded in nucleomorph genomes. Evol. Bioinforma. *2*, 23–34.

Neumann, N., Lundin, D., and Poole, A.M. (2010). Comparative genomic evidence for a complete nuclear pore complex in the last eukaryotic common ancestor. PLoS One *5*, e13241.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.

Nock, A., Ascano, J.M., Barrero, M.J., and Malik, S. (2012). Mediator-regulated transcription through the +1 nucleosome. Mol. Cell *48*, 837–848.

Nowack, E.C.M., Price, D.C., Bhattacharya, D., Singer, A., Melkonian, M., and Grossman, A.R. (2016). Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of Paulinella chromatophora. Proc. Natl. Acad. Sci. U. S. A. *113*, 12214–12219.

Nozaki, H., Takano, H., Misumi, O., Terasawa, K., Matsuzaki, M., Maruyama, S., Nishida, K., Yagisawa, F., Yoshida, Y., Fujiwara, T., et al. (2007). A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga *Cyanidioschyzon merolae*. BMC Biol. *5*, 28.

Obado, S.O., Brillantes, M., Uryu, K., Zhang, W., Ketaren, N.E., Chait, B.T., Field, M.C., and

Rout, M.P. (2016). Interactome mapping reveals the evolutionary history of the nuclear pore complex. PLoS Biol. *14*, e1002365.

Okamoto, N., Horák, A., and Keeling, P.J. (2012). Description of two species of early branching dinoflagellates, *Psammosa pacifica* n. g., n. sp. and *P. atlantica* n. sp. PLoS One *7*, e34900.

Olins, A.L., and Olins, D.E. (1990). Identification of 10 nm non-chromatin filaments in the macronucleus of *Euplotes eurystomus*. Chromosoma *99*, 205–211.

Olins, D.E., and Olins, A.L. (1993). Inhibition of DNA synthesis in the macronuclear replication band of *Euplotes eurystomus*. J. Eukaryot. Microbiol. *40*, 459–467.

Olins, D.E., and Olins, A.L. (1994). The replication band of ciliated protozoa. Int. Rev. Cytol. *153*, 137–170.

Olins, A.L., Olins, D.E., Franke, W.W., Lipps, H.J., and Prescott, D.M. (1981). Stereo-electron microscopy of nuclear structure and replication in ciliated protozoa (Hypotricha). Eur. J. Cell Biol. *25*, 120–130.

Olins, D.E., Olins, A.L., Cacheiro, L.H., and Tan, E.M. (1989). Proliferating cell nuclear antigen/cyclin in the ciliate *Euplotes eurystomus*: Localization in the replication band and in micronuclei. J. Cell Biol. *109*, 1399–1410.

Olins, D.E., Olinsl, A.L., Herrmann, A., Lin, R., Allis, C.D., and Robert-Nicoud, M. (1991). Localization of acetylated histone H4 in the macronucleus of *Euplotes*. Chromosoma *100*, 377–385.

Onuma, R., Mishra, N., and Miyagishima, S.Y. (2017). Regulation of chloroplast and nucleomorph replication by the cell cycle in the cryptophyte *Guillardia theta*. Sci. Rep. *7*, 1–12.

Ori, A., Banterle, N., Iskar, M., Andrés-Pons, A., Escher, C., Khanh Bui, H., Sparks, L., Solis-Mezarino, V., Rinner, O., Bork, P., et al. (2013). Cell type-specific nuclear pores: A case in point for context-dependent stoichiometry of molecular machines. Mol. Syst. Biol. *9*, 648.

Paine, P.L., Moore, L.C., and Horowitz, S.B. (1975). Nuclear envelope permeability. Nature *254*, 109–114.

Palenik, B., Grimwood, J., Aerts, A., Rouze, P., Salamov, A., Putnam, N., Dupont, C., Jorgensen, R., Derelle, E., Rombauts, S., et al. (2007). The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. Proc. Natl. Acad. Sci. U. S. A. *104*, 7705–7710.

Pan, X., Bourland, W.A., and Song, W. (2013). Protargol synthesis: An in-house protocol. J. Eukaryot. Microbiol. *60*, 609–614.

Pappas, G.D. (1956). The fine structure of the nuclear envelope of *Amoeba proteus*. J. Biophys. Biochem. Cytol. *2*, 431–434.

Parisis, N., Krasinska, L., Harker, B., Urbach, S., Rossignol, M., Camasses, A., Dewar, J., Morin, N., and Fisher, D. (2017). Initiation of DNA replication requires actin dynamics and formin activity. EMBO J. *36*, 3212–3231.

Parky, D., Morrisy, A.R., Battenhouse, A., and Iyer, V.R. (2014). Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements. Nucleic Acids Res. *42*, 3736–3749.

Paschka, A.G., Jönsson, F., Maier, V., Möllenbeck, M., Paeschke, K., Postberg, J., Rupprecht, S., and Lipps, H.J. (2003). The use of RNAi to analyze gene function in spirotrichous ciliates. Eur. J. Protistol. *39*, 449–454.

Percipalle, P., and Visa, N. (2006). Molecular functions of nuclear actin in transcription. J. Cell Biol. *172*, 967–971.

Petersen, T.N., Brunak, S., Von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat. Methods *8*, 785–786.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.

Pirkl, N., Maier, K.C., Schenk, A., Failmezger, H., Tresch, A., and Cramer, P. (2013). Article global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. Mol. Cell *52*, 52–62.

Pittis, A.A., and Gabaldón, T. (2016). Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. Nature *531*, 101–104.

Postberg, J., Alexandrova, O., Cremer, T., and Lipps, H.J. (2005). Exploiting nuclear duality of ciliates to analyse topological requirements for DNA replication and transcription. J. Cell Sci. *118*, 3973–3983.

Pringle, J.R., Preston, R.A., Adams, A.E.M., Stearns, T., Drubin, D.G., Haarer, B.K., and Jones, E.W. (1989). Fluorescence microscopy methods for yeast. Methods Cell Biol. *31*, 357–435.

Pringsheim, E.G. (1964). Pure cultures of algae, their preparation and maintenance. (New York and London: Hafner Publishing Co).

Prioleau, M.-N., and MacAlpine, D.M. (2016). DNA replication origins - where do we begin? Genes Dev. *30*, 1683–1697.

Prochnik, S.E., Umen, J., Nedelcu, A.M., Hallmann, A., Miller, S.M., Nishii, I., Ferris, P., Kuo, A., Mitros, T., Fritz-laylin, L.K., et al. (2010). Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. Science *329*, 223–226.

Ptak, C., and Wozniak, R.W. (2016). Nucleoporins and chromatin metabolism. Curr. Opin. Cell Biol. *40*, 153–160.

Puerta, C., Hernandez, F., Gutierrez, C., Pineiro, M., Lopez-Alarcon, L., and Palacian, E. (1993). Efficient transcription of a DNA template associated with histone (H3·H4)2 tetramers. J. Biol. Chem. *268*, 26663–26667.

Pyrihová, E., Motyčková, A., Voleman, L., Wandyszewska, N., Fišer, R., Seydlová, G., Roger, A., Kolísko, M., and Doležal, P. (2018). A single Tim translocase in the mitosomes of *Giardia intestinalis* illustrates convergence of protein import machines in anaerobic eukaryotes. Genome Biol. Evol. *10*, 2813–2822.

Quang, L.S., Gascuel, O., and Lartillot, N. (2008). Empirical profile mixture models for phylogenetic reconstruction. Bioinformatics *24*, 2317–2323.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R.

(2005). InterProScan: Protein domains identifier. Nucleic Acids Res. *33*, W116–W120.

Raices, M., and D'Angelo, M.A. (2017). Nuclear pore complexes and regulation of gene expression. Curr. Opin. Cell Biol. *46*, 26–32.

Ramachandran, S., Zentner, G.E., and Henikoff, S. (2015). Asymmetric nucleosomes flank promoters in the budding yeast genome. Genome Res. *25*, 381–390.

Rambaut, A. (2012). FigTree, a graphical viewer of phylogenetic trees. See http://tree.bio.ed.ac.uk/Software/Figtree.

Rando, O.J., and Winston, F. (2012). Chromatin and transcription in yeast. Genetics *190*, 351–387.

Raoult, D., and Forterre, P. (2008). Redefining viruses: Lessons from Mimivirus. Nat. Rev. Microbiol. *6*, 315–319.

Reis, C.C., and Campbell, J.L. (2007). Contribution of Trf4/5 and the nuclear exosome to genome stability through regulation of histone mRNA levels in *Saccharomyces cerevisiae*. Genetics *175*, 993–1010.

Ren, Y., Seo, H.-S., Blobel, G., and Hoelz, A. (2010). Structural and functional analysis of the interaction between the nucleoporin Nup98 and the mRNA export factor Rae1. Proc. Natl. Acad. Sci. U. S. A. *107*, 10406–10411.

Rhind, N., and Gilbert, D.M. (2013). DNA replication timing. Cold Spring Harb. Perspect. Med. *5*, a010132.

Riaz, S., Niaz, Z., Khan, S., Liu, Y., and Sui, Z. (2019). Detection, characterization and expression dynamics of histone proteins in the dinoflagellate *Alexandrium pacificum* during growth regulation. Harmful Algae *87*, 101630.

Riley, J.L., and Katz, L.A. (2001). Widespread distribution of extensive genome fragmentation in ciliates. Mol. Biol. Evol. *18*, 1372–1377.

Rizzo, P.J., and Noodén, L.D. (1972). Chromosomal proteins in the dinoflagellate alga *Gyrodinium cohnii*. Science *176*, 796–797.

Robinson, M.D., Grigull, J., Mohammad, N., and Hughes, T.R. (2002). FunSpec: A web-based cluster interpreter for yeast. BMC Bioinformatics *3*, 35.

Rotterová, J., Salomaki, E., Pánek, T., Bourland, W., Žihala, D., Táborský, P., Edgcomb, V.P., Beinart, R.A., Kolísko, M., and Čepička, I. (2020). Genomics of new ciliate lineages provides insight into the evolution of obligate anaerobiosis. Curr. Biol. *30*, 2037-2050.e6.

Roure, B., Rodriguez-Ezpeleta, N., and Philippe, H. (2007). SCaFoS: A tool for selection, concatenation and fusion of sequences for phylogenomics. BMC Evol. Biol. *7*, 52.

Rowley, M.J., and Corces, V.G. (2016). The three-dimensional genome: Principles and roles of long-distance interactions. Curr. Opin. Cell Biol. *40*, 8–14.

Roy, S., and Morse, D. (2012). A full suite of histone and histone modifying genes are transcribed in the dinoflagellate *Lingulodinium*. PLoS One *7*, e34340.

Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T.C., Robins, A.J., Dalton, S., and Gilbert, D.M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Res. *20*, 761–770.

Sagulenko, E., Morgan, G.P., Webb, R.I., Yee, B., Lee, K.C., and Fuerst, J.A. (2014). Structural studies of planctomycete *Gemmata obscuriglobus* support cell compartmentalisation in a bacterium. PLoS One *9*, e91344.

Sagulenko, E., Nouwens, A., Webb, R.I., Green, K., Yee, B., Morgan, G., Leis, A., Lee, K.C., Butler, M.K., Chia, N., et al. (2017). Nuclear pore-like structures in a compartmentalized bacterium.

Sanjuán, R., Nebot, M.R., Chirico, N., Mansky, L.M., and Belshaw, R. (2010). Viral mutation rates. J. Virol. *84*, 9733–9748.

Sazer, S., Lynch, M., and Needleman, D. (2014). Deciphering the evolutionary history of open and closed mitosis. Curr. Biol. *24*, R1099–R1103.

Schwaiger, M., Kohler, H., Oakeley, E.J., Stadler, M.B., and Schübeler, D. (2010).

Heterochromatin protein 1 (HP1) modulates replication timing of the *Drosophila* genome. Genome Res. *20*, 771–780.

Schwelm, A., Fogelqvist, J., Knaust, A., Jülke, S., Lilja, T., Bonilla-Rosso, G., Karlsson, M., Shevchenko, A., Dhandapani, V., Choi, S.R., et al. (2015). The *Plasmodiophora brassicae* genome reveals insights in its life cycle and ancestry of chitin synthases. Sci. Rep. *5*, 11153.

Shin, M.K., Hwang, U.W., Kim, W., Wright, A.D.G., Krawczyk, C., and Lynn, D.H. (2000). Phylogenetic position of the ciliates *Phacodinium* (Order Phacodiniida) and *Protocruzia* (Subclass Protocruziidia) and systematics of the spirotrich ciliates examined by small subunit ribosomal RNA gene sequences. Eur. J. Protistol. *36*, 293–302.

Shiratori, T., Suzuki, S., Kakizawa, Y., and Ishida, K. ichiro (2019). Phagocytosis-like cell engulfment by a planctomycete bacterium. Nat. Commun. *10*, 5529.

Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., Takeuchi, T., Hisata, K., Tanaka, M., Fujiwara, M., et al. (2013). Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. Curr. Biol. *23*, 1399–1408.

Shutt, T.E., and Gray, M.W. (2006). Bacteriophage origins of mitochondrial replication and transcription proteins. Trends Genet. *22*, 90–95.

Siegel, T.N., Hekstra, D.R., Kemp, L.E., Figueiredo, L.M., Lowell, J.E., Fenyo, D., Wang, X., Dewell, S., and Cross, G.A.M. (2009). Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. Genes Dev. *23*, 1063–1076.

Sikorski, R.S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics *122*, 19–27.

da Silva-Neto, I.D., da Silva Paiva, T., Pedroso Dias, R.J., Alexandre Campos, C.J., and Migotto, A.E. (2012). Redescription of *Licnophora chattoni* Villeneuve-Brachon, 1939 (Ciliophora, Spirotrichea), associated with *Zyzzyzus warreni* Calder, 1988 (Cnidaria, Hydrozoa). Eur. J. Protistol. *48*, 48–62.

da Silva Neto, I.D. (1993). Structural and ultrastructural observations of the ciliate *Phacodinium metchnicoffi* Certes, 1891 (Heterotrichea, Phacodiniida). Eur. J. Protistol. *29*, 209–218.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E.M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212.

Sirri, V., Urcuqui-Inchima, S., Roussel, P., and Hernandez-Verdun, D. (2008). Nucleolus: The fascinating nuclear body. Histochem. Cell Biol. *129*, 13–31.

Solovei, I., Thanisch, K., and Feodorova, Y. (2016). How to rule the nucleus: Divide et impera. Curr. Opin. Cell Biol. *40*, 47–59.

Soyer-Gobillard, M.O., Gillet, B., Géraud, M.L., and Bhaud, Y. (1999). Dinoflagellate chromosome behaviour during stages of replication. Int. Microbiol. *2*, 93–102.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

Stephens, A.D., Banigan, E.J., and Marko, J.F. (2019a). Chromatin's physical properties shape the nucleus and its functions. Curr. Opin. Cell Biol. *58*, 76–84.

Stephens, A.D., Liu, P.Z., Kandula, V., Chen, H., Almassalha, L.M., Herman, C., Backman, V., O'Halloran, T., Adam, S.A., Goldman, R.D., et al. (2019b). Physicochemical mechanotransduction alters nuclear shape and mechanics via heterochromatin formation. Mol. Biol. Cell *30*, 2320–2330.

Sterner, D.E., Grant, P.A., Roberts, S.M., Duggan, L.J., Belotserkovskaya, R., Pacella, L.A., Winston, F., Workman, J.L., and Berger, S.L. (1999). Functional organization of the yeast SAGA Complex: Distinct components involved in structural integrity, nucleosome acetylation, and TATA-binding protein interaction. Mol. Cell. Biol. *19*, 86–98.

Stewart, M. (2007). Molecular mechanism of the nuclear protein import cycle. Nat. Rev. Mol. Cell Biol. *8*, 195–208.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41–45.

Strambio-De-Castillia, C., Niepel, M., and Rout, M.P. (2010). The nuclear pore complex:

bridging nuclear transport and gene regulation. Nat. Rev. Mol. Cell Biol. *11*, 490–501.

Sun, H.Q., Yamamoto, M., Mejillano, M., and Yin, H.L. (1999). Gelsolin, a multifunctional actin regulatory protein. J. Biol. Chem. *274*, 33179–33182.

Sun, J., Shi, Y., and Yildirim, E. (2019). The nuclear pore complex in cell type-specific chromatin structure and gene regulation. Trends Genet. *35*, 579–588.

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. PLoS One *6*, e21800.

Sutton, R.E., and Boothroyd, J.C. (1986). Evidence for trans splicing in trypanosomes. Cell *47*, 527–535.

Suzuki, S., Ishida, K.I., and Hirakawa, Y. (2016). Diurnal transcriptional regulation of endosymbiotically derived genes in the chlorarachniophyte *Bigelowiella natans*. Genome Biol. Evol. *8*, 2672–2682.

Svenson, S.B., and Karlström, O.H. (1976). Bacteriophage T4-induced shut-off of host-specific translation. J. Virol. *17*, 326–334.

Swart, E.C., Bracht, J.R., Magrini, V., Minx, P., Chen, X., Zhou, Y., Khurana, J.S., Goldman, A.D., Nowacki, M., Schotanus, K., et al. (2013). The *Oxytricha trifallax* macronuclear genome: A complex eukaryotic genome with 16,000 tiny chromosomes. PLoS Biol. *11*, e1001473.

Taddei, A., Van Houwe, G., Hediger, F., Kalck, V., Cubizolles, F., Schober, H., and Gasser, S.M. (2006). Nuclear pore association confers optimal expression levels for an inducible yeast gene. Nature *441*, 774–778.

Takemura, M. (2001). Poxviruses and the origin of the eukaryotic nucleus. J. Mol. Evol. *52*, 419–425.

Talbert, P.B., and Henikoff, S. (2012). Chromatin: Packaging without nucleosomes. Curr. Biol. *22*, R1040–R1043.

Talbert, P.B., Meers, M.P., and Henikoff, S. (2019). Old cogs, new tricks: the evolution of gene expression in a chromatin context. Nat. Rev. Genet. *20*, 283–297.

Taylor, F.J.R. (1976). Autogenous theories for the origin of eukaryotes. Taxon *25*, 377–390.

Tesar, M., and Marquardt, O. (1990). Foot-and-mouth disease virus protease 3C inhibits cellular transcription and mediates cleavage of histone H3. Virology *174*, 364–374.

Teves, S.S., and Henikoff, S. (2014). Transcription-generated torsional stress destabilizes nucleosomes. Nat. Struct. Mol. Biol. *21*, 88–94.

Tillo, D., and Hughes, T.R. (2009). G+C content dominates intrinsic nucleosome occupancy. BMC Bioinformatics *10*, 442.

Timmis, J.N., Ayliff, M.A., Huang, C.Y., and Martin, W. (2004). Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. *5*, 123–135.

Timmons, L., and Fire, A. (1998). Specific interference by ingested dsRNA. Nature *395*, 854.

Tong, A.H.Y., and Boone, C. (2006). Synthetic genetic array analysis in *Saccharomyces cerevisiae*. Methods Mol. Biol. *313*, 171–192.

Tong, A.H.Y., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Page, N., Robinson, M., Raghibizadeh, S., Andrews, B., Tyers, M., et al. (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science *294*, 2364–2368.

Tromer, E.C., van Hooff, J.J.E., Kops, G.J.P.L., and Snel, B. (2019). Mosaic origin of the eukaryotic kinetochore. Proc. Natl. Acad. Sci. U. S. A. *116*, 12873–12882.

Uniprot Consortium (2015). UniProt: A hub for protein information. Nucleic Acids Res. *43*, D204–D212.

Venkatesh, S., Smolle, M., Li, H., Gogol, M.M., Saint, M., Kumar, S., Natarajan, K., and Workman, J.L. (2012). Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. Nature *489*, 452–455.

Vergnolle, M.A.S., and Taylor, S.S. (2007). Cenp-F Links kinetochores to Ndel1/Nde1/Lis1/dynein microtubule motor complexes. Curr. Biol. *17*, 1173–1179.

Vogelauer, M., Rubbi, L., Lucas, I., Brewer, B.J., and Grunstein, M. (2002). Histone acetylation

regulates the time of replication origin firing. Mol. Cell *10*, 1223–1233.

Vosseberg, J., and Snel, B. (2017). Domestication of self-splicing introns during eukaryogenesis: The rise of the complex spliceosomal machinery. Biol. Direct *12*, 30.

Wang, Y., Sherrard, A., Zhao, B., Melak, M., Trautwein, J., Kleinschnitz, E.-M., Tsopoulidis, N., Fackler, O.T., Schwan, C., and Grosse, R. (2019). GPCR-induced calcium transients trigger nuclear actin assembly for chromatin dynamics. Nat. Commun. *10*, 5271.

Wedel, C., Förstner, K.U., Derr, R., and Siegel, T.N. (2017). GT-rich promoters can drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes. EMBO J. *36*, 2581–2594.

Wei, M., Fan, X., Ding, M., Li, R., Shao, S., Hou, Y., Meng, S., Tang, F., Li, C., and Sun, Y. (2020). Nuclear actin regulates inducible transcription by enhancing RNA polymerase II clustering. Sci. Adv. *6*, eaay6515.

Williams, E., Place, A., and Bachvaroff, T. (2017). Transcriptome analysis of core dinoflagellates reveals a universal bias towards "GC" rich codons. Mar. Drugs *15*, 125.

Williams, T.A., Cox, C.J., Foster, P.G., Szöllősi, G.J., and Embley, T.M. (2020). Phylogenomics provides robust support for a two-domains tree of life. Nat. Ecol. Evol. *4*, 138–147.

Wilson, K.L., and Dawson, S.C. (2011). Functional evolution of nuclear structure. J. Cell Biol. *195*, 171–181.

Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., et al. (1999). Functional characterization of the *S . cerevisiae* genome by gene deletion and parallel analysis. Science *285*, 901–907.

Wisecaver, J.H., and Hackett, J.D. (2011). Dinoflagellate genome evolution. Annu. Rev. Microbiol. *65*, 369–387.

Wunderlich, F. (1972). The macronuclear envelope of *Tetrahymena pyriformis* GL in different physiological states - V. Nuclear pore complexes - A controlling system in protein biosynthesis? J. Membr. Biol. *7*, 220–230.

Wyrick, J.J., Holstege, F.C., Jennings, E.G., Causton, H.C., Shore, D., Grunstein, M., Lander,

E.S., and Young, R.A. (1999). Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. Nature *402*, 418–421.

Xia, Y., Ivanovska, I.L., Zhu, K., Smith, L., Irianto, J., Pfeifer, C.R., Alvey, C.M., Ji, J., Liu, D., Cho, S., et al. (2018). Nuclear rupture at sites of high curvature compromises retention of DNA repair factors. J. Cell Biol. *217*, 3796–3808.

Yamagishi, T., Kurihara, A., and Kawai, H. (2015). A ribbon-like structure in the ejective organelle of the green microalga *Pyramimonas parkeae* (Prasinophyceae) consists of core histones and polymers containing N-acetyl-glucosamine. Protist *166*, 522–533.

Young, B.P., and Loewen, C.J.R. (2013). Balony: a software package for analysis of data generated by synthetic genetic array experiments. BMC Bioinformatics *14*, 354.

Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, Di., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., et al. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature *541*, 353–358.

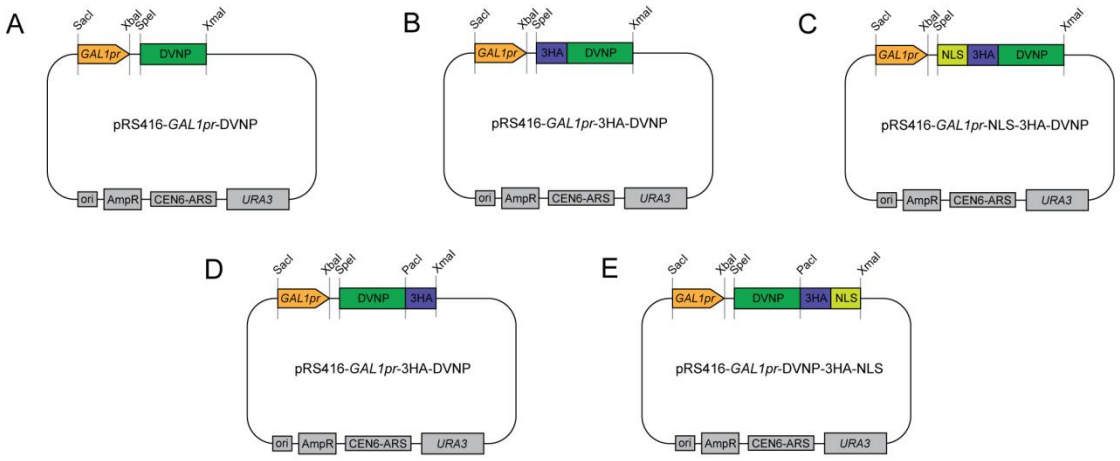
Zhang, H., Hou, Y., Miranda, L., Campbell, D.A., Sturm, N.R., Gaasterland, T., and Lin, S. (2007). Spliced leader RNA trans-splicing in dinoflagellates. Proc. Natl. Acad. Sci. U. S. A. *104*, 4618–4623.

Zhang, H., Campbell, D.A., Sturm, N.R., Dungan, C.F., and Lin, S. (2011). Spliced leader RNAs, mitochondrial gene frameshifts and multi-protein phylogeny expand support for the genus perkinsus as a unique group of alveolates. PLoS One *6*, e19933.

Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: A fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics *30*, 614–620.

Appendices

Appendix A: Supplementary information for Chapter 2



F

```
DVNP            M-----------------------------------AAMKKAMKVKKSAKKSAKKSGKKGGMKKKAKRVSKVARGKRAKSSVFRGTKERTSGGLTKNSLV
3HA-DVNP        M--------YPYDVPDYAYPYDVPDYAYPYDVPDYAAAMKKAMKVKKSAKKSAKKSGKKGGMKKKAKRVSKVARGKRAKSSVFRGTKERTSGGLTKNSLV
NLS-3HA-DVNP    MPPKKKRKVYPYDVPDYAYPYDVPDYAYPYDVPDYAAAMKKAMKVKKSAKKSAKKSGKKGGMKKKAKRVSKVARGKRAKSSVFRGTKERTSGGLTKNSLV
DVNP-3HA        M-----------------------------------AAMKKAMKVKKSAKKSAKKSGKKGGMKKKAKRVSKVARGKRAKSSVFRGTKERTSGGLTKNSLV
DVNP-3HA-NLS    M-----------------------------------AAMKKAMKVKKSAKKSAKKSGKKGGMKKKAKRVSKVARGKRAKSSVFRGTKERTSGGLTKNSLV

DVNP            KNKQGRVVSKKQSEHGKKIFKKHGLQKWIDAVTKARKALGIKGFQAVGGSSAKGKILLAKSRSFYKK--------------------------------
3HA-DVNP        KNKQGRVVSKKQSEHGKKIFKKHGLQKWIDAVTKARKALGIKGFQAVGGSSAKGKILLAKSRSFYKK--------------------------------
NLS-3HA-DVNP    KNKQGRVVSKKQSEHGKKIFKKHGLQKWIDAVTKARKALGIKGFQAVGGSSAKGKILLAKSRSFYKK--------------------------------
DVNP-3HA        KNKQGRVVSKKQSEHGKKIFKKHGLQKWIDAVTKARKALGIKGFQAVGGSSAKGKILLAKSRSFYKKLINIFYPYDVPDYAGYPYDVPDYAGSYPYDVPD
DVNP-3HA-NLS    KNKQGRVVSKKQSEHGKKIFKKHGLQKWIDAVTKARKALGIKGFQAVGGSSAKGKILLAKSRSFYKKLINIFYPYDVPDYAGYPYDVPDYAGSYPYDVPD

                                                                                                    Mw (kDa)
DVNP            -------------                                                                       14.4
3HA-DVNP        -------------                                                                       17.6
NLS-3HA-DVNP    -------------                                                                       18.6
DVNP-3HA        YAAQC--------                                                                       18.7
DVNP-3HA-NLS    YAAQCPPKKKRKV                                                                      19.7
```

Figure A1. DVNP expression vectors. A-E. Schematic representations of expression vectors used in this study. GAL1pr, GAL1 promoter; 3HA, three hemaglutinin epitope tags; NLS, nuclear localization signal; ori, bacterial origin of replication, AmpR, ampicillin resistance gene; CEN6-ARS, centromeric yeast autonomous replicating sequence. **F.** Amino acid sequences of all DVNP constructs used in this study. Molecular weights for each construct are shown. DVNP, 3HA tags, and NLS are shown in green, blue, and yellow, respectively.

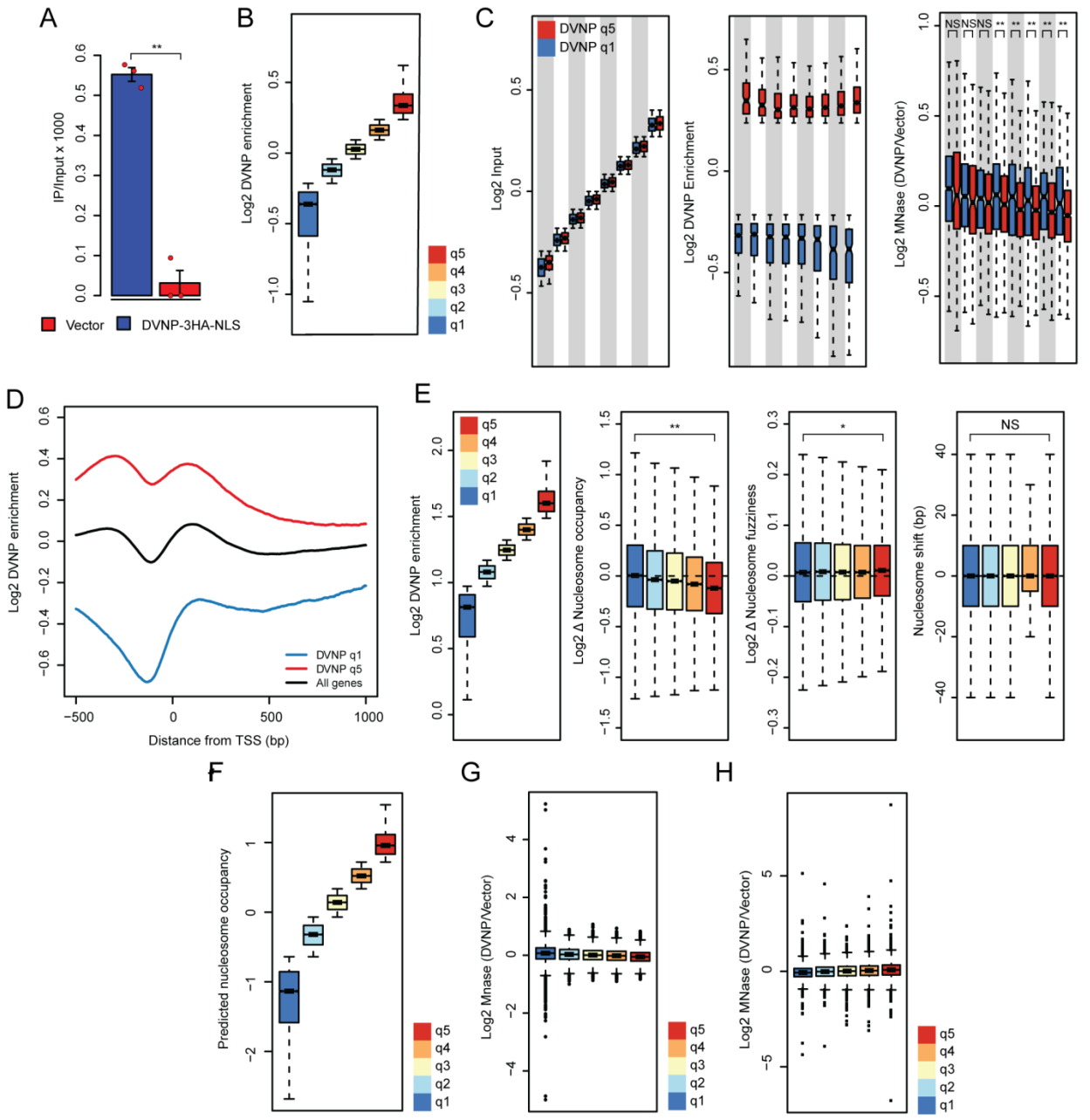


Figure A2. Supplemental information for DVNP ChIP and MNase sequencing. A. ChIP efficiency from DVNP-3HA-NLS expressing and vector control cells (*n* = 3 biological replicates). Error bars represent SEM. This experiment was repeated twice with the same results. **B.** DVNP enrichment within the 20% DVNP quintiles shown in Figure 5D. **C.** Genome wide association between DVNP enrichment and differential nucleosome occupancy in 10% input quintiles. The 2nd to 9th input quintiles (first panel) were subdivided into top and bottom 20% DVNP quintiles (q1,5, second panel) (n = 966). The ratio between DVNP-3HA-NLS and control

MNase occupancy (third panel) is shown in these bins. **D.** DVNP enrichment within the 20% DVNP quintiles shown in Figure 5D. **E.** DANPOS calculated changes in nucleosome occupancy (2nd panel), fuzziness (3rd panel), and position (4th panel), binned into 20% DVNP quintiles (q1-q5, 1st panel) (n = 14,557). **F.** Predicted nucleosome occupancy within the 20% quintiles shown in Figure 5E. **G-H.** Figure 5C, E showing outliers. *P*-values were obtained by two sided Welch's t-tests. Box plot wedges represent an estimate of the 95% confidence interval of the median. *, $P < 0.05$; **, $P < 10^{-3}$; NS, not significant.

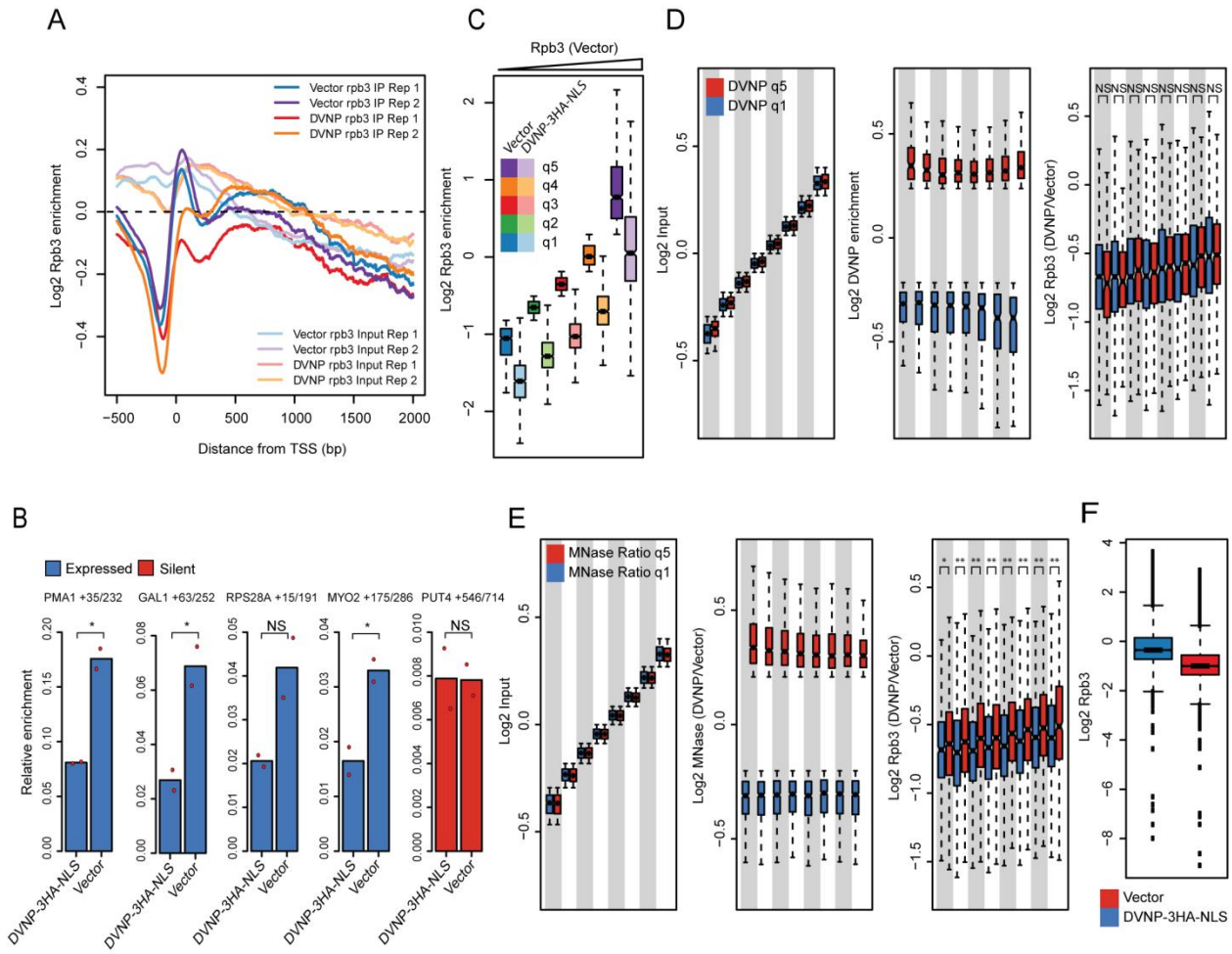


Figure A3. Supplemental information for Rpb3 ChIP sequencing. A. Average gene plot showing the relative enrichment of Rpb3 chromatin immunoprecipitates (ChIPs), and inputs from DVNP-3HA-NLS-expressing and control cells without adjusting for spike-in controls. Two biological replicates are shown for the ChIPs and inputs. **B.** Rpb3 ChIP-qPCR enrichment at five loci in DVNP3HA-NLS expressing and control cells (n = 2 biological replicates). Expressed and silent genes are shown in red and blue, respectively. Values were made relative by normalizing to an arbitrary input. **C.** Rpb3 loss and transcriptional state are not associated. Genes were binned into 20% vector Rpb3 ChIP quintiles to reflect transcriptional level. Rpb3 enrichment from DVNP-3HA-NLS-expressing (faded colours) and control cells (bold colours) are plotted in these bins. **D-E.** Genome wide association between DVNP enrichment, differential nucleosome occupancy, and Rpb3 loss in 10% input quintiles. The 2nd to 9th input quintiles (first panels) were subdivided into top and bottom 20% DVNP quintiles (D) or MNase ratio quintiles (E,

second panels). The change in Rpb3 is shown in these bins (third panels) (n = 966). **F.** Figure 6B showing outliers. *P*-values were obtained by two-sided Welch's t-tests. Box plot wedges represent an estimate of the 95% confidence interval of the median. *, $P < 0.05$; **, $P <$ 10-5 ; NS, not significant.

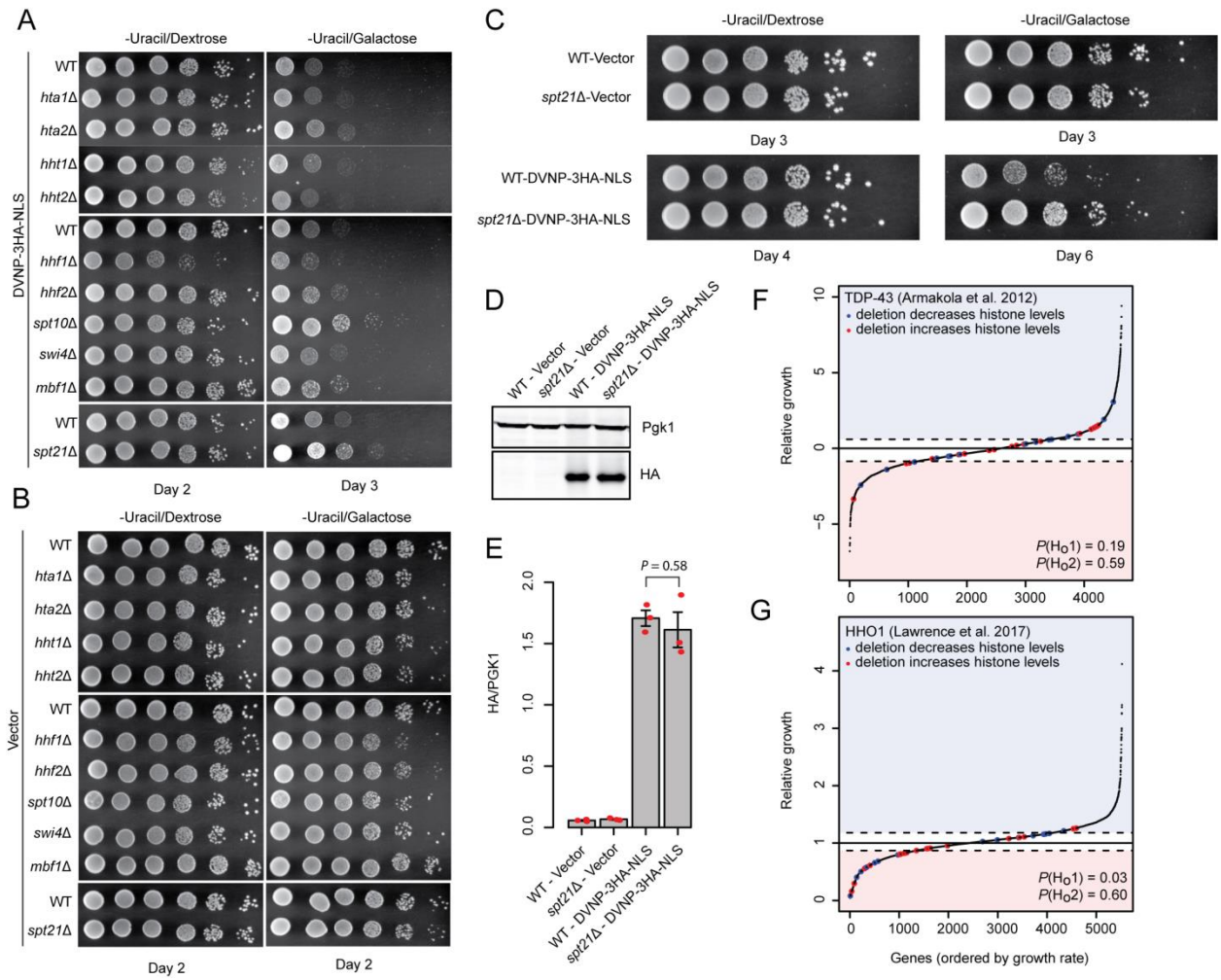


Figure A4. Supplemental information for the SGA. **A, B.** Serial dilution growth assays for DVNP-3HA-NLS expressing (A) and control cells (B). Histone and histone activator mutants from the SGA screen were obtained and freshly transformed with constructs prior to assessing growth. Cells were grown on selective media lacking uracil in the presence of either dextrose or galactose. This experiment was repeated twice with the same results. **C.** Serial dilution growth assays for DVNP-3HA-NLS expressing and control cells in a wildtype (WT) and newly generated spt21Δ mutant in a different strain background than the deletion collection. Cells were grown on selective media lacking uracil in the presence of either dextrose or galactose. This experiment was repeated three times with the same result. **D.** Immunoblot on total protein extracts following galactose induction. **E.** Quantified immunoblot data showing similar levels of HA (DVNP) in the WT and newly generated spt21Δ strain (n = 3 biological replicates). Error bars represent the SEM. The P-value was obtained by Welch's t-test. This experiment was

repeated twice with similar results. **F, G.** Previously published data showing the relative growth of ~4,500 non-essential deletion strains expressing TDP-43 (F) or HHO1 (G) from a GAL1 promoter. Gene deletions known to increase and decrease histone expression are shown in red and blue, respectively. Two null hypotheses could not be rejected by $\chi 2$ test: Ho1: gene deletions that affect histone levels are randomly distributed and, Ho2: gene deletions that increase and decrease histones are enriched below and above the growth thresholds by chance. Dashed lines denote positive and negative growth thresholds.

Table A1. Yeast strains used in Chapter 2

Strain name	Mating type	Genotype	Reference
FY602	Mat a	*his3 Δ200 leu2 Δ1 lys2-128δ ura3-52 trp1 Δ63*	(Sterner et al., 1999)
Y7092	Mat α	*can1Δ::STE2pr-Sp_his5 lyp1Δ his3 Δ1 leu2 Δ0 ura3 Δ0 met15 Δ0 trp1::NATMX*	(Tong and Boone, 2006)
BY4730	Mat a	*his3 Δ1 leu2 Δ0 met15 Δ0 ura3 Δ0*	(Winzeler et al., 1999)
*hta1*Δ	Mat a	*his3 Δ1 leu2 Δ0 met15 Δ0 ura3 Δ0 hta1::KANMX*	(Winzeler et al., 1999)
*hta2*Δ	Mat a	*his3 Δ1 leu2 Δ0 met15 Δ0 ura3 Δ0 hta2::KANMX*	(Winzeler et al., 1999)
*hht1*Δ	Mat a	*his3 Δ1 leu2 Δ0 met15 Δ0 ura3 Δ0 hht1::KANMX*	(Winzeler et al., 1999)
*hht2*Δ	Mat a	*his3 Δ1 leu2 Δ0 met15 Δ0 ura3 Δ0 hht2::KANMX*	(Winzeler et al., 1999)
*hhf1*Δ	Mat a	*his3 Δ1 leu2 Δ0 met15 Δ0 ura3 Δ0 hhf1::KANMX*	(Winzeler et al., 1999)
*hhf2*Δ	Mat a	*his3 Δ1 leu2 Δ0 met15 Δ0 ura3 Δ0 hhf2::KANMX*	(Winzeler et al., 1999)
*spt10*Δ	Mat a	*his3 Δ1 leu2 Δ0 met15 Δ0 ura3 Δ0 spt10::KANMX*	(Winzeler et al., 1999)
*spt21*Δ	Mat a	*his3 Δ1 leu2 Δ0 met15 Δ0 ura3 Δ0 spt21::KANMX*	(Winzeler et al., 1999)
*swi4*Δ	Mat a	*his3 Δ1 leu2 Δ0 met15 Δ0 ura3 Δ0 swi4::KANMX*	(Winzeler et al., 1999)
*mbf1*Δ	Mat a	*his3 Δ1 leu2 Δ0 met15 Δ0 ura3 Δ0 mbf1::KANMX*	(Winzeler et al., 1999)
FY602-*spt21*Δ	Mat a	*his3 Δ200 leu2 Δ1 lys2-128δ ura3-52 trp1 Δ63 spt21::HIS3MX6*	This thesis

Table A2. qPCR primers used in Chapter 2.

Primer Name	Sequence
PMA1 +85	CTTACGATGACGCTGCATC
PMA1 +232	CCTCTGGAACTGGTCTAGC
GAL1 +63	CCAAGACCATTGGCCGAAAAG
GAL1 +252	CAAAACTTTGACGGCGCAAAGC
RPS28A +15	CCAGTCACTTTAGCCAAGGTC
RPS28A +191	CGAGCTTCACGTTCAGATTCC
MYO2 +175	CTCTACCGCTTCTTAGAAACCCTC
MYO2 +286	GAGAATAGCGCTGTTTGATGGC
PUT4 +546	CTGGTCACTAGGTACGTTGAC
PUT4 +714	CACGCATAGAAAGATCGTGATCC

Appendix B: Supplementary information for Chapter 3

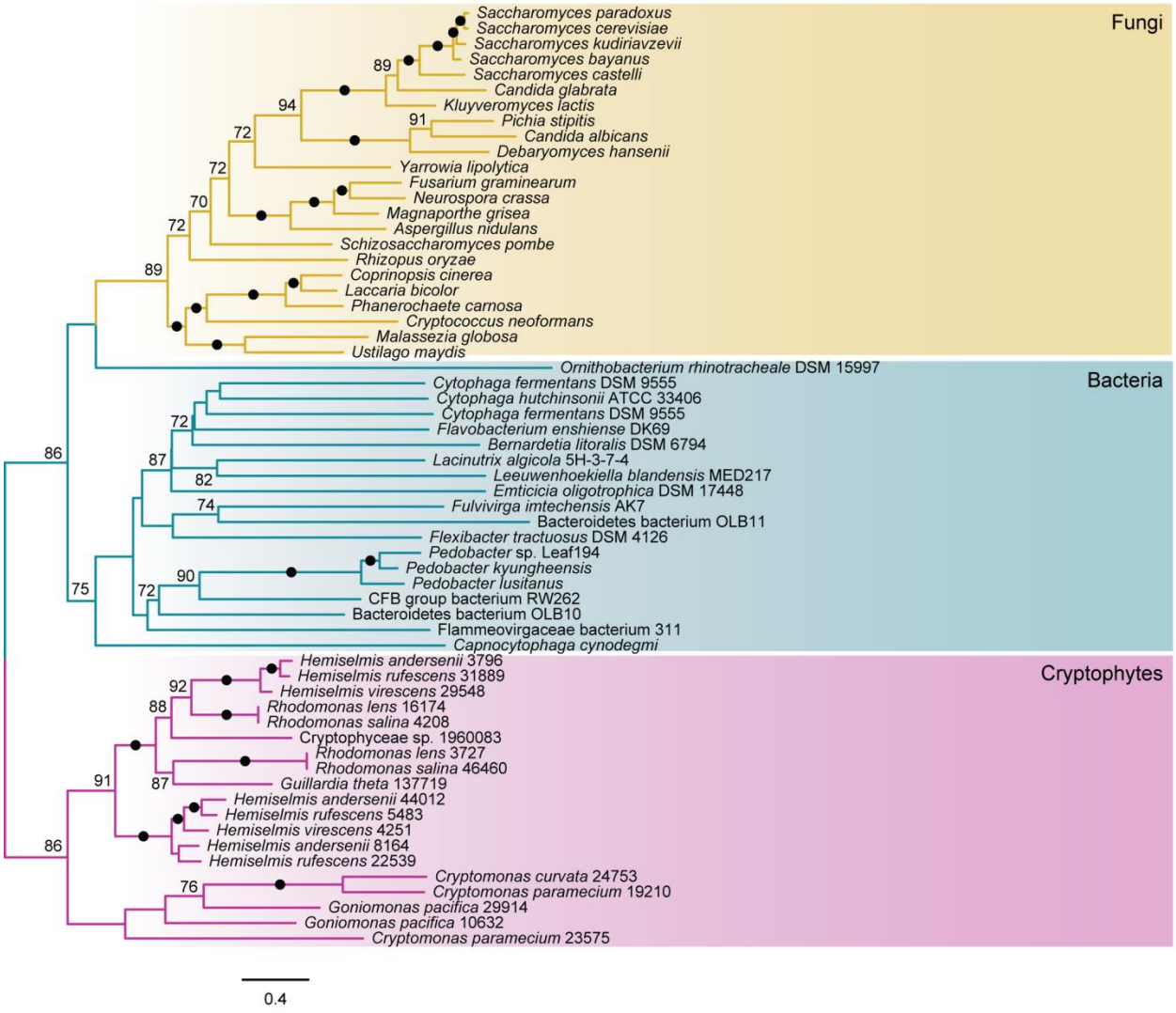


Figure B1. Phylogenetic analysis of Pom152. A maximum likelihood tree generated in IQ-Tree using the LG+F+R5 substitution model as selected by ModelFinder based on Bayesian Information Criteria. Statistical support was assessed using 1000 ultrafast bootstraps and values above 95 are represented as black circles whereas those below 70 are not shown. Fungal, bacterial, and cryptistan sequences have been highlighted in yellow, blue, and pink, respectively. The tree was rooted at the midpoint.

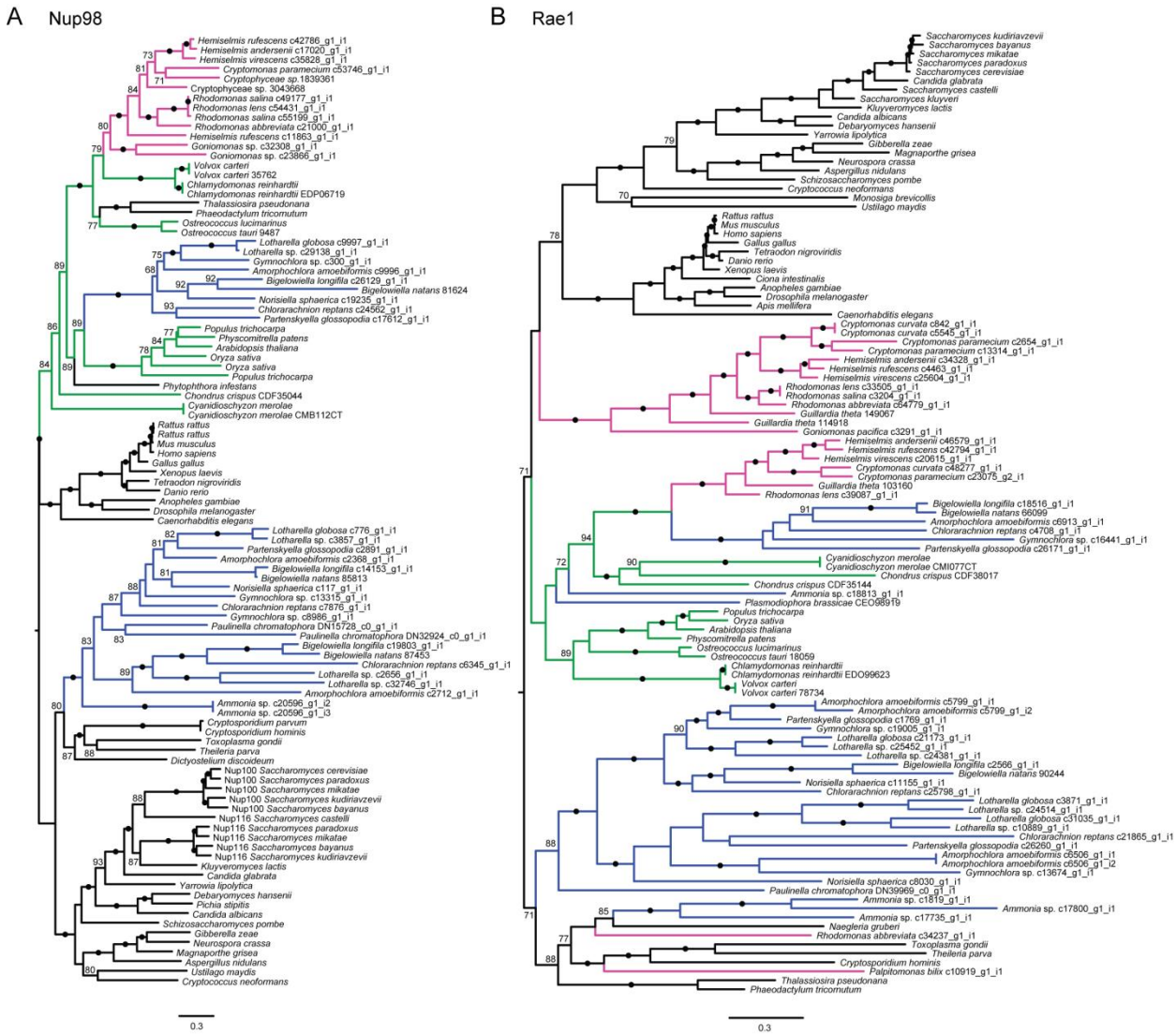


Figure B2. Fully annotated Nup98 and Rae1 phylogenies. The same maximum likelihood trees for Nup98 (**A**) and Rae1 (**B**) seen in Figure 11 but fully annotated with species names and protein prediction IDs for sequences identified in this study. Sequences lacking IDs are reference sequences. Values below 70 are not shown and black circles represent values equal to or greater than 95. Rhizarians, cryptistan, and algae and plants are labeled with blue, pink, and green, respectively.

Appendix C: Supplementary information for Chapter 4

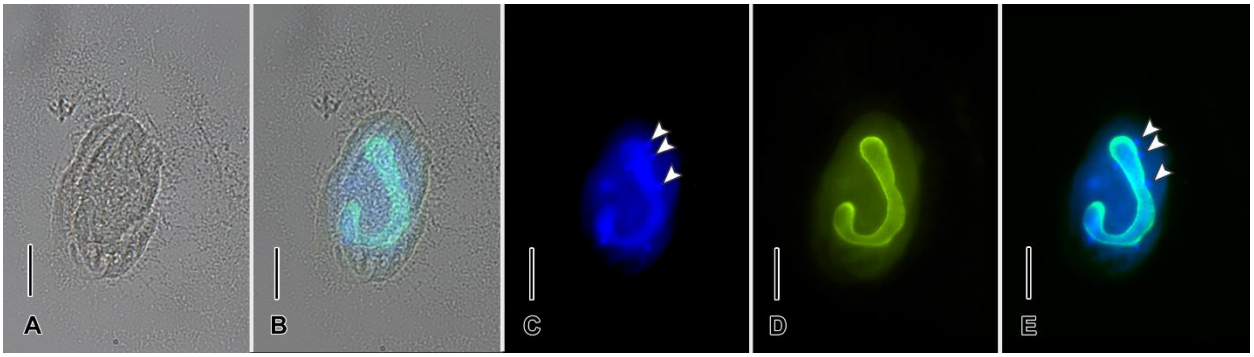


Figure C1. *Phacodinium metchnikoffi* in early S-phase.

Brightfield (A), brightfield image merged with Hoechst and Cy3-EdU signal (B), Hoechst (C), Cy3-EdU (D), and merged Hoechst and Cy3-EdU signals (E). Arrowheads (C, E) indicate three of the micronuclei. Micronuclei are not visible in (D) because they have not yet entered S-phase and therefore have not incorporated EdU. Scale bar: 25 µm.

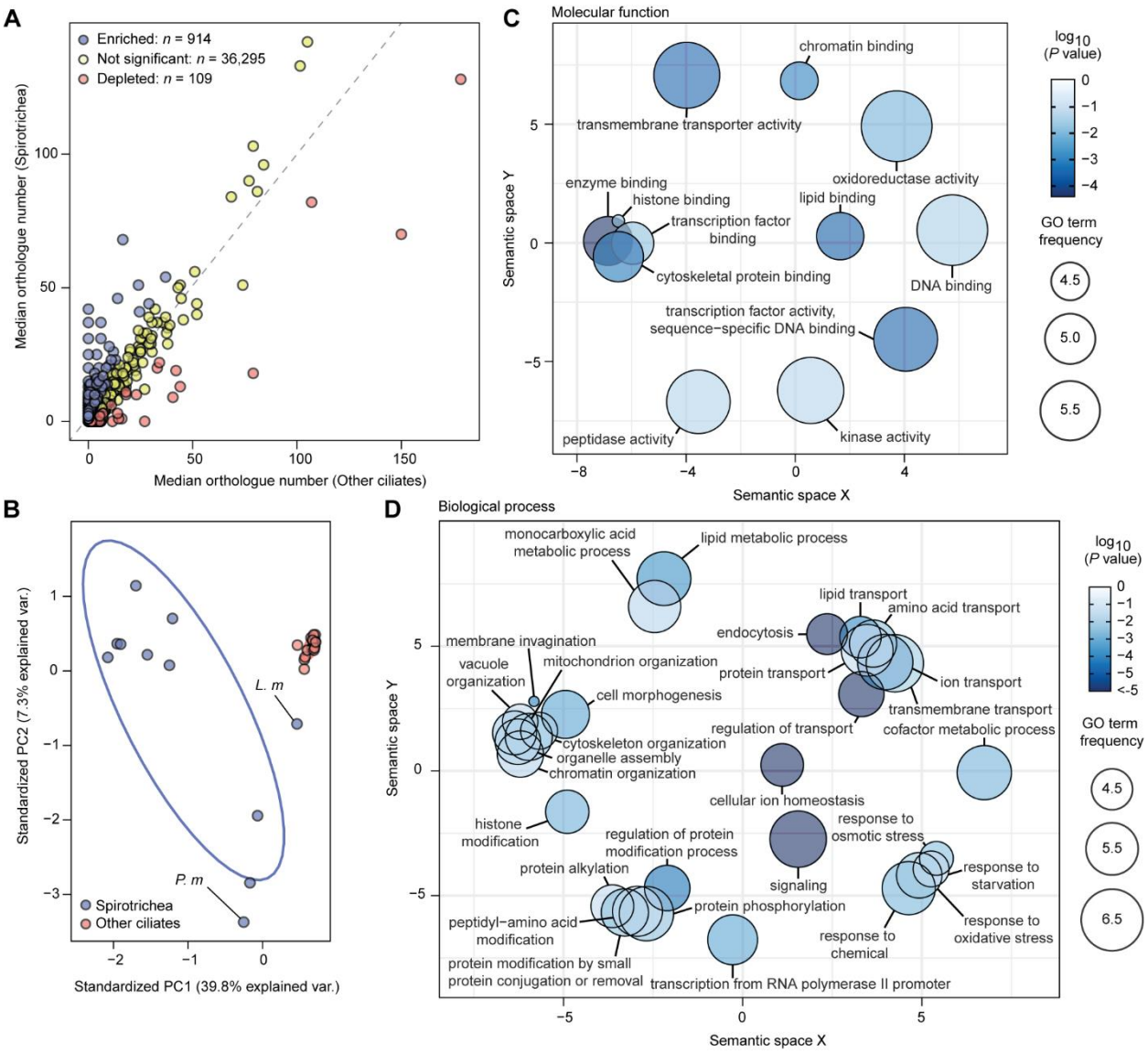


Figure C2. Spirotrich-associated protein families and their functional enrichments.

A. Scatter plot depicting the median number of homologues per protein family for the Spirotrichea and other ciliate groups. Significantly enriched and depleted families are denoted in blue and red, respectively ($P < 0.05$, permutation test, $n = 10{,}000$). **B.** Principal component analysis based on the number of homologues in enriched and depleted protein families for taxa included in the analysis. *P.m*, *Phacodinium metchnikoffi*; *L. m*, *Licnophora macfarlandi*. Counts were normalized by the mean count of the respective species for clarity. **C.** Semantic similarity-based scatter plot generated using REVIGO that displays significant molecular function gene ontology (GO) terms following summarization. Point colour and size are dependent on the *P*-

value (permutation test, n = 100,000) and the frequency of the GO-term within UniProt. **D.** The same scatter plot as in **C**, but displaying biological process GO-terms.

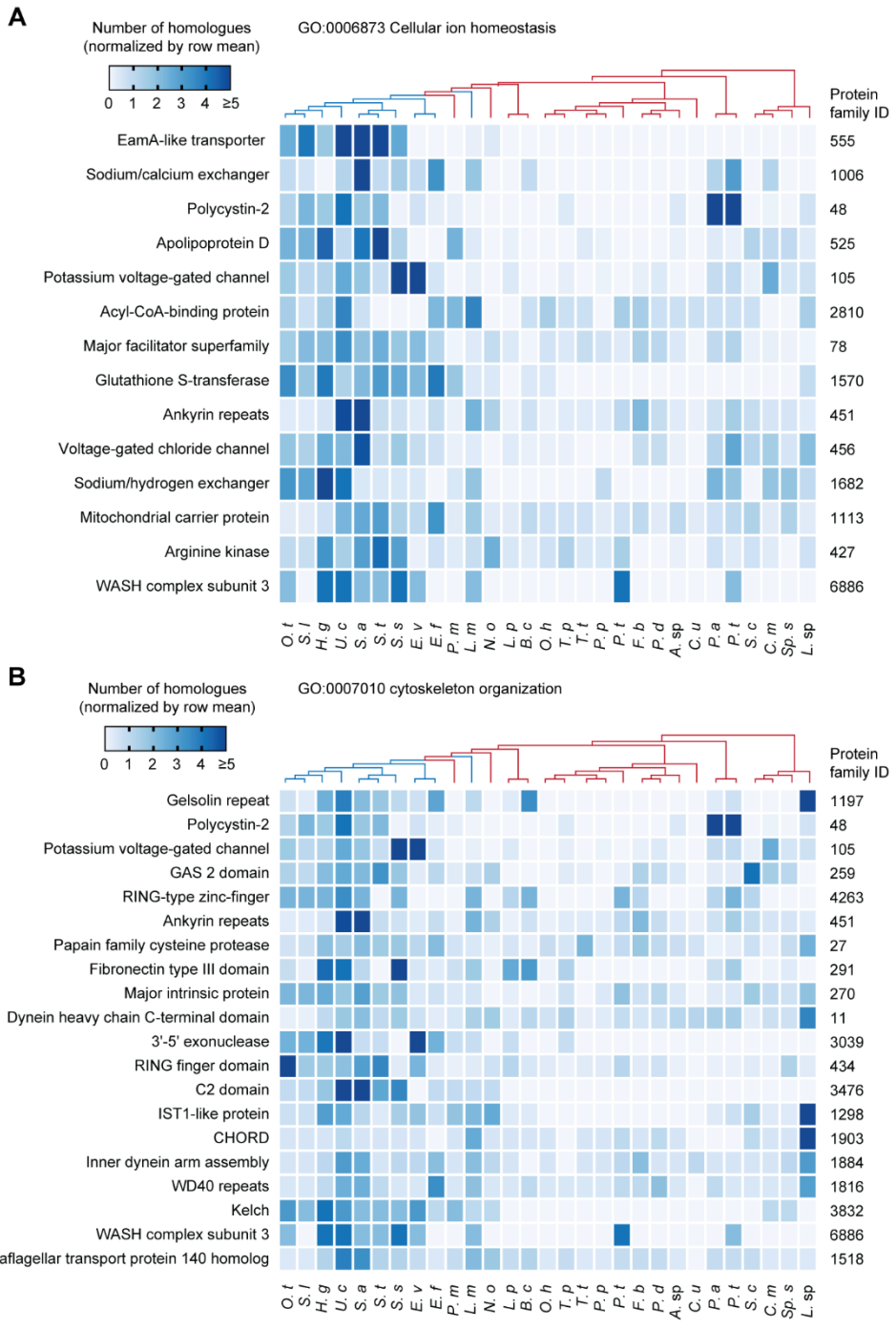


Figure C3. Replication band-associated protein families with functional enrichments in cellular ion homeostasis and cytoskeleton organization.

Heatmaps displaying the frequency of the RB-associated protein families ($P < 1x10^{-5}$, permutation test, $n = 10,000$) that were annotated as functioning in either cellular ion homeostasis (**A**, GO:0006873) or cytoskeleton organization (**B**, GO:0007010) across examined

taxa that had representation from both *Licnophora macfarlandi* and *Oxytricha trifallax*. The phylogeny at the top is based on Figure 14 and abbreviated species names are found at the bottom. Protein identifiers and functional annotations are listed on the right and left, respectively. Values have been normalized by the mean of each row. Note that families that were annotated as genes that function in these processes may not have been assigned these specific GO-terms due to divergence and therefore may not be included here (see Table C2).

Table C1. Datasets used for phylogenetic and comparative genomic analyses.

Species	Data type	Accession	Used in protein family analysis?
Oxytricha trifallax	Genome	NA	Yes
Sterkiella histriomuscorum	Genome	SRA000256	No
Stylonychia lemnae	Genome	NA	Yes
Halteria grandinella	Transcriptome	SRR5514623	Yes
Uroleptopsis citrina	Transcriptome	SRR7662950	Yes
Pseudokeronopsis sp. Brazil	Transcriptome	MMETSP1396	No
Pseudokeronopsis sp.	Transcriptome	MMETSP0211	No
Strombidinopsis acuminata	Transcriptome	MMETSP0126	Yes
Strombidinopsis sp.	Transcriptome	MMETSP0463	No
Schmidingerella taraikaensis	Transcriptome	MMETSP0436	Yes
Strombidium sulcatum	Transcriptome	SRR3985423	Yes
Strombidium inclinatum	Transcriptome	MMETSP0208	No
Euplotes vannus	Transcriptome	SRR7662949	Yes
Euplotes crassus	Transcriptome	MMETSP1380	No
Euplotes focardii	Transcriptome	MMETSP0205	Yes
Euplotes harpa	Transcriptome	MMETSP0213	No
Phacodinium metchnikoffi	Transcriptome	SRR12376981	Yes
Licnophora macfarlandi	Transcriptome	SRR12376982	Yes
Nyctotherus ovalis	Transcriptome	SRR10355984	Yes
Litonotus pictus	Transcriptome	MMETSP0209	Yes
Balantidium ctenopharyngodoni	Transcriptome	SRR5896119	Yes
Opistohnecta henneguyi	Transcriptome	NA	Yes
Vorticella microstoma	Transcriptome	NA	No
Scyphidia ubiquita	Transcriptome	NA	No
Trichodina pediculus	Transcriptome	NA	Yes
Urceolaria parakorschelti	Transcriptome	NA	No
Tetrahymena thermophila	Genome	NA	Yes
Icthyopthirius multifiliis	Genome	NA	No
Pseudocohnilembus persalinus	Transcriptome	SRR1768438	Yes

Paramecium tetraurelia	Genome	NA	Yes
Nassula variabilis	Transcriptome	SRR6754446	No
Furgasonia blochmanni	Transcriptome	SRR6754448	Yes
Pseudomicrothorax dubius	Transcriptome	SRR6754450	Yes
Platyophrya macrostoma	Transcriptome	MMETSP0127	No
Aristerostoma sp.	Transcriptome	MMETSP0125	Yes
Chilodonella uncinata	Transcriptome	PRJNA161607	Yes
Protocruzia adherens	Transcriptome	MMETSP0216	Yes
Protocruzia tuzeti	Transcriptome	SRR7662948	Yes
Stentor coeruleus	Genome	NA	Yes
Condylostoma magnum	Transcriptome	MMETSP0210	Yes
Spirostomum semivirescens	Transcriptome	SRR7141205	Yes
Loxodes sp.	Transcriptome	SRR10512999	Yes
Babesia bovis	Genome	NA	No
Theileria annulata	Genome	NA	No
Plasmodium falciparum	Genome	NA	No
Toxoplasma gondii	Genome	NA	No
Cryptosporidium muris	Genome	NA	No
Chromera velia	Genome	GCA_000585135.1	No
Vitrella brassicaformis	Genome	GCA_001179505.1	No
Alexandrium tamarense	Transcriptome	MMETSP0382	No
Polarella glacialis	Transcriptome	MMETSP0227	No
Karenia brevis	Transcriptome	MMETSP0027	No
Amphidinium carterae	Transcriptome	MMETSP0398	No
Perkinsus marinus	Genome	GCA_000006405.1	No

Table C2. Replication band-associated protein families.

Cluster[1]	Type	*P*-value[2]	Top Annotation[3]	%withTopAnnot	AnnotSeqs	Seqs
12334	enriched	0	Nuclear distribution protein nudE homolog 1	100.00%	5	8
1197	enriched	0	Gelsolin repeat	82.05%	78	83
1892	enriched	0	no annotation	0%	0	57
1006	enriched	0	Sodium/calcium exchanger protein	52.22%	90	96

1336	enriched	0	Methyltransferase domain	98.53%	68	76
3921	enriched	0	Alpha/beta hydrolase family	91.67%	12	28
3276	enriched	0	Dual 3',5'-cyclic-AMP and -GMP phosphodiesterase 11	66.67%	3	34
1120	enriched	0	Telomeric single stranded DNA binding POT1/CDC13	100.00%	55	88
1008	enriched	0	EF-hand domain pair	87.10%	31	96
2113	enriched	0	no annotation	0%	0	52
280	enriched	0	Ion transport protein	58.97%	78	252
555	enriched	0	EamA-like transporter family	98.02%	101	152
66	enriched	0	Transient receptor potential (TRP) ion channel	90.91%	11	859
48	enriched	0	Ion transport protein	89.01%	446	972
1767	enriched	0	Vesicle transport v-SNARE protein N-terminus	100.00%	12	60
525	enriched	0	Lipocalin-like domain	84.67%	150	158
105	enriched	0	Cyclic nucleotide-binding domain	64.13%	407	576
552	enriched	0	Transmembrane amino acid transporter protein	100.00%	123	153
450	enriched	0	Ankyrin repeats (3 copies)	67.06%	170	172
482	enriched	0	Ion transport protein	64.10%	39	166
1029	enriched	0	Ribosomal protein S5, C-terminal domain	60.00%	90	94
2810	enriched	0	Acyl CoA binding protein	100.00%	39	39
211	enriched	0	Furin-like repeat, cysteine-rich	100.00%	9	307
115	enriched	0	Cytosolic domain of 10TM putative phosphate transporter	61.36%	44	527
259	enriched	0	Growth-Arrest-Specific Protein 2 Domain	100.00%	176	265
78	enriched	0	Major Facilitator Superfamily	54.64%	474	696
4263	enriched	0.0001	Zinc finger, C3HC4 type (RING finger)	87.50%	24	25
2683	enriched	0.0001	Proprotein convertase subtilisin/kexin type 5	50.00%	4	41
3715	enriched	0.0001	Kelch motif	93.33%	15	29
2743	enriched	0.0001	Syntaxin-71	100.00%	13	40
1909	enriched	0.0001	Extracellular matrix protein FRAS1	42.31%	26	56
1570	enriched	0.0001	Glutathione S-transferase, N-terminal domain	74.07%	54	66
8479	enriched	0.0002	Tetratricopeptide repeat	80.00%	10	11
4606	enriched	0.0002	no annotation	0%	0	23
1891	enriched	0.0002	no annotation	0%	0	57

451	enriched	0.0002	Ankyrin repeats (3 copies)	63.80%	163	172
4111	enriched	0.0003	no annotation	0%	0	26
4121	enriched	0.0003	Inhibitor of apoptosis-promoting Bax1	100.00%	22	26
1537	enriched	0.0003	Extracellular matrix protein FRAS1	65.52%	29	67
27	enriched	0.0003	Papain family cysteine protease	66.11%	1307	1315
4424	enriched	0.0004	no annotation	0%	0	24
402	enriched	0.0005	Granulin	100.00%	1	189
3794	enriched	0.0006	no annotation	0%	0	29
291	enriched	0.0006	Fibronectin type III domain	80.36%	112	244
20128	enriched	0.0007	no annotation	0%	0	5
3562	enriched	0.0008	no annotation	0%	0	31
270	enriched	0.0008	Major intrinsic protein	99.58%	236	259
3315	enriched	0.001	Protein kinase domain	71.43%	7	33
3372	enriched	0.001	CCDC81 eukaryotic HU domain 2	100.00%	1	33
2081	enriched	0.001	Protein of unknown function (DUF2838)	97.56%	41	52
420	enriched	0.001	Golgi CORVET complex core vacuolar protein 8	78.00%	50	181
1382	enriched	0.0011	Activator of Hsp90 ATPase, N-terminal	100.00%	20	74
3206	enriched	0.0012	Serine/threonine-protein phosphatase 4 regulatory subunit 4	100.00%	17	34
16272	enriched	0.0013	Nuclear fragile X mental retardation-interacting protein 1	100.00%	6	6
6576	enriched	0.0013	no annotation	0%	0	15
3339	enriched	0.0013	no annotation	0%	0	33
1022	enriched	0.0014	Domain of unknown function (DUF3342)	80.95%	63	95
987	enriched	0.0014	no annotation	0%	0	97
13417	enriched	0.0015	Eukaryotic-type carbonic anhydrase	100.00%	1	7
5881	enriched	0.0017	Ferric reductase NAD binding domain	100.00%	1	17
11	enriched	0.0017	Dynein heavy chain C-terminal domain	15.45%	2634	3112
2858	enriched	0.0018	Thioredoxin-like	93.33%	15	39
2981	enriched	0.0018	Cyclic nucleotide-binding domain	100.00%	6	37
3039	enriched	0.0019	3'-5' exonuclease	96.15%	26	36
5863	enriched	0.0021	no annotation	0%	0	17
5578	enriched	0.0023	Glutathione peroxidase	83.33%	6	18
5612	enriched	0.0025	no annotation	0%	0	18

13857	enriched	0.0026	no annotation	0%	0	7
5103	enriched	0.0027	Coatomer epsilon subunit	100.00%	16	20
4947	enriched	0.003	VRR-NUC domain	93.33%	15	21
2586	enriched	0.0031	2OG-Fe(II) oxygenase superfamily	72.22%	36	43
5362	enriched	0.0032	no annotation	0%	0	19
12322	enriched	0.0034	no annotation	0%	0	8
4934	enriched	0.0034	no annotation	0%	0	21
11683	enriched	0.0035	ATPase family associated with various cellular activities (AAA)	100.00%	3	8
434	enriched	0.0036	Ring finger domain	78.02%	91	176
12063	enriched	0.0037	no annotation	0%	0	8
539	enriched	0.004	CHAT domain	32.00%	50	156
456	enriched	0.0045	Voltage gated chloride channel	78.06%	155	172
285	enriched	0.0048	Major Facilitator Superfamily	90.10%	192	247
4607	enriched	0.005	no annotation	0%	0	23
4563	enriched	0.005	no annotation	0%	0	23
10969	enriched	0.0053	no annotation	0%	0	9
10671	enriched	0.0056	no annotation	0%	0	9
4587	enriched	0.0056	no annotation	0%	0	23
10778	enriched	0.0059	Zinc finger, C3HC4 type (RING finger)	87.50%	8	9
2999	enriched	0.0059	Replication protein A C terminal	91.67%	12	37
1071	enriched	0.0059	Ribosomal protein L11, N-terminal domain	50.55%	91	91
2267	enriched	0.0063	EF-hand domain pair	89.19%	37	48
4257	enriched	0.0069	Colon cancer-associated protein Mic1-like	100.00%	20	25
4224	enriched	0.0077	Zinc finger, C3HC4 type (RING finger)	57.14%	7	25
333	enriched	0.0077	MAPEG family	99.48%	192	218
4115	enriched	0.0079	Phosphatidate cytidylyltransferase, mitochondrial	100.00%	16	26
1443	enriched	0.008	Protein of unknown function (DUF2475)	85.71%	7	71
4002	enriched	0.0085	PH domain	63.16%	19	27
9766	enriched	0.0088	no annotation	0%	0	10
2151	enriched	0.0088	no annotation	0%	0	51
819	enriched	0.0094	SNARE associated Golgi protein	100.00%	95	111
9759	enriched	0.0098	no annotation	0%	0	10
3676	enriched	0.0109	no annotation	0%	0	30
1560	enriched	0.0112	TMC domain	100.00%	17	66

8873	enriched	0.0119	no annotation	0%	0	11
8760	enriched	0.012	no annotation	0%	0	11
8922	enriched	0.0123	no annotation	0%	0	11
8765	enriched	0.0127	Tim10/DDP family zinc finger	100.00%	3	11
3476	enriched	0.0134	C2 domain	100.00%	5	32
1298	enriched	0.0136	Regulator of Vps4 activity in the MVB pathway	100.00%	63	78
267	enriched	0.0138	von Willebrand factor type A domain	47.14%	70	260
1039	enriched	0.0148	IPT/TIG domain	98.55%	69	93
1254	enriched	0.015	Tetratricopeptide repeat	61.11%	18	80
3077	enriched	0.0159	Sperm-tail PG-rich repeat	100.00%	5	36
1903	enriched	0.0161	CHORD	96.43%	56	56
1884	enriched	0.0164	Coiled-coil domain-containing protein 39	100.00%	47	57
8182	enriched	0.0167	Trypsin-like peptidase domain	100.00%	8	12
50757	enriched	0.0173	no annotation	0%	0	2
8137	enriched	0.0173	no annotation	0%	0	12
8102	enriched	0.0181	no annotation	0%	0	12
8103	enriched	0.0186	no annotation	0%	0	12
1176	enriched	0.0188	Enoyl-CoA hydratase/isomerase	100.00%	81	84
7877	enriched	0.0189	Protein tyrosine kinase	63.64%	11	12
8170	enriched	0.0195	no annotation	0%	0	12
1198	enriched	0.0196	Adenosine/AMP deaminase	95.83%	72	83
1816	enriched	0.0203	Region in Clathrin and VPS	36.36%	11	59
2188	enriched	0.0206	no annotation	0%	0	50
7536	enriched	0.0227	no annotation	0%	0	13
4116	enriched	0.023	Adenylate kinase	100.00%	4	26
7479	enriched	0.0235	no annotation	0%	0	13
7609	enriched	0.0235	Ribosome-binding factor A	100.00%	4	13
7515	enriched	0.0237	no annotation	0%	0	13
7506	enriched	0.0238	no annotation	0%	0	13
7660	enriched	0.0238	no annotation	0%	0	13
7438	enriched	0.0244	no annotation	0%	0	13
4047	enriched	0.0248	Cornichon protein	100.00%	22	27
7491	enriched	0.0251	COMM domain	90.91%	11	13
7547	enriched	0.0251	Sas10 C-terminal domain	58.33%	12	13
33970	enriched	0.0253	Tim10/DDP family zinc finger	100.00%	2	3
7448	enriched	0.0253	DnaJ domain	100.00%	6	13
7504	enriched	0.0253	no annotation	0%	0	13
7334	enriched	0.0258	Tim10/DDP family zinc finger	66.67%	3	13

7568	enriched	0.0258	Protein tyrosine phosphatase-like protein, PTPLA	100.00%	9	13
1682	enriched	0.0258	Sodium/hydrogen exchanger family	100.00%	25	63
7387	enriched	0.0259	Regulator of chromosome condensation (RCC1) repeat	100.00%	9	13
7498	enriched	0.0259	no annotation	0%	0	13
7499	enriched	0.026	no annotation	0%	0	13
1113	enriched	0.0263	Mitochondrial carrier protein	100.00%	82	88
7532	enriched	0.0267	no annotation	0%	0	13
4020	enriched	0.0272	no annotation	0%	0	27
3832	enriched	0.0277	Galactose oxidase, central domain	65.22%	23	28
33959	enriched	0.0279	no annotation	0%	0	3
6996	enriched	0.0295	no annotation	0%	0	14
427	enriched	0.0302	ATP:guanido phosphotransferase, C-terminal catalytic domain	59.26%	162	178
3702	enriched	0.0307	Kinesin motor domain	100.00%	6	29
6952	enriched	0.0311	Zinc finger, ZZ type	72.73%	11	14
736	enriched	0.0319	GCN5-like protein 1 (GCN5L1)	100.00%	1	121
7099	enriched	0.0323	Repeat domain in Vibrio, Colwellia, Bradyrhizobium and Shewanella	100.00%	1	14
6886	enriched	0.0325	Subunit CCDC53 of WASH complex	100.00%	13	14
6951	enriched	0.0326	Vps23 core domain	85.71%	14	14
6958	enriched	0.0331	no annotation	0%	0	14
6970	enriched	0.034	no annotation	0%	0	14
6984	enriched	0.0347	no annotation	0%	0	14
6924	enriched	0.0349	Signal recognition particle 14kD protein	100.00%	13	14
6882	enriched	0.035	SPRY domain	54.55%	11	14
6801	enriched	0.0351	Ankyrin repeats (3 copies)	84.62%	13	14
6961	enriched	0.0351	no annotation	0%	0	14
6922	enriched	0.036	Galactose oxidase, central domain	50.00%	2	14
6983	enriched	0.0373	no annotation	0%	0	14
23557	enriched	0.0374	SPOC domain	75.00%	4	4
3379	enriched	0.038	Mga helix-turn-helix domain	33.33%	3	33
24079	enriched	0.0381	no annotation	0%	0	4
3268	enriched	0.0397	Uncharacterised protein family (UPF0172)	100.00%	34	34
1518	enriched	0.04	Tetratricopeptide repeat	66.67%	9	68
968	enriched	0.041	Ribosomal L29 protein	100.00%	95	98

789	enriched	0.0418	CoA binding domain	51.55%	97	114
3209	enriched	0.0423	Carboxypeptidase regulatory-like domain	55.56%	9	34
1449	enriched	0.0434	YEATS family	56.06%	66	71
6399	enriched	0.0436	no annotation	0%	0	15
6532	enriched	0.0437	no annotation	0%	0	15
20149	enriched	0.0448	no annotation	0%	0	5
276	enriched	0.0451	FtsX-like permease family	100.00%	173	254
6581	enriched	0.0452	no annotation	0%	0	15
19014	enriched	0.0452	IBR domain, a half RING-finger domain	100.00%	1	5
1451	enriched	0.0452	WD domain, G-beta repeat	100.00%	18	71
6477	enriched	0.0454	Phosphatidylinositol N-acetylglucosaminyltransferase subunit Y	100.00%	13	15
19851	enriched	0.0464	no annotation	0%	0	5
20130	enriched	0.0479	no annotation	0%	0	5
657	enriched	0.0481	Homocysteine S-methyltransferase	32.08%	106	132
8	depleted	0	WD domain, G-beta repeat	91.21%	2911	3438
32	depleted	0	Zinc finger, C2H2 type	90.16%	1016	1144
195	depleted	0	AAA domain (dynein-related subfamily)	67.31%	52	330
34	depleted	0.0002	Myb-like DNA-binding domain	99.80%	1000	1088
905	depleted	0.0008	HAD-hyrolase-like	65.31%	98	103
111	depleted	0.0017	Armadillo/beta-catenin-like repeat	50.28%	360	555
755	depleted	0.0073	DENN (AEX-3) domain	71.05%	76	118
774	depleted	0.0075	PLD-like domain	55.81%	86	116
1136	depleted	0.0101	GAF domain	100.00%	1	87
609	depleted	0.0139	Zinc knuckle	66.92%	130	142
28	depleted	0.0172	Ion transport protein	41.74%	563	1303
361	depleted	0.0203	Cyclin	99.01%	203	206
94	depleted	0.025	Thioredoxin	77.55%	579	631
1146	depleted	0.0281	Protein kinase domain	98.00%	50	86
23	depleted	0.0282	ABC transporter	71.84%	1179	1398
400	depleted	0.0284	Sec7 domain	44.72%	161	189
1265	depleted	0.0318	Phosphatidylinositol-specific phospholipase C, Y domain	34.21%	76	80
554	depleted	0.037	Snf7	100.00%	145	153
219	depleted	0.0496	no annotation	0%	0	303

[1] Protein families in bold were tested with RNA interference in *Oxytricha trifallax*.

[2] *P*-values derived from permutation tests (n = 10,000)

Abbreviations: %withTopAnnot, percent of sequences with the top annotation; AnnotSeqs, number of annotated sequences; Seqs, total number of sequences in the orthogroup

[3] Annotations were determined using either EggNOG, SWISS-PROT, or PFAM databases.

Table C3. Gene ontology analysis of enriched replication band-associated protein families.

term ID	*P*-value[1]	Fold Enrichment	Test set count[2]	Background count[3]	Expected count[4]	GO class	Description
GO:0006873	0	3.72	14	328	3.76	biological process	cellular ion homeostasis
GO:0006970	0.00041	2.75	14	444	5.09	biological process	response to osmotic stress
GO:0006811	0.00065	2.16	20	808	9.26	biological process	ion transport
GO:0051049	0.0008	2.07	21	887	10.17	biological process	regulation of transport
GO:0055085	0.00267	2.01	18	780	8.94	biological process	transmembrane transport
GO:0023052	0.00392	1.46	37	2210	25.33	biological process	signaling
GO:0042221	0.00406	1.42	40	2456	28.15	biological process	response to chemical
GO:0007010	0.008	1.74	20	1004	11.51	biological process	cytoskeleton organization
GO:0009408	0.00891	2.37	10	368	4.22	biological process	response to heat
GO:0061025	0.01509	2.31	9	340	3.90	biological process	membrane fusion
GO:0016050	0.01696	2.27	9	346	3.97	biological process	vesicle organization
GO:0000902	0.018	1.84	14	664	7.61	biological process	cell morphogenesis
GO:0051604	0.01965	2.21	9	355	4.07	biological process	protein maturation
GO:0006979	0.02008	1.92	12	545	6.25	biological process	response to oxidative stress
GO:0015031	0.02069	1.50	24	1398	16.03	biological process	protein transport
GO:0070925	0.02382	1.68	16	829	9.50	biological process	organelle assembly

GO:0048284	0.04313	2.35	6	223	2.56	biological process	organelle fusion
GO:0016570	0.04975	1.87	9	420	4.81	biological process	histone modification
GO:0005576	1.00E-05	2.67	22	720	8.25	cellular component	extracellular region
GO:0031410	1.00E-05	2.18	28	1119	12.83	cellular component	cytoplasmic vesicle
GO:0012505	1.00E-05	1.74	40	2002	22.95	cellular component	endomembrane system
GO:0005783	0.00017	2.13	24	982	11.26	cellular component	endoplasmic reticulum
GO:0016020	0.00035	1.42	50	3062	35.10	cellular component	membrane
GO:0005737	0.00397	1.17	68	5087	58.31	cellular component	cytoplasm
GO:0005886	0.00951	1.37	39	2485	28.49	cellular component	plasma membrane
GO:0005794	0.01615	1.76	16	792	9.08	cellular component	Golgi apparatus
GO:0005856	0.02334	1.55	21	1183	13.56	cellular component	cytoskeleton
GO:0005773	0.03868	1.58	16	881	10.10	cellular component	vacuole
GO:0022857	0.00025	2.62	17	566	6.49	molecular function	transmembrane transporter activity
GO:0019899	0.00242	1.67	27	1412	16.19	molecular function	enzyme binding
GO:0016791	0.03228	2.30	7	266	3.05	molecular function	phosphatase activity
GO:0008092	0.03366	1.85	11	520	5.96	molecular function	cytoskeletal protein binding

[1] *P*-values calculated from permutation tests (*n* = 100,000).

[2] Number of enriched protein families (that contained representatives from both *Licnophora macfarlandi* and *Oxytricha trifallax*) annotated with the respective GO-term

[3] Total number of ciliate protein families annotated with the respective GO-term

[4] Expected test count based on random sampling of the background set