MASTERARBEIT | MASTER'S THESIS

Titel | Title

The functional diversification of nucleosomes across eukaryotes

verfasst von | submitted by

Darya Kripkiy Casado BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 220
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Joint-Masterstudium Evolutionary Genomics and Systems Biology
Betreut von | Supervisor:	Assoz. Prof. Dr. Silvia Bulgheresi Privatdoz.

Acknowledgements

First and foremost, I would like to express my gratitude for Dr. Nicholas Irwin for giving me the opportunity to work in his lab on this project. His mentorship allowed me to immensely grow on both scientific and personal level. I would like to thank him for the immense amount of time and effort that he put in advising me on experimental design, results interpretation and scientific presentation. I could not have wished for a more supportive mentor.

I would also like to thank Dr. Frédéric Berger for his mentorship and expertise in this project. I would also like to thank all the current and past members of the Irwin lab for their scientific and personal support. Furthermore, would like to express my gratitude to Assoz. Prof. Silvia Bulgheresi for her expertise and advise. I am thankful to her for taking time of her busy schedule to provide me guidance.

Lastly, I would like to thank my partner Marco Schindler for always being there for me when I need it.

Abstract

Nucleosomes are the fundamental unit of chromatin organization across eukaryotes. A nucleosome particle consists of two copies of each core histone - H2A, H2B, H3 and H4, which constitute a protein complex which wraps approximately 146 base pairs of DNA. Nucleosomes regulate DNA-dependent processes such as DNA replication and gene expression through functional dynamics influenced by post-translational modifications, chromatin remodelers, chaperones and histone variants. Given their important functionality, histones are some of the most conserved eukaryotic proteins. This conservation has led to the assumption that the functional diversification of chromatin across eukaryotes results primarily from variation in chromatin modification machinery rather than changes to the intrinsic properties of nucleosomes themselves. However, to our knowledge, no comprehensive comparative functional studies of nucleosomes across eukaryotes have been conducted. To investigate the functional diversity of the nucleosome, we applied a novel synthetic method that produces assembled nucleosomes in the bacterium *Escherichia coli* through ectopic histone expression. This synthetic model provides a tractable and high-throughput chassis for comparing intrinsic nucleosome properties from diverse eukaryotes. We expressed histones from 13 different eukaryotes from across the eukaryotic tree of life to generate synthetic nucleosomes and identified that most can form nucleosome-like structures which exhibit both comparable and distinct characteristics from one another. Using MNase-seq, we profiled nucleosomes across the *E. coli* genome and found that nucleosomes formed by histones from diverse eukaryotes exhibit diverse binding properties. This work provides crucial insights into understanding the role of histone sequence diversity in chromatin evolution. Additionally, here we further expand and validate the novel method of ectopic expression of self-assembling nucleosomes, which is a powerful tool for functional chromatin studies.

Zusammenfassung

Nukleosomen sind die grundlegende Einheit das Chromatin in Eukaryoten. Ein Nukleosom besteht aus zwei Kopien jedes Kernhistons – H2A, H2B, H3 und H4. Die bilden einen Proteinkomplex, der etwa 146 Basenpaare der DNA umhüllt. Nukleosomen regulieren DNA-abhängige Prozesse wie die DNA-Replikation und die Genexpression durch funktionelle Dynamiken, die durch posttranslationale Modifikationen, Chromatin-Remodeler, Chaperone und Histonvarianten beeinflusst werden. Aufgrund ihrer wichtigen Funktionalität gehören Histone zu den am besten konservierten eukaryotischen Proteinen. Diese Konservierung hat zu der Annahme geführt, dass die funktionelle Diversifizierung des Chromatins in Eukaryoten findet hauptsächlich durch Variation im Chromatin-Modifikationsmechanismus statt und nicht durch Veränderungen der intrinsischen Eigenschaften der Nukleosomen. Unseres Wissens wurden jedoch noch keine umfassenden vergleichenden Funktionsstudien zu Nukleosomen bei Eukaryoten durchgeführt. Um die funktionelle Vielfalt des Nukleosoms zu untersuchen, haben wir eine neuartige synthetische Methode angewendet, mit der durch ektopische Histon-Expression in der Bakterium *Escherichia coli* assemblierte Nukleosomen hergestellt werden. Dieses synthetische Modell bietet eine handhabbare und hochdurchsatzfähige Plattform für den Vergleich der intrinsischen Eigenschaften von Nukleosomen aus verschiedenen Eukaryoten. Wir haben Histone aus 13 verschiedenen Eukaryoten aus dem gesamten eukaryotischen Stammbaum exprimiert, um synthetische Nukleosomen zu erzeugen, und festgestellt, dass die meisten davon Nukleosom ähnliche Strukturen bilden können, die sowohl vergleichbare als auch unterschiedliche Eigenschaften aufweisen. Mithilfe von MNase-seq haben wir Nukleosomen im gesamten *E. coli*-Genom profiliert und festgestellt, dass Nukleosomen, die aus Histonen verschiedener Eukaryoten gebildet wurden, unterschiedliche Bindungseigenschaften aufweisen. Diese Arbeit liefert wichtige Erkenntnisse zum Verständnis der Rolle der Histonsequenzvielfalt in der Chromatin-Evolution. Darüber hinaus erweitern und validieren wir hier die neuartige Methode der ektopischen Expression selbstorganisierender Nukleosomen, die ein leistungsfähiges Werkzeug für funktionelle Chromatin-Studien darstellt.

Abbreviations

ANOVA	Analysis of Variance
ARS	Autonomous replicating sequence
Asf1	Anti-silencing function 1
Chl	Chloramphenicol
Cryo-EM	Cryo-electron microscopy
DNA	Deoxyribonucleic acid
IPOD-HR	*In vivo* protein occupancy display – high resolution
GFP	Green fluorescent protein
H	hour
H1	Histone H1
H2A	Histone H2A
H2B	Histone H2B
H3	Histone H3
H4	Histone H4
HMM	Hidden Markov models
HmfA	Histone *Methanothermus fervidus* A
HmfB	Histone *Methanothermus fervidus B*
LB	Lysogeny broth
LECA	Last eukaryotic common ancestor
μL	Microliter
mL	Milliliter
min	Minute
MNase	Micrococcal nuclease
MNase-seq	Micrococcal nuclease sequencing
NAP	Nucleoid-associated protein
NCP	Nucleosome core particle
PTM	Post-translational modification
RBS	Ribosome binding site
RNA-seq	Ribonucleic acid sequencing
s	Seconds
Snf2	Sucrose non-fermenting 2

SOC	Super Optimal broth with Catabolite repression
PCR	Polymerase chain reaction
OD_{600}	Optical density at 600 nm
TukeyHSD	Tukey's Honestly Significant Difference
xg	Times gravity

Table of contents

1 Introduction

1.1 Nucleosome, a fundamental unit of chromatin

A hallmark of eukaryotic organisms is the organization of their genomes into chromatin – a structure formed between nucleic acids and proteins that organizes and compacts DNA inside the cell nucleus. A core unit of chromatin is the nucleosome, a complex between ~146 base pairs of DNA and two pairs of each of the core histones H2A, H2B, H3 and H4, each of which is a small highly basic protein. Over the genome, nucleosomes are flanked by stretches of unbound DNA with variable lengths that can associate with linker histone H1, which is structurally and evolutionary distinct from core histones (Kasinsky et al., 2001). Each of the core histones possesses a characteristic histone fold domain, comprising three alpha helices connected by two loops that enable them to dimerize in an antiparallel way. Additionally, H2A-H2B dimers form an acidic patch, which serves as an interface for interaction with other proteins (Luger et al., 1997; McGinty & Tan, 2016). It has been shown that nucleosome assembly occurs sequentially in a chaperone-mediated manner via a replication-dependent or replication-independent pathway (Ahmad & Henikoff, 2002). Histones H3 and H4 form a complex with cytosolic chaperone Asf1, which promotes their import into nucleus, where two H3-H4 heterodimers tetramerize in the presence of DNA. This tetramer is required for the following incorporation of two H2A-H2B heterodimers (Brown & McMurray, 2025; J. Y. Lee et al., 2015). Similar stepwise assembly process occurs during *in vitro* nucleosome reconstitution, where equimolar ratios of purified core histones form a complex with DNA following salt dialysis. In this case, however, no chaperons are required (Flaus, 2011).

An important characteristic of the nucleosome is its position on the DNA strand which is determined by intrinsic sequence preferences and additional factors which define its position. There is consensus that nucleosome positioning is highly dependent on the GC-richness of a given DNA sequence, with some models indicating that the GC content alone can account for 50% of nucleosome positioning *in vivo* (Peckham et al., 2007; Segal et al., 2006). The most robust predictor of nucleosome positioning is the 10-base-pair periodic enrichment of the dinucleotides GC, TA, TT and AA, which has been observed by various studies (Segal et al., 2006; Van Der Heijden et al., 2012). However, the sequence preference of nucleosomes *in vivo* appears to be less well defined compared to what is observed in *vitro*. Nucleosome positioning models based on *in vitro* data show a pronounced preference for dinucleotide (AA/AT/TA/TT and GG/GC/CG/CC) periodicity (Kaplan et al., 2009), which establishes this preference as an intrinsic nucleosome property. While nucleosome positioning i*n vivo* still retains this general periodicity

pattern, it is less consistent (Kaplan et al., 2009; Zhang et al., 2009). The decreased importance of intrinsic nucleosome sequence preference for positioning *in vivo* is attributed to the fact that in the eukaryotic cell the positioning is also influenced by chromatin remodelers, which will guide nucleosomes to specific position as well as nuclear factors such as DNA binding proteins (e.g. RNA polymerase II) that can prevent the nucleosome from binding at specific locations (Y. Zhang et al., 2009).

Positioning of nucleosomes is important because it allows them to perform functions like DNA compaction into nucleus, structural organization of chromatin and gene expression (Li et al., 2007). For instance, in yeast, DNA replication occurs at autonomous replicating sequences (ARS), which are nucleosome-free. Additionally, ARS are flanked by well-positioned nucleosomes containing the histone variant H2A.Z (Eaton et al., 2010). While the *S. cerevisiae* genome contains multiple ARS consensus sequences, not all of them are functionally active. Strong associations between precise nucleosome positioning and origins of replication suggests that nucleosome positioning may be involved in the selection replication origins (Bell & Labib, 2016; Eaton et al., 2010). Another important function of nucleosome is transcriptional regulation. Generally, actively transcribed genes are depleted of nucleosomes, and RNA Polymerase II preferentially stalls near nucleosome-bound regions of DNA (Kornberg & Lorch, 1999, 2020). Additionally, several amino acid residues on histone tails can be subject to post-translational modifications (PTMs), which provide additional layer of transcriptional regulation: for example, acetylation is usually associated with active transcription (Li et al., 2007).

1.2 Histones are among the most conserved proteins across eukaryotes

Given the crucial role of the nucleosome in various cellular processes, it is unsurprising that the remarkable evolutionary conservation of histones was noticed long ago, with H4 hypothesized to have nearly-perfect conservation across eukaryotic diversity (DeLange et al., 1969; Thatcher & Gorovsky, 1994). To this day not a single species across eukaryotic diversity is known to have lost either one of the four core histones (Soo & Warnecke, 2021). All of the core histones, including histone variants have conserved their signature histone fold motif, which is an indication of the conservation of their key architecture and functionality (Hocher & Warnecke, 2024). In fact, eukaryotic histones likely have archaeal ancestry, as most archaeal groups contain histones that stack into hypernucleosomes as opposed to the individually spaced nucleosomes observed in eukaryotes (Maruyama et al., 2013). There is evidence that the four core histone proteins were present in the last eukaryotic common ancestor (LECA) (Hocher & Warnecke, 2024). Phylogeny of the histone H3 shows its high degree of conservation, which is especially true for the variant

H3.3 – the difference between *Arabidopsis thaliana* (bikonta) and *Homo sapiens* (unikonta) is only 9 amino acids, which corresponds to 93.3% sequence identity. H3.3 is therefore considered to be ancestral to eukaryotes. Most other H3 histone variants with the exception of CenH3 have derived from H3.3 (Postberg et al., 2010). Like H3, other core histones have retained a similar degree of conservation. While H2A has undergone more diversification and split into multiple lineage-specific specialized variants like animal macroH2A associated with inactive X chromosome compaction (Osakabe & Molaro, 2023), its ancient variants H2A.Z and core H2A are both present in most eukaryotic lineages (Zlatanova & Thakar, 2008).

Despite the high degree of conservation across histones, there are known cases of core histone divergence in diverse eukaryotes. One of the most striking examples are trypanosomes, a group of unicellular parasites within the supergroup Metamonada (Guhl & Vallejo, 2003). The sequences of core trypanosomatid histones were found to be some of the most divergent across eukaryotes, with low degrees of conservation observed not only in the N- and C-terminal tails, but also in the central histone fold (Deák et al., 2023).This sequence divergence also appears to have a functional impact. A recently resolved structure obtained by cryo-electron microscopy (Cryo-EM) of the *Trypanosoma brucei* nucleosome revealed that it displays some important differences when compared to the *Homo sapiens* nucleosome (Deák et al., 2023).Divergent amino acid residues in H3 disturb the H3-H4 binding interface as well as electrostatic interactions in the H3-H3 dimer, which results in weaker histone-DNA interaction across the nucleosome core particle (NCP). Trypanosome nucleosomes were also found to be less thermostable and undergo an alternative disassembly pathway in one step instead of conventional two. In contrast, the H2A-H2B dimer interface is reinforced and provides stronger contact with DNA. This is further confirmed by in vivo micrococcal nuclease digestion, since enzymatic digestion stalls at the points of DNA contact with H2A-H2B dimer. Additionally, due to amino acid divergence in H2A and H3 tails, DNA entry and exit points appear to be flexible, with only 126 base pairs (bp) out of the normal146 bp resolved in the structure (Deák et al., 2023).

Importantly, there are other parasitic protists with functionally diversified nucleosomes. A similar case is observed in diplomonads, a group of unicellular flagellated organisms that can be both parasitic and free-living in low-oxygen environments (Simpson & Čepička, 2019). Protein alignments of histones from the parasitic diplomonad *Giardia intestinalis* showed that they encode some of the most divergent histones found across eukaryotes (Wu et al., 2000), and the genes encoding for linker histone H1 are absent (Yee et al., 2007). A Cryo-EM structure of the *Giardia intestinalis* nucleosome revealed some structural differences to *Homo sapiens* nucleosome. For

instance, in a similar manner to the *Trypanosoma brucei* nucleosome, only 125 bp were resolved in the nucleosome structure. This DNA entry point flexibility can be explained by divergence in H2A C-terminus as hydrophobic residues which normally interact with the N-terminal tail of H3, have been replaced by hydrophilic residues (Sato et al., 2021). Similar to the trypanosome nucleosome, the *Giardia* nucleosome is less thermostable than the *Homo sapiens* nucleosome and disassembles in a single step. In the *Giardia* nucleosome, however, acidic amino acid insertions near alpha helices 2 and 3 in H2A and H2B cause alterations in the H2A-H2B dimer resulting in a distinct acidic patch. This indicates that *Giardia* nucleosomes have unique properties that likely impact their interaction with chromatin modifying machinery (Sato et al., 2021).

Similar to parasites, histone divergence can also be observed among free living eukaryotes such as dinoflagellates, a group of unicellular algae best known for their roles as coral endosymbionts (Lin, 2024). Dinoflagellates have a unique form of DNA compaction. Unlike other eukaryotes, dinoflagellates associate their DNA with viral nucleoproteins as well as bacterial histone-like proteins (Gornik et al., 2012; Rizzo & Morris, 1983). A consequence of this is that dinoflagellate chromosomes are condensed into a liquid crystalline state (Wong, 2019). Histones, on the other hand, are only lowly expressed, and given that they are less important for dinoflagellate chromatin stability and gene regulation, they have significantly diversified, both in comparison to other eukaryotes and within the dinoflagellate group (Marinov & Lynch, 2016). A unique feature of dinoflagellate histones is their prominent elongation. While in some species H2A, H2B, H3 and H4 can be several amino acids longer than the average eukaryotic histone, Symbiodinium *microadriaticum* H2A sequences are triple the size of *Homo sapiens* H2A. Despite high heterogeneity across species, a group of putative H2A.Z orthologs could be identified, as well as H2A proteins containing the characteristic C-terminal H2A.X phosphorylation motif (SQEF/Y) (Marinov & Lynch, 2016).

The above listed examples show that despite their strong conservation, histone sequence can diversify leading to functional differences in the nucleosome. However, given the essentiality of histone function and their overall high conservation, even small differences in amino acid sequence can translate into significant phenotypic difference. This can be observed in the H2A.X variant in *A. thaliana*, which is different from canonical H2A variant in its SQEF/Y C-terminal motif that can be phosphorylated. The SQEF/Y motif accounts for the recruitment of H2A.X as a DNA damage response variant (Osakabe et al., 2018), making it functionally distinct from the canonical H2A variant. Given that within a single species small change in amino acid sequence cause

significant functional differences, I hypothesize that subtle differences between diverse eukaryotic histones have driven the diversification of core nucleosome function across eukaryotic diversity.

1.3 Nucleosome properties are impacted by diverse chromatin architecture

In addition to variation in intrinsic nucleosome properties, nucleosome function is also impacted by chromatin-modifying proteins. For example, chromatin remodelers alter chromatin landscapes by interacting with nucleosomes. Chromatin remodelers belong to the Snf2 superfamily characterized by the helicase-like ATPase domain (Cairns, 2009), which can be divided into four main families – ISWI, SWI/SNF, SWR1 (INO80), and CHD (Cairns, 2009; Flaus, 2006). Chromatin remodelers, similarly to histones, are highly conserved proteins that were likely present in LECA and are currently found in all eukaryotic groups (Flaus, 2006; Grau-Bové et al., 2022). Some diversity, however, can be observed in Snf2 protein number variation across species. For instance, in *Homo sapiens* 53 Snf2 proteins are present (Rippe et al., 2007), in *Arabidopsis thaliana* – 41 (Wang et al., 2023). Meanwhile *Plasmodium falciparum* has only 10 (Watzlowik et al., 2025), and *Giardia intestinalis* – 6 (Flaus, 2006). This drastic difference could imply a greater dependence on chromatin remodeling in multicellular organisms. Some degree of functional diversification in chromatin remodelers also exists, as observed in the ISWI family. ISWI is involved in nucleosome positioning, ensuring that nucleosomes are separated by the correct linker distance (Kagalwala et al., 2004). *Plasmodium falciparum,* however, does not have a canonical ISWI remodeler, as PfSnf2L protein, which is the most closely related *Plasmodium* remodeler to the ISWI subfamily, has retained only 30% similarity to *Homo sapiens* SNF2L, perhaps as an adaptation to the highly AT-rich *P. falciparum* genome (Watzlowik et al., 2025).

In addition to remodelers, post-translational modifications (PTMs) and their functional readouts are also variable across species. Another layer of chromatin regulation is mediated via post-translational modifications that are deposited onto histone tails. PTMs are deposited, read and removed by chromatin readers, writers and erasers respectively (Dan & Chen, 2023). These proteins show diversity in copy number across eukaryotic lineages, but all eukaryotic lineages retain acetylases, methyltransferases and reader proteins (Grau-Bové et al., 2022). The most common histone PTMs are acetylation, methylation, phosphorylation and ubiquitination (Strahl & Allis, 2000). Histone acetylation is known to impact nucleosome stability and positioning by destabilizing histone-DNA interaction leading to partial nucleosome unwrapping (Bowman & Poirier, 2015) and is universally associated with transcriptional activity (Navarrete et al., 2025). Histone methylation marks, on the other hand, are thought to work as an interface for other chromatin-binding proteins and combinations of different methylation marks are associated with

different transcriptional states, producing patterns sometimes referred to as the histone code (Strahl & Allis, 2000). However, methylation marks have varying degrees of conservation across eukaryotes. Residues H3K79 and H3K27 are well-known methylation target sites in animals, but no modifications are present at these sites in the *Discoba,* which includes *Trypanosoma brucei*, *Leishmania major* and *Naegleria gruberi* (De Lima et al., 2020; Navarrete et al., 2025). Another drastic example of histone mark loss can be observed in the dinoflagellate lineage. In most dinoflagellate species at least one variant has lost all common H3 residues subject to PTMs including H3K9, H3K27, H3K36, H3K79 (Marinov & Lynch, 2016). Even when the PTM is conserved between two organisms, sometimes its functionality is not. For example, in the amoeba species *Dictyostelium discoideum* and *Acanthoamoeba castellanii*, H3K79me3 is associated with transcription repression, despite being an active transcription mark in animals. This is likely caused by divergence in reader proteins. Lineage specific PTMs can also be observed in different eukaryotic groups. For example, trypanosomatid genes are organized in large unidirectional clusters transcribed by RNA Pol II as polycistronic transcripts (Palenchar & Bellofatto, 2006). The transcription start sites in trypanosomes are marked with hyperacetylation of H2A histones, a modification that has not been found in other lineages (Maree et al., 2022). This indicates that chromatin dynamics and regulation are not uniform across eukaryotes (Navarrete et al., 2025), and, in the case of dinoflagellates, PTMs can lose some of their importance.

While chromatin-modifying proteins highly impact chromatin architecture, nucleosome affinity for specific dinucleotide periodicity indicates that nucleosome assembly is DNA sequence-driven. Interestingly, genome organization and GC content has significantly diversified across eukaryotic tree, and even though GC content is pivotal for intrinsic nucleosome positioning, most eukaryotes form nucleosomes across their genomes. For instance, among unicellular eukaryotes, the GC content of *Saccharomyces cerevisiae* is on average 38.2%, and this value is evenly distributed across the genome (Costantini & Musto, 2017). These numbers are very different for the green alga *Ostreococcus tauri*, which has an average GC content as high as 59%, excluding chromosomes 2 and 19, both of which have slightly lower GC content (Derelle et al., 2006) and *Chlamydomonas reinhardtii* which has a GC content of 62% (Naya et al., 2001). The malaria parasite *Plasmodium falciparum* has an extremely AT-rich genome with a GC content of only 19.2%, hence *P. falciparum* nucleosome positioning landscape is especially intriguing (Kensche et al., 2016). For multicellular organisms, however, it is common to have heterogeneous GC content across the genome. In mammals and birds, the GC content across the whole genome and intron regions is lower than in exonic regions and promoters. Plants, as well as invertebrate animals, display similar patterns, except for promoter GC content, which is comparable with the

whole genome (Jiang et al., 2014). Such GC content heterogeneity within genomes and across eukaryotic diversity indicates that diverse nucleosomes can be positioned on a variety of sequences despite their general preference towards GC-richness and GC dinucleotide periodicity. While some chromatin remodelers may be adapted for unique linage-specific genome organization, such as the *P. falciparum* ISWI-like protein PfSnf2L (Watzlowik et al., 2025), diverse nucleosomes may also display different intrinsic binding properties as an adaptation for the genome organization of the species. To disseminate the complex relation between chromatin modifying machinery, underlying genome and histone diversity, robust comparative experimental approaches to study chromatin are needed.

1.4 Classic methods in chromatin biology

Traditionally, nucleosome properties have been studied using both *in vivo* and *in vitro* methods. One classic way to gain insight into chromatin structure and function *in vivo* is a micrococcal nuclease (MNase) digestion assay. MNase is an endo-exonuclease enzyme isolated from *Staphylococcus aureus* that is unable to cut nucleosomal DNA, enabling the isolation of nucleosome-protected chromatin regions. Isolated nuclei or permeabilized cells are treated with varying concentrations of MNase to generate nucleosome arrays or mononucleosomes. The resulting nucleosomal DNA can be purified and resolved using gel electrophoresis (Zaret, 1999). Coupling this technique with next-generation sequencing (MNase-seq) allows for precise mapping of nucleosome positions on the genome as well as quantifying the probability of the nucleosome to be present on a certain genomic region, i.e. nucleosome occupancy (Kaplan et al., 2010; Z. Zhang & Pugh, 2011). Another way to map nucleosome locations on a genome is chromatin immunoprecipitation (ChIP), a technique that consists of crosslinking DNA-bound proteins, often histones, followed by incubation with a specific antibody to isolate the DNA region of interest (O'Neill & Turner, 1996). This method, also followed by next generation sequencing (ChIP-seq), allows to precisely map, quantify and compare specific histone variants, PTMs and histone modifiers genome wide (Hino et al., 2023).

Since chromatin is impacted by multiple factors, such as chromatin remodelers, histone PTMs and histone variant diversity, it is often difficult to study the impact of individual chromatin components using *in vivo* assays or genetic studies. Additionally, since chromatin modifying factors also show variation across eukaryotic diversity, comparative studies of nucleosome properties *in vivo* are not possible. Therefore, many advances in the field have resulted from chromatin reconstitution *in vitro*. For reconstitution, histones are isolated from a eukaryotic organism or expressed in the bacterium *Escherichia coli* and purified. Recombinant histones form

inclusion bodies, which are denatured in 8M guanidine hydrochloride and refolded into histone octamers via salt dialysis from 2M NaCl in equimolar concentration (Y.-T. Lee et al., 2015). Refolded histones are mixed with DNA under 2M NaCl concentration and then the salt is dialyzed away, which leads to stepwise nucleosome assembly (Ura & Kaneda, 2002). *In vitro* reconstituted nucleosomes are often used for structural studies, which provide an insight into nucleosome properties and can connect them to specific amino acid residues (Deák et al., 2023; Luger et al., 1997; Sato et al., 2021). I*n vitro* reconstitution can also be used to assess nucleosome stability (Taguchi et al., 2014) and positioning (Kaplan et al., 2010). Additionally, this method can enable the study of the impact of individual chromatin modifiers on nucleosome positioning (Flaus & Owen-Hughes, 2003).

Despite their significant advantages, there are caveats to *in vitro* assembled nucleosomes. Firstly, *in vitro* assays will provide only a limited model of chromatin behavior, since reconstituted nucleosome will be isolated from chromatin remodeling machinery. I*n vitro* methods also do not reflect the cellular processes such as gene expression or DNA replication, which are known to influence nucleosome positioning. This poses limitations on how relevant in vitro observations are for understanding nucleosome binding properties in living organisms. Additionally, the protocols for *in vitro* reconstitution are often laborious and require complex protocols to obtain singular samples of nucleosome-bound DNA (Y.-T. Lee et al., 2015). This makes large-scale comparative chromatin studies highly challenging to conduct. Therefore, comparative chromatin studies require an alternative experimental model that would be easily scalable while maintaining a neutral comparable environment.

1.5 Ectopic expression of histones in *Escherichia coli* as a novel method in chromatin biology

A novel alternative to *in vitro* reconstitution of nucleosomes is the ectopic expression of histones in *E. coli,* which is sufficient for nucleosome assembly inside the bacterial cell. This method for studying chromatin was pioneered in 2019, when Rojec and colleagues expressed histones HmfA and HmfB from the archaea *Methanothermus fervidus* in *E. coli*. They demonstrated that archaeal histones can assemble into nucleosome inside the living bacteria without strongly disturbing bacterial growth or gene expression (Rojec et al., 2019). Histone-containing *E. coli* cells were profiled using MNase digestion and showed DNA footprints reminiscent of the ones obtained from *M. fervidus* culture. Archaeal histones were not found to have detectable effects on most bacterial nucleoid-associated proteins (NAPs). This method has been further leveraged to reconstruct viral nucleosomes in *E. coli* and demonstrate their stacking into archaeal-like oligomers, highlighting a

role of viruses in nucleosome evolution (Irwin & Richards, 2024). Additionally, eukaryotic nucleosome have been assembled in *E. coli* by expressing *X. laevis* histones (Irwin & Richards, 2024; Zhou et al., 2024). Cryo-EM analysis confirmed that *Xenopus laevis* nucleosomes assembled in *E. coli* are *bona fide* nucleosomes that show high structural similarity to in *vitro* reconstituted *X. laevis* nucleosomes (Ho et al., 2026). This work also revealed the formation of hexasomes (non-canonical nucleosome structures consisting of an H3-H4 tetramer and a single H2A-H2B dimer complexed with DNA) as well as previously unreported di-hexasomes (Ho et al., 2026). These studies show that eukaryotic histones are capable of assembling into nucleosomes on a naïve genome *in vivo* without histone chaperones. Given that *E. coli* is a well-established model organism with very short doubling time and a high availability of genetic and genomic tools, ectopic expression of eukaryotic histones constitutes an easy, fast and physiologically relevant method to characterize nucleosomes outside of their native context. Additionally, it provides a suitable chassis for comparative nucleosome studies. Nucleosome-free organism such as *E. coli* could ensure that functional studies with nucleosomes from diverse species are comparable since the investigated cell is devoid of chromatin modifying proteins similar to *in vitro* approach. At the same time, the observed properties may be more biologically relevant, since *E. coli* cells are metabolically active and undergo gene expression, cell division and other processes that could impact nucleosome positioning.

1.6 Questions and aims

Across eukaryotes, nucleosomes are ubiquitous, despite the organism differing in their cell number, genome architecture and chromatin remodeling machinery. Given the known cases of histone and nucleosome structure divergence and how significant even small changes in histone amino acid sequence are, I hypothesize that intrinsic nucleosome function is different across eukaryotes. To test this, I first sought to determine how conserved histone sequences and nucleosome structures are across eukaryotic diversity. To address this, I have surveyed histone sequence and structure similarity across 45 eukaryotic species to assess the degree of histone divergence across the eukaryotic tree. As a next step, I sought to determine whether nucleosome binding properties differ between diverse eukaryotes. To answer this, I have compared nucleosome properties of 13 selected species. Using ectopic histone expression in *E. coli* followed by MNase treatment, I have detected that nucleosomes display variation in fragment size, intensity and pattern. We used MNase-seq to demonstrate that diverse nucleosomes vary in sequence preference, stability and mean linker distance. Overall, our results indicate that diverse nucleosomes vary in their intrinsic properties, suggesting that nucleosome function has diversified across eukaryotic lineages.

2 Methods

2.1 Histone sequence identification and analysis of nucleosome divergence

Since histones have diverged into variants across eukaryotes, the least divergent histone per histone family per species was selected for comparison. To achieve this, histones H2A, H2B, H3 and H4 were identified and annotated in proteomes from representative species across eukaryotic diversity using histone hidden Markov models (HMMs) (Irwin & Richards, 2024). The resulting hits were aligned using MAFFT (v. 7.526) (Katoh, 2002), trimmed using trimal (v. 1.4.1) (Capella-Gutiérrez et al., 2009) and filtered to remove sequences with less than 75% alignment coverage. The resulting alignments were used to build phylogenies using IQ-Tree (v.2.1.2) (Minh et al., 2020). Each phylogeny was inspected with FigTree (v.1.4) (https://tree.bio.ed.ac.uk/software/figtree/) and was redone after the removal of long branching sequences. For species missing histones, histone sequences were identified from transcriptomes. Publicly available RNA-seq datasets were obtained using SRA toolkit (v. 3.2.1) (Wheeler, 2006) and transcriptomes were generated using SPAdes (v. 4.2.0) (Vasilinetc et al., 2015). Proteins were predicted with TransDecoder (v. 5.7.1) (Haas, BJ., 2023) and the resulting proteomes were searched as described previously. After histone sequences had been collected, the least divergent paralog of histone H3, H2B, H4, and H2AZ were selected for each species.

Selected histones were re-aligned and used to generate histone phylogenies from which histone sequence conservation was assessed based on branch length distances. Isoelectric points for each protein sequence were calculated in Python (v.3.12) with Bio.SeqUtils package from Biopython (v.1.85) (Cock et al., 2009).

Nucleosome structures were predicted from the selected histone sequences. Interactions between two copies of H2A, H2B, H3 and H4 from each species were modelled alongside a fragment of the *E. coli* genome with GC content of 56% (AGGCGAAGTGGCGTTCTTTATCGCCAAGCGTCTACGATCTAACGTACGTGAGCTGGAAGG GGCGCTGAACCGCGTTATTGCTAACGCCAACTTTACCGGACGGGCGATCACCATCGACTTC GTGCGTGAGGCGCTGCGCGACTTGC and reverse complement strand) using Alphafold3 (Abramson et al., 2024). The resulting models were locally aligned by individual chains using FoldSeek (v. 10.941cc33) (Van Kempen et al., 2024) and the bit-scores for each chain were used to calculate pairwise median bit-scores for nucleosomes from each species in Python (v.3.12).

2.2 Plasmids

2.2.1 Histone expression plasmids

For each species, each of the core histones was synthesized and inserted into a pCC1 plasmid by GenScript. Histone protein-coding sequences were codon-optimized for prokaryotic translation using the GenScript optimization tool. Since free histones are known to be toxic for cells (Xu et al., 2009), low-copy plasmid pCC1 was selected as an expression vector. pCC1 provided contains a chloramphenicol (Chl) resistance cassette, oriV origin of replication and T7 promoter. T7 promoters are known to have high basal expression (Dubendorf & Studier, 1991). To ensure tighter expression control, a rhamnose-inducible promoter was introduced in the vector. Lastly, histone protein coding sequences were introduced downstream of the rhamnose promoter. Each histone sequence is preceded by an identical ribosome binding site (RBS) (AATAATTTTGTTTAACTTTAAGAAGGAGATATACAT) to ensure the expression and translation of all four histone genes from a polycistronic transcript.

2.2.2 Cloning of pCC1-GFP

To generate a negative control for the expression of histones as well as a control to account for MNase digestion bias, a pCC1 plasmid expressing green fluorescent protein (GFP) was created. To generate a GFP expression vector, a fragment containing an RBS and GFP was amplified with Q5 High-Fidelity DNA polymerase (NEB, M0491S) from the pD861-GFP plasmid (Irwin & Richards, 2024) with forward primer cgcCTCGAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGTGAGCAAAGGC and reverse primer gcgcGGATCCTTAATGATGGTGGTGGTG, designed to introduce XhoI and BamHI restriction enzymes digestion sites. The thermocycler program used for fragment amplification was 98°C for 30s, 35 cycles of 98°C for 10s, 72°C for 60s and final step of 72°C for 2 min. The PCR product was isolated with an in-house PCR purification kit (MBS, IMP) and eluted in 50 μL Mono-Q grade H_2O. Isolated PCR product as well as the pCC1-*Homo sapiens* vector were digested with restriction enzymes BamHI (NEB, R136S) and XhoI (NEB, R0146S) at 37°C for 2h in a water bath and subsequently purified with an in-house PCR purification kit (MBS, IMP) before being eluted in 30 μL Mono-Q grade H_2O. The insert and vector were ligated in a 3:1 ratio with T4 ligation enzyme (NEB, M0202) for 20 min at room temperature and transformed into chemically competent *E. coli* strain DH5α cells (MBS, IMP). Competent *E. coli* cells were thawed on ice for 20 min before 1 μL of plasmid was gently mixed with 50 μL of competent cells and incubated on ice for 20 min. Cells were heat-shocked for 45s at 42°C in a water bath, then recovered on ice for 2 min prior to the addition of 400 μL of Super Optimal broth with Catabolite repression (SOC)

media. Cells were recovered at 37°C while shaking at 180 rpm for 1h. Transformed competent cells were then plated on lysogeny broth-agar (LB-agar) plates with Chl selection (25 μL/ng) and incubated at 37°C overnight. Colonies were screened with Taq PCR mastermix (MBS, IMP) using forward primer AATCTCGAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACAT and reverse primer CTAGTTATTGCTCAGCGGTGGCGG. The thermocycling program for fragment amplification was 96°C for 3 min, 35 cycles of 96°C for 30 s, 52°C for 30 s, 72°C for 1 min and the final step of 72°C for 5 min. Positive colonies were grown overnight in LB+Chl (25 μg/mL) shaking at 37°C. Plasmids were isolated with an in-house miniprep kit (MBS, IMP) and sequenced by Plasmidsaurus.

2.3 Histone expression

2.3.1 Escherichia coli transformation

pCC1 histone plasmids were transformed into *Escherichia coli* strain BL21 (DE3). Competent *E. coli* cells (MBS, IMP) were thawed on ice for 20 min before 1 μL of plasmid was gently mixed with 50 μL of competent cells and incubated on ice for 20 min. Cells were heat-shocked for 45s at 42°C in a water bath, then recovered on ice for 2 min prior to the addition of 400 μL of SOC media. Cells were recovered while shaking at 180 rpm for 1h. Transformed bacteria was plated on LB-agar plates with and incubated for two days. For long-term storage, individual colonies were transferred to 5 mL of LB media and shaken at 180 rpm overnight. Resulting cultures were stored in 25% glycerol solution at -70°C. All described steps were done in presence of Chl selection (25 μl/ng) and at 30°C.

2.3.2 Histone expression

For expression, glycerol-preserved stock cultures were streaked onto LB-agar plates and incubated overnight for revival. The resulting cultures were then grown overnight in 5 mL fresh LB. Overnight cultures were diluted 1:50 into 60 ml fresh LB and grown until the OD_{600} (Optical density at 600 nm) reached 0.6. All previous steps were done in presence of Chl selection (25 μg/mL) and at 30°C. Prior to induction, cells were cold-shocked on ice for 30 min at 125 rpm. The inducer, L-rhamnose monohydrate (Sigma-Aldrich) was added to 10 mM and cultures were induced for 20 hours at 21°C shaking at 180 rpm. After induction, the OD_{600} for all cultures was measured.

2.3.3 Immunoblotting

Immunoblots were used to confirm the expression of eukaryotic histones. 150 μL of each induced culture was sampled, spun down at 21300xg for 2 min and the supernatant was discarded. The

pellet was resuspended in Laemmli Buffer (50 mM Tris pH 6.8, 2% SDS, 10% glycerol, 100 mM β-mercaptoethanol, 1x protease inhibitor (Roche, 11697498001)) in a volume corresponding to 1/16 of the OD_{600} measured post-induction. Samples were incubated for 5 min at 95°C while shaking at 700 rpm and spun down at 21300xg for 2 min. 5 μL of each protein extract was loaded into a 4-20% gradient polyacrylamide precast gel for SDS-PAGE (Bio-Rad, #4561096EDU) and electrophoresed for 60 minutes at 120V in SDS-PAGE buffer (2.5 mM Tris, 19.2 mM glycine, 0.01% SDS, pH 8.3). Afterwards, the proteins were transferred onto a 0.2 μm pore size nitrocellulose membrane at room temperature for 60 min at constant 400 mA in transfer buffer (2.5 mM Tris, 19.2 mM glycine, 10% ethanol, pH 8.3) and stained using Pierce™ Reversible Protein Stain Kit for Nitrocellulose membranes (ThermoFisher, 24580). After destaining, membranes were blocked for 45 minutes in cow milk solution (1xPBS, 1% tween-20 with 5% cow milk powder). Blocked membranes were incubated for 1 hour at room temperature with 0.01 mg/ml of a corresponding antibody (table 2.1) diluted in 10 mL cow milk solution. Prior to secondary antibody incubation, membranes were washed three times for 10 minutes each in 10 mL of cow milk solution. Membranes were incubated with 0.005 mg/mL of anti-rabbit IgG antibody diluted in 10 mL of cow milk solution at room temperature for 1 hour. The membranes were then washed two times for 10 minutes each with 1xPBS containing 1% tween-20 and once with 1xPBS. To develop the membrane, 1 mL of luminol and 1 mL of stable peroxidase buffer from Pierce™ Western Blotting Substrate (Thermo Scientific, 32106) were mixed, and the membrane was placed onto the solution face down and incubated for 1 min in a dark cabinet. Membranes were imaged using an iBright FL1500 imaging system (Thermo Fisher).

Antibody	**Catalog number**	**Type**	**Source**
Anti-histone H3	ab1791	Primary	Rabbit
Anti-histone H4	ab223875	Primary	Rabbit
Anti-rabbit IgG H+L HRP conjugate	W4011	Secondary	Goat

Table 1. Antibodies used for immunoblotting

2.3.4 Micrococcal nuclease digestion and sequencing

To assess nucleosome formation, *E. coli* BL21 cultures were induced as described before, fixed, lysed and digested with micrococcal nuclease according to previously published methods (Irwin

& Richards, 2024; Rojec et al., 2019). To prevent heterogeneous digestion, prior to fixation, the densities of the induced *E. coli* cultures were normalized to a total volume of 50 mL according to a calculated normalization factor:

$$f_{sample} = \frac{OD_{600 minimal}}{OD_{600 sample}}$$

Here, $OD_{600_{minimal}}$ is the lowest measured OD_{600} of the experiment and $OD_{600_{sample}}$ is the OD_{600} of the normalized sample. Cultures were diluted with fresh LB to 50 ml total. Cells were fixed with 1% formaldehyde for 15 min while shaking at 180 rpm. Fixation was quenched with 125 mM glycine for 5 min while shaking at 180 rpm. Cells were then pelleted for 2 min at 21300xg, the supernatant was removed, and the pellet was washed with PBS once before being resuspended in 1 ml PBS and spun down again in a microcentrifuge tube. After removing the supernatant, the cells were frozen and stored at -70°C.

For micrococcal nuclease digestion, cells were thawed on ice, then treated for 1 hour on ice with freshly made lysozyme buffer (120 mM Tris-HCl pH 8, 50 mM EDTA, 4 mg/ml lysozyme). After treatment, cells were spun down at 15,000xg for 3 minutes at 4°C, then resuspended and lysed with fresh lysis buffer (10 mM NaCl, 10 mM Tris-HCl pH 7.5, 3 mM $MgCl_2$, 0.5% NP-40,1x protease inhibitor (Roche, , 11697498001), 0.15 mM spermine (Sigma-Aldrich, S3256), 0.5 mM spermidine (Sigma-Aldrich, S2626)) for 20 min on ice. After lysis, samples were spun at 15,000xg for 10 minutes at 4°C and washed with -CA buffer (15 mM NaCl, 10 mM Tris-HCl pH 7.5, 60 mM KCl, 1x protease inhibitor (Roche, 11697498001), 0.15 mM spermine (Sigma-Aldrich, S3256), 0.5 mM spermidine (Sigma-Aldrich, S2626)). Samples were then spun at 15,000xg for 5 min at 4°C and resuspended in +CA buffer (15 mM NaCl, 10 mM Tris-HCl pH 7.5, 60 mM KCl, 5 mM $CaCl_2$, 1x protease inhibitor (Roche, 11697498001), 0.15 mM spermine (Sigma-Aldrich, S3256), 0.5 mM spermidine (Sigma-Aldrich, S2626)) before being incubated for at least 5 min at 37°C. Since micrococcal nuclease digestion assays are conducted as a digestion time course, 100 µl of the culture was taken as time point zero. Approximately 222 U/mL of micrococcal nuclease (Thermo Fisher, 88216) were added to samples which were then briefly vortexed and incubated at 37°C. Samples of 100 µL were taken at 15 min, 30 min and 45 min and immediately transferred into tubes containing STOP solution (200 mM EDTA pH 8, 200 mM EGTA pH 8). Each time point was treated with 0.05 mg of RNase A (Thermo Fisher) for 30 min at 37°C. Afterwards samples were treated with 0.1 mg of Proteinase K (Thermo Fisher) and 1% SDS for 20 h at 65°C in a water bath. DNA was then isolated using an in-house PCR purification kit (MBS, IMP), eluted in 50 µL

Mono-Q grade H_2O and quantified by Nanodrop. 150 ng of the resulting DNA were mixed with 1x Orange loading dye (NEB, B7022SVIAL) and size-separated via gel electrophoresis for 20 min at 150V on 4% agarose gel with 1x SYBR™ Safe DNA Gel stain (Invitrogen, S33102).

Nucleosomal DNA was size selected via purification with nucleic acid-binding magnetic beads (MBS, IMP). 20 μL of DNA was incubated with 3.5x bead ratio and the supernatant discarded. After two washes with 70% EtOH the beads were eluted in 30 μL Mono-Q grade H_2O and quantified with a Quibt dsDNA broad-range assay kit (Invitrogen, Q33231). The resulting DNA was sequenced using short-read DNA sequencing. DNA libraries were prepared in-house by the VBC NGS core facilities and sequenced in two separate runs as flowcell spike-ins using pair-end 150 bp reads on a NovaSeqX 10B XP. This resulted in a total of 30M reads for first run and 26M reads for the second run. Each condition was done in biological duplicate.

2.4 Sequencing analysis

2.4.1 MNase-seq mapping and calculation of pairwise correlation

Paired-end reads were trimmed with fastp (v.1.0.1) (S. Chen, 2025) and mapped to the *E. coli* K12 MG1655 genome assembly (NCBI RefSeq GCF_000005845.2) using bwa (v. 0.7) (H. Li & Durbin, 2009). Alignments were sorted and indexed with samtools (v. 1.22.1) (Danecek et al., 2021) and deduplicated with Picard (v. 3.2.0) (Broad Institute, 2019). All further analysis was done only with reads sized between 135 and 161 base pairs (95 to 130 for *G. intestinalis* based on smaller DNA footprint resolved with gel electrophoresis) to remove background signal of digested free DNA. Read coverage over the genome was calculated over 10-base pair bins and normalized by reads per kilobase per million mapped reads with deepTools (v.3.5.6) (Ramírez et al., 2016). Pairwise Spearman correlations were calculated over 500 base-pair windows. Since all sample replicates had Spearman correlations of 0.8 or higher, the reads were merged and processed as described above. To account for MNase digestion bias resulting from the tendency of MNase to cleave AT-rich regions, coverage from a non-histone GFP control was scaled and subtracted from the rest of the samples. The scaling factor was calculated as:

$$f_{sample} = \frac{\frac{c_{sample_1} + c_{sample_2}}{2}}{\frac{c_{GFP_1} + c_{GFP_2}}{2}}$$

Here, $c_{sample1}$ and $c_{sample2}$ are the measured DNA concentration of the two replicates of a normalized sample after MNase digestion. c_{GFP1} and c_{GFP2} are DNA concentrations of both GFP replicates 1 and 2 respectively. GFP coverage was divided by the resulting normalization factor and subtracted from the corresponding sample in R (v. 4.5.0) using Bioconductor package

GenomicRanges (v. 1.6.0) (Lawrence et al., 2013). Normalized coverage was used to recalculate Spearman pairwise correlation over 500 base-pair bins.

2.4.2 Analysis of nucleosome binding properties

To examine nucleosome binding properties, MNase-seq peaks were called with DANPOS (v. 3.1.1) (K. Chen et al., 2013). To assess the correlation between MNase-seq peak intensity and *E. coli* gene expression, a publicly available RNA-seq dataset (SRR2954253) was obtained using the SRA toolkit. Reads were mapped to the *E. coli K12* MG1655 genome and processed and normalized by reads per kilobase per million mapped reads as described above. Spearman correlations between RNA-seq coverage and nucleosome occupancy as well as Spearman correlation between MNase-seq peak intensity and corresponding GC content of the underlying sequences were determined. To assess linker distances, I defined pseudo-linker distances as the distance between two called MNase-seq peaks if both peaks had occupancies above the 25th percentile. All calculations were done in Python (v.3.12) with packages Biopython (v. 1.85), NumPy (v1.26.4) (Harris et al., 2020) and Pandas (v.2.3.0) (McKinney, 2010). Testing for statistical significance were done using one-way ANOVA and post-hoc TukeyHSD test in R (v. 4.5.0). To assess the correlations between the genomic GC content of studied species and MNase-peak GC preference, a Spearman correlation was calculated between genomic GC content obtained from the literature (Franzén et al., 2009; He et al., 2024; Herold et al., 2008; Jung et al., 2011; Kensche et al., 2016b; Merchant et al., 2007; Naya et al., 2001; Piganeau et al., 2011; Tyler et al., 2006; Wuitschick & Karrer, 1999) and the correlation between MNase-seq peak intensity and GC content of the underlying sequence.

To determine dinucleotide frequency preferences, the occurrence of 2-mers across each sequence was calculated for MNase-seq peaks with intensity over the 20th percentile. Abundance of individual 2-mers was normalized by the MNase-seq peak bin length (140 bp). GG/GC/CG/CC (GC-rich 2-mers) and AA/AT/TA/TT 2-mers were summed up (AT-rich 2-mers). To calculate the sum of differences between GC-rich and AT-rich 2-mers, the frequency of GC-rich and AT-rich dinucleotides was calculated, and the frequency of AT-rich dinucleotides were subtracted from GC-rich dinucleotide frequency. Frequency differences for each position of the MNase-seq peak bin were summed up. This calculation was done for MNase-seq peaks grouped by percentile windows. All dinucleotide frequency analyses were done on well-positioned nucleosomes with maximal fuzziness cutoff at the 20th percentile. Calculations were done in R (v. 4.5.0) using Bioconductor package GenomicRanges (v. 1.6.0) (Lawrence et al., 2013)

To assess correlations between MNase-seq coverage and endogenous nucleoid-associated proteins occupancy in *E. coli,* publicly available normalized *in vivo* protein occupancy display – high resolution (IPOD-HR) coverage data for E. *coli* str. K12 MG1655 was obtained (GEO accession GSE142291). MNase-seq read coverage was recalculated for 5-base-pair windows and normalized by reads per kilobase per million mapped reads with deepTools (v.3.5.6). Spearman correlation was calculated between MNase-seq read coverage and IPOD-HR coverage over 500-base-pair windows using deepTools (v.3.5.6).

3 Results

3.1 Core histone sequences exhibit variation across eukaryotes

To assess histone sequence divergence across eukaryotes, we identified core histones with hidden Markov models (Irwin & Richards, 2024) from 45 selected species across eukaryotic diversity (Fig. 1A). The most conserved ortholog per histone family per species was identified using histone phylogenies constructed using IQ-Tree (v.2.1.2). Analysis of evolutionary distance of H2A, H2B, H3 and H4 showed that while in general all four core histone families show a high degree of conservation (Fig. 1B), some variation can be observed. Firstly, histone families H3 and H4 display an overall higher degree of conservation relative to H2A and H2B. Despite this, in all four histone families we identified outlier species that showed higher sequence divergence. Among the species with the most divergent core histones were parasitic protists such as *Giardia intestinalis, Trypanosoma brucei* and *Plasmodium falciparum.* Interestingly, divergent histone sequences were not ubiquitous across unicellular eukaryotes. For instance, parasitic species *Toxoplasma gondii,* which is closely related to *P. falciparum*, as well as *Phytophthora sojae* display average histone divergence. Histone divergence of all four histone families is to some degree also present in the free-living protists like the ciliates *Tetrahymena thermophila* and *Oxytricha trifallax.* Additionally, a high degree of divergence is present in the H2A and H2B sequences of dinoflagellates *Fugacium kawagutii* and *Symbiodinium microadriaticum*, while dinoflagellate histones H3 show lesser degree of divergence, and H4 are highly conserved.

To understand how sequence divergence may affect histone properties, we calculated protein length and isoelectric point for all four histone families using Biopython (v.1.85) (Fig. 1C). Protein lengths showed that in the case of histones H2A and H2B, some degree of sequence length divergence exists, with the histone length median being 138 (stdev = 2.3, min = 130, max = 146) and 125 (stdev = 1.7, min = 99, max = 110) amino acids, respectively. *S. microadriaticum* and *F. kawagutii* H2A and H2B sequences constitute strong outliers. For instance, the H2A sequences of the two dinoflagellates are 425 and 417 amino acids, respectively, which is three times higher than the median of 138. H2B sequences of dinoflagellates correspond to 286 amino acids and 228 accordingly, which is also significantly above the median of 125. This extreme size deviation is explained by the prominent tails characteristic to dinoflagellates, which contribute to the high evolutionary distance of dinoflagellate H2A and H2B (Marinov & Lynch, 2016). Additionally, the H2A sequences of the discobans *Naegleria gruberi*, *Leishmania major* and *Trypanosoma brucei* are noted as outliers. In contrast to H2A and H2B, H3 and H4 show extreme conservation in sequence length. H3 and H4 medians are 136 and 103, and most of the surveyed histone H3 and

H4 lengths of the species surveyed are identical to the corresponding median. However, several protists such as the diplomonads *G. intestinalis* and *S. salmoncida* (Fig. 7C) constitute an exception and show a variation of up to 4 amino acids in size for H4 and up to 7 for H3.

In addition to length, another important property of histones is their high isoelectric point. Since high isoelectric point is caused by the high abundance of positively charged amino acids that are often involved in interaction with DNA, differences in isoelectric point could indicate possible differences in nucleosome binding properties. Our analysis found that H3 and H4 histones from many of the protist species, such as diplomonads *G intestinalis* and *Spironucleus salmonida*, discobans *Leishmania major* and *Trypanosoma brucei* as well as the ciliate *Tetrahymena thermophila* and oomycete *P. sojae* have lower isoelectric point (Fig. 7C). Curiously, H4 isoelectric point divergence is also observed in animals, as the H4 isoelectric point from the nematode *Caenorhabditis elegans* is 11.1, which is over one standard deviation below the median of 11.4 (stdev = 0.2, min =10.6, max = 11.5). Isoelectric points of H2A and H2B show higher variation (stdev = 0.27 and 0.3 respectively). H2A histones of S*. microadriaticum* and *F. kawagutii* are 9.6 and 9.8 respectively, which is different from median of 10.4. Interestingly, some species also display histones H2A and H2B with isoelectric points higher than the median. For instance, an isoelectric point of *P. sojae* H2A is 11.1, and isoelectric point of t *Leishmania major* H2B is 11.2.

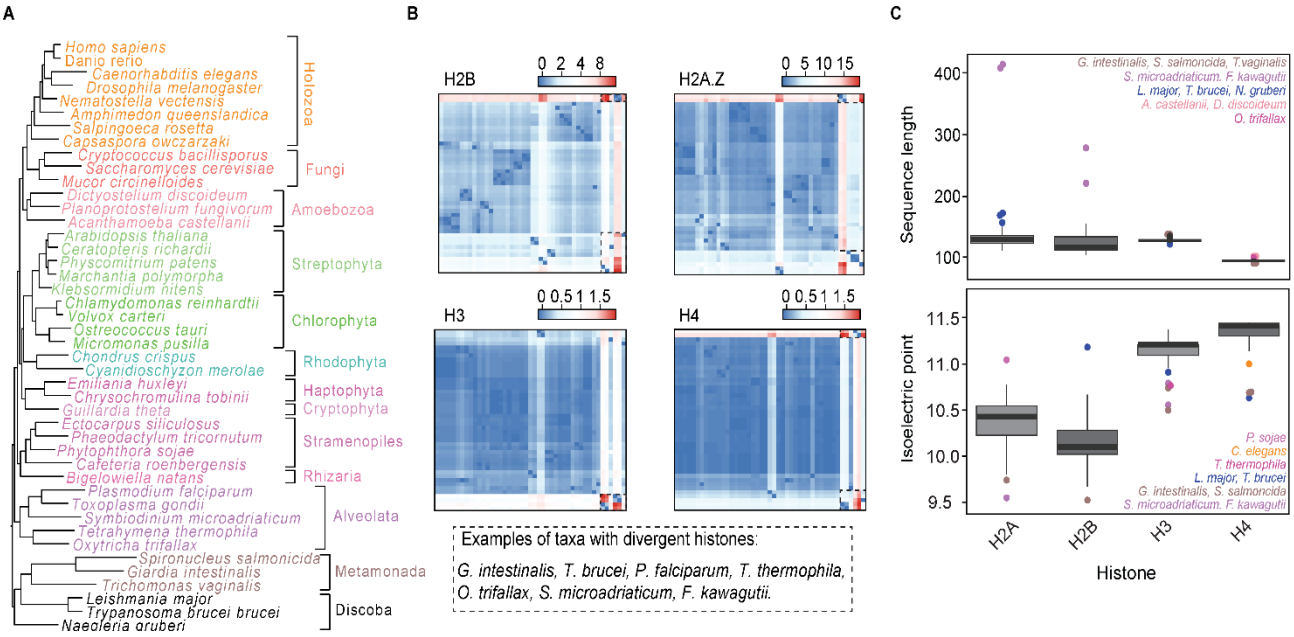

Figure 1. Histone sequence comparison. A) A phylogenetic tree of species used for the analysis. B) Heatmap of evolutionary distance between histones H2B, H2A.Z, H3 and H4 across selected species. C) Boxplot of the sequence length and isoelectric point across species.

3.2 Predicted nucleosome structures exhibit variation across eukaryotes

To assess whether sequence-level histone divergence corresponds to structural differences in the nucleosomes of different species, we sought to compare predicted nucleosome structure among exemplary species. Since resolved structures are available for only a few eukaryotic species, we predicted nucleosome structures based on previously selected histone sequences and an *E. coli* DNA sequence with ~56% GC content using AlphaFold3. To compare predicted structures, we used FoldSeek to produce local alignments. Each nucleosome was aligned against all others histone by histone. The resulting bit-scores described similarities of individual histone structures. To summarize these comparisons, I calculated median bit-scores representing structure similarity between each histone compared across different nucleosomes. Similar to sequence level divergence, most predicted structures showed a high degree of similarity (Fig.2). However, several parasitic and free-living protist species showed higher divergence. For example, several species with the most divergent nucleosome structures such as diplomonads *G. intestinalis* and *S. salmoncida* as well as trypanosomatids *T. brucei* and *L. major* also showed higher divergence at a sequence level and in the isoelectric points of their individual histones. Together with recent structural studies that found specific amino acid residues in *G. intestinalis* and *T. brucei* histones that likely cause divergence in their nucleosome structure, our analysis suggests that differences in sequence or structure may lead to strong differences in nucleosome function.

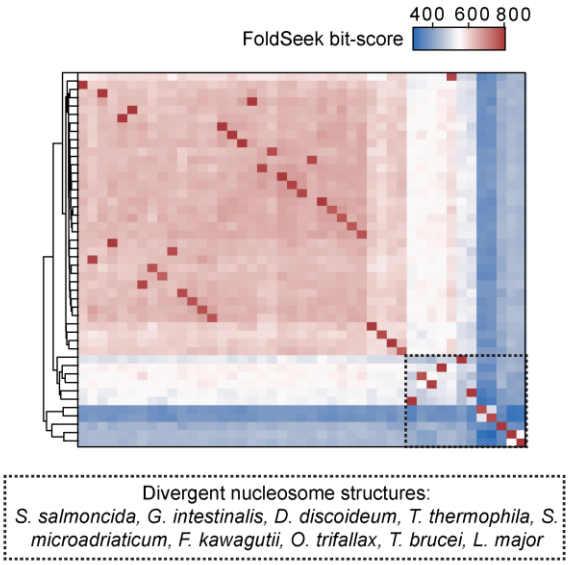

Figure 2. Nucleosome structure comparison. Clustered heatmap of similarity between predicted nucleosome structures. The structures were predicted using AlphaFold3 and aligned locally by each histone using Foldseek (v. 10.941cc33). Reported score is a calculated median of individual bit-scores per histone.

3.3 Recombinant eukaryotic histones from diverse eukaryotes form nucleosomes with distinct characteristics in *E. coli*

Given the observed sequence and structural variation, we next sought to determine whether this variation results in functional differences. To compare nucleosomes in a neutral, nucleosome-free environment, we used the novel method of ectopic nucleosome formation in *E. coli*. In 2024 it was shown that when expressed in *E. coli*, *X. laevis* histones are capable to self-assemble into nucleosomes on *E. coli* genomic DNA (Irwin & Richards, 2024; Zhou et al., 2024). For this, an expression vector containing the sequences of each core histone with individual ribosome binding sites (RBS) is transformed into *E. coli* (Fig. 3A). Using this approach, we expressed histones from 13 species across eukaryotic diversity in *E. coli* str. BL21 (DE3). Green fluorescent protein (GFP) was included as expression control. To confirm the expression of histones, we performed immunoblots against H3 and H4 (Fig. 3B). The high degree of conservation across H3 and H4 histone families allows detection of all histones belonging to the family with the same antibody. H3 and H4 production was detected in all tested samples with various intensities apart from GFP. Varying immunoblot intensity could be caused by varying antibody specificity for different protein sequences or could reflect different histone expression levels.

To assess whether expressed histones assemble into nucleosomes, we performed MNase digestion assays and inspected the resulting digested genomic DNA of *E. coli* with gel electrophoresis. I hypothesized that nucleosome formation in *E. coli* would result in partial protection of DNA against MNase digestion. Each of the samples from *E. coli* expressing histones from diverse species produced DNA fragments sized around 150 base pairs, which were not present in the sample from GFP-expressing bacteria (Fig. 3C). These results indicate that diverse nucleosomes can form on the *E. coli* genome when histones are expressed, which is consistent with previous studies that expressed eukaryotic and archaeal histones in *E. coli* (Irwin & Richards, 2024; Rojec et al., 2019; Zhou et al., 2024).

Surprisingly, this MNase digestion assay also detected distinct patterns associated with nucleosomes originating from different species. For instance, while most samples from *E. coli* expressing eukaryotic histones show distinct 150 base-pair fragments, most of them are surrounded by different amounts of smearing resulting from accumulation of DNA fragments of various size, which may be the products of ongoing DNA digestion, although these often persisted following prolonged digestion. This indicates that protection against MNase digestion varies across tested samples. However, samples from *E. coli* expressing histones of *D. rerio, S. cerevisiae* and *C. reinhardtii* contained less of these non-specific DNA fragments. Notably,

samples from *E. coli* expressing *A. thaliana, T. thermophila* and *O. trifallax* histones produced a second DNA fragment of approximately 100 base pairs that is not observed in the rest of the samples. These fragments may correspond to protection against MNase digestion by non-canonical nucleosome structures. These structures are likely to be hexasomes, since hexasomes were previously shown to protect ~100 base pairs of DNA from MNase digestion (Ho et al., 2026). 150 base-pair DNA fragments were also detected in a sample from *E. coli* expressing *F. kawagutii* (dinoflagellate) histones. Since histones are normally very lowly expressed in dinoflagellates, detection of nucleosome-bound DNA in dinoflagellates was previously unsuccessful (Gornik et al., 2012). This indicates that despite high degree of divergence in dinoflagellate histones H2A and H2B and the existence of an alternative DNA packaging system in dinoflagellates, dinoflagellate core histones can assemble into functional nucleosome. This implies that histones may still play a role in dinoflagellate chromatin biology.

Additionally, samples from *E. coli* expressing *G. intestinalis* histones produced DNA fragments sized at approximately 120 base pairs, which is consistent with a previously reported Cryo-EM structure of a *G. intestinalis* nucleosome (Sato et al., 2021). In the same sample, 200 base-pair fragments were also detected, suggesting the possible presence of higher order nucleosome structures. These results indicate that core nucleosomes from diverse eukaryotes can display distinct characteristics when assessed in a comparable nucleosome-free environment, such as *E. coli*.

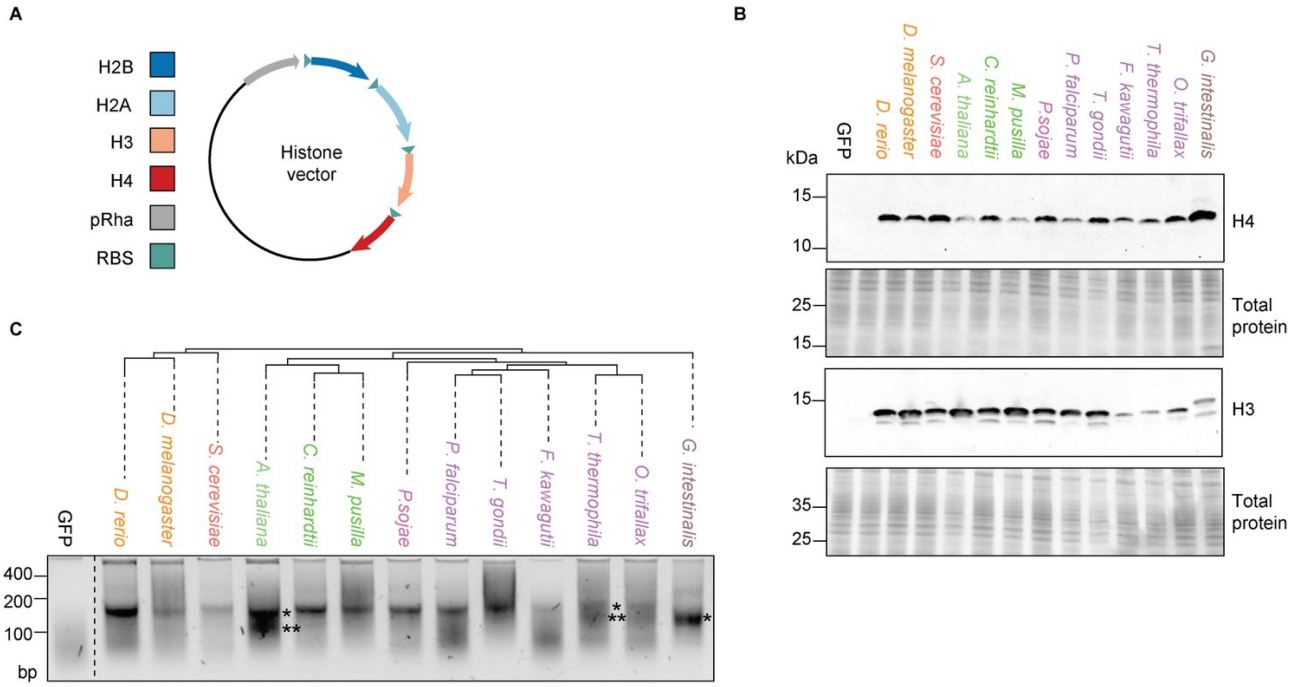

Figure 3. Recombinant histones form nucleosomes with distinct characteristics in *E. coli.* A) Schematic representation of a recombinant histone expression vector in *E. coli*. B) Immunoblots against H3 and H4 on whole-cell extracts from induced *E. coli.* C) Gel electrophoresis of genomic DNA from *E. coli* after 30 minutes of MNase digestion. 120 bp DNA fragment of *G.intestinalis* (*) and 100 bp fragments of *A. thaliana, T. thermophila* and *O. trifallax* (**) are noted. These data are representative of multiple biological replicates derived from two independent experiments. -

3.4 Genome-wide profiling of diverse nucleosomes in *E. coli* reveal functional similarities and differences

To determine whether diverse nucleosomes show distinct characteristics at a genomic level, we sequenced MNase-protected DNA from histone expressing *E. coli*. To enrich for the DNA fragments resulting from nucleosomal protection, the samples in duplicates were partially depleted of small fragments using magnetic beads and sequenced with short-read paired-end sequencing (Illumina). The reads were mapped to the *E. coli* K12 MG1655 reference genome and the similarity of the duplicates were assessed over 500 base-pair windows with Spearman correlations. All duplicates showed a high level of correlation ($r \geq 0.8$) and were merged to increase the sequencing depth. We assessed the read size distribution across samples to precisely determine the DNA fragment sizes resulting from MNase protection (Fig.4). Firstly, DNA fragments from *E. coli* expressing GFP peak at around 70 base pairs. This indicates that most of the DNA has been digested by MNase, excluding fragments that possibly correspond to DNA bound to bacterial nucleoid-associated proteins (NAPs) or small GC-rich fragments that are left

intact by the MNase. DNA fragments from *E. coli* expressing histones from *D. rerio, D. melanogaster, S. cerevisiae, C. reinhardtii, P. sojae, F. kawagutii* and *M. pusilla* have a clearly detectable peak at 140-150 base pairs, which corresponds to the DNA protected by canonical nucleosome. Similarly, the DNA from *E. coli* expressing *G. intestinalis* histones peaks around 110-120 base pairs, which corresponds to the expected size of DNA protected by *G. intestinalis* nucleosome. Additionally, DNA from *E. coli* expressing histones from *A. thaliana* and *T. thermophila* additionally peak at around 100 base pairs, which likely corresponds to the protection by non-canonical nucleosomal structures previously detected with gel electrophoresis. Lastly, samples from *E. coli* expressing histones from *P. falciparum*, *T. gondii* and *O. trifallax* do not show a pronounced peak, but rather a more extended distribution of reads that includes the length of around 150 base pairs. This is unsurprising, since samples were only partially depleted from smaller DNA fragments, and factors like nucleosome stability can contribute to different MNase digestion profiles. Various quantities of smaller fragments can be detected in all samples. Therefore, for the following analysis only the reads sized 146±15 base pairs were used. For *G. intestinalis*, reads sized 110±15 base pairs were used.

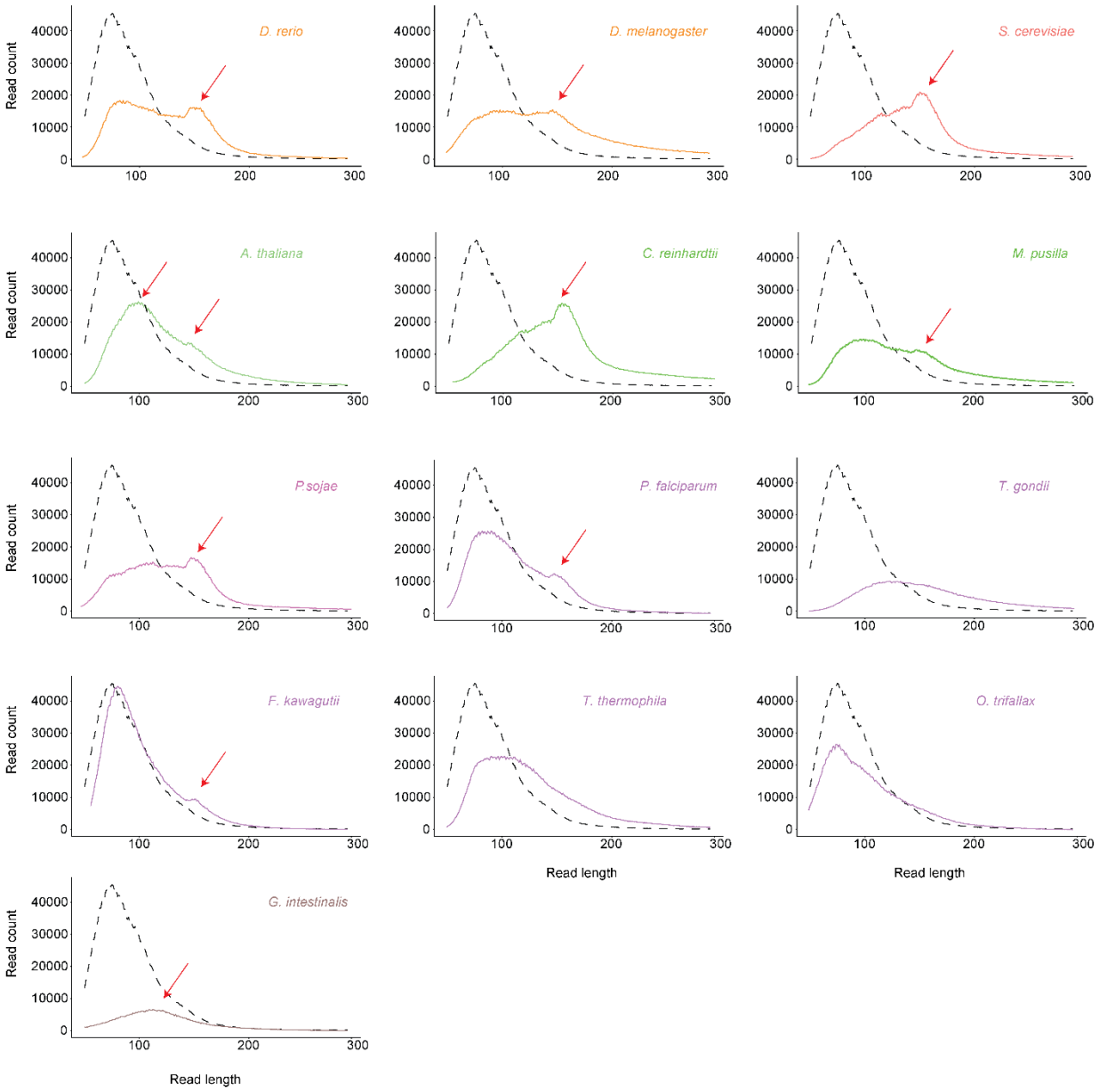

Figure 4. Diverse nucleosomes produce DNA fragments of distinct size after MNase digestion. Read length plot of each tested sample relative to the GFP negative control (dashed line). Peaks corresponding to DNA fragments resulting from nucleosomal protection against MNase are noted (red arrows).

We next sought to estimate the genome coverage similarities between samples from diverse histone-expressing *E. coli*. We found that most of the *E. coli* genome is covered by MNase reads, even though the coverage is not uniform across the genome (Fig. 5A). Closer genomic inspection revealed that the intensity and distribution of coverage is comparable across all samples except for *G. intestinalis.* Additionally, MNase coverage of all samples constitute a group of peaks flanked by small valleys, which suggests that nucleosomes may organize into arrays on the *E. coli* DNA

(Fig. 5A). To assess the overall similarity of the MNase coverage between the samples, we calculated pairwise correlations. Since MNase is known to favor AT-rich DNA sequences and therefore introduces bias, we attempted to account for it by subtracting normalized GFP signal from samples, which represents MNase digestion bias in the absence of nucleosomal protection. For GFP-subtracted samples, we calculated Spearman correlations over 500 base-pair bins (Fig. 5B). Overall, most samples showed a high degree of correlation ($r \geq 0.62$), indicating that their properties are similar and comparable. Clustering analysis highlighted species with similar characteristics. For instance, nucleosomes from *P. sojae, C. reinhardtii, P. falciparum* and *D. rerio* form a cluster together, which corresponds to the samples with a distinct peak at around 150 base pairs (Fig. 4). Another cluster is constituted by *A. thaliana*, *T. thermophila* and *O. trifallax,* which corresponds to the samples that form non-canonical nucleosomal structures (Fig. 3C). Additionally, *F. kawagutii* seemed to not group together with any of the other samples. These results suggest that nucleosomes from diverse eukaryotes exhibit similarities in their binding profiles. It is worth mentioning, however, that clustering branches were not statistically supported. A major exception was *G. intestinalis*, which showed only a weak correlation with the rest of the samples ($r \leq 0.46$), further indicating that nucleosomes formed with *G. intestinalis* histones likely have very different properties from those derived from other eukaryotes.

To explore which nucleosome binding properties differ across histone-expressing *E. coli*, we analyzed the relationship between nucleosome occupancy and GC content as well as wildtype *E. coli* gene expression. To do this, we first called 140 base-pair MNase peaks across the *E. coli* genome with DANPOS (v. 3.1.1), which will be further referred to as nucleosomes. We then calculated Spearman correlations between nucleosome occupancy and the GC content of the 140 base-pair bins corresponding to each nucleosome (Fig. 4C). The resulting correlation was high for most samples ($r \geq 0.67$), indicating that overall nucleosome formation from diverse histones may be dependent on genomic GC content. An exception to this are *G. intestinalis* nucleosomes, which showed a correlation of only 0.29, suggesting a decreased dependency on GC for nucleosome formation. Initial comparisons of the calculated GC correlation to the genomic GC content of the studied species revealed no association (Fig. 5D). However, after excluding *G. intestinalis* and *C. reinhardtii* as outliers we found that some positive correlation ($r = 0.43$) can be observed. This correlation, however, was not statistically significant (Spearman $\rho = 0.18$), which suggests that the relation between nucleosome GC preference and genomic GC content should be explored further. To determine whether nucleosome occupancy correlates with gene expression, we calculated Spearman correlations between nucleosome occupancy and RNA-seq

coverage from a publicly available *E. coli* RNA-sequencing dataset (SRR2954253) (Fig. 5C). Only weak correlations were observed across diverse nucleosomes (r_{min} = -0.14, r_{max} = 0.16), indicating that there is no direct association between nucleosome formation in *E. coli* and actively transcribed genes. However, the disparity of the correlation between samples is intriguing. Nucleosome positioning *in vivo* is known to be impacted by statistical positioning factors, such asproteins that are capable of dislocating nucleosomes when binding to DNA, e.g. RNA polymerase. Therefore, nucleosomes showing negative correlations with RNA-seq coverage, such as *D. rerio, S. cerevisiae, C. reinhardtii* and *P. sojae* may be more easily disrupted by transcriptional machinery.

While *E. coli* is naturally devoid of nucleosomes, its DNA contains plethora of nucleoid-associated proteins (NAPs), which take part in DNA compaction, cellular processes and transcriptional regulation (Schwab & Dame, 2025). *In vivo* protein occupancy display – high resolution (IPOD-HR) is a method used to accurately assess NAP occupancy across the bacterial genome while accounting for RNA Polymerase binding sites and bacterial gene expression (Freddolino et al., 2021). To assess how ectopically expressed histones in *E. coli* interact with endogenous NAPs, we obtained a publicly available normalized IPOD-HR dataset (GEO accession GSE142291). We calculated MNase coverage of our samples over 5 base-pair windows and calculated Spearman correlation between MNase coverage and IPOD-HR coverage across 500 base-pair windows (Fig. 5C). All MNase samples showed weak negative correlations with IPOD-HR dataset (r_{min} = -0.29, r_{max} = -0.22). This indicates that some of the NAPs may prevent nucleosomes from forming on the *E. coli* DNA, but this effect is not strong and is comparable across diverse nucleosomes.

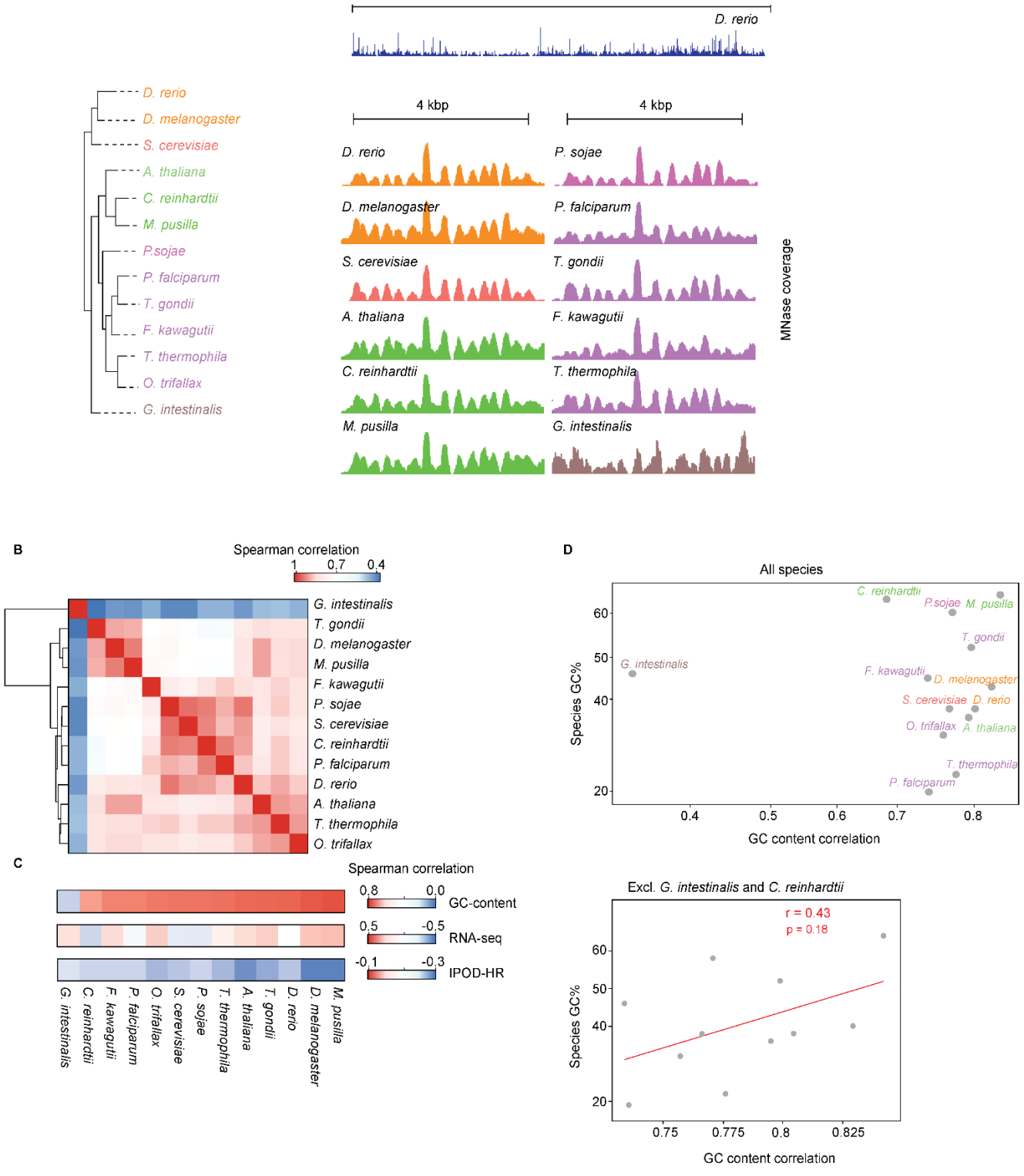

Figure 5. Genome-wide profiling reveals that diverse nucleosomes display distinct and comparable characteristics. A) Smoothed MNase coverage over the whole *E. coli* genome (above) and over a 4 kilobase-pair window. B) Clustered heatmap of pairwise Spearman correlations across 500 base-pair windows of normalized, size-selected, GFP-subtracted samples. C) Heatmaps of Spearman correlation between nucleosome occupancy called by

DANPOS and GC content of the corresponding genomic bin, RNA-seq coverage, and IPOD-HR coverage. D) Scatterplot demonstrating relation between MNase GC content correlation and genomic GC content of species included in the study (above). Scatterplot demonstrating correlation with two outlier samples excluded (below).

3.5 Diverse nucleosomes display differences in sequence preference and spacing

To assess the sequence preference of diverse nucleosomes expressed in *E. coli*, we calculated dinucleotide frequencies for well-positioned nucleosomes with occupancy above the 25th percentile. Well-positioned nucleosomes were defined as those with fuzziness below 20th percentile. Similar to previous reports based on *in vitro* and *in vivo* experiments, we detected a high preference for GG/GC/CG/CC dinucleotides, which would peak approximately every 10-11 base pairs (Fig. 6A). Interestingly, the observed patterns display more similarity to dinucleotide frequencies preferences observed in vivo which are also impacted by chromatin factors, rather than what is observed *in vitro* ones which are thought to only reflect intrinsic nucleosomal sequence preference (Kaplan et al., 2009). This pattern was observed across all tested species to various degrees. This includes *G. intestinalis,* despite the overall lower affinity for GG/GC/CG/CC dinucleotides. To assess whether sequence preference remains the same for nucleosomes with different occupancy levels, we calculated the dinucleotide dependency as the sum of dinucleotide frequency differences (GG/GC/CG/CC – AA/AT/TA/TT) across different percentile windows of nucleosome occupancy (Fig. 6B). We found that high nucleosome occupancy is directly connected to the dinucleotide enrichment of the underlying sequence, while nucleosomes with low occupancy tend to have AT-rich sequences, a pattern observed across all tested nucleosomes. Curiously, *G. intestinalis* nucleosomes with weak occupancy are also associated with GC-poor sequences, but as nucleosome occupancy increases, periodicity preference grows only moderately. This data indicates that nucleosomes can form across any available DNA sequence, but GC-rich sequences are strongly preferred for nucleosome formation in all species. The degree of this sequence preference, however, varies across species and in the case of *G. intestinalis*, enrichment for GC dinucleotides stops being preferential for nucleosome formation past a certain threshold.

To determine whether diverse nucleosomes can form array-like structures on the *E. coli* genome, we also assessed the distance between nucleosomes within 300 base pairs one of eachother. We found that all diverse nucleosomes form array-like structures indicated by consistent spacing between nucleosomes (Fig. 6C). Additionally, we calculated the distance separating nucleosome peaks, also known as a linker distance. We found that on average nucleosome peaks are separated by 56-73 base pairs (Fig. 6C). The differences between samples were found to be

significant by one-way ANOVA (Analysis of Variance) (F = 2.879, Pr(>F) = 2.04e-06). To explore which samples significantly differ from one another, we applied a TukeyHSD (Tukey's Honestly Significant Difference) post-hoc test and found that samples with similar MNase profiles tend to fall into the same statistical groups. For instance, *S. cerevisiae, C. reinhardtii, D. rerio* and *P. sojae* nucleosomes*,* which exhibited similarities in previous analyses (Fig. 5B-C), share a statistical group *b,* which means that the variance between these groups is not significant. This can also be observed between *A. thaliana, T. thermophila* an *O. trifallax*, which share group *e.* Curiously, average linker distance observed across histone-expressing *E. coli* is significantly larger than the genome-wide linker distance observed *in vivo* for *S. cerevisiae* (Wiese et al., 2019). This disparity could indicate that on each individual DNA strand, nucleosomes are not forming throughout the entire genomic sequence. Alternatively, this increased spacing could be an effect specific to nucleosome assembly in *E. coli*.

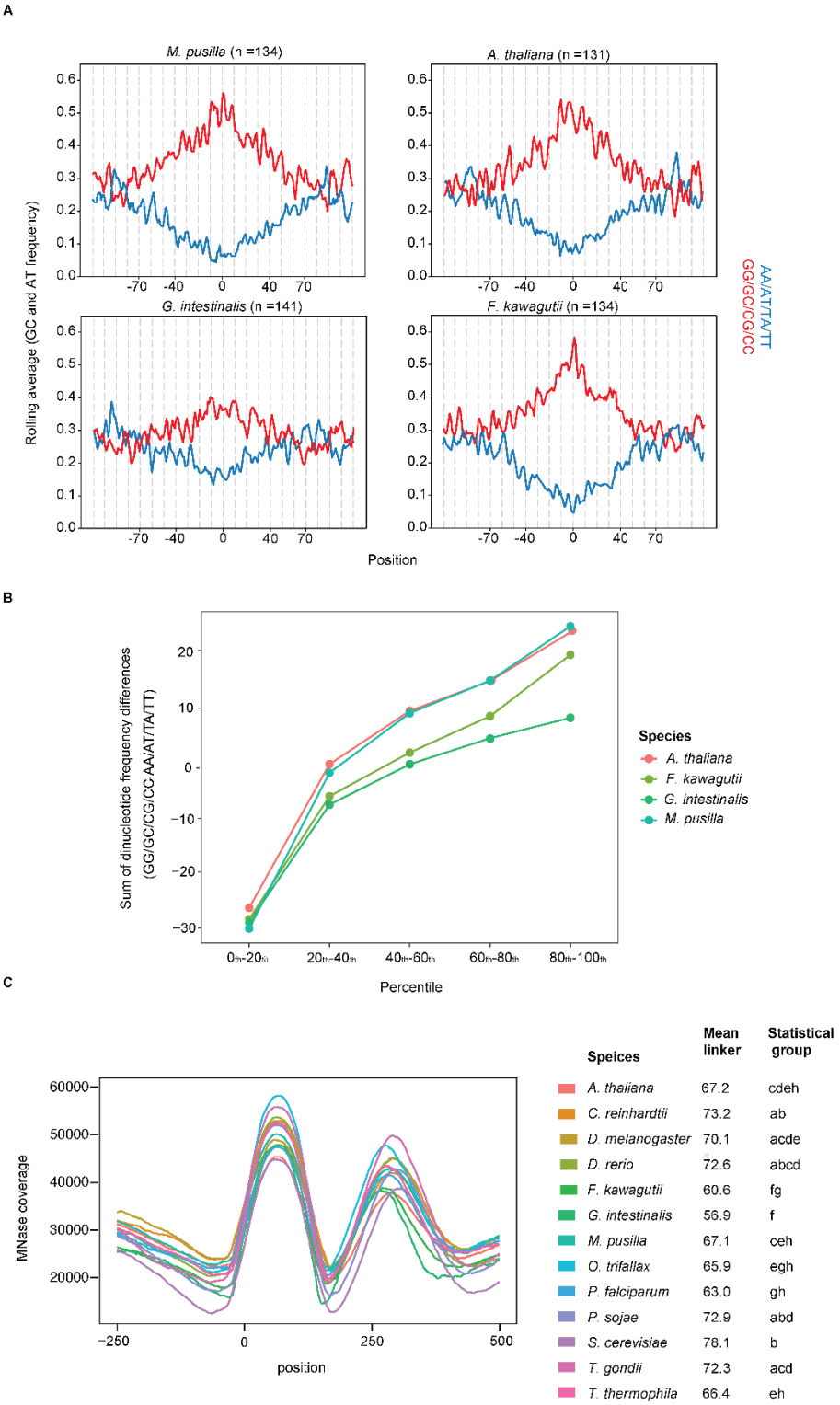

Figure 6. Diverse nucleosomes display variation in sequence preference and spacing. A) Dinucleotide frequency plots of well-positioned nucleosomes for selected samples. B) Sum of dinucleotide frequency differences across increasing nucleosome occupancy percentile windows for selected samples. C) Average plot of MNase coverage over regions with adjacent nucleosomes. Calculated mean linker distance and the results of the ANOVA test followed by Tukey's HSD are shown.

4 Discussion

4.1 Core nucleosomes from diverse eukaryotes display variation

Histones are some of the most conserved eukaryotic proteins. However, there are known cases of unicellular organisms with divergent core histone sequences which impact nucleosome structure and possibly nucleosome function. To our knowledge, no comprehensive comparative studies on intrinsic nucleosome properties have been done before. Here, we demonstrated that histones from eukaryotic diversity are capable of assembly into nucleosome on a naïve *E. coli* genome and display distinct characteristics detectable by MNase assays. Additional genomic profiling by MNase-seq revealed that diverse nucleosomes differ in their stability, sequence preference, linker distance and nucleosome particle size (Fig. 7). Based on our analysis, we were able to identify a group of species with well-defined nucleosomes that are characterized by the accumulation of specific150 base-pair DNA fragments not digested by MNase, anticorrelation with gene expression, and similarities in sequence preference and positioning. This study has also detected DNA fragments of ~120 base pairs in samples expressing *G. intestinalis* histones, which is in line with observations from the *G. intestinalis* nucleosome Cryo-EM structure (Sato et al., 2021). This confirms that the canonical nucleosome wrapping ~146 base pairs of DNA is not a universal eukaryotic trait. Additionally, published Cryo-EM structures of the *T. brucei* nucleosome resolved only 126 base pairs of DNA (Deák et al., 2023) and power spectrum analysis of the *Trichomonas vaginalis* genome followed by MNase digestion assay suggests the possible existence of ~120 bp nucleosome repeats (K. Chen et al., 2008). Together, this evidence suggests that nucleosome particles of alternative size may be a recurring trait across different parasitic lineages.

During this study, we also were able to detect nucleosome formation in *E. coli* expressing *F. kawagutii* histones. *F. kawagutii* belongs to the dinoflagellate linage, which is known to compact DNA primarily with viral and bacterial nucleoproteins, which makes their chromatin stand out among other eukaryotes. Low levels of histone expression across dinoflagellates complicates the study of dinoflagellate histones *in vivo*. However, ectopic expression of dinoflagellate histones in *E. coli* reveals their capacity to assemble into nucleosome with binding properties similar to other eukaryotic species, raising important questions about the role of these nucleosomes in this unique lineage.

Additionally, we detected non-canonical nucleosome structures forming in samples from *E. coli* expressing *A. thaliana, O. trifallax* and *T. thermophila* histones. Based on the DNA fragment size, these structures are likely to be hexasomes. However, structures would need to be isolated and

analyzed to confirm their actual protein composition. Similar structures were also observed in samples from *E. coli* expressing *X. laevis* histones, and their identity as hexasomes were confirmed with Cryo-EM analysis (Ho et al., 2026). *In vivo*, hexasomes are suggested to be involved in regulation of transcription, since loss of H2A-H2B dimers increases nucleosome accessibility (Kaur et al., 2024). While the biological meaning of hexasome formation over *E. coli* genome is unclear, its inconsistent presence across diverse eukaryotes indicates a potential difference in H2A-H2B dimer stability across eukaryotic lineages. This is further supported by the reccurring presence of hexasomes in samples expressing ciliate histones (*O. trifallax*, *T. thermpohila*), suggesting that this property may be shared across the ciliate lineage. Since *A. thaliana* is the only streptophyte that was included in this study, it is unknown whether histones of other land plant species readily produce hexasomes. MNase profiling of *E. coli* expressing histones from across plant diversity may reveal more insights about nucleosome particle stability across eukaryotes.

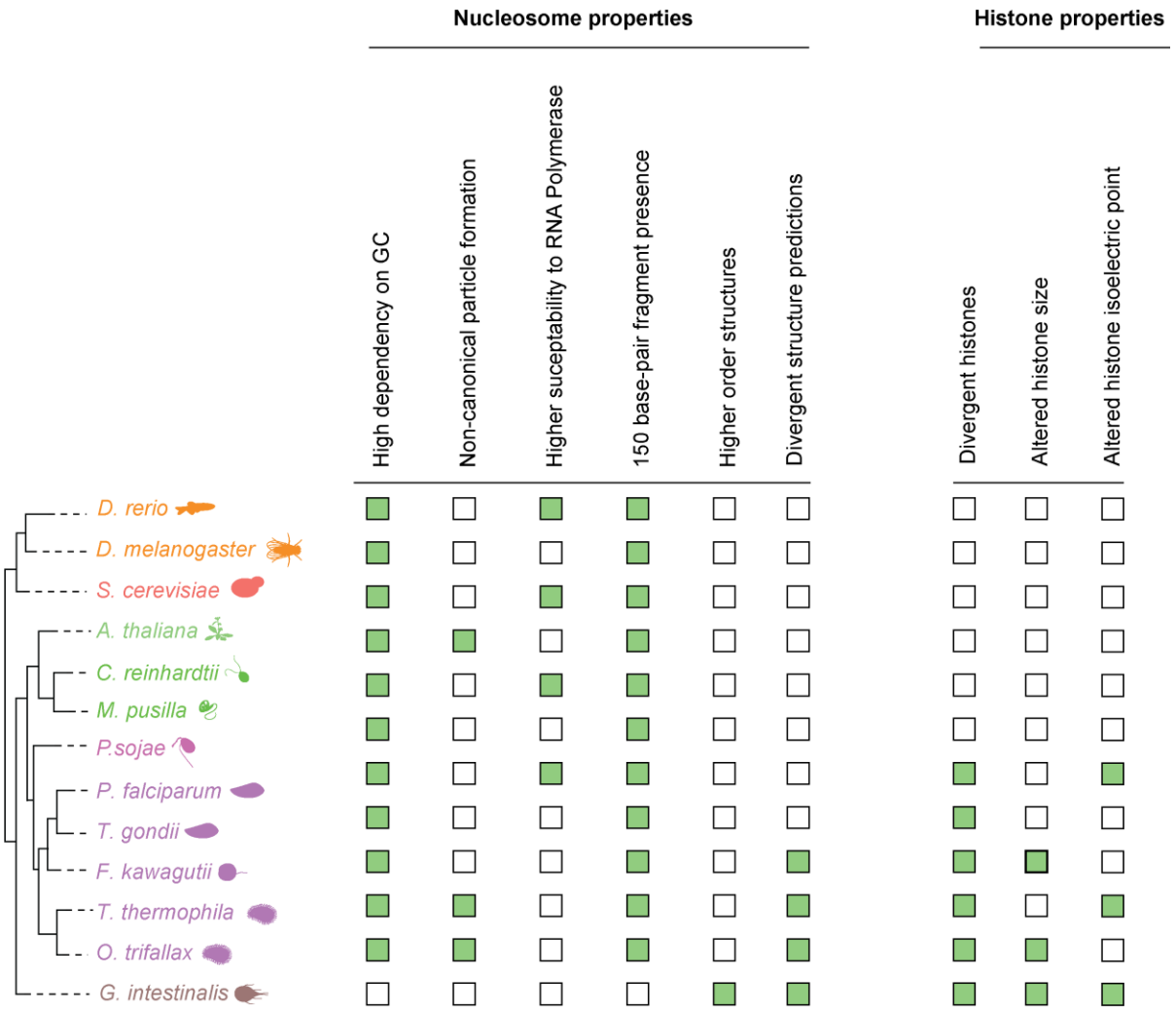

Figure 7. Schematic overview of intrinsic properties of nucleosomes and histones from diverse eukaryotes.

Somewhat surprisingly, similarities in binding profiles of diverse nucleosomes do not reflect phylogenetic relationships between species (Fig. 7). This can be partially explained by the small subset of distantly related species selected for this study. Increasing the sample size of related eukaryotic lineages may reveal similarities in nucleosome properties between related species. Additionally, increasing sequencing depth may help capture more subtle effects that were difficult to disseminate in this study, such as the correlation between nucleosome GC-preference and genomic GC content.

Another observation was that the average linker distance across all samples appeared to be significantly larger than the linker distance reported for *S. cerevisiae* nucleosomes *in vivo* (Wiese et al., 2019). One possible explanation would be that any given *E. coli* genome is not fully covered

by nucleosomes, since MNase-seq coverage enables estimations of average nucleosome positioning across population of cells but does reveal the positioning of nucleosomes on an individual DNA molecule. Alternatively, this disparity may be a result of the nucleosome assembly process in *E. coli*. The opposite effect can usually be observed *in vitro* as reconstituted nucleosomes tend to stack into oligomers with shorter linker distance (Y. Zhang et al., 2009). To resolve this, future work will need to assess nucleosome positioning on a single copy of the *E. coli* genome using long-read technologies such as chromatin fiber sequencing (Stergachis et al., 2020). Additionally, introducing chromatin remodelers which regulate nucleosome positioning (e.g. ISWI) into histone-expressing *E. coli* may elucidate differences between nucleosome positioning in *E. coli* versus *in vivo*.

Despite these differences, we found that most samples from *E. coli* expressing histones of diverse species showed a high correlation between MNase coverage and DNA GC content. The only exception were nucleosomes from *G. intestinalis,* a parasitic species with divergent histone sequences. Previously reported Cryo-EM analysis showed that *G. intestinalis* nucleosomes have weaker interactions with DNA in comparison to *H. sapiens* nucleosomes. Since GC dinucleotide periodicity is generally required for nucleosome stability, the loss for GC affinity in *G. intestinalis* nucleosomes may be connected to their weaker DNA binding strength. No correlation between GC preference of the diverse nucleosomes and genomic GC content of the studied species was found at first. After exclusion of *G. intestinalis* and *C. reinahardtii,* which showed the weakest correlations between nucleosome occupancy and GC content, we saw a weak positive correlation. However, this correlation was not significant. Since this analysis included only 11 species, this effect could be caused by low sample number. Including more species may help increase statistical power. Additionally, genomic GC content found in the literature was chosen for comparison. This approach is somewhat arbitrary, since the GC content may have been calculated with different methods (e.g. over different genomic windows). It is also worth mentioning that some species have heterogenous GC content across different genomic elements, such as promoters and intergenic regions. Therefore, interpreting the correlation between nucleosome sequence preference and genomic GC content will require further investigations. Nonetheless, unraveling the connection between genomic architecture and intrinsic properties of diverse nucleosomes may enable us to connect our heterologous observations with the native biology of the organisms from which these nucleosomes were derived.

4.2 Ectopic expression of histones in *E. coli* may serve as a complementary or alternative method to *in vitro* nucleosome reconstitution

In scope of this study, we observed that some of the nucleosome binding properties detected in samples from histone-expressing *E. coli* showed more similarity to the properties observed in nucleosomes formed *in vivo* rather than *in vitro.* For instance, in our analysis of sequence preference of nucleosomes formed over the *E. coli* genome we found that diverse nucleosomes maintain previously reported dinucleotide periodicity (GG/GC/CG/CC and AA/AT/TA/TT). Interestingly, dinucleotide periodicity observed in samples from *E. coli* expressing histones showed more resemblance to the periodicity patterns occurring *in vivo*. It is widely accepted that *in vivo* dinucleotide periodicity patterns are impacted by chromatin remodelers, while *in vitro* dinucleotide periodicity reflect intrinsic nucleosome sequence preferences. Even though *E. coli* is devoid of chromatin remodelers, it still constitutes a living system with active cellular processes such as DNA replication, gene expression, etc., therefore nucleosome positioning over *E. coli* genome will be additionally affected by statistical factors resulting from cellular processes. However, nucleosome positioning data from *E. coli* may also provide data that is more relevant for an organismal context than that originating from *in vitro* experiments, since it puts histones in a physiological context. For example, it is worth noting that *in vitro* nucleosome reconstitution requires salt dialysis starting at 2M NaCl. While it is widely assumed that *in vitro* reconstituted nucleosomes reflect intrinsic nucleosomal properties, it is possible that such high NaCl concentrations would bias towards only strong histone interactions with DNA. In fact, it was previously shown that histones do not initially assemble at their most thermodynamically favorable sites during *in vitro* assembly and undergo thermal repositioning later (Flaus & Owen-Hughes, 2003). Therefore, while *in vitro* methods are indispensable for structural characterization of histones, ectopic expression of histones in *E. coli* may be a more suitable model for studying nucleosome binding properties.

Despite the enhanced physiological relevance of heterologous nucleosome production in *E.* coli. these experiments have some key limitations. Although histones are expressed polycistronically, there is a possibility that histone expression stoichiometry is not equal between constructs, which complicates comparison of binding properties between samples. However, assessing expression stoichiometry by immunochemistry (e.g. immunoblot, immunoprecipitation followed by mass spectrometry) would not elucidate differences in expression between histones constructs due to differences in antibody specificity. However, it is possible to assess whether different histone expression levels cause differences in nucleosome properties. This can be achieved by testing different inducer concentrations for *E. coli* histone expression to determine whether different

histone expression levels have an impact on nucleosome properties. Additionally, integrating histone genes into *E. coli* genome instead of using an expression vector would promote more stable and uniform histone expression across diverse samples.

4.3 Conclusions and future directions

So far, we have managed to demonstrate that diverse nucleosomes can be heterologously assembled in *E.* coli where they display different binding properties when compared in nucleosome-free environment. Expanding this study to more diverse eukaryotes would allow us to characterize the differences between eukaryotic species more precisely and eventually disseminate lineage-specific nucleosome traits. Previously mentioned studies of *T. brucei* and *G. intestinalis* nucleosome structures revealed divergent amino acid residues in histones that lead to structural differences of nucleosome when compared to *Homo sapiens* nucleosome. Using point mutagenesis across histone sequences of diverse eukaryotes, we could characterize which amino acids impact binding properties of diverse nucleosomes in a high-throughput manner using ectopic expression in *E. coli*. Moving forward, this approach could enable us to build a comprehensive overview of the evolution and functional diversification of the nucleosome across the eukaryotic tree of life, challenging the prevailing view that the core nucleosome is evolutionarily static.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., … Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, *630*(8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w

Ahmad, K., & Henikoff, S. (2002). The Histone Variant H3.3 Marks Active Chromatin by Replication-Independent Nucleosome Assembly. *Molecular Cell*, *9*(6), 1191–1200. https://doi.org/10.1016/S1097-2765(02)00542-7

Bell, S. P., & Labib, K. (2016). Chromosome Duplication in *Saccharomyces cerevisiae*. *Genetics*, *203*(3), 1027–1067. https://doi.org/10.1534/genetics.115.186452

Bowman, G. D., & Poirier, M. G. (2015). Post-translational modifications of histones that influence nucleosome dynamics. *Chemical Reviews*, *115*(6), 2274–2295. https://doi.org/10.1021/cr500350x

Broad Institute. (2019). *Picard Toolkit* (Version 3.2.0) [Computer software].

Brown, M. T., & McMurray, M. A. (2025). Stepwise order in protein complex assembly: Approaches and emerging themes. *Open Biology*, *15*(1), 240283. https://doi.org/10.1098/rsob.240283

Cairns, B. R. (2009). The logic of chromatin architecture and remodelling at promoters. *Nature*, *461*(7261), 193–198. https://doi.org/10.1038/nature08450

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chen, K., Meng, Q., Ma, L., Liu, Q., Tang, P., Chiu, C., Hu, S., & Yu, J. (2008). A novel DNA sequence periodicity decodes nucleosome positioning. *Nucleic Acids Research*, *36*(19), 6228–6236. https://doi.org/10.1093/nar/gkn626

Chen, K., Xi, Y., Pan, X., Li, Z., Kaestner, K., Tyler, J., Dent, S., He, X., & Li, W. (2013). DANPOS: Dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Research*, *23*(2), 341–351. https://doi.org/10.1101/gr.142067.112

Chen, S. (2025). fastp 1.0: An ultra-fast all-round tool for FASTQ data quality control and preprocessing. *iMeta*, *4*(5), e70078. https://doi.org/10.1002/imt2.70078

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & De Hoon, M. J. L. (2009). Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, *25*(11), 1422–1423. https://doi.org/10.1093/bioinformatics/btp163

Costantini, M., & Musto, H. (2017). The Isochores as a Fundamental Level of Genome Structure and Organization: A General Overview. *Journal of Molecular Evolution*, *84*(2–3), 93–103. https://doi.org/10.1007/s00239-017-9785-9

Dan, J., & Chen, T. (2023). Writers, erasers, and readers of DNA and histone methylation marks. In *Epigenetic Cancer Therapy* (pp. 39–63). Elsevier. https://doi.org/10.1016/B978-0-323-91367-6.00012-X

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

De Lima, L. P., Poubel, S. B., Yuan, Z.-F., Rosón, J. N., Vitorino, F. N. D. L., Holetz, F. B., Garcia, B. A., & Da Cunha, J. P. C. (2020). Improvements on the quantitative analysis of Trypanosoma cruzi histone post translational modifications: Study of changes in epigenetic marks through the parasite's metacyclogenesis and life cycle. *Journal of Proteomics*, *225*, 103847. https://doi.org/10.1016/j.jprot.2020.103847

Deák, G., Wapenaar, H., Sandoval, G., Chen, R., Taylor, M. R. D., Burdett, H., Watson, J. A., Tuijtel, M. W., Webb, S., & Wilson, M. D. (2023). Histone divergence in trypanosomes results in unique alterations to nucleosome structure. *Nucleic Acids Research*, *51*(15), 7882–7899. https://doi.org/10.1093/nar/gkad577

DeLange, R. J., Fambrough, D. M., Smith, E. L., & Bonner, J. (1969). Calf and Pea Histone IV. *Journal of Biological Chemistry*, *244*(2), 319–334. https://doi.org/10.1016/S0021-9258(18)94434-0

Derelle, E., Ferraz, C., Rombauts, S., Rouzé, P., Worden, A. Z., Robbens, S., Partensky, F., Degroeve, S., Echeynié, S., Cooke, R., Saeys, Y., Wuyts, J., Jabbari, K., Bowler, C., Panaud, O., Piégu, B., Ball, S. G., Ral, J.-P., Bouget, F.-Y., … Moreau, H. (2006). Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proceedings of the National Academy of Sciences*, *103*(31), 11647–11652. https://doi.org/10.1073/pnas.0604795103

Dubendorf, J. W., & Studier, F. W. (1991). Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *Journal of Molecular Biology*, *219*(1), 45–59. https://doi.org/10.1016/0022-2836(91)90856-2

Eaton, M. L., Galani, K., Kang, S., Bell, S. P., & MacAlpine, D. M. (2010). Conserved nucleosome positioning defines replication origins. *Genes & Development*, *24*(8), 748–753. https://doi.org/10.1101/gad.1913210

Flaus, A. (2006). Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Research*, *34*(10), 2887–2905. https://doi.org/10.1093/nar/gkl295

Flaus, A. (2011). Principles and practice of nucleosome positioning *in vitro*. *Frontiers in Life Science*, *5*(1–2), 5–27. https://doi.org/10.1080/21553769.2012.702667

Flaus, A., & Owen-Hughes, T. (2003). Dynamic Properties of Nucleosomes during Thermal and ATP-Driven Mobilization. *Molecular and Cellular Biology*, *23*(21), 7767–7779. https://doi.org/10.1128/MCB.23.21.7767-7779.2003

Franzén, O., Jerlström-Hultqvist, J., Castro, E., Sherwood, E., Ankarklev, J., Reiner, D. S., Palm, D., Andersson, J. O., Andersson, B., & Svärd, S. G. (2009). Draft Genome Sequencing of Giardia intestinalis Assemblage B Isolate GS: Is Human Giardiasis Caused by Two Different Species? *PLoS Pathogens*, *5*(8), e1000560. https://doi.org/10.1371/journal.ppat.1000560

Freddolino, L., Amemiya, H. M., Goss, T. J., & Tavazoie, S. (2021). Dynamic landscape of protein occupancy across the Escherichia coli chromosome. *PLOS Biology*, *19*(6), e3001306. https://doi.org/10.1371/journal.pbio.3001306

Gornik, S. G., Ford, K. L., Mulhern, T. D., Bacic, A., McFadden, G. I., & Waller, R. F. (2012). Loss of Nucleosomal DNA Condensation Coincides with Appearance of a Novel Nuclear Protein in Dinoflagellates. *Current Biology*, *22*(24), 2303–2312. https://doi.org/10.1016/j.cub.2012.10.036

Grau-Bové, X., Navarrete, C., Chiva, C., Pribasnig, T., Antó, M., Torruella, G., Galindo, L. J., Lang, B. F., Moreira, D., López-Garcia, P., Ruiz-Trillo, I., Schleper, C., Sabidó, E., & Sebé-Pedrós, A. (2022). A phylogenetic and proteomic reconstruction of eukaryotic chromatin evolution. *Nature Ecology & Evolution*, *6*(7), 1007–1023. https://doi.org/10.1038/s41559-022-01771-6

Guhl, F., & Vallejo, G. A. (2003). Trypanosoma (Herpetosoma) rangeli Tejera, 1920: An updated review. *Memórias Do Instituto Oswaldo Cruz*, *98*(4), 435–442. https://doi.org/10.1590/S0074-02762003000400001

Haas, BJ. (2023). *TransDecoder* (Version 5.7.1) [Computer software]. https://github.com/TransDecoder/TransDecoder

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

He, J., Huang, Y., Li, L., Lin, S., Ma, M., Wang, Y., & Lin, S. (2024). Novel Plastid Genome Characteristics in *Fugacium kawagutii* and the Trend of Accelerated Evolution of

Plastid Proteins in Dinoflagellates. *Genome Biology and Evolution*, *16*(1), evad237. https://doi.org/10.1093/gbe/evad237

Herold, J., Kurtz, S., & Giegerich, R. (2008). Efficient computation of absent words in genomic sequences. *BMC Bioinformatics*, *9*(1), 167. https://doi.org/10.1186/1471-2105-9-167

Hino, S., Sato, T., & Nakao, M. (2023). Chromatin Immunoprecipitation Sequencing (ChIP-seq) for Detecting Histone Modifications and Modifiers. In I. Hatada & T. Horii (Eds.), *Epigenomics* (Vol. 2577, pp. 55–64). Springer US. https://doi.org/10.1007/978-1-0716-2724-2_4

Ho, C.-H., Kobayashi, Y., Ogasawara, M., Takizawa, Y., & Kurumizaka, H. (2026). A method for cryo-EM analysis of eukaryotic nucleosomes reconstituted in bacterial cells. *iScience*, *29*(1), 114453. https://doi.org/10.1016/j.isci.2025.114453

Hocher, A., & Warnecke, T. (2024). Nucleosomes at the Dawn of Eukaryotes. *Genome Biology and Evolution*, *16*(3), evae029. https://doi.org/10.1093/gbe/evae029

Irwin, N. A. T., & Richards, T. A. (2024). Self-assembling viral histones are evolutionary intermediates between archaeal and eukaryotic nucleosomes. *Nature Microbiology*, *9*(7), 1713–1724. https://doi.org/10.1038/s41564-024-01707-9

Jiang, N., Wang, L., Chen, J., Wang, L., Leach, L., & Luo, Z. (2014). Conserved and divergent patterns of DNA methylation in higher vertebrates. *Genome Biology and Evolution*, *6*(11), 2998–3014. https://doi.org/10.1093/gbe/evu238

Jung, S., Swart, E. C., Minx, P. J., Magrini, V., Mardis, E. R., Landweber, L. F., & Eddy, S. R. (2011). Exploiting Oxytricha trifallax nanochromosomes to screen for non-coding

RNA genes. *Nucleic Acids Research*, *39*(17), 7529–7547. https://doi.org/10.1093/nar/gkr501

Kagalwala, M. N., Glaus, B. J., Dang, W., Zofall, M., & Bartholomew, B. (2004). Topography of the ISW2–nucleosome complex: Insights into nucleosome spacing and chromatin remodeling. *The EMBO Journal*, *23*(10), 2092–2104. https://doi.org/10.1038/sj.emboj.7600220

Kaplan, N., Hughes, T. R., Lieb, J. D., Widom, J., & Segal, E. (2010). Contribution of histone sequence preferences to nucleosome organization: Proposed definitions and methodology. *Genome Biology*, *11*(11), 140. https://doi.org/10.1186/gb-2010-11-11-140

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., & Segal, E. (2009a). The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, *458*(7236), 362–366. https://doi.org/10.1038/nature07667

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., & Segal, E. (2009b). The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, *458*(7236), 362–366. https://doi.org/10.1038/nature07667

Kasinsky, H. E., Lewis, J. D., Dacks, J. B., & Ausló, J. (2001). Origin of H1 linker histones. *The FASEB Journal*, *15*(1), 34–42. https://doi.org/10.1096/fj.00-0237rev

Katoh, K. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Kaur, U., Muñoz, E. N., & Narlikar, G. J. (2024). Hexasomal particles: Consequence or also consequential? *Current Opinion in Genetics & Development*, *85*, 102163. https://doi.org/10.1016/j.gde.2024.102163

Kensche, P. R., Hoeijmakers, W. A. M., Toenhake, C. G., Bras, M., Chappell, L., Berriman, M., & Bártfai, R. (2016a). The nucleosome landscape of *Plasmodium falciparum* reveals chromatin architecture and dynamics of regulatory sequences. *Nucleic Acids Research*, *44*(5), 2110–2124. https://doi.org/10.1093/nar/gkv1214

Kensche, P. R., Hoeijmakers, W. A. M., Toenhake, C. G., Bras, M., Chappell, L., Berriman, M., & Bártfai, R. (2016b). The nucleosome landscape of *Plasmodium falciparum* reveals chromatin architecture and dynamics of regulatory sequences. *Nucleic Acids Research*, *44*(5), 2110–2124. https://doi.org/10.1093/nar/gkv1214

Kornberg, R. D., & Lorch, Y. (1999). Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome. *Cell*, *98*(3), 285–294. https://doi.org/10.1016/S0092-8674(00)81958-3

Kornberg, R. D., & Lorch, Y. (2020). Primary Role of the Nucleosome. *Molecular Cell*, *79*(3), 371–375. https://doi.org/10.1016/j.molcel.2020.07.020

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., & Carey, V. J. (2013). Software for Computing and Annotating Genomic Ranges.

PLoS Computational Biology, *9*(8), e1003118. https://doi.org/10.1371/journal.pcbi.1003118

Lee, J. Y., Lee, J., Yue, H., & Lee, T.-H. (2015). Dynamics of Nucleosome Assembly and Effects of DNA Methylation. *Journal of Biological Chemistry*, *290*(7), 4291–4303. https://doi.org/10.1074/jbc.M114.619213

Lee, Y.-T., Gibbons, G., Lee, S. Y., Nikolovska-Coleska, Z., & Dou, Y. (2015). One-pot refolding of core histones from bacterial inclusion bodies allows rapid reconstitution of histone octamer. *Protein Expression and Purification*, *110*, 89–94. https://doi.org/10.1016/j.pep.2015.02.007

Li, B., Carey, M., & Workman, J. L. (2007). The Role of Chromatin during Transcription. *Cell*, *128*(4), 707–719. https://doi.org/10.1016/j.cell.2007.01.015

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Lin, S. (2024). A decade of dinoflagellate genomics illuminating an enigmatic eukaryote cell. *BMC Genomics*, *25*(1), 932. https://doi.org/10.1186/s12864-024-10847-5

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, *389*(6648), 251–260. https://doi.org/10.1038/38444

Maree, J. P., Tvardovskiy, A., Ravnsborg, T., Jensen, O. N., Rudenko, G., & Patterton, H.-G. (2022). *Trypanosoma brucei* histones are heavily modified with combinatorial post-translational modifications and mark Pol II transcription start regions with

hyperacetylated H2A. *Nucleic Acids Research*, *50*(17), 9705–9723. https://doi.org/10.1093/nar/gkac759

Marinov, G. K., & Lynch, M. (2016). Diversity and Divergence of Dinoflagellate Histone Proteins. *G3 Genes|Genomes|Genetics*, *6*(2), 397–422. https://doi.org/10.1534/g3.115.023275

Maruyama, H., Harwood, J. C., Moore, K. M., Paszkiewicz, K., Durley, S. C., Fukushima, H., Atomi, H., Takeyasu, K., & Kent, N. A. (2013). An alternative beads-on-a-string chromatin architecture in *Thermococcus kodakarensis*. *EMBO Reports*, *14*(8), 711–717. https://doi.org/10.1038/embor.2013.94

McGinty, R. K., & Tan, S. (2016). Recognition of the nucleosome by chromatin factors and enzymes. *Current Opinion in Structural Biology*, *37*, 54–61. https://doi.org/10.1016/j.sbi.2015.11.014

McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., Salamov, A., Fritz-Laylin, L. K., Maréchal-Drouard, L., Marshall, W. F., Qu, L.-H., Nelson, D. R., Sanderfoot, A. A., Spalding, M. H., Kapitonov, V. V., Ren, Q., Ferris, P., Lindquist, E., … Grossman, A. R. (2007). The *Chlamydomonas* Genome Reveals the Evolution of Key Animal and Plant Functions. *Science*, *318*(5848), 245–250. https://doi.org/10.1126/science.1143609

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for

Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Navarrete, C., Montgomery, S. A., Mendieta, J., Lara-Astiaso, D., & Sebé-Pedrós, A. (2025). *Diversity and evolution of chromatin regulatory states across eukaryotes*. Evolutionary Biology. https://doi.org/10.1101/2025.03.17.643675

Naya, H., Romero, H., Carels, N., Zavala, A., & Musto, H. (2001). Translational selection shapes codon usage in the GC-rich genome of *Chlamydomonas reinhardtii*. *FEBS Letters*, *501*(2–3), 127–130. https://doi.org/10.1016/S0014-5793(01)02644-8

O'Neill, L. P., & Turner, B. M. (1996). Immunoprecipitation of chromatin. In *Methods in Enzymology* (Vol. 274, pp. 189–197). Elsevier. https://doi.org/10.1016/S0076-6879(96)74017-X

Osakabe, A., Lorković, Z. J., Kobayashi, W., Tachiwana, H., Yelagandula, R., Kurumizaka, H., & Berger, F. (2018). Histone H2A variants confer specific properties to nucleosomes and impact on chromatin accessibility. *Nucleic Acids Research*, *46*(15), 7675–7685. https://doi.org/10.1093/nar/gky540

Osakabe, A., & Molaro, A. (2023). Histone renegades: Unusual H2A histone variants in plants and animals. *Seminars in Cell & Developmental Biology*, *135*, 35–42. https://doi.org/10.1016/j.semcdb.2022.05.001

Palenchar, J. B., & Bellofatto, V. (2006). Gene transcription in trypanosomes. *Molecular and Biochemical Parasitology*, *146*(2), 135–141. https://doi.org/10.1016/j.molbiopara.2005.12.008

Peckham, H. E., Thurman, R. E., Fu, Y., Stamatoyannopoulos, J. A., Noble, W. S., Struhl, K., & Weng, Z. (2007). Nucleosome positioning signals in genomic DNA. *Genome Research*, *17*(8), 1170–1177. https://doi.org/10.1101/gr.6101007

Piganeau, G., Grimsley, N., & Moreau, H. (2011). Genome diversity in the smallest marine photosynthetic eukaryotes. *Research in Microbiology*, *162*(6), 570–577. https://doi.org/10.1016/j.resmic.2011.04.005

Postberg, J., Forcob, S., Chang, W.-J., & Lipps, H. J. (2010). The evolutionary history of histone H3 suggests a deep eukaryotic root of chromatin modifying mechanisms. *BMC Evolutionary Biology*, *10*(1), 259. https://doi.org/10.1186/1471-2148-10-259

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*(W1), W160–W165. https://doi.org/10.1093/nar/gkw257

Rippe, K., Schrader, A., Riede, P., Strohner, R., Lehmann, E., & Längst, G. (2007). DNA sequence- and conformation-directed positioning of nucleosomes by chromatin-remodeling complexes. *Proceedings of the National Academy of Sciences*, *104*(40), 15635–15640. https://doi.org/10.1073/pnas.0702430104

Rizzo, P. J., & Morris, R. L. (1983). Some properties of the histone-like protein from crypthecodinium cohnii (HCc). *Biosystems*, *16*(3–4), 211–216. https://doi.org/10.1016/0303-2647(83)90005-9

Rojec, M., Hocher, A., Stevens, K. M., Merkenschlager, M., & Warnecke, T. (2019). Chromatinization of Escherichia coli with archaeal histones. *eLife*, *8*. https://doi.org/10.7554/elife.49038

Sato, S., Takizawa, Y., Hoshikawa, F., Dacher, M., Tanaka, H., Tachiwana, H., Kujirai, T., Iikura, Y., Ho, C.-H., Adachi, N., Patwal, I., Flaus, A., & Kurumizaka, H. (2021). Cryo-EM structure of the nucleosome core particle containing *Giardia lamblia* histones. *Nucleic Acids Research*, *49*(15), 8934–8946. https://doi.org/10.1093/nar/gkab644

Schwab, S., & Dame, R. T. (2025). Identification, characterization and classification of prokaryotic nucleoid-associated proteins. *Molecular Microbiology*, *123*(3), 206–217. https://doi.org/10.1111/mmi.15298

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I. K., Wang, J.-P. Z., & Widom, J. (2006). A genomic code for nucleosome positioning. *Nature*, *442*(7104), 772–778. https://doi.org/10.1038/nature04979

Simpson, A. G. B., & Čepička, I. (2019). Amitochondriate Protists (Diplomonads, Parabasalids and Oxymonads). In *Reference Module in Life Sciences* (p. B9780128096338208881). Elsevier. https://doi.org/10.1016/B978-0-12-809633-8.20888-1

Soo, V. W. C., & Warnecke, T. (2021). Slaying the last unicorn: Discovery of histones in the microalga *Nanochlorum eucaryotum*. *Royal Society Open Science*, *8*(2), rsos.202023, 202023. https://doi.org/10.1098/rsos.202023

Stergachis, A. B., Debo, B. M., Haugen, E., Churchman, L. S., & Stamatoyannopoulos, J. A. (2020). Single-molecule regulatory architectures captured by chromatin fiber

sequencing. *Science*, *368*(6498), 1449–1454. https://doi.org/10.1126/science.aaz1646

Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, *403*(6765), 41–45. https://doi.org/10.1038/47412

Taguchi, H., Horikoshi, N., Arimura, Y., & Kurumizaka, H. (2014). A method for evaluating nucleosome stability with a protein-binding fluorescent dye. *Methods*, *70*(2–3), 119–126. https://doi.org/10.1016/j.ymeth.2014.08.019

Thatcher, T. H., & Gorovsky, M. A. (1994). Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. *Nucleic Acids Research*, *22*(2), 174–179. https://doi.org/10.1093/nar/22.2.174

Tyler, B. M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R. H. Y., Aerts, A., Arredondo, F. D., Baxter, L., Bensasson, D., Beynon, J. L., Chapman, J., Damasceno, C. M. B., Dorrance, A. E., Dou, D., Dickerman, A. W., Dubchak, I. L., Garbelotto, M., Gijzen, M., Gordon, S. G., … Boore, J. L. (2006). *Phytophthora* Genome Sequences Uncover Evolutionary Origins and Mechanisms of Pathogenesis. *Science*, *313*(5791), 1261–1266. https://doi.org/10.1126/science.1128796

Ura, K., & Kaneda, Y. (2002). Reconstitution of Chromatin In Vitro. In A. Ward (Ed.), *Genomic Imprinting* (pp. 309–325). Humana Press. https://doi.org/10.1385/1-59259-211-2:309

Van Der Heijden, T., Van Vugt, J. J. F. A., Logie, C., & Van Noort, J. (2012). Sequence-based prediction of single nucleosome positioning and genome-wide nucleosome

occupancy. *Proceedings of the National Academy of Sciences*, *109*(38). https://doi.org/10.1073/pnas.1205659109

Van Kempen, M., Kim, S. S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C. L. M., Söding, J., & Steinegger, M. (2024). Fast and accurate protein structure search with Foldseek. *Nature Biotechnology*, *42*(2), 243–246. https://doi.org/10.1038/s41587-023-01773-0

Vasilinetc, I., Prjibelski, A. D., Gurevich, A., Korobeynikov, A., & Pevzner, P. A. (2015). Assembling short reads from jumping libraries with large insert sizes. *Bioinformatics*, *31*(20), 3262–3268. https://doi.org/10.1093/bioinformatics/btv337

Wang, J., Sun, Z., Liu, H., Yue, L., Wang, F., Liu, S., Su, B., Liu, B., Kong, F., & Fang, C. (2023). Genome-Wide Identification and Characterization of the Soybean Snf2 Gene Family and Expression Response to Rhizobia. *International Journal of Molecular Sciences*, *24*(8), 7250. https://doi.org/10.3390/ijms24087250

Watzlowik, M. T., Silberhorn, E., Das, S., Singhal, R., Venugopal, K., Holzinger, S., Stokes, B., Schadt, E., Sollelis, L., Bonnell, V. A., Gow, M., Klingl, A., Marti, M., Llinás, M., Meissner, M., & Längst, G. (2025). Plasmodium blood stage development requires the chromatin remodeller Snf2L. *Nature*, *639*(8056), 1069–1075. https://doi.org/10.1038/s41586-025-08595-x

Wheeler, D. L. (2006). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *34*(90001), D173–D180. https://doi.org/10.1093/nar/gkj158

Wiese, O., Marenduzzo, D., & Brackley, C. A. (2019). Nucleosome positions alone can be used to predict domains in yeast chromosomes. *Proceedings of the National Academy of Sciences*, *116*(35), 17307–17315. https://doi.org/10.1073/pnas.1817829116

Wong, J. T. Y. (2019). Architectural Organization of Dinoflagellate Liquid Crystalline Chromosomes. *Microorganisms*, *7*(2), 27. https://doi.org/10.3390/microorganisms7020027

Wu, G., McArthur, A. G., Fiser, A., Šali‡, A., Sogin, M. L., & M, M. (2000). Core Histones of the Amitochondriate Protist, Giardia lamblia. *Molecular Biology and Evolution*, *17*(8), 1156–1163. https://doi.org/10.1093/oxfordjournals.molbev.a026398

Wuitschick, J. D., & Karrer, K. M. (1999). Analysis of Genomic G + C Content, Codon Usage, Initiator Codon Context and Translation Termination Sites In *Tetrahymena Thermophila*. *Journal of Eukaryotic Microbiology*, *46*(3), 239–247. https://doi.org/10.1111/j.1550-7408.1999.tb05120.x

Xu, J., Zhang, X., Pelayo, R., Monestier, M., Ammollo, C. T., Semeraro, F., Taylor, F. B., Esmon, N. L., Lupu, F., & Esmon, C. T. (2009). Extracellular histones are major mediators of death in sepsis. *Nature Medicine*, *15*(11), 1318–1321. https://doi.org/10.1038/nm.2053

Yee, J., Tang, A., Lau, W.-L., Ritter, H., Delport, D., Page, M., Adam, R. D., Müller, M., & Wu, G. (2007). Core histone genes of Giardia intestinalis: Genomic organization, promoter structure, and expression. *BMC Molecular Biology*, *8*(1), 26. https://doi.org/10.1186/1471-2199-8-26

Zaret, K. (1999). Micrococcal Nuclease Analysis of Chromatin Structure. *Current Protocols in Molecular Biology*, *45*(1). https://doi.org/10.1002/0471142727.mb2101s45

Zhang, Y., Moqtaderi, Z., Rattner, B. P., Euskirchen, G., Snyder, M., Kadonaga, J. T., Liu, X. S., & Struhl, K. (2009). Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions *in vivo*. *Nature Structural & Molecular Biology*, *16*(8), 847–852. https://doi.org/10.1038/nsmb.1636

Zhang, Z., & Pugh, B. F. (2011). High-Resolution Genome-wide Mapping of the Primary Structure of Chromatin. *Cell*, *144*(2), 175–186. https://doi.org/10.1016/j.cell.2011.01.003

Zhou, X., Zhang, N., Gong, J., Zhang, K., Chen, P., Cheng, X., Ye, B.-C., Zhao, G., Jing, X., & Li, X. (2024). *In vivo* assembly of complete eukaryotic nucleosomes and (H3-H4)-only non-canonical nucleosomal particles in the model bacterium *Escherichia coli*. *Communications Biology*, *7*(1), 1510. https://doi.org/10.1038/s42003-024-07211-4

Zlatanova, J., & Thakar, A. (2008). H2A.Z: View from the Top. *Structure*, *16*(2), 166–179. https://doi.org/10.1016/j.str.2007.12.008

List of tables

List of figures