Published online 17 November 2021

NAR Genomics and Bioinformatics, 2021, Vol. 3, No. 4 **1**
https://doi.org/10.1093/nargab/lqab107

Identification and characterization of histones in *Physarum polycephalum* evidence a phylogenetic vicinity of Mycetozoans to the animal kingdom

Axel Poulet [1], Laxmi Narayan Mishra[2], Stéphane Téletchéa[3], Jeffrey J. Hayes [2], Yannick Jacob [1], Christophe Thiriet [4] and Céline Duc [4,*]

[1]Department of Molecular, Cellular and Developmental Biology, Faculty of Arts and Sciences, Yale University, New Haven, CT 06520-8103, USA, [2]Department of Biochemistry and Biophysics, University of Rochester Medical Center, Rochester 14620 NY, USA, [3]Conception de protéines in silico, Université de Nantes, CNRS, UFIP, UMR 6286, Nantes, France and [4]Epigénétique et dynamique de la chromatine, Université de Nantes, CNRS, UFIP, UMR 6286, Nantes, France

Received August 23, 2021; Revised October 12, 2021; Editorial Decision October 20, 2021; Accepted November 10, 2021

ABSTRACT

***Physarum polycephalum* belongs to Mycetozoans, a phylogenetic clade apart from the animal, plant and fungus kingdoms. Histones are nuclear proteins involved in genome organization and regulation and are among the most evolutionary conserved proteins within eukaryotes. Therefore, this raises the question of their conservation in *Physarum* and the position of this organism within the eukaryotic phylogenic tree based on histone sequences. We carried out a comprehensive study of histones in *Physarum polycephalum* using genomic, transcriptomic and molecular data. Our results allowed to identify the different isoforms of the core histones H2A, H2B, H3 and H4 which exhibit strong conservation of amino acid residues previously identified as subject to post-translational modifications. Furthermore, we also identified the linker histone H1, the most divergent histone, and characterized a large number of its PTMs by mass spectrometry. We also performed an in-depth investigation of histone genes and transcript structures. Histone proteins are highly conserved in *Physarum* and their characterization will contribute to a better understanding of the polyphyletic Mycetozoan group. Our data reinforce that *P. polycephalum* is evolutionary closer to animals than plants and located at the crown of the eukaryotic tree. Our study provides new insights in the evolutionary history of *Physarum* and eukaryote lineages.**

INTRODUCTION

Chromatin consists of nucleosome arrays, with each nucleosome being composed of a central tetramer of H3/H4 histones flanked by two dimers of histones H2A-H2B. The core histone octamer is wrapped by ∼147 bp of DNA to form the nucleosome core (1). A fifth histone, the linker histone or H1, binds to linker DNA entering/exiting the nucleosome core and is involved in higher-order structures of chromatin to compact and organize DNA into the nuclear space (2). Core histone proteins share a structured domain, the histone fold domain, which provides an extensive dimerization interface between histone protein pairs and defines contacts to DNA. Core histone proteins also contain unstructured N-terminal tails that mediate interactions within and between nucleosomes (3) and are decorated by a broad range of post-translational modifications (PTMs) that serve as epigenetic modulators of transcription.

Histones are highly conserved through eukaryotic evolution. However, within the same organism, each histone type can display several isoforms presenting differences in their sequence, such as punctual amino acid changes, insertions or deletions. The different isoforms are typically classified as canonical histones and variants; this classification is mainly defined based on their temporal expression during the cell cycle. Indeed, canonical histones are largely synthesized during the S-phase in conjunction with DNA replication while the histone variants are typically produced throughout the cell cycle (4). The targeted incorporation of histone isoforms into nucleosomes as well as the presence of PTMs on histone tails contribute to define the epigenetic landscape. Since stoichiometric amounts of each histone type are required, their synthesis, transport and incorporation need to be finely tuned.

*To whom correspondence should be addressed. Tel: +33 2 51 12 56 34; Email: celine.duc@univ-nantes.fr
Present address: Laxmi Narayan Mishra, Department of Cell Biology, Albert Einstein College of Medicine, Bronx, New York 10461, USA.

This control can be achieved at the level of 3′ end processing of histone transcripts. Messenger RNAs undergo 3′ end processing with an endonucleolytic cleavage of the 3′-untranslated region (UTR) downstream of the polyadenylation signal (PAS) followed by the addition of a poly(A) tail. All eukaryotic mRNAs follow this two-step processing, except those of the animal canonical histones whose 3′ end processing is achieved via a stem-loop (SL) structure recognized by the stem-loop binding protein (SLBP) and via a histone downstream element (HDE) recognized by the U7 small nuclear RNA (snRNA) by base-pairing (5). However, this 3′ end processing mechanism of canonical histone transcripts is not found in yeast and plants (5).

The Mycetozoans include a large variety of slime molds and represent a divergent eukaryotic lineage from plants, animals and fungi. The Myxomycetes are categorized in five orders with Physarales (which include *Physarum polycephalum*) constituting the largest order. *Physarum polycephalum* displays a sophisticated life cycle comprising a vegetative diploid stage and a haploid reproductive stage as well as several other forms depending on the environmental conditions (6). The vegetative stage is characterized by a syncytium called plasmodium containing millions of nuclei. This organism has been used in a wide range of life science fields of studies, such as cognition (reviewed in (7)), biophysics (8) and epigenetics (9) and was extensively used in the 1980s for the study of the cell cycle (10) as well as in cell and developmental biology (11). These studies have brought to ask the fore important questions regarding the position of *P. polycephalum* within the phylogenic tree.

Here, we performed the first comprehensive study of *P. polycephalum* histones. Although histones from this organism were first isolated in 1969 from nuclei (12), they were only partially defined more than a decade later (13). Indeed, while similarities were reported between *Physarum* histones and histones from other model systems such as calf thymus (13,14), these studies were lacked identification of the various histone isoforms in *Physarum*. To address this, we carried out a comprehensive analysis of the *P. polycephalum* genome and transcriptomes to identify the various histone isoforms. Phylogenetic analyses comprising histone proteins from different organisms (animals, plants and unicellular eukaryotes) allowed us to establish the position of *P. polycephalum* in the tree of life. Moreover, comparative analyses of histone protein sequences allowed the identification of conserved residues harboring PTMs. In addition, we determined transcript abundance at precise cell cycle stages for all the genes encoding the *Physarum* putative histone isoforms while a detailed analysis of histone transcript sequences to identify signatures for 3′ end processing revealed that most histone mRNAs in *Physarum* contained a stem-loop structure as well as a PAS. Our comprehensive characterization of histones provides evidence that *Physarum polycephalum* is positioned closer to animals in the eukaryotic tree than previously appreciated, although this organism also displays characteristics specific to the plant kingdom.

MATERIALS AND METHODS

Identification of Physarum genes, transcripts and proteins for histones and other proteins

Genomic and transcriptomic sequences from *Physarum polycephalum* were obtained from www.physarum-blast.ovgu.de and from further published data (6,15). Protein sequences from *Homo sapiens, Mus musculus, Danio rerio, Drosophila melanogaster, Xenopus laevis, Caenorhabditis elegans, Amborella trichopoda, Arabidopsis thaliana, Zea mays, Physcomitrella patens, Tetrahymena thermophila, Dictyostelium discoideum* and *Saccharomyces cerevisiae* were obtained from NCBI. Protein sequences for yeast and plant H2A used were the ones described in (16). For H2B, retrieved proteins for an organism were cleaned from proteins identical at 92–99% on the whole length. For H1, retrieved proteins for an organism were cleaned from proteins identical at 92–99% on the whole length and from truncated isoforms. Protein homologs in *P. polycephalum* were identified with TBLASTN from NCBI (17). All used protein sequences were reported in Supplementary Table S3. Transcripts identified as histone-coding were reported in Supplementary Figure S9. Sequences for transcripts and proteins were deposited in Genbank, numbers being indicated in Supplementary Table S3. To identify histone-coding genes, mRNAs coding for histone proteins were identified from the *Physarum* reference transcriptomes (6,15) by a local BLASTn (blast 2.6.0) and then aligned them on the *Physarum* reference genome (6) to identify expressed coding genes. A tBLASTn search was finally performed to identify genes coding histones but not expressed in the reference transcriptomes.

Phylogenetic analyses and protein sequence alignments

To generate phylogenetic trees, selected sequences were first aligned with MUSCLE (18), a multiple sequence alignment tool, using default parameters. For H3 trees, H3v2 from *D. discoideum* was not included in the study since it was reported to localize to the cytoplasm (19). For H1, alignment was then refined using Gblocks (v0.91b) (20). FastTree (v2.1.8) (21) was further applied with default parameters for the construction of the phylogenetic trees. Fast-Tree enabled to infer approximately-maximum-likelihood phylogenetic trees from alignments. Finally, phylogenetic trees were drawn using the ITOL (Interactive Tree Of Life) tool (22). In phylogenetic trees, plant, animal and *Physarum* proteins were, respectively, depicted in green, blue and red. For yeast, *D. discoideum* and *T. thermophila*, they were displayed in black. To compare identified *Physarum* homologs with known proteins, multiple alignments were performed with the Clustal Omega program (23). Conserved domains for SLBP and LSm proteins were identified with the InterPro webserver (www.ebi.ac.uk/interpro/search/sequence/). The CPH models-3.2 server was used to create the 3D protein modelling and the Chimera software (24) for superimposition.

Searches for stem-loop structures in histone mRNAs and of U7 snRNA

To identify the SL structures in histone mRNAs, the RNA-fold webserver was used to predict stem loops (rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi). The logo of conserved features for histone SLs was generated with the WebLogo web server (25). Presence of two U7 snRNAs were reported (26), but their sequences were not available. A Perl script was made to screen *Physarum* genome with the *Mus musculus* sequence of the U7 snRNA to identify U7 snRNA genes (according to (27)). The analysis was performed also for the U7 reverse complement. The U7 sequence was shortened until a match was found against the *Physarum* genome. No *bona fide* U7 snRNA was identified with this method. Moreover, we were not able to retrieve U7 snRNAs in the available Physarum poly(A)-selected transcriptomes since snRNAs are generally not poly(A)-tailed. Thus, our determination of a potential HDE signature relied on the presence of a purine-rich sequence downstream of the stem-loop structure.

Physarum material

Physarum polycephalum strain TU291 from was cultured as asynchronous plasmodia, as described in (28). For analyses at specific cell cycle stages, mitosis was monitored on mitotically synchronous plasmodia by phase contrast microscopy observations of explants (28).

Experimental RNA analysis procedures

Asynchronous plasmodia were briefly centrifuged and washed in water before freezing in liquid nitrogen. Three mitotically synchronous plasmodia were prepared on three independent dates. Each mitotically synchronous plasmodium was cut in six equal fragments. Each fragment corresponded to one stage of the cell cycle which was harvested following a wash in 5 mM EDTA and quickly frozen in liquid nitrogen. The fragments were harvested ∼10 min before mitosis 2 (late G2-phase), 2 min after mitosis (beginning of S-phase), 1 h after mitosis (mid S-phase), 2.5 h after mitosis (late S-phase) and 5.5 h after mitosis (beginning of G2-phase). RNAs were extracted with Tri-Reagent (Ambion) from frozen samples according to the manufacturer's instructions followed by a phenol-chloroform extraction prior to a DNase I (NEB) treatment and a second purification by phenol-chloroform extraction. Polyadenylated-enriched RNA samples were prepared from 10 μg of total RNAs with the NEBNext® Poly(A) mRNA Magnetic Isolation Module (NEB). The cDNAs were synthesized with iScript™ Reverse Transcription Supermix for RT-qPCR (BioRad) containing a blend of oligo(dT) and random primers. They were diluted 1:3 and 5 μl were used in PCR (DreamTaq DNA Polymerase, Thermofisher) or in quantitative PCR with the SyberGreen qPCR master mix kit (ThermoScientific) on a Biorad Cycler. Transcript abundances (in copy number of transcripts produced by the analyzed gene) for each gene were calculated as follow in asynchronous plasmodia: Transcript Level $= 10^{11} \times E^{-Ct}$. E is the efficiency for each primer set and was calculated with the following formula: $E = 10(-1/\text{-slope})$, the slope being determined from the standard curve obtained with serial dilutions of cDNAs. For histone mRNA abundance analyses during the cell cycle in mitotically synchronous plasmodia, relative transcript levels (no unit) were calculated as follow: $10^6 \times E^{-Ct[\text{histone gene}]}/E^{-Ct[19S]}$. RT-qPCR histograms show means of transcript levels ±SE obtained for two independent PCR amplifications of three biological replicates. Primers used in this study are listed in Supplementary Table S4.

Experimental protein analysis procedures

Histones were prepared either from asynchronous plasmodia (Figure 1B) or from mitotically synchronous plasmodia harvested at specific cell cycle stages (Figure 6A and Supplementary Table S2). Nuclei were isolated from plasmodia accordingly to (29). Total histones were prepared by 20% trichloroacetic acid (TCA) precipitation with 0.4 M HCl to solubilize nuclear proteins. Similarly, linker histone H1 was prepared by 20% TCA precipitation with 5% $HClO_4$ to solubilize nuclear proteins. To determine the H1 PTMs, a *Physarum* H1 synthetic gene was cloned in the pET3a plasmid and the *Physarum* H1 protein was expressed in bacteria and extracted by a precipitation of the cellular lysate with 5% $HClO_4$ and 20% TCA. This unmodified *Physarum* H1 produced in bacteria was used as a reference to determine the *in vivo Physarum* H1 modifications at specific cell cycle stages. Mass spectrometry analyses were carried out by MS-Bioworks and the protein sequence coverage in G2-phase, S-phase and mitosis were, respectively, of 94%, 100% and 94%.

RESULTS

Identification of core histones from the Physarum genome and transcriptomes

In order to identify core histones, we used the *Physarum* reference genome and transcriptomes (6,15). We identified 15 histone-coding genes (one gene for H1, 3 genes for H2A, 3 genes for H2B, 5 genes for H3 and 3 genes for H4; see Materials and Methods) and 12 distinct transcripts. Our analysis revealed that each histone isoform is encoded by a single gene, except one *Physarum* H4 protein which is encoded by two different genes (Table 1). In animals, genes encoding histone proteins are clustered together and present in multiple copies (i.e. ∼100 copies in Drosophila on chromosome 2 (30)) while they are present in one-to-a few copies in plants (31) and *S. cerevisiae* (32). To decipher the genomic distribution of histone genes, we mapped them onto the *Physarum* genome and observed that histones are encoded by a few gene copies dispersed throughout the *Physarum* genome, similarly to plants. We monitored the expression of these 15 histone genes in *Physarum* asynchronous plasmodia. Most histone genes exhibit an elevated abundance of transcripts compared to the 19S ribosomal transcript level. Indeed, *PpHTT4* transcripts are around ∼100 times less abundant that the 19S ribosomal transcripts, ribosomal RNAs representing ∼80% of RNAs found in cells (Figure 1A). Moreover, *PpHTA2* and *PpHTA3* transcripts are around ∼1000 times less abundant, *PpHTB1* and *PpHTT2* and

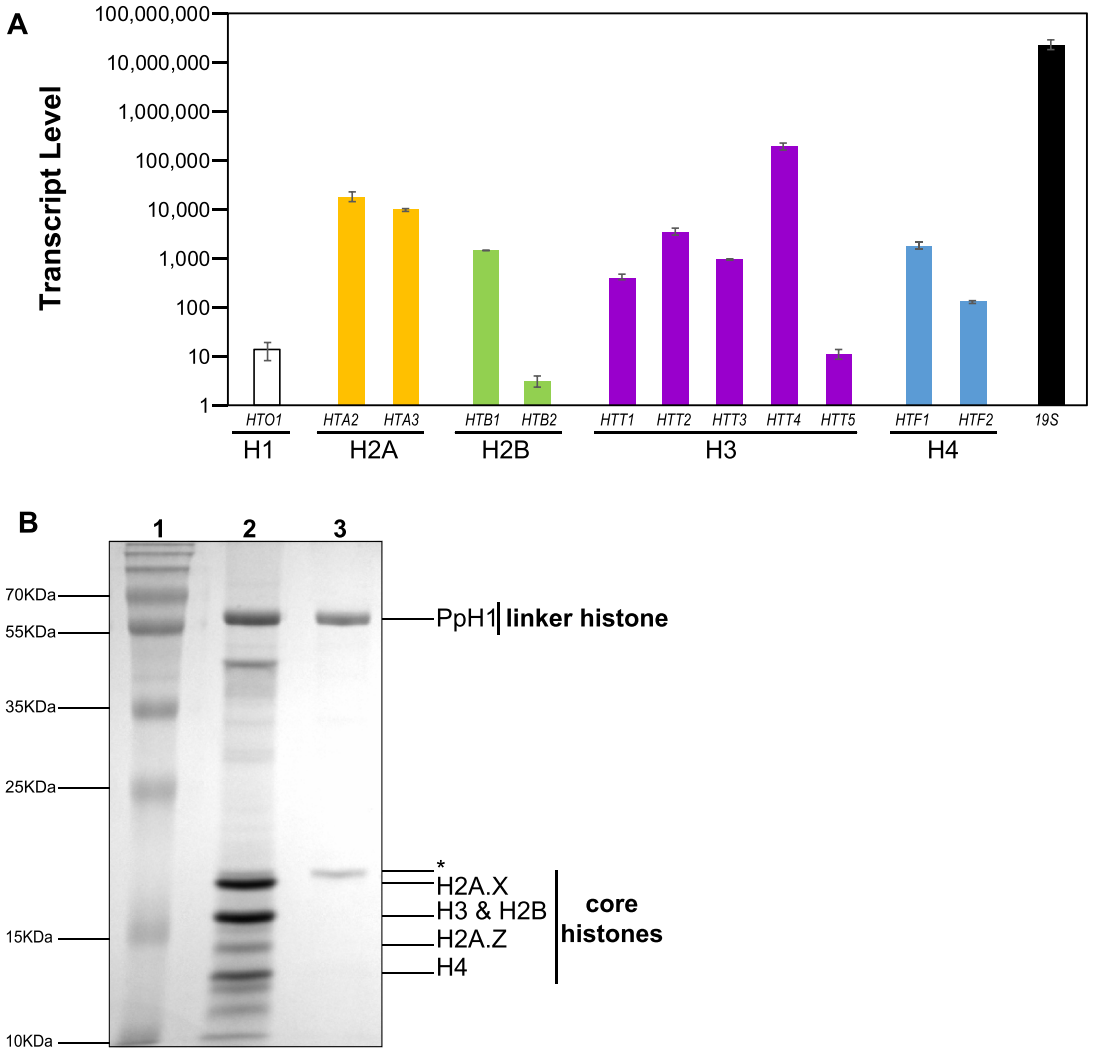

Figure 1. Analysis of *Physarum* histones at the RNA and protein levels. (**A**) Mean transcript level for histone genes displayed with a logarithmic scale and measured by qRT-PCR on three biological replicates consisting of independent asynchronous plasmodia cultures. The 19S rRNA was used as a control. (**B**). SDS-PAGE analysis of histones proteins from *Physarum* nuclear extracts. The lane #1 is the molecular weight marker, the lane #2 corresponds to a nuclear preparation of total histones and the lane #3 corresponds to a nuclear preparation of histone H1. The asterisk indicates a protein present in the extract besides histones.

PpHTF1 (∼10 000 times), *PpHTT3* (∼25 000), *PpHTT1* (∼60 000 times) and *PpHTF2* (∼200 000) transcript abundance being lower (Figure 1A). Finally, *PpHTO1*, *PpHTB2* and *PpHTT5* genes produce much less transcripts since their abundance is ∼2 (*PpHTO1* and *PpHTT5*) or ∼10 (*PpHTB2*) million times smaller than the 19S transcript abundance. In addition, *PpHTA1, PpHTB3* and *PpHTF3* (*Physarum polycephalum* HisTone 2A 1, HisTone 2B 3 and HisTone Four 3) genes, which do not appear to be transcribed in sporulation-competent or sporulating plasmodia (15,33), were not detectably expressed in plasmodia and were not found in the genome of the TU291 strain we used for your study (Supplementary Figure S1A).

We then analyzed the predicted histone proteins. Previously reported molecular weights of *Physarum* histones (13) were consistent with our bioinformatic (Table 1) and SDS-PAGE analyses (Figure 1B and Supplementary Figure S1B). After this initial identification of *Physarum* histone genes, transcripts and proteins, we performed a comprehensive analysis for the identified histone protein isoforms—predicted from the analysis of the *Physarum* genome and transcriptomes—to infer the position of Mycetozoa and more specifically *Physarum* in the eukaryotic tree of life. Hence, we carried out a phylogenetic analysis using several species among animals (*H. sapiens and M. musculus* for mammals, *D. rerio* for fishes, *D. melanogaster* for insects, *C. elegans* for nematodes), plants (*A. thaliana* for eudicots and *Z. mays* for monocots and *A. trichopoda* that is frequently used for plant phylogeny since it is considered as the most basal angiosperm and *P. patens* for mosses) and unicellular eukaryotes (the ciliated protozoan *T. thermophila*, the yeast *S. cerevisiae* and a Dictyostelid *D. discoideum* that also belongs to the Mycetozoa).

Table 1. List of *Physarum* histone-coding genes.

Histone gene ID	Isoform	Histone protein ID	Protein length (AA)	Estimated molecular weight (kDa)	Scaffold	Transcript ID	Number of introns	Stem-loop	Distance between stop and stem-loop (bp)	HDE	PAS
					Histone H1						
PpHT01	canonical	PpH1	359	36.6	Scaffold153	Phypoly_transcript_12627	0	Yes	30	Yes	AAUGAA
					Histone H2A						
PpHTA1	canonical ?	PpH2A.1	178/131	18.8/13.7	Scaffold12319	/	0	Yes	51	Yes	-
PpHTA2	variant	PpH2A.X	148	15.9	Scaffold67	Phypoly_transcript_11156	1	Yes	31	Yes	AAUAAA
PpHTA3	variant	PpH2A.Z	119	12.9	Scaffold829	Phypoly_transcript_18931	4	No	-	No	AAUAAA
					Histone H2B						
PpHTB1	NA	PpH2B.1	136	14.9	Scaffold10	Phypoly_transcript_20418	1	Yes	32	Yes	AAUAAA
PpHTB2	NA	PpH2B.2	136	14.8	Scaffold1031	Phypoly_transcript_18509	1	Yes	58	Yes	AAUAAA
PpHTB3	NA	PpH2B.3	106	11.7	Scaffold21517	/	0	No	-	No	-
					Histone H3						
PpHTT1	canonical	PpH3.1	136	15.2	Scaffold83	Phypoly_transcript_24983 & Phypoly_transcript_26166	1	Yes	2	Yes	AAUUAA
PpHTT2	centromeric variant	PpCenH3	280	31.5	Scaffold299	Phypoly_transcript_13632	2	No	-	No	AAUAUU
PpHTT3	variant	PpH3.3	136	15.2	Scaffold1263	Phypoly_transcript_17849	1	Yes	48	Yes	AAUAAU
PpHTT4	variant	PpH3.4	136	15.3	Scaffold3245	Phypoly_transcript_08192	3	No	-	No	AAUAAA
PpHTT5	variant	PpH3.5	134	15.2	Scaffold1409	Phypoly_transcript_27948	-	Yes	15	No	AAUUAA
					Histone H4						
PpHTF1	canonical	PpH4.1	102	11.3	Scaffold15	Phypoly_transcript_28492	1	Yes	29	Yes	ACUAAA
PpHTF2	canonical	PpH4.1			Scaffold15	Phypoly_transcript_15849	1	Yes	26	Yes	AAUUAA
PpHTF3	NA	PpH4.2	102	11.2	Scaffold9510	/	0	Yes	42	Yes	AAUAAA

There are two putative translation starts for the PpH2A.1 protein so both predicted lengths and molecular weights are indicated. *PpHTF1* and *PpHTF2* were previously named *H41* and *H42*, respectively (Wilhelm *et al.* 1987). The sequence of the *PpHTT5* gene was localized on Scaffold1409, but the presence of a *N* stretch prevented to have the detailed number of introns.

Evolution and characterization of histones in *Physarum*

Histones H2A. The histone H2A family consists of several groups: (i) H2A.W is a plant specific H2A isoform (16), (ii) macroH2A and H2A.Bbd are restricted to animals and mammals, respectively, (iii) H2A.X, which has diverged several times through evolution, constitutes a polyphyletic group and harbors a SQEY motif, the serine and tyrosine residues being phosphorylated during the DNA damage response (34,35), (iv) H2A.Z constitutes a monophyletic group and displays a 15 amino acid motif (LEYLTAEVLELAGNA), which was proposed as a signature for the H2A.Z variant (36), and (v) canonical H2A diverged repeatedly (37). *Physarum* possesses the H2A.1, H2A.X and H2A.Z isoforms, each one being encoded by a single gene (Table 1). The *PpHTA1* (HisTone 2A 1, no intron, Table 1) gene encodes a protein without any typical H2A-variant signature that was thus named H2A.1 and further referred as the *Physarum* canonical H2A (Figure 2A). The *PpHTA2* gene (one intron, Table 1) encodes a protein with the highly conserved SQEY motif in its C-terminus (Figure 2A and Supplementary Figure S2A) that was named H2A.X. The protein encoded by *PpHTA3* (4 introns, Table 1) contains the LEYLTAEVLELAGNA motif (Figure 2A and Supplementary Figure S2B) and was named H2A.Z. The *Physarum* H2A.X and H2A.Z proteins share 63% identity while they are more divergent from H2A.1 that harbors a longer N-terminus (Figure 2A and Supplementary Table S1). Various PTMs have been identified in the human H2A and are found on a large number of amino acid residues (38), thus we restricted our predictions to lysine methylation and acetylation, arginine methylation and serine/threonine phosphorylation. Sequence comparison with HsH2A (Supplementary Figure S2A,B) disclosed that most residues harboring PTMs in human were conserved in at least one of the three different *Physarum* H2A isoforms (14 out of 23 residues, Figure 2A). Our phylogenetic analysis revealed that the *Physarum* H2A.1 diverged from the other H2A proteins since it displays a long branch in the unrooted phylogenetic tree, similarly to the *D. discoideum* H2A canonical proteins (Figure 2B). The *Physarum* H2A.X belongs to the same clade as ScH2A and animal H2A and H2A.X proteins (Figure 2B). As expected, the *Physarum* H2A.Z belongs to the monophyletic group composed of the various eukaryotic H2A.Z (Figure 2B). The natural synchrony of *Physarum* plasmodia provides the opportunity to study histone transcript levels at precise stages of the cell cycle, which comprises of a 0.5 h closed mitosis, 3 h S-phase and 6 h G2-phase. Importantly, *Physarum* plasmodia lack a G1-phase (39). We thus took advantage of this perfect synchrony to assess *PpHTA2* and *PpHTA3* mRNA levels during the cell cycle, the *PpHTA1* gene being absent from our *Physarum* strain. As in asynchronous plasmodia (Figure 1A), *PpHTA2* exhibits at a higher transcript level than *PpHTA3* throughout the cell cycle (Figure 2C) and is expressed at a constant level, except for a ∼4-fold increased expression in late G2-phase. This finding suggests that the putative H2A.X protein encoded by *PpHTA2* might be produced at higher levels during the cell cycle than the putative H2A.Z protein encoded by *PpHTA3*. Indeed, we found that these mRNA levels are consistent with the relative abundance of H2A.X and H2A.Z proteins (Figure 1B) and the lower abundance of H2A.Z in chromatin of most organisms. Moreover, the *PpHTA3* mRNA level increases during the S-phase and in late G2-phase (Figure 2C). To conclude, *Physarum* possesses three putative types of H2A proteins with conserved features compared to the other organisms but with specific variations in sequences.

Histones H2B. The histone H2B family contains variants but their dedicated functions remain largely unknown. We identified three genes that encode three putative H2B isoforms. The putative H2B.1 and H2B.2 proteins, respectively, encoded by *PpHTB1* and *PpHTB2* (with one intron each, Table 1), shares 87% identity between each other but only ∼46% identity with H2B.3 (Supplementary Table S1). Indeed, H2B.3 is shorter than the two other isoforms by 30 amino acids with a truncated N-terminus (Figure 3A). Similar to H2A PTMs, we performed a sequence comparison between the putative *Physarum* H2B proteins and human H2B (Supplementary Figure S3A). It revealed that most residues harboring PTMs in the human H2B (38) were conserved in at least one of the three *Physarum* H2B isoforms (27 out of 33 residues, Figure 3A). Our phylogenetic analysis of the *Physarum* H2B proteins showed that the *Physarum* H2B.1 and H2B.2 constitute a monophyletic group—related to the *D. discoideum* H2B.3—while the *Physarum* H2B.3 belongs to a clade composed of divergent H2B proteins from *Tetrahymena, Arabidopsis* and maize and displays a long branch (Figure 3B and Supplementary Figure S3B). Moreover, the coding sequence of *PpHTB1* and *PpHTB2* mRNAs are 89% identical while the *PpHTB3* transcript is highly divergent. Both findings suggest that a duplication event gave rise to H2B.1 and H2B.2 paralogs while H2B.3 has a different evolutionary history. Finally, we used mitotically synchronous plasmodia to assess *Physarum PpHTB1* and *PpHTB2* mRNA levels during the cell cycle, the *PpHTB3* gene being absent from our *Physarum* strain. As in asynchronous cells (Figure 1A), the *PpHTB1* mRNA level was higher than that of the *PpHTB2* one during the entire cell cycle (>100 times more abundant, Figure 3C). Moreover, the *PpHTB1* transcript level increased during S-phase and in late G2-phase while the *PpHTB2* transcript level remained fairly constant during the cell cycle (Figure 3C). To conclude, of the three H2B-coding genes found in the *Physarum* reference genome, we identified the H2B.1 and H2B.2 proteins in our strain, with the former being the major plasmodial protein isoform. Of note, all three predicted *Physarum* H2B proteins are characterized by a strong divergence from other species.

Histones H3. The histone H3 family contains three major groups: (i) the replication-dependent H3.1 defined as canonical H3 and expressed only during S-phase of the cell cycle, the replication-independent (ii) H3.3 and (iii) cenH3 (the centromeric H3 isoform) variants that are expressed during the whole cell cycle. We identified five H3-coding genes (named *PpHTT1* to *5)* in the *Physarum* reference genome and found that all five genes were expressed in our strain (Figure 1A and (6)) (Table 1). Four H3 genes encode putative proteins of highly similar sequences whereas *PpHTT2* encodes a much more divergent protein that we

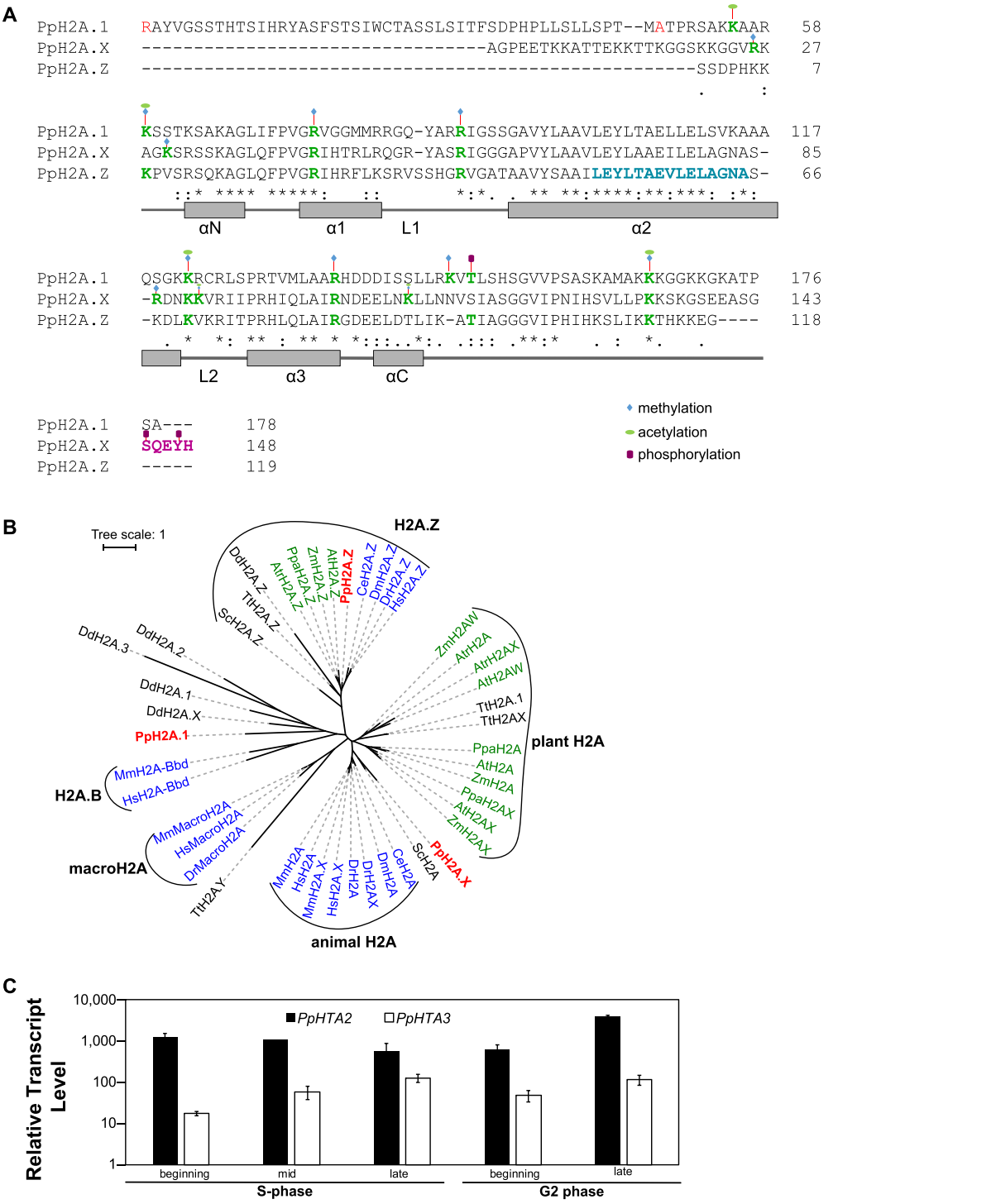

Figure 2. Identification and analysis of PpH2A. (**A**) Protein sequence alignment of *Physarum* H2A isoforms. PTMs are shown based on the sequence alignment of the canonical human H2A (Supplementary Figure S2A,B) (38) and conserved residues potentially harboring those PTMs are displayed in green (except for the S and Y residues in the SQEY motif of the PpH2A.X protein that are phosphorylated upon double-strand break damage in most eukaryotes). Positions refer to the mature protein without the initial methionine. There are two putative translation starts for PpH2A.1 that are indicated in red (residues R1 and A48). Only lysine methylation and acetylation as well as S/T phosphorylation are reported. The various helices and loops (L1 and L2) are indicated. The signature of H2A.X is depicted in pink and the H2A.Z one in cyan. Asterisk indicates a fully conserved residue. Colon and dot indicate conservation between residues of, respectively, strongly and weak similar properties. (**B**) Phylogenetic tree of H2A proteins from *P. polycephalum* and various eukaryotes. The mammal-specific H2A.L was not included in the analysis. Atr: *Amborella trichopoda;* At: *Arabidopsis thaliana*; Ce: *Caenorhabditis elegans;* Dd: *Dictyostellium discoideum*; Dr: *Danio rerio;* Dm: *Drosophila melanogaster*; Hs: *Homo sapiens*; Mm: *Mus musculus*; Pp: *Physarum* polycephalum; Ppa: *Physcomitrella patens*, Sc: *Saccharomyces cerevisiae;* Tt: *Tetrahymena thermophila*; Zm: *Zea mays*. (**C**) Mean relative transcript level for *Physarum* histone H2A genes displayed with a logarithmic scale and measured by qRT-PCR on three biological replicates consisting of independent mitotically synchronous plasmodia cultures harvested at different times of the cell cycle.

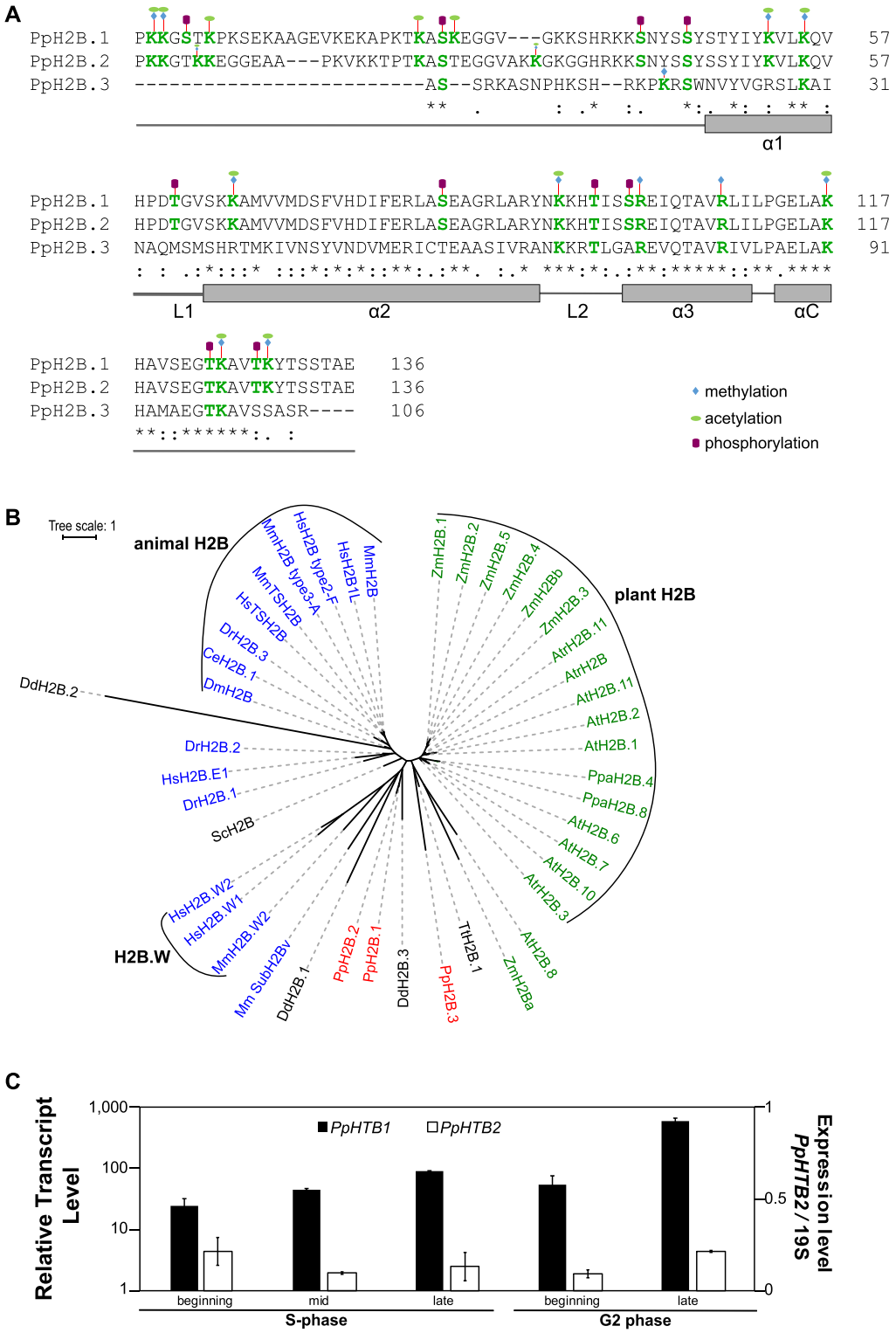

Figure 3. Identification and analysis of PpH2B. (**A**) Protein sequence alignment of *Physarum* H2B isoforms. PTMs are shown based on the sequence alignment of the canonical human H2B (Supplementary Figure S3A) (38) and conserved residues potentially harboring those PTMs are displayed in green. Positions refer to the mature protein without the initial methionine. Only lysine methylation and acetylation as well as S/T phosphorylation are reported. The various helices and loops (L1 and L2) are indicated. Asterisk indicates a fully conserved residue. Colon and dot indicate conservation between residues of, respectively, strongly and weak similar properties. (**B**) Phylogenetic tree of H2B proteins from *P. polycephalum* and various eukaryotes. Atr: *Amborella trichopoda;* At: *Arabidopsis thaliana*; Ce: *Caenorhabditis elegans;* Dd: *Dictyostellium discoideum*; Dr: *Danio rerio;* Dm: *Drosophila melanogaster*; Hs: *Homo sapiens*; Mm: *Mus musculus*; Pp: *Physarum* polycephalum; Ppa: *Physcomitrella patens*, Sc: *Saccharomyces cerevisiae;* Tt: *Tetrahymena thermophila*; Zm: *Zea mays*. (**C**) Mean relative transcript level for *Physarum* histone H2B genes displayed with a logarithmic scale and measured by qRT-PCR on three biological replicates consisting of independent mitotically synchronous plasmodia cultures harvested at different times of the cell cycle.

identified as the H3 centromeric variant and named it CenH3 (Supplementary Table S1). In order to discriminate between canonical and variant isoforms of H3, we analyzed the transcript levels of the *PpHTT* genes during the cell cycle in mitotically synchronous plasmodia. While the mRNA level of *PpHTT1* is elevated at the beginning and mid S-phase, it decreased later in the cell cycle before an increase in late G2-phase (Figure 4A), suggesting that *PpHTT1* encodes the canonical H3 of *Physarum*, H3.1. The *PpHTT5* transcripts were present at a low level during the cell cycle whereas *PpHTT4* mRNAs were found at a consistently high level throughout the cell cycle. The *PpHTT3* mRNA level strongly increased in late G2-phase while *PpHTT3* mRNAs presented a low level during the rest of the cell cycle (Figure 4A). We thus proposed that *PpHTT3, PpHTT4* and *PpHTT5* encode *Physarum* variants that we named H3.3, H3.4 and H3.5, respectively. Noteworthy, *PpHTT2* which encodes the *Physarum* CenH3 exhibited a constant mRNA level during most of the cell cycle, but with an increased mRNA abundance in late G2-phase (Figure 4A).

Alignment of proteins encoded by *PpHTT1* and *PpHTT3* revealed that they differed by four amino acids at positions 31 (S/T), 59 (E/D), 78 (Y/F) and 135 (A/T) (Figure 4B) while human H3.1 and H3.3 differ by five amino acid at positions 31 (A/S), 87 (S/A), 89–90 (VM/IG) and 96 (C/S) (40). The S31 residue is characteristic of the human H3.3 and can be phosphorylated, the phosphorylated form localizing in genomic regions adjacent to centromeres (Hake *et al*. 2005): this residue was found only in the *Physarum* H3.3, supporting our assignment of this protein as a H3.3 variant. The H3.4 and PpH3.5 variants present more sequence variations and some insertions/deletions (Figure 4B) that do not affect their overall structure (Supplementary Figure S4A). As above, we carried out a sequence comparison between the putative *Physarum* histones and the human H3 proteins (Supplementary Figure S4B) and found that all 32 residues harboring analyzed PTMs in the human H3 protein (38) were conserved in at least one of the four *Physarum* non-centromeric isoforms (either in H3.1, H3.3, H3.4 or H3.5, Figure 4B), consistently with the presence of some of these PTMs experimentally identified in *Physarum* (41). We performed a phylogenetic analysis separately for cenH3 and the other H3 proteins since centromeric variants are clearly separable from other H3s (Supplementary Figure S4C and Talbert *et al*. 2012). While H3.1 and H3.3 belong to the same clade, H3.4 and H3.5 are more divergent with longer branches (Figure 4C), suggesting that they arose independently. Moreover, the *Physarum* H3.4 belongs to the same clade than the *D.* discoideum H3.3 proteins. Regarding the centromeric variant, cenH3 proteins from several organisms harbor long branches in the phylogenetic tree (Figure 4D), confirming that this centromeric-specific variant has arisen multiple times during evolution. The closest relative of the *Physarum* CenH3 was the zebra fish protein (Figure 4D) with 74% identity on 36% of the protein sequence while the centromeric variants from the Mycetozoans *P.* polycephalum and *D. discoideum* harbored an independent evolutionary history. In addition, similar to *D. discoideum* CenH3 (19), the *Physarum* PpCenH3 harbors a much longer N-terminal tail than most centromeric variants such as the human CENP-A (Figure 4E and Supplementary Figure S4D) but does contain the characteristic CATD (CENP-A targeting domain) domain (Figure 4E), such that the overall structure of the *Physarum* CenH3 is similar to the human CENP-A (Supplementary Figure S4D). Similar to above analyses of histone proteins, we aligned the *Physarum* CenH3 with its human counterpart CENP-A (Figure 4E and Supplementary Figure S4E) and looked for conserved residues with known PTMs. Indeed, the human centromeric variant CENP-A displays specific PTMs mainly located in its N-terminal tail, such as S/T phosphorylation (3 residues out of 9 being conserved in the *Physarum* CenH3) but also the R41 methylation and the K124 methylation and acetylation (42), both residues being conserved in the *Physarum* CenH3 (Figure 4E). To conclude, our analyses identified one canonical putative histone H3 in *Physarum*, three H3 variants and a centromeric isoform; these various *Physarum* H3 isoforms have conserved features to other organisms although they appeared independently from other organisms during evolution.

Histones H4. Contrary to other histones, the H4 protein family does not typically include variants, with the exception of the H4G variant expressed in breast cancer cells (43), a H4 variant in rice (44) and the H4V variant in *T. brucei* (45). It is believed that the histone H4 is among the proteins with the slowest evolution rate. In the *Physarum* genome, we identified three H4-coding genes, *PpHTF1* and *PpHTF2* (with one intron each, Table 1) both coding the H4.1 protein and *PpHTF3* (no intron, Table 1) coding H4.2 (Table 1). The H4.1 and H4.2 proteins differ by five amino acids, three of them being localized in the α2 helix (Figure 5A). The *PpHTF1* and *PpHTF2* genes were localized ∼4 kb apart on the same chromosome and correspond to the two H4 genes previously described in a study based on Southern Blot analysis (46), in which *PpHTF1* was identified as the *H41* gene and *PpHTF2* as the *H42* gene (47). Sequence comparison between *Physarum* H4s and the human H4 (Supplementary Figure S5A) showed that all residues harboring PTMs in human (38) are conserved in both *Physarum* proteins (Figure 5A), except for K77 which is only present in H4.2. This observation was consistent with the identification of these conserved epigenetic marks in *Physarum* H4 (41). Since H4 proteins are identical at ∼95%, our phylogenetic analysis revealed little divergence between *Physarum* H4 proteins and the studied eukaryotic H4s, except for those of *T. thermophila, D. discoideum* and yeast that are much more divergent (Figure 5B). Surprisingly, the *Physarum* H4.2 was found to be 100% identical to the *C. elegans* protein (Supplementary Figure S5A). After the analysis of *Physarum* H4 proteins, we performed a study of *Physarum* H4 transcripts. As previously reported (48), RNA coding sequences (ATG-STOP) of *PpHTF1* and *PpHTF2* are 87% identical, with 5′ and 3′UTR regions being more divergent, while RNA coding sequence of *PpHTF3* transcript is, respectively, 78% and 79% identical to those of *PpHTF1* and *PpHTF2*. We then used mitotically synchronous plasmodia to assess *PpHTF1* and *PpHTF2* mRNA abundance during the cell cycle, the *PpHTF3* gene being absent from our *Physarum* strain. Transcripts of *PpHTF1* and *PpHTF2*

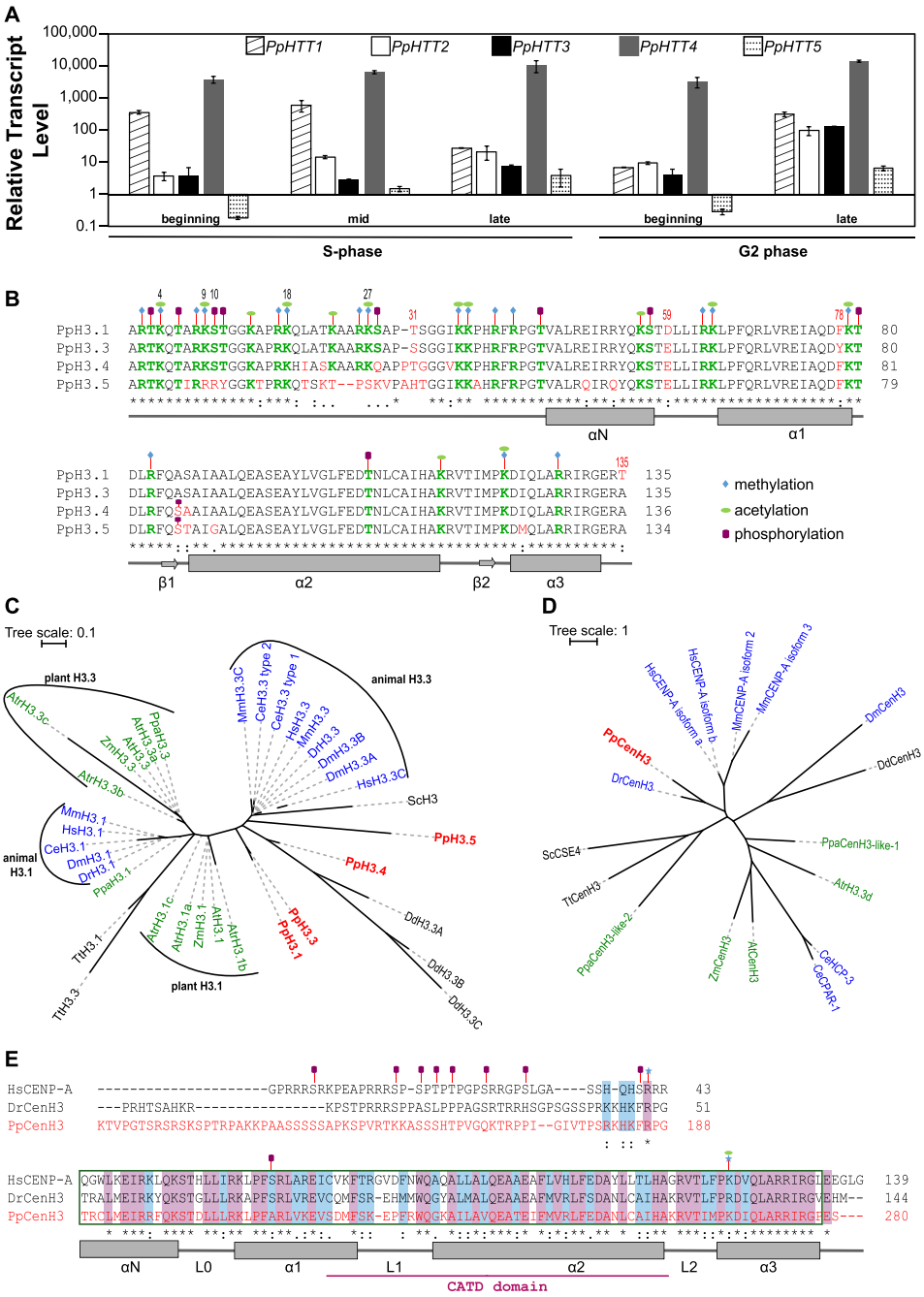

Figure 4. Identification and specificities of canonical histone H3 and their variants in *Physarum*. (**A**) Mean relative transcript level for *Physarum* histone H3 genes displayed with a logarithmic scale and measured by qRT-PCR on three biological replicates consisting of independent mitotically synchronous plasmodia cultures harvested at different times of the cell cycle. The *PpHTT1* and *PpHTT3-5* genes, respectively, encoded the PpH3.1, PpH3.3-PpH3.5 proteins and *PpHTT2* the centromeric variant PpCenH3. (**B**) Protein sequence alignment of the *Physarum* H3 isoforms. The PTMs are shown based on the sequence alignment of the human H3 (Supplementary Figure S4B) (38) and conserved residues potentially harboring those PTMs are displayed in green; positions for those with experimental records in *Physarum* (41) are displayed in black. Positions refer to the mature protein without the initial methionine. Only lysine methylation and acetylation as well as S/T phosphorylation are reported. The various helices are indicated and divergent residues between the different *Physarum* H3 isoforms are labeled in red. Asterisk indicates a fully conserved residue. Colon and dot indicate conservation between residues of, respectively, strongly and weak similar properties. (**C** and **D**) Phylogenetic tree of H3.1 and H3.3 proteins (in C) and cenH3 proteins (in D) from *P. polycephalum* and various eukaryotes. Atr: *Amborella trichopoda*; At: *Arabidopsis thaliana*; Ce: *Caenorhabditis elegans*; Dd: *Dictyostellium discoideum*; Dr: *Danio rerio*; Dm: *Drosophila melanogaster*; Hs: *Homo sapiens*; Mm: *Mus musculus*; Pp: *Physarum* polycephalum; Ppa: *Physcomitrella patens*, Sc: *Saccharomyces cerevisiae*; Tt: *Tetrahymena thermophila*; Zm: *Zea mays*. (**E**) Protein sequence alignment of cenH3 from human (HsCENP-A isoform a, HsCENP-A), zebra fish (DrCenH3) and *Physarum* (PpCenH3). The various helices are indicated and the CATD domain is labeled with strong pink line and the histone fold domain by a dark green frame. The *Physarum* protein PpCenH3 appears in red. Asterisk indicates a fully conserved residue which is labeled with a pale pink rectangle. Colon and dot indicate conservation between residues of, respectively, strongly and weak similar properties, the corresponding residue being labeled with a pale blue rectangle. Known PTMs from CENP-A are shown (38). Positions refer to the mature protein without the initial methionine. Only lysine methylation and acetylation as well as S/T phosphorylation are reported and depicted as in Figure 4B.

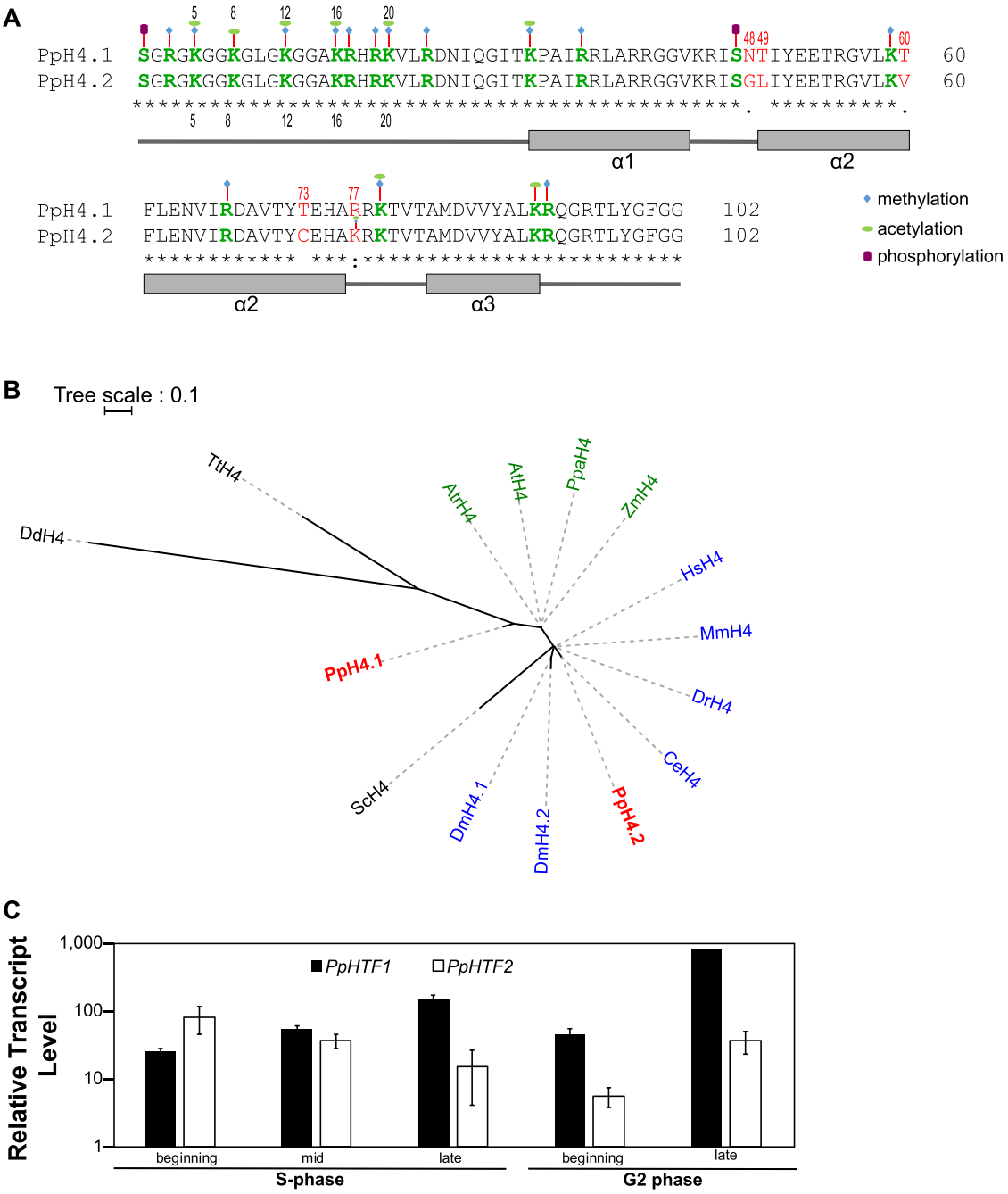

Figure 5. Identification of H4 variants in *Physarum*. (**A**) Protein sequence alignment of the PpH4.1 and PpH4.2 proteins. The PTMs are shown based on the sequence alignment of the human H4 (38) and conserved residues potentially harboring those PTMs are displayed in green. Positions for those with experimental records in *Physarum* (41) are displayed in black and divergent residues between the two *Physarum* H4 isoforms are labeled in red. Positions refer to the mature protein without the initial methionine. Only lysine methylation and acetylation as well as S/T phosphorylation are reported. The various helices are indicated and divergent residues between both H4 isoforms are labeled in red. (**B**) Phylogenetic tree of H4 proteins from *P. polycephalum* and various eukaryotes. Atr: *Amborella trichopoda;* At: *Arabidopsis thaliana*; Ce: *Caenorhabditis elegans;* Dd: *Dictyostellium discoideum;* Dr: *Danio rerio;* Dm: *Drosophila melanogaster*; Hs: *Homo sapiens*; Mm: *Mus musculus*; Pp: *Physarum* polycephalum; Ppa: *Physcomitrella patens*, Sc: *Saccharomyces cerevisiae;* Tt: *Tetrahymena thermophila*; Zm: *Zea mays*. (**C**) Mean relative transcript level for *Physarum* histone H4 genes displayed with a logarithmic scale and measured by qRT-PCR on three biological replicates consisting of independent mitotically synchronous plasmodia cultures harvested at different times of the cell cycle.

present distinct patterns. *PpHTF1* mRNAs exhibit an increased level through the S-phase while *PpHTF2* mRNAs display a decreased level at the same stages of the cell cycle. For both genes, the transcript levels significantly increase in late G2 phase (Figure 5C). These analyses are consistent with previous northern blot analyses (49). To conclude, *Physarum* seems to present two putative H4 isoforms which are strongly conserved compared to known H4 proteins.

Histones H1. We found one H1-coding gene, *PpHTO1* (no intron, Table 1), which was expressed and encodes the *Physarum* H1 protein. The *Physarum* H1 protein has 358 amino acids with a predicted size of 36.6 kDa and is lysine-rich (91 lysines out of 359 residues i.e. 25%). Consistently, the *Physarum* H1 was reported to have a size of ∼30.7 kDa (13) and to contain 330 ± 30 residues (14). The H1 linker proteins do not contain the histone fold motif, which is also the case for the *Physarum* H1. These proteins are rather characterized by a tripartite structure that consists of a conserved globular domain—composed of three helices and responsible for H1 binding to nucleosomes—as well as N- and C-terminal tails that are less structured (50). Indeed, the *Physarum* H1 displays a short N-terminal tail of 81 amino acids and a long C-terminal tail of 220 amino acids which is highly enriched in basic residues (64 out of 220 amino acids i.e. ∼30%; Figure 6A), as previously reported (14). The globular domain of H1 contains two DNA-binding sites (51): five of the seven basic amino acids of these DNA binding sites are conserved while two are missing in Site II (Figure 6A and Supplementary Figure S6A). Moreover, two regions in the C-terminal tail of the Drosophila H1 have been involved in nucleosome binding (52) and both sites—the ASAKKEK sequence located immediately after the globular domain and the lysine-rich sequence KKPKAKKAVAT in the middle of the C-terminal tail—are highly conserved in the *Physarum* H1 (Figure 6A and Supplementary Figure S6A,B). Regarding PTMs, the number of *Physarum* H1 phosphorylated residues was reported to vary during *Physarum* cell cycle (53) and *Physarum* H1 phosphorylation was linked to regulation of DNA replication timing (54,55). However, most residues harboring PTMs in *T. thermophila* (which is devoid of a globular domain), *D. melanogaster* or human (reviewed in (56)) are not conserved in *Physarum* H1 (Supplementary Figure S6B–D). We thus decided to further investigate the specific PTMs (lysine methylation and acetylation, S/T phosphorylation) of *Physarum* H1 during the cell cycle in mitotically synchronous plasmodia. We carried out mass spectrometry analyses on samples harvested in mitosis, mid S-phase and mid G2-phase. Surprisingly, we did not detect any PTMs in the *Physarum* H1 globular domain while the N- and C-terminal tails present a large number of residues harboring PTMs (Figure 6A). We identified 28 phosphorylation sites on S/T residues (Figure 6A and Supplementary Table S2) that either correspond to conserved residues phosphorylated in other organisms (Supplementary Figure S6B–D) or to other non-conserved S/T residues (Figure 6A and Supplementary Table S2). Regarding lysine methylation, we identified 54 sites with various methylation patterns (Supplementary Table S2). Lysine acetylation was identified at 13 positions. Furthermore, two acetylated regions are specific to the cell cycle stage. Indeed, acetylation of K71, K72 and K77 are exclusive of mitosis, while acetylation of K263, K264, K270, K272 are detected only in S-phase (Supplementary Table S2). Regarding phylogenetic analyses, a high diversity of H1 proteins was observed as expected and with long branch sizes notably for *Physarum* H1 (Figure 6B); animal and plant H1 proteins constituting distinct clades as expected (37). The globular domain is the most conserved H1 region among different species: *Physarum* H1 globular domain presents, respectively, 31% and 60% identity with the DmH1 and HsH1.1 domains (Supplementary Figure S6B–D). Besides the analyses of H1 PTMs during the cell cycle, we also performed an analysis of *PpHTO1* mRNA abundance in mitotically synchronous plasmodia. *PpHTO1* transcript abundance was maximal in early S-phase before decreasing during the cell cycle and increasing in late G2-phase to reach a level comparable to the early S-phase (Figure 6C). Hence, *Physarum* displays only one linker histone with high divergence compared to other organisms and specific epigenetic marks that could vary during the cell cycle.

Histone genes and mRNA structure

Our analyses of *Physarum* histone proteins revealed that they mostly had an independent evolutionary history but could present features shared with animal proteins. Thus, we wanted to know whether various features of *Physarum* histones genes and transcripts were shared with other kingdoms. Animal genes encoding canonical histones are intron-less while metazoan variant genes present introns like those coding both plant histone isoforms (31). In *Physarum,* the *PpHTO1, PpHTA1, PpHTB3* and *PpHTF3* are devoid of intron while the other histone-coding genes present at least one (Table 1 and Supplementary Figure S7A). Thus, genes coding canonical histones such as *PpHTT1* present introns in *Physarum*. Besides histone gene structure, animal and plant kingdoms differ by the 3′ end processing of their histone transcripts. Metazoan histone replication-dependent mRNAs have a conserved stem-loop structure composed of a six base-pair stem and a four base-pair loop. Such SL structures were retrieved in most *Physarum* histone mRNAs, not only in canonical but also in the variant transcripts (Figure 7A and Supplementary Figure S8A, Table 1). Moreover, some transcripts present all the invariant elements in the SL structure while some deviations were observed for others (Figure 7A) but, generally, the loop is U-rich and the base of the stem harbors a G-C pair (Figure 7B) similar to what was reported in other species (57). Besides the SL structure, the purine-rich HDE sequence which is also involved in this 3′end processing mechanism was found in all SL-containing transcripts, except for the *PpHTT5* mRNA (Supplementary Figure S8A and Table 1). Furthermore, we looked in *Physarum* histone transcripts for the presence of the PAS signal which consists of the AAUAAA-like motif and its variants (58). The PAS signature was found in all histone mRNAs, except for the *PpHTA1* one (Supplementary Figure S8A–C, Table 1). Thus, most histone transcripts contain signals for both 3′end processing mechanisms (Supplementary Figure S8C), suggesting that most *Physarum* histone mRNAs could be processed either by one or the other 3′end mech-

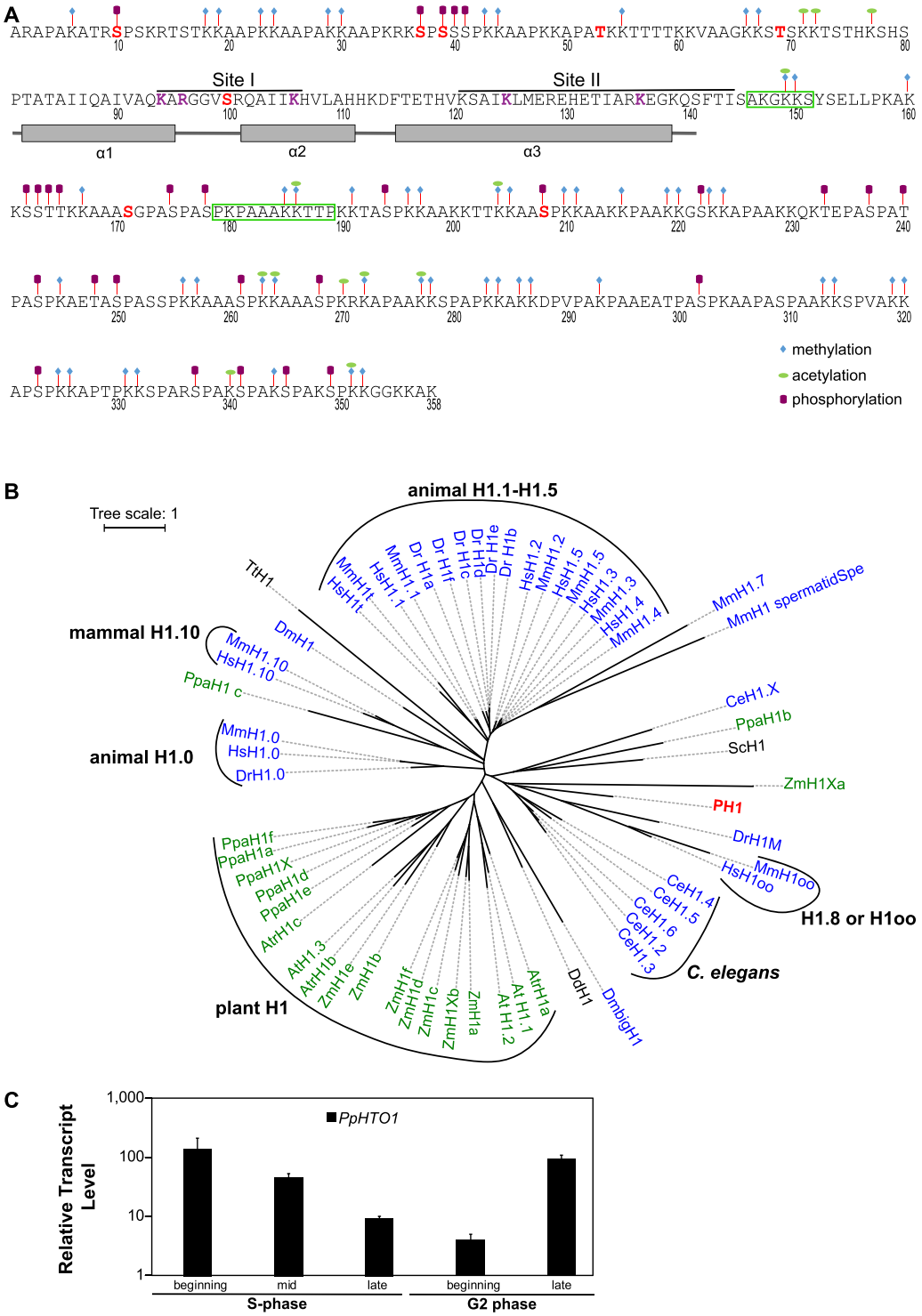

Figure 6. Conservation and divergence of PpH1. (**A**) Protein sequence of the PpH1 protein. The PTMs are shown based on the mass spectrometry analysis. Only lysine methylation and acetylation as well as S/T phosphorylation were analyzed. Positions refer to the mature protein without the initial methionine. The various helices are indicated. Conserved residue known to be phosphorylated in *T. thermophila*, in *S. cerevisiae*, in *D. melanogaster* or in *H. sapiens* are depicted in red. Details for identified PTMs in PpH1 are reported in Supplementary Table S2. DNA-binding sites I and II are indicated by a black line and basic residues located in both sites are highlighted in dark purple. Regions shown to be involved in nucleosome binding for DmH1 (52) are labeled by a green frame. (**B**) Phylogenetic tree of H1 proteins from *P. polycephalum* and various eukaryotes. H1oo: oocyte-specific H1 variant also named H1.8. The H1.0 is an animal-specific clade. Atr: *Amborella trichopoda;* At: *Arabidopsis thaliana*; Ce: *Caenorhabditis elegans;* Dd: *Dictyostellium discoideum*; Dr: *Danio rerio;* Dm: *Drosophila melanogaster*; Hs: *Homo sapiens*; Mm: *Mus musculus*; Pp: *Physarum* polycephalum; Ppa: *Physcomitrella patens*, Sc: *Saccharomyces cerevisiae;* Tt: *Tetrahymena thermophila*; Zm: *Zea mays*. (**C**) Mean relative transcript level for the *Physarum* histone H1-coding gene displayed with a logarithmic scale and measured by qRT-PCR on three biological replicates consisting of independent mitotically synchronous plasmodia cultures harvested at different times of the cell cycle.

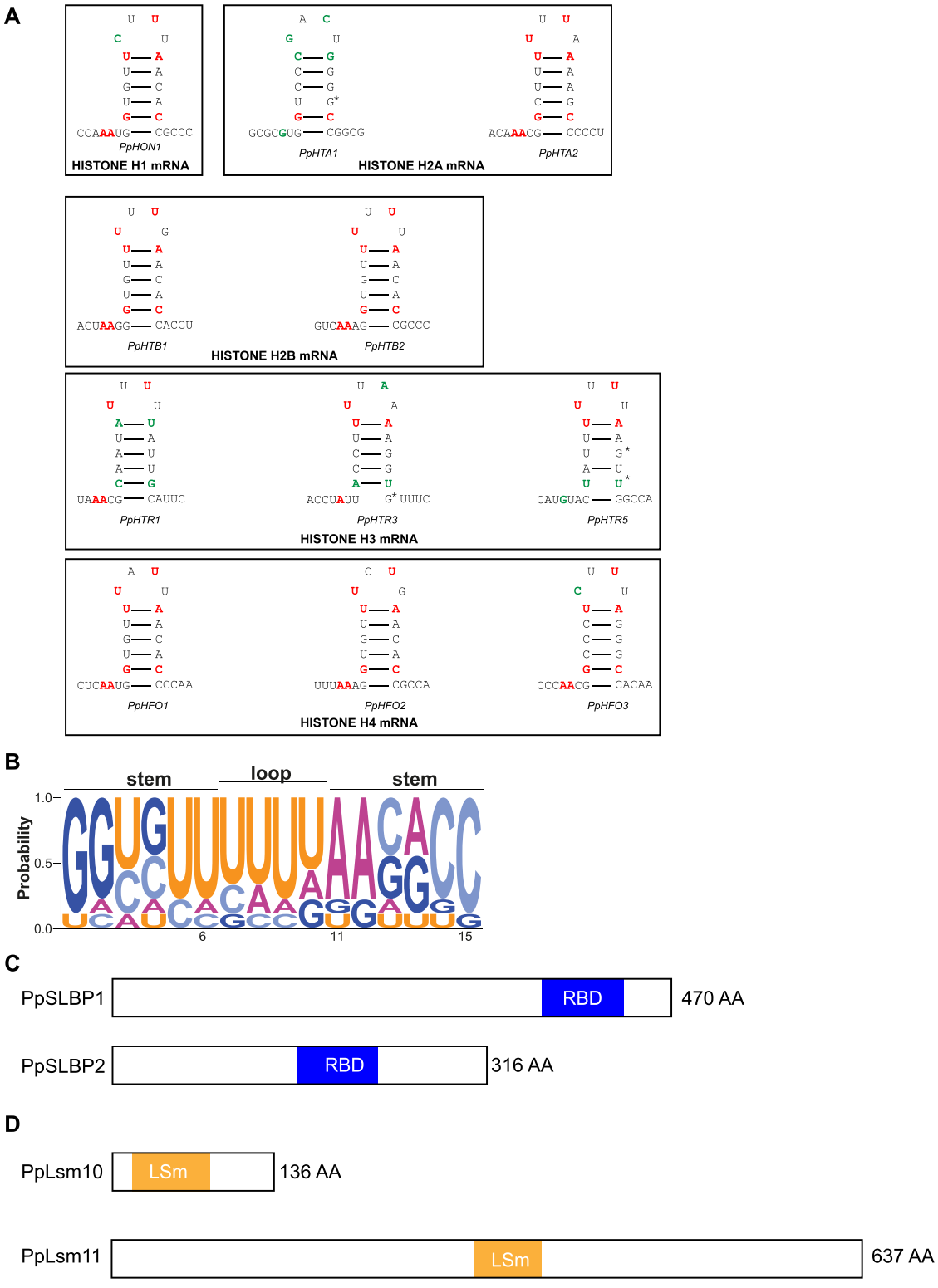

Figure 7. Analysis of 3′end processing of histone transcripts. (**A**) Stem-loop (SL) structures identified in *Physarum* histone mRNAs. Nucleotide positions identified as crucial for SLBP binding are depicted in red, deviations from those positions are displayed in green. Asterisks indicate lack of Watson–Crick base-pairing. (**B**) Logo for nucleotides composing the stem-loop structure of *Physarum* histone mRNAs. The nucleotide height represents its relative frequency at that position. (**C** and **D**) Functional domains of PpSLBP1, PpSLBP2, PpLSm10 and PpLSm11 proteins. RBD: RNA hairpin-binding domain; LSm: Like-Sm domain.

anism or by both as previously reported in various organisms ranging from mice to *D. discoideum* (57). We further confirmed the presence of polyadenylated forms of histone transcripts for all histone genes expressed in plasmodia, by the selection of poly(A)+ mRNAs (Supplementary Figure S8D). *Physarum* H4 mRNAs were reported to be mainly not polyadenylated (49). However, the presence of poly(A)+ RNAs seemed to have been underestimated since published transcriptomes of *Physarum* (6,15) were produced from poly(A)-selected transcripts and enabled the identification of H4 transcripts and we also showed presence of poly(A)+ RNAs for *PpHTF1* and *PpHTF2* mRNAs. Our comparison of the relative abundance of poly(A)+ histone mRNAs showed a similar abundance between each other when compared to the total RNAs (Figure 1A), with the exception of *PpHTA2* and *PpHTT4*. Indeed, they both present more abundant poly(A)+ mRNAs when compared to total RNAs, suggesting that these two genes might produce more poly(A)+ mRNAs than SL-structured RNAs.

After the analysis of histone transcript structure for processing, we investigated the presence of some proteins involved in 3′end processing of replication-dependent mRNAs in *Physarum,* i.e. SLBP and the LSm10 and LSm11 (Like Sm10 and 11) proteins, which are both required for U7 binding. We found two SLBP proteins (Figure 7C): (i) PpSLBP1, a protein of 470 amino acids encoded by Phypoly_transcript_05700; (ii) PpSLBP2, a protein of 316 amino acids encoded by Phypoly_transcript_13501. Both proteins contained the characteristic histone RNA-hairpin-binding domain (RBD) which is centrally located in PpSLBP2, similarly to the human protein, and in the C-terminus of PpSLBP1 (Figure 7C). The N-terminus of PpSLBP1 (361 amino acids) is much longer than the PpSLBP2 one (155 amino acids) and their RBD domains share 60% identity between them and, respectively, 50% or 48% to the human one (Supplementary Figure S9A). The alignment of RBD from several species showed a strong conservation (Supplementary Figure S9B). Indeed, most RBD-located residues (9 out of 11) necessary for SLBP-binding to the SL structure in human (59) are conserved in PpSLBP1 and PpSLBP2 (Supplementary Figure S9B). The overall structure of the RBD domain i.e. the three helices (60) are also conserved in both *Physarum* proteins: PpSLBP2 presents a structure similar to HsSLBP but PpSLBP1 presents a differently structured region between helices αB and αC (Supplementary Figure S9C). Besides *Physarum, X. laevis* and *D. rerio* also present two SLBP proteins; *Xenopus* SLBP proteins have specialized functions: XlSLBP1 participates in pre-mRNA processing in the nucleus (61) while the cytoplasm-located and oocyte-specific XlSLBP2 protein binds stored histone mRNAs (61) and inhibits their translation (62). Our phylogenetic analysis revealed that each SLBP emerged independently during evolution in these three organisms and that SLBPs display a high diversity given the long branches of the phylogenetic tree (Supplementary Figure S9D). We observed that *PpSLBP2* transcripts are far more abundant in plasmodia than *PpSLBP1* transcripts (Supplementary Figure S9E), suggesting that *Physarum* SLBP proteins could have distinct functions that may be related to the life stage of *Physarum* and/or to control histone mRNA processing and translation initiation. Besides SLBP, we also identified the two U7 snRNP-specific Sm-like proteins that we named PpLSm10 and PpLSm11 (Figure 7D, respectively encoded by Phypoly_transcript_22825 and Phypoly_transcript_04064).

To conclude, *Physarum* possesses the major components of 3′end processing by SL structure and its histone transcripts harbor elements for 3′end processing through SL structure and polyadenylation, suggesting that these two mechanisms could operate in *Physarum*. This suggests that regarding gene and transcript structures, this organism shares features of both animal and plant kingdoms.

DISCUSSION

Our study opened new perspectives on the evolutionary history of eukaryotes and Mycetozoans, which is still under debate. The mono- or polyphyly of Mycetozoans has been based on several phylogenetic trees (63,64), mainly constructed with ribosomal RNA sequences which were suggested to introduce a bias because of accelerated evolution of *Physarum* for ribosomal RNAs (65). Our phylogenetic trees use histone proteins which harbor a slow rate of sequence evolution and thus constitute a well-suited tool to perform an accurate phylogenetic analysis of distantly related organisms. Phylogenetic analyses of *Physarum* histones confirmed that H2A.Z isoforms (66) and H4 arise from a monophyletic group while other *Physarum* histones (H1, H2A.1, the 3 H2B and the 5 H3 proteins) appeared independently during evolution. In addition, we found that the *Physarum* H2A.X shared a common ancestor with yeast H2A, which has been proposed as a first step towards evolution of the repair-specific function for H2A.X of mammals (67), strengthening the hypothesis that the ancestral eukaryote harbored the H2A.Z variant and only one other H2A that further evolved in the canonical H2A and H2A.X variant.

The investigation of the histone genes within the *Physarum* genome pinpointed that three histone genes (*PpHTA1, PpHTB3, PpHTF3*) failed to be amplified from genomic DNA and cDNA from vegetative growth of the TU291 strain. Furthermore, analyses of published transcriptomes showed that these three genes were not expressed in the various tested stages of the life cycle (6,15). These discrepancies might be due to sequencing issues, spurious assembly or sequence variations between strains, although the TU291 strain was shown to have one of the largest *Physarum* genome size (68).

There is a great diversity of histone gene organization and structure within eukaryotes. Animals have multiple canonical histone gene copies organized in clusters while plants have few copies dispersed within the genome (31). Moreover, contrary to the plant kingdom, canonical-coding genes in mammals are intron-less. Regarding these features, *Physarum* was found closer to plants since few histone genes are dispersed within the *Physarum* genome and genes coding for canonical histones can contain introns. Importantly, our analyses revealed that *Physarum* exhibited a unique profile of histone gene expression. Indeed, whereas animal and plant kingdoms present an abundant transcription of histone genes in S-phase (69), we found that all expressed histone genes exhibit an elevated mRNA abundance in late

G2-phase, in agreement with what was previously reported for H4 genes (70). It is likely that this unusual timing of histone gene transcription enables the constitution of a pool of histone mRNAs in preparation for the next S-phase, since *Physarum* lacks a G1 phase.

The cell-cycle regulation of histone mRNA abundance is achieved not only at the transcriptional level but also at the post-transcriptional level by 3′end processing of histone mRNAs in mammalian cells (69,71) and plants (72). Indeed, translation initiation and efficiency of histone mRNAs can be modulated by their 3′end structure (73). Analyses of a reporter mRNA containing a 3′ stem-loop structure revealed that the SL structure increased mRNA translation efficiency and stability in Chinese hamster ovary cells but not in plant protoplasts (74). Indeed, in plants, the post-transcriptional regulation of histone mRNA abundance is achieved by histone mRNA polyadenylation but not by the SL structure (72,75). Our analyses of the histone 3′end mRNA structure in *P. polycephalum* showed that most histone transcripts display *cis*-elements for both 3′end processing mechanisms i.e. by the SL structure and polyadenylation (Figure 7A and Supplementary Figure S8A–C). The unique structural features of the 3′end of *Physarum* histone mRNAs led us to propose two mutually non-exclusive hypotheses to explain how *Physarum* could coordinate 3′end processing of histone transcripts by both mechanisms: (i) each 3′end processing mechanism occurs at specific phases of the life cycle, similarly to the human H2A.X (76); (ii) cryptic PAS located downstream of the HDE signature could constitute a fail-safe mechanism for cleavage and 3′end processing. These coordinated mechanisms might contribute *in fine* to control histone mRNA stability and translation efficiency to produce a stoichiometric amount of histones to reconstitute chromatin after completion of DNA replication.

Our comprehensive study of *P. polycephalum* histones demonstrates that this organism shares ancient and universal features of histones at the gene, transcript and protein levels. However, we also identified features specific to either animals or plants although our various phylogenic trees suggested an evolution closer to animals than plants, as previously reported thanks to phylogenomic analyses (77). Most surprisingly, we found that most *Physarum* histone mRNAs presented two mechanisms of 3′end processing for histones mRNAs, which are believed to be exclusive of the animal or plant kingdoms. Our unprecedented analyses of *Physarum* histones combined with the unique characteristics of the biology of this organism (i.e. giant cells with millions of synchronous nuclei) demonstrate that *Physarum polycephalum* is a valuable organism for deciphering a broad range of epigenetic mechanisms.

DECLARATIONS

Availability of data and materials

All *Physarum* transcript and protein data described in the present study were deposited at GenBank. All public data (genome and transcriptomes of *Physarum*) used in this study have been listed in Materials and Methods.

SUPPLEMENTARY DATA

Supplementary Data are available at NARGAB Online.

ACKNOWLEDGEMENTS

We thank Pascal Gervier for the technical assistance as well as Augustin Moreau.
Authors' Contributions: C.D., A.P. and C.T. conceived the study. C.D., L.N.M., A.P. and C.T. designed and performed the experiments. C.D. and A.P. analyzed the available genomic and RNA-Seq data. C.D., A.P. and S.T. performed the bioinformatics analyses. C.D., L.N.M., A.P. and C.T. analyzed the data. J.J.H., Y.J., L.N.M. and S.T. participated in the discussions about the project and critically read the manuscript. C.D., A.P. and C.T. wrote the article. All authors read and approved the final manuscript.

FUNDING

University of Nantes (to C.D.); Région Pays de la Loire (PULSAR) (to C.D.); Ligue contre le Cancer (Grand Ouest) (to C.T.); National Institutes of Health [R35GM128661 (to Y.J.); R01GM052426 (to J.J.H.)].
Conflict of interest statement. None declared.

REFERENCES

1. Luger,K., Mäder,A.W., Richmond,R.K., Sargent,D.F. and Richmond,T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251–260.
2. Thoma,F. and Koller,T. (1977) Influence of histone H1 on chromatin structure. *Cell*, **12**, 101–107.
3. Hansen,J.C., Connolly,M., McDonald,C.J., Pan,A., Pryamkova,A., Ray,K., Seidel,E., Tamura,S., Rogge,R. and Maeshima,K. (2018) The 10-nm chromatin fiber and its relationship to interphase chromosome organization. *Biochem. Soc. Trans.*, **46**, 67–76.
4. Marzluff,W.F. and Duronio,R.J. (2002) Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. *Curr. Opin. Cell Biol.*, **14**, 692–699.
5. Marzluff,W.F., Wagner,E.J. and Duronio,R.J. (2008) Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat. Rev. Genet.*, **9**, 843–854.
6. Schaap,P., Barrantes,I., Minx,P., Sasaki,N., Anderson,R.W., Benard,M., Biggar,K.K., Buchler,N.E., Bundschuh,R., Chen,X. *et al.* (2015) The physarum polycephalum genome reveals extensive use of prokaryotic two-component and metazoan-type tyrosine kinase signaling. *Genome Biol. Evol.*, **8**, 109–125.
7. Boussard,A., Fessel,A., Oettmeier,C., Briard,L., Döbereiner,H.G. and Dussutour,A. (2021) Adaptive behaviour and learning in slime moulds: the role of oscillations. *Philos. Trans. R. Soc. B Biol. Sci.*, **376**, 20190757.
8. Alim,K., Andrew,N., Pringle,A. and Brenner,M.P. (2017) Mechanism of signal propagation in *physarumpolycephalum*; *Proc. Natl. Acad. Sci.*, **114**, 5136–5141.
9. Ejlassi,A., Menil-Philippot,V., Galvani,A. and Thiriet,C. (2017) Histones H3 and H4 require their relevant amino-tails for efficient nuclear import and replication-coupled chromatin assembly *in vivo*. *Sci. Rep.*, **7**, 1–10.
10. Schofield,P.N. and Walker,I.O. (1982) Control of histone gene expression in *physarumpolycephalum*. I. Protein synthesis during the cell cycle. *J. Cell Sci.*, **57**, 139–150.
11. Burland,T.G., Solnica-Krezel,L., Bailey,J., Cunningham,D.B. and Dove,W.F. (1993) Patterns of inheritance, development and the mitotic cycle in the protist *physarumpolycephalum*. *Adv. Microb. Physiol.*, **35**, 1–69.
12. Mohberg,J. and Rusch,H.P. (1969) Isolation of the nuclear histones from the myxomycete, *physarumpolycephalum*. *Arch. Biochem. Biophys.*, **134**, 577–589.

13. Mende,L.M., Waterborg,J.H., Mueller,R.D. and Matthews,H.R. (1983) Isolation, identification, and characterization of histones from plasmodia of the true slime mold *physarumpolycephalum* using extraction with guanidine hydrochloride. *Biochemistry*, **22**, 38–51.
14. Cary,P.D., Carpenter,B.G. and Foote,A.M. (1985) Physical studies by NMR and circular dichroism determining three structurally different domains in *physarumpolycephalum* histone h1. *Eur. J. Biochem.*, **151**, 579–589.
15. Glöckner,G. and Marwan,W. (2017) Transcriptome reprogramming during developmental switching in physarum polycephalum involves extensive remodeling of intracellular signaling networks. *Sci. Rep.*, **7**, 1–12.
16. Yelagandula,R., Stroud,H., Holec,S. and Zhou,K. (2014) The histone variant H2A. W defines heterochromatin and promotes chromatin condensation in *arabidopsis*. *Cell*, **158**, 98–109.
17. Gertz,E.M., Yu,Y.-K., Agarwala,R., Schäffer,A.A. and Altschul,S.F. (2006) Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol.*, **4**, 41.
18. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
19. Dubin,M., Fuchs,J., Gräf,R., Schubert,I. and Nellen,W. (2010) Dynamics of a novel centromeric histone variant cenh3 reveals the evolutionary ancestral timing of centromere biogenesis. *Nucleic Acids Res.*, **38**, 7526–7537.
20. Castresana,J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.*, **17**, 540–552.
21. Price,M.N., Dehal,P.S. and Arkin,A.P. (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One*, **5**, e9490.
22. Letunic,I. and Bork,P. (2021) Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.*, **49**, W293–W296.
23. Madeira,F., Park,Y.M., Lee,J., Buso,N., Gur,T., Madhusoodanan,N., Basutkar,P., Tivey,A.R.N., Potter,S.C., Finn,R.D. *et al.* (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.*, **47**, W636–W641.
24. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
25. Crooks,G.E., Hon,G., Chandonia,J.-M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
26. Adams,D.S., Noonan,D., Burn,T.C. and Skinner,H.B. (1987) A library of trimethylguanosine-capped small RNAS in *physarumpolycephalum*. *Gene*, **54**, 93–103.
27. Marz,M., Mosig,A., Stadler,B.M.R. and Stadler,P.F. (2007) U7 snRNAs: a computational survey. *Genomics Proteomics Bioinformatics*, **5**, 187–195.
28. Thiriet,C. and Hayes,J.J. (1999) Histone proteins in vivo: cell-cycle-dependent physiological effects of exogenous linker histones incorporated into *physarumpolycephalum*. *Methods*, **17**, 140–150.
29. Thiriet,C. (2004) Analysis of chromatin assembled *in vivo* using exogenous histones in physarum polycephalum. *Methods*, **33**, 86–92.
30. Lifton,R.P., Goldberg,M.L., Karp,R.W. and Hogness,D.S. (1978) The organization of the histone genes in drosophila melanogaster: functional and evolutionary implications. *Cold Spring Harb. Symp. Quant. Biol.*, **42**, 1047–1051.
31. Probst,A. V, Desvoyes,B. and Gutierrez,C. (2020) Similar yet critically different: the distribution, dynamics and function of histone variants. *J. Exp. Bot.*, **71**, 5191–5204.
32. Kurat,C.F., Recht,J., Radovani,E., Durbic,T., Andrews,B. and Fillingham,J. (2014) Regulation of histone gene transcription in yeast. *Cell. Mol. Life Sci.*, **71**, 599–613.
33. Glöckner,G., Golderer,G., Werner-Felmayer,G., Meyer,S. and Marwan,W. (2008) A first glimpse at the transcriptome of physarum polycephalum. *BMC Genomics*, **9**, 1–11.
34. Rogakou,E.P., Pilch,D.R., Orr,A.H., Ivanova,V.S. and Bonner,W.M. (1998) DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.*, **273**, 5858–5868.
35. Xiao,A., Li,H., Shechter,D., Ahn,S.H., Fabrizio,L.A., Erdjument-Bromage,H., Ishibe-Murakami,S., Wang,B., Tempst,P., Hofmann,K. *et al.* (2009) WSTF regulates the H2A.X DNA damage response via a novel tyrosine kinase activity. *Nature*, **457**, 57–62.
36. Jiang,W., Guo,X. and Bhavanandan,V.P. (1998) Histone H2A.F/Z subfamily: the smallest member and the signature sequence. *Biochem. Biophys. Res. Commun.*, **245**, 613–617.
37. Talbert,P., Ahmad,K., Almouzni,G., Ausió,J., Berger,F., Bhalla,P.L., Bonner,W.M., Cande,W.Z., Chadwick,B.P., Chan,S.W. *et al.* (2012) A unified phylogeny-based nomenclature for histone variants. *Epigenetics Chromatin*, **5**, 7.
38. Barnes,C.E., English,D.M. and Cowley,S.M. (2019) Acetylation & co: an expanding repertoire of histone acylations regulates chromatin and transcription. *Essays Biochem.*, **63**, 97–107.
39. Kubbies,M. and Pierron,G. (1983) Mitotic cell cycle control in *physarum*. Unprecedented insights via flow-cytometry. *Exp. Cell Res.*, **149**, 57–67.
40. Szenker,E., Ray-Gallet,D. and Almouzni,G. (2011) The double face of the histone variant H3.3. *Cell Res.*, **21**, 421–434.
41. Arnaud,M.-C. and Thiriet,C. (2013) Histone tail modifications of H3 and H4 during the physarum polycephalum cell cycle. *Online J. Biol. Sci.*, **12**, 54–61.
42. Srivastava,S. and Foltz,D.R. (2018) Posttranslational modifications of CENP-A: marks of distinction. *Chromosoma*, **127**, 279–290.
43. Long,M., Sun,X., Shi,W., Yanru,A., Leung,S.T.C., Ding,D., Cheema,M.S., MacPherson,N., Nelson,C.J., Ausio,J. *et al.* (2019) A novel histone H4 variant H4G regulates rDNA transcription in breast cancer. *Nucleic Acids Res.*, **47**, 8399–8409.
44. Hu,Y. and Lai,Y. (2015) Identification and expression analysis of rice histone genes. *Plant Physiol. Biochem.*, **86**, 55–65.
45. Siegel,T.N., Hekstra,D.R., Kemp,L.E., Figueiredo,L.M., Lowell,J.E., Fenyo,D., Wang,X., Dewell,S. and Cross,G.A.M. (2009) Four histone variants mark the boundaries of polycistronic transcription units in trypanosoma brucei. *Genes Dev.*, **23**, 1063–1076.
46. Wilhelm,M.L. and Wilhelm,F.X. (1984) A transposon-like DNA fragment interrupts a *physarumpolycephalum* histone H4 gene. *FEBS Lett.*, **168**, 249–254.
47. Wilhelm,M.L. and Wilhelm,F. (1987) Both histone H4 genes of *physarumpolycephalum* are interrupted by an intervening sequence. *Nucleic Acids Res.*, **15**, 5478.
48. Wilhelm,M.L. and Wilhelm,F.X. (1989) Histone genes in *physarumpolycephalum*: transcription and analysis of the flanking regions of the two H4 genes. *J. Mol. Evol.*, **28**, 322–326.
49. Wilhelm,M.L., Toublan,B., Fujita,R.A. and Wilhelm,F.X. (1988) Histone H4 mRNA is stored as a small cytoplasmic RNP during the G2 phase in *physarumpolycephalum*. *Biochem. Biophys. Res. Commun.*, **153**, 162–171.
50. Cutter,A.R. and Hayes,J.J. (2017) Linker histones: novel insights into structure-specific recognition of the nucleosome. *Biochem. Cell Biol.*, **95**, 171–178.
51. Goytisolo,F.A., Gerchman,S.E., Yu,X., Rees,C., Graziano,V., Ramakrishnan,V. and Thomas,J.O. (1996) Identification of two DNA-binding sites on the globular domain of histone h5. *EMBO J.*, **15**, 3421–3429.
52. Zhou,B.R., Feng,H., Kato,H., Dai,L., Yang,Y., Zhou,Y. and Bai,Y. (2014) Structural insights into the histone H1-nucleosome complex. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 19390–19395.
53. Mueller,R.D., Yasuda,H. and Bradbury,E.M. (1985) Phosphorylation of histone H1 through the cell cycle of *physarumpolycephalum*. 24 sites of phosphorylation at metaphase. *J. Biol. Chem.*, **260**, 5081–5086.
54. Bradbury,E.M., Inglis,R.J. and Matthews,H.R. (1974) Control of cell division by very lysine rich histone (F1) phosphorylation. *Nature*, **247**, 257–261.
55. Thiriet,C. and Hayes,J.J. (2009) Linker histone phosphorylation regulates global timing of replication origin firing. *J. Biol. Chem.*, **284**, 2823–2829.
56. Andrés,M., García-Gomis,D., Ponte,I., Suau,P. and Roque,A. (2020) Histone h1 post-translational modifications: update and future perspectives. *Int. J. Mol. Sci.*, **21**, 1–22.
57. López,M.D. and Samuelsson,T. (2008) Early evolution of histone mRNA 3′ end processing. *RNA*, **14**, 1–10.
58. Sherstnev,A., Duc,C., Cole,C., Zacharaki,V., Hornyik,C., Ozsolak,F., Milos,P.M., Barton,G.J. and Simpson,G.G. (2012) Direct sequencing of arabidopsis thaliana RNA reveals patterns of cleavage and polyadenylation. *Nat. Struct. Mol. Biol.*, **19**, 845–852.

59. Dominski,Z., Erkmann,J.A., Greenland,J.A. and Marzluff,W.F. (2001) Mutations in the RNA binding domain of stem-loop binding protein define separable requirements for RNA binding and for histone pre-mRNA processing. *Mol. Cell. Biol.*, **21**, 2008–2017.
60. Tan,D., Marzluffl,W.F., Dominski,Z. and Tong,L. (2013) Structure of histone mRNA stem-loop, human stem-loop binding protein, and 3′hExo ternary complex. *Science (80-.)*, **339**, 318–321.
61. Wang,Z.F., Ingledue,T.C., Dominski,Z., Sanchez,R. and Marzluff,W.F. (1999) Two xenopus proteins that bind the 3′ end of histone mRNA: implications for translational control of histone synthesis during oogenesis. *Mol. Cell. Biol.*, **19**, 835–845.
62. Sánchez,R. and Marzluff,W.F. (2004) The oligo(A) tail on histone mRNA plays an active role in translational silencing of histone mRNA during xenopus oogenesis. *Mol. Cell. Biol.*, **24**, 2513–2525.
63. Baldauf,S.L. and Doolittle,W.F. (1997) Origin and evolution of the slime molds (Mycetozoa). *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 12007–12012.
64. Nandipati,S.C.R., Haugli,K., Coucheron,D.H., Haskins,E.F. and Johansen,S.D. (2012) Polyphyletic origin of the genus physarum (Physarales, myxomycetes) revealed by nuclear rDNA mini-chromosome analysis and group i intron synapomorphy. *BMC Evol. Biol.*, **12**, 1.
65. Fiore-Donno,A.M., Nikolaev,S.I., Nelson,M., Pawlowski,J., Cavalier-Smith,T. and Baldauf,S.L. (2010) Deep phylogeny and evolution of slime moulds (Mycetozoa). *Protist*, **161**, 55–70.
66. Thatcher,T.H. and Gorovsky,M.A. (1994) Phylogenetic analysis of the core histones H2A, H2B, H3, and h4. *Nucleic Acids Res.*, **22**, 174–179.
67. House,N.C., Polleys,E.J., Quasem,I., De la Rosa Mejia,M., Joyce,C.E., Takacsi-Nagy,O., Krebs,J.E., Fuchs,S.M. and Freudenreich,C.H. (2019) Distinct roles for s. cerevisiae H2A copies in recombination and repeat stability, with a role for H2A.1 threonine 126. *Elife*, **8**, e53362.
68. Kubbies,M., Wick,R., Hildebrandt,A. and Sauer,H.W. (1986) Flow cytometry reveals a high degree of genomic size variation and mixoploidy in various strains of the acellular slime mold physarum polycephalum'. *Cytometry*, **7**, 491–485.
69. Osley,M.A. (1991) The regulation of histone synthesis in the cell cycle. *Annu. Rev. Biochem.*, **60**, 827–861.
70. Wilhelm,M.L., Toublan,B., Jalouzot,R. and Wilhelm,F.X. (1984) Histone H4 gene is transcribed in s phase but also late in G2 phase in physarum polycephalum. *EMBO J.*, **3**, 2659–2662.
71. Harris,M.E., Bohni,R., Schneiderman,M.H., Ramamurthy,L., Schumperli,D. and Marzluffl,W.F. (1991) Regulation of histone mRNA in the unperturbed cell cycle : evidence suggesting control at two posttranscriptional steps. *Mol. Cell. Biol.*, **11**, 2416–2424.
72. Reichheld,J.P., Gigot,C. and Chaubet-gigot,N. (1998) Multilevel regulation of histone gene expression during the cell cycle in tobacco cells. *Nucleic Acids Res.*, **26**, 3255–3262.
73. Curry,S., Kotik-Kogan,O., Conte,M.R. and Brick,P. (2009) Getting to the end of RNA: structural analysis of protein recognition of 5′ and 3′ termini. *Biochim. Biophys. Acta - Gene Regul. Mech.*, **1789**, 653–666.
74. Gallie,D.R., Lewis,N.J. and Marzluff,W.F. (1996) The histone 3′-terminal stem-loop is necessary for translation in chinese hamster ovary cells. *Nucleic Acids Res.*, **24**, 1954–1962.
75. Chaubet,N., Chaboute,M.E., Clément,B., Ehling,M., Philipps,G. and Gigot,C. (1988) The histone H3 and H4 mRNAs are polyadenylated in maize. *Nucleic Acids Res.*, **16**, 1295–1304.
76. Mannironi,C., Bonner,W.M. and Hatch,C.L. (1989) H2A.X. a histone isoprotein with a conserved C-terminal sequence, is encoded by a novel mRNA with both DNA replication type and polyA 3′ processing signals. *Nucleic Acids Res.*, **17**, 9113–9126.
77. Keeling,P.J. and Burki,F. (2019) Progress towards the tree of eukaryotes. *Curr. Biol.*, **29**, R808–R817.