Research Article

Evolution of a histone variant involved in compartmental regulation of NAD metabolism

Iva Guberovic[1], Sarah Hurtado-Bagès[1], Ciro Rivera-Casas[2], Gunnar Knobloch[3,4], Roberto Malinverni[1], Vanesa Valero[1], Michelle M. Leger[5], Jesús García[6], Jerome Basquin[7], Marta Gómez de Cedrón[8], Marta Frigolé-Vivas[6], Manjinder S. Cheema[9], Ainhoa Pérez[1], Juan Ausió[9], Ana Ramírez de Molina[8], Xavier Salvatella[6,10,11], Iñaki Ruiz-Trillo[5,11], Jose M. Eirin-Lopez[2], Andreas G. Ladurner[3,4], Marcus Buschbeck[1,12,*]

Affiliations:

1 Cancer and Leukaemia Epigenetics and Biology Program, Josep Carreras Leukaemia Research Institute (IJC), Campus Can Ruti, Badalona, Spain.

2 Environmental Epigenetics Laboratory, Institute of Environment, Florida International University, Miami FL, USA.

3 Biomedical Center (BMC), Department of Physiological Chemistry, LMU Munich, 82152 Martinsried Planegg, Germany.

4 Eisbach Bio GmbH, Am Klopferspitz 19, 82152 Martinsried Planegg, Germany.

5 Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain.

6 Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Spain

7 Max Planck Institute of Biochemistry (MPI), Department of Structural Cell Biology, 82152 Martinsried Planegg, Germany.

8 Molecular Oncology Group, IMDEA Food Institute, CEI UAM + CSIC, Madrid, Spain.

9 Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada.

10 Joint BSC-IRB Research Programme in Computational Biology, Barcelona, Spain.

11 ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain.

12 Program for Predictive and Personalized Medicine of Cancer, Germans Trias i Pujol Research Institute (PMPPC-IGTP), Badalona, Spain.

*Correspondence: Marcus Buschbeck, email: mbuschbeck@carrerasresearch.org

Keywords:
NAD metabolism, macroH2A, compartmentalization, macrodomains, chromatin
Short-title: Conserved role of a histone in nuclear NAD metabolism

SUMMARY

NAD metabolism is essential for all forms of life. Compartmental regulation of NAD^+ consumption, especially between the nucleus and the mitochondria, is required for energy homeostasis. However, how compartmental regulation evolved remains unclear. Here, we investigated the evolution of the macrodomain-containing histone variant macroH2A1.1, an integral chromatin component that limits nuclear NAD^+ consumption by inhibiting PARP1 in vertebrate cells. We found that macroH2A originated in pre-metazoan protists. The crystal structure of the macroH2A macrodomain from the protist *Capsaspora* allowed us to identify highly conserved principles of ligand binding and pinpoint key residue substitutions, selected for during the evolution of the vertebrate stem lineage. Metabolic characterization of the *Capsaspora* life cycle suggested that the metabolic function of macroH2A was associated with non-proliferative stages. Taken together, we provide insight into the evolution of a chromatin element involved in compartmental NAD regulation, relevant for understanding of its metabolism and potential therapeutic applications.

INTRODUCTION

NAD metabolism plays an essential role in all domains of life[1,2]. NAD functions as a redox cofactor and a signalling molecule. As a redox cofactor in catabolic reactions, it enables ATP production in mitochondria[3]. As a signalling molecule in its oxidized form, NAD^+ serves as donor of ADP-ribose moieties for effector enzymes such as PARPs, sirtuins and CD38 family of hydrolases[4]. The biosynthetic enzymes and effectors involved in NAD metabolism are differentially distributed in the cell, making NAD metabolism highly compartmentalized[5,6]. In particular, the balance between NAD^+ consumption in the nucleus and its availability for cytosolic and mitochondrial redox reactions is essential for energy homeostasis[6,7]. A major challenge for the field is to understand how is NAD metabolism regulated on the compartmental level, and to dissect the relevance of this regulation for health and disease[5].

PARP1 is the major NAD^+ consuming enzyme in the nucleus, best known for its function as a sensor in the DNA damage response[8]. The inhibition of PARP1 results in increased global NAD^+ levels in both cultured cells and mice, indicating that PARP1 also consumes NAD^+ under basal physiological conditions[9–11]. Conversely, nuclear NAD^+ levels regulate PARP1 activity[12,13]. A picture is emerging in which the nuclear NAD^+ consumption by PARP1 in differentiated cells needs to be kept low to maintain NAD-dependent functions in other compartments[14]. In particular, during myogenic differentiation this is achieved by the transcriptional downregulation of PARP1 expression[15] and the simultaneous upregulation of its endogenous inhibitor, macrodomain-containing histone variant macroH2A1.1[11].

Macrodomains are ancient globular protein modules that have emerged as key players in NAD^+-dependent ADP-ribose signalling[16]. They bind ADP-ribose as a free molecule, as an oligomer, or when covalently bound to proteins as a post-translational modification[17–19], while in some cases they have hydrolysing activity[20,21]. The macroH2A family of histone variants are the only structural chromatin components containing macrodomains. In vertebrates, two genes and one event of alternative splicing give rise to three macroH2A proteins that differ in their macrodomains[22]. The capacity to bind ADP-ribose is limited to the macrodomain of the splice variant macroH2A1.1[23,24]. As a consequence, macroH2A1.1, but not macroH2A1.2 or macroH2A2, binds auto-ADP-ribosylated PARP1[19] and contributes to its enrichment on specific chromatin regions[25]. MacroH2A1.1 can inhibit PARP1 activity and thus interfere with PARP1-dependent processes[24,26]. In differentiated muscle cells, macroH2A1.1 reduces nuclear NAD^+ consumption by PARP1 and increases the availability of NAD^+ in other compartments, thus indirectly promoting respiration and ATP production in mitochondria[11].

Hence, macroH2A1.1 is a chromatin component that takes part in the compartmentalized regulation of NAD metabolism in vertebrates. Here, we addressed the question on when did macroH2A histone variants emerge on an evolutionary scale and whether their implication in metabolism was an ancient

trait. For this, we performed a phylogenetic analysis to determine the order of events in the evolution of macroH2A histone variants and characterized the function of a macroH2A macrodomain from one of the most divergent macroH2A-containing species in comparison to vertebrates.

RESULTS

MacroH2A first appeared in unicellular protists

The amino acids encoded by the mutually exclusive exon 5 determine the capacity of macroH2A1.1 to bind ADP-ribose, inhibit PARP1 and thus affect NAD metabolism (Figure 1a). To understand when did this role of macroH2A1.1 evolve, we first aimed to determine when did the fusion between the histone-fold and the metabolite-binding macrodomain occur. For this, we analyzed genomic and transcriptomic sequencing data representing a wide diversity of eukaryotes, and identified a *MACROH2A* gene in 330 of them (Supplementary Data File S1 and S2). Previously, we reported the presence of macroH2A in a unicellular filasterean, *Capsaspora*[27]. Together with animals and fungi, filastereans belong to the group of opisthokonts[28]. Here, we report the presence of macroH2A sequences in protists which diverged earlier than opisthokonts, such as the haptist *Choanocystis* sp. and the breviate *Pygsuia biforma* (Figure 1b, Supplementary Data File S3). Interestingly, we did not find macroH2A in *Fungi,* nor in other unicellular opisthokonts. The presence of macroH2A became more prevalent with the emergence of animal multicellularity. Indeed, we found macroH2A-encoding genes in diverse species of most animal phyla and all vertebrates. On the other hand, and as previously reported[16,27], macroH2A was absent from several non-vertebrate species, such as *Drosophila melanogaster* and *Caenorhabditis elegans,* as well as in tunicates (Figure 1b), which is indicative of lineage-specific losses. The second macroH2A gene corresponding to human *MACROH2A2* appeared after the whole-genome duplication in the last common ancestor of vertebrates, followed by the appearance of the alternative splicing variant of the ancestral *MACROH2A1* in a common ancestor of jawed vertebrates (Figure 1b). All three vertebrate macrodomains show a significant level of conservation that placed them closer to the highly conserved, replication-coupled histone H2B than the fast-evolving histone variant H2A.Bbd (Figure 1c). Comparison of the amino acid sequences corresponding to exon 5 indicated that non-vertebrate macroH2A is most similar to macroH2A1.1 (Figure 1d). Importantly, this included a high level of conservation of amino acids known to be required for ADP-ribose binding in human macroH2A1.1 such as aspartate 203 (D203) and glycine 224 (G224)[17,23].

Taken together, these results allowed us to describe for the first time the evolutionary order of events that resulted in the three different macroH2A histone variants present in vertebrates (Figure 1e). Importantly, we found that macroH2A is much older than previously reported and originated in pre-metazoan protists. Excitingly, the first macroH2A gene is similar to vertebrate macroH2A1.1. This suggests that the function of macroH2A histone variant in nuclear NAD^+ metabolism might be ancient. The high conservation of the macrodomains of macroH2A1.2 and macroH2A2 suggest that they have acquired novel and yet unknown binding functions.

A protist macrodomain has higher ADP-ribose affinity

Given the origin of macroH2A before metazoans (animals), we sought to determine its potential ancestral metabolic implication by determining the biochemical properties of macroH2A in one of the protist organisms. As a model system we used *Capsaspora owczarzaki* (referred to as *Capsaspora*), one of the closest unicellular relatives of animals[29]. The macrodomain of *Capsaspora* interacted with ADP-ribose in a similar fashion as the murine macroH2A1.1 macrodomain as suggested by NMR-based binding spectra (Figure 2a, Extended Data Figure 1a,b). However, the ADP-ribose binding by *Capsaspora* macrodomain was eight times stronger than the macrodomain of mouse macroH2A1.1, with equilibrium dissociation constants of 1,3 µM and 11,3 µM, respectively (Figure 2b, Extended Data Figure 1c-e). The thermodynamic profiles indicated that the two macrodomains bind ADP-ribose using a different binding mode. *Capsaspora* macrodomain bound ADP-ribose through favourable enthalpic and entropic contributions, whereas the mouse macroH2A1.1 macrodomain bound it mainly in an enthalpy-driven manner, which was partially compensated by an unfavourable entropic contribution (Figure 2c). Furthermore, *Capsaspora* macroH2A macrodomain showed high selectivity for ADP-ribose binding, since its affinity towards ADP was ~50-fold lower, and no interaction was observed with related nucleotides (ATP, AMP, GDP), nor ribose (Extended Data Figure 1f).

To characterize the *Capsaspora* macroH2A macrodomain and its interactions with ADP-ribose at the atomic level, we solved the structure of the protein in the presence and the absence of ADP-ribose by protein X-ray crystallography (Figure 2d,e, Supplementary Data File S4). The unliganded macroH2A macrodomain crystallized in the space group $P12_11$ and could be refined to 1.4 Å resolution (PDB ID: 7NY6), while the ADP-ribose bound protein crystallized in the space group $P3_221$ and was refined to 2.0 Å resolution (PDB ID: 7NY7). The obtained globular structures with seven central β-sheets in the characteristic 1276354 order surrounded by α-helices (Extended Data Figure 2a,b) showed high structural similarity to the previously described macrodomains[17,23,30]. Although the C-terminal α-helix was not resolved in the ADP-ribose bound form of *Capsaspora* its root mean square deviation (RMSD) was only 0.5 Å to the human ADP-ribose bound macroH2A1.1 macrodomain (PDB ID: 3IID). The electron density of ADP-ribose clearly revealed that the ligand situates within the canonical binding pocket of macroH2A macrodomain (Figure 2f). The residues important for ligand binding included aspartate 203 (D203), which established a hydrogen bond (H-bond) with the adenine amino-group, and phenylalanine 352 (F352), which stabilized the adenine ring via π-electron stacking of the aromatic rings. Further, the amino group of the side chain of asparagine 316 (N316) established an H-bond with the distal ribose of ADP-ribose (Extended Data Figure 2c), which can explain the selectivity for ADP-ribose over ADP (Extended Data Figure 1f). Upon ADP-ribose binding, the side chains of glutamine 225 (Q225) and N316 move towards each other and establish an H-bond, resulting in a conformation that encloses the central diphosphate moiety of ADP-ribose (Figure 2g).

In summary, the capacity of the macroH2A macrodomain to bind ADP-ribose is conserved in the protist *Capsaspora*. Indeed, the *Capsaspora* macroH2A macrodomain bound ADP-ribose stronger than its vertebrate counterpart. This raises the possibility that evolution has selected for decreased ADP-ribose affinity along the vertebrate stem lineage.

Two evolutionary divergent residues close the binding pocket

Next, we determined the importance of the conservation of the amino acid sequence for ADP-ribose binding and for the structural integrity of the macrodomain. *Capsaspora* macroH2A and human macroH2A1.1 macrodomains share only 50% identity at the level of the amino acid sequence (Extended Data Figure 3a). However, the multiple sequence alignment of more than 300 macroH2A macrodomain sequences delineated two well conserved regions of the protein, one located toward the N-terminus, overlapping with region encoded by exon 5, and one toward the C-terminus of the macrodomain fold (Figure 3a). Interestingly, with the exception of serine 275, all ADP-ribose-interacting residues were located in the conserved N- and C-terminal regions. Comparison of these two regions between *Capsaspora* macroH2A and human macroH2A1.1 indicated that most amino acids involved in ADP-ribose binding were invariant or structurally related, although the *Capsaspora* macroH2A macrodomain established a lower number of bonds with ADP-ribose (Figure 3a, Extended Data Figure 3b, Supplementary Data File S5). Mapping the conservation rate on the crystal structure of the apo-form of *Capsaspora* macroH2A macrodomain indicated a high conservation of the inner part of the ADP-ribose binding pocket (Figure 3b), while the surface regions were more variable (Figure 3c, Extended Data Figure 3c, Supplementary Data File S6). The only two residues involved in ADP-ribose binding and with a low degree of conservation were Q225 and N316 in the *Capsaspora* macroH2A, which correspond to glutamate 225 (E225) and arginine 315 (R315) in human macroH2A1.1. Both pairs of residues close the binding pocket upon ADP-ribose binding. However, the polar uncharged side chains of Q225 and N316 establish an H-bond (Figure 2g), while the corresponding residues E225 and R315 in human macroH2A1.1 establish a salt bridge to close the binding pocket upon ADP-ribose binding[23]. Despite the slight difference in the orientation of these side chains in the *Capsaspora* apo form from the human apo form (Figure 3d), both macrodomains adopt an almost identical conformation upon ADP-ribose binding (Figure 3e).

Taken together, the low *a priori* conservation at the level of primary sequence contrasts with the highly conserved three-dimensional structure of the *Capsaspora* and human macrodomains. The residues forming the inner ADP-ribose binding pocket are highly conserved between *Capsaspora* and vertebrates. Two non-conserved residues corresponding to *Capsaspora* Q225 and N316 close the binding pocket of the macroH2A macrodomain after an ADP-ribose induced fit.

Succeeding substitutions of Q225 and N316 reduce affinity

We hypothesized that the specific substitution of amino acids Q225 and N316 drove the evolution towards the decreased ADP-ribose binding affinity from *Capsaspora* to vertebrates. To test this, we reconstructed the protein phylogeny with a subset of 260 sequences and inferred the ancestral states of these two sites (Figure 4a, Extended Data Figure 4, Supplementary Data File S1). The obtained results revealed that the replacement of the protist-characteristic residue at position 225 (Q225E) occurred early during macroH2A evolution, most likely in the ancestor of bilaterians and cnidarians, and was subsequently maintained in most animals (Figure 4a). On the other hand, the transition to arginine at position 316 (N316R) shows more variability, as revealed by its prevalence only at specific protostome lineages, while becoming almost ubiquitous in hemichordates and chordates (Figure 4a, Extended Data Figure 5a-c).

To understand the physiological consequence of this evolutionary course, we individually introduced Q225E and N316R mutations in the *Capsaspora* macrodomain and confirmed that the mutant proteins were folded (Extended Data Figure 5d). Interestingly, the Q225E substitution resulted in a 7-fold decrease of the affinity towards ADP-ribose compared to wild-type *Capsaspora* protein (Figure 4b). Thus, the affinity and the thermodynamic properties of ADP-ribose binding by the Q225E *Capsaspora* mutant were more similar to those of the mouse macroH2A1.1 macrodomain (Extended Data Figure 5e,f). Strikingly, the N316R substitution resulted in loss of ADP-ribose (Figure 4b, Extended Data Figure 5e). Interestingly, we observed alternative splicing of exon 7 affecting position 316 in *Mollusca* (Figure 4c, Extended Data Figure 6a). The alternatively spliced exon 7 was present in more than 25% of analyzed species (Figure 4c, Extended Data Figure 6b) and strikingly determined an either protist- or vertebrate-characteristic configuration of the amino acid 316 (Figure 4d, Extended Data Figure 6c).

Taken together, these results shed light on the evolutionary events which affected the affinity of macroH2A macrodomains towards ADP-ribose. The exchange of Q225E proved to be a determining factor for the decrease in ADP-ribose affinity over the course of evolution and preceded the epistatic change of N316R.

Dynamic regulation of *Capsaspora* metabolism

To understand how is the protist macroH2A related with metabolism, we used *Capsaspora* as a model organism. *Capsaspora* has a dynamic life cycle composed of three stages that are filopodial, aggregative and cystic stage (Figure 5a). As a response to environmental cues, single cells from the filopodial stage transition to the aggregative stage, one of the simple forms of multicellularity. In both filopodial and aggregative stage, cells are proliferative. In less advantageous environments, *Capsaspora* can transition to a spore-like cystic stage, which represents a resistance form with much smaller, non-proliferative cells[29]. Using previously generated stage-specific RNA sequencing data[29], we analyzed the expression levels of a curated list of more than 400 metabolic genes including genes encoded by mitochondria (Supplementary Data File S7). The large majority of these metabolic

genes were differentially expressed in at least one of the three life stages. Differentially expressed genes could be grouped in two larger and two smaller clusters by unsupervised clustering (Figure 5b, Extended Data Figure 7a). The larger groups contained 103 and 192 genes that were specifically up- or down-regulated, respectively, in the cystic stage in comparison to filopodial and aggregative stages. We refer to these groups as Cys high and Cys low. 25 and 29 metabolic genes formed the smaller clusters that were differentially expressed between filopodial and aggregative stages, placing the cystic stage in an intermediate position. These results suggested that the three life-stages differ in their metabolic activity, with the cystic stage having the most divergent metabolic state.

Interestingly, almost one fifth of the genes in the Cys high cluster were not related to anabolism or catabolism, rather to other metabolic processes including autophagy-related V-ATPase proton pumps (Figure 5c, Extended Data Figure 7b). Although the genes involved in catabolism had a relatively lower representation in the Cys high cluster than in the Cys low (Figure 5c), they seemed to be more directed towards efficient mitochondrial ATP generation. For instance, mitochondrial genes encoding for components of the respiratory chain were exclusively found in the Cys high (Figure 5d). In terms of anabolism, the relative proportion of anabolism-related genes in the Cys high and the Cys low gene clusters was the same (Figure 5c), but they had a different metabolic implication. NAD and NADP biosynthesis pathways were strikingly overrepresented in the Cys high cluster (Figure 5d). Intrigued by this observation, we examined the dynamic expression of an extended set of genes related with NAD^+ metabolism including effectors such as PARP1 and macroH2A. Unsupervised clustering distinguished 4 clusters related to their differential expression in the three life stages (Figure 5e). Cluster 1 and 2 contained genes with increased expression in the two proliferative stages, and decreased expression in the non-proliferative cystic stage (Figure 5e). In line with the proliferative characteristics, they included most of the genes encoding for PARP enzymes and other components of the DNA repair machinery, which might allow cells to cope with replication-associated DNA damage. Clusters 3 and 4 contained genes more prominently expressed in aggregative and cystic stages. This included genes encoding components of biosynthetic NAD pathways such as the salvage enzyme NAMPT and the *de novo* synthesis enzyme QPRT (Figure 5e).

Noteworthy, the RNA expression patterns of PARP1 and macroH2A in cluster 2 were highly similar, suggesting co-regulation at the transcriptional level (Figure 5e). On the protein level, PARP1 was readily detected in both proliferative stages by immunoblot, but its levels dropped in the cystic stage (Figure 5e, Extended Data Figure 7c). For macroH2A, we detected a doublet at the expected size, which collapsed into a single band in the cystic stage (Figure 5f, Extended Data Figure 7d). To confirm this change in ratio, we extracted the information on PARP1 and macroH2A from available shotgun mass proteomic data[31]. Although this approach is not quantitative on the absolute level, it demonstrated that the relative ratio of macroH2A and PARP1 was the highest in the cystic stage (Figure 5g), consistent with our western blot results.

Taken together, gene expression data indicated that metabolism is dynamically regulated between the three life stages of the unicellular filasterean *Capsaspora*. It is possible that the cells in cystic stage use salvaged and *de novo* synthesized NAD^+ for life-sustaining oxidative phosphorylation and ATP production primarily fuelled by a catabolic metabolism. The relative ratio of macroH2A and PARP1 was the highest in the cystic stage suggesting that the macroH2A-dependent inhibition of PARP1's nuclear NAD^+ consumption is most likely to occur in the non-proliferative stage. Together, these metabolic adaptations can explain how *Capsaspora* survives in adverse conditions.

Affinity correlates with PARP1 inhibition and respiration

The inhibition of PARP1 by macroH2A1.1 in vertebrate cells reduces nuclear consumption of NAD^+, thereby increasing NAD^+ availability in the mitochondria necessary for respiration[11] (Figure 6a). In accordance with its increased affinity towards ADP-ribose, the macrodomain of *Capsaspora* macroH2A had an increased inhibitory capacity towards PARP1 *in vitro* than the murine macroH2A1.1 macrodomain. This was abolished by mutations in the ADP-ribose-binding pocket, such as G224E and G314E (Figure 6b).

Next, we sought to determine the impact of ADP-ribose binding by the *Capsaspora* macrodomain on metabolism *in vivo*. Currently, there are no available tools that would allow for the genetic manipulation of *Capsaspora*. Therefore, we used an orthogonal approach and introduced the *Capsaspora* protein into human HepG2 cells, stably depleted of all macroH2A isoforms[32]. To avoid confounding influences caused by any differences in histone-fold or linker sequences, we fused the wild-type and mutant macrodomains of *Capsaspora* macroH2A to the histone-fold and linker region of mouse macroH2A1.1 (Figure 6c). The expression levels of the WT reached approximately half of the level of full-length mouse macroH2A1.1 and was in the range of the endogenous levels of macroH2A proteins in parental HepG2 cells (Figure 6d, Extended Data Figure 8a-c). We confirmed that the GFP-tagged chimeric and murine proteins were fully incorporated into chromatin (Extended Data Figure 8d) and in contact with PARP1 (Extended Data Figure 8e). Key metabolic genes were largely unaffected by the expression of the different macroH2A constructs (Figure 6e). Furthermore, the mitochondrial content was similar between the four cell lines (Figure 6f). However, when analysing the oxygen consumption, we found that both the basal and the maximal mitochondrial respiration and increased in the presence of mouse macroH2A1.1 (Figure 6g). Strikingly, despite its lower expression level, this was even more pronounced in the case of the chimeric protein containing the *Capsaspora* wild-type macrodomain and translated into an increased calculated ATP production (Figure 6g). The G224E mutant macrodomain was inert, further substantiating the requirement for a functional and intact ADP-ribose pocket. In sum, our results show that the increased ADP-ribose affinity of *Capsaspora* macroH2A translated into stronger PARP1 inhibitory capacity and a more pronounced impact on the mitochondrial respiration, when compared to mouse macroH2A1.1.

Taken together, the results of our study suggest that the capacity of macroH2A to bind ADP-ribose, inhibit PARP1 and dampen its nuclear NAD^+ consumption is an ancient trait that was already present in protists. During evolution, a reduction in ADP-ribose affinity, mediated by changes in the residues that close the binding pocket upon ligand binding, fine-tuned the stringency of this mechanism.

DISCUSSION

The origin of macrodomain-containing histone variants

The macrodomain is the defining feature of all macroH2A histone variants. By focusing our evolutionary analysis on the amino acid sequence of macroH2A macrodomains, we were able to delineate the evolutionary history of this atypical histone variant. MacroH2A first appeared in protists ancestral to modern animals, filastereans and breviates, with the original gene resembling that of macroH2A1.1. The presence of the macroH2A gene was retained and further diversified in vertebrates, while it was sporadically lost in some invertebrates with accelerated evolution, such as *Drosophila*. Interestingly, the loss of macroH2A in these species correlates with a reduced number of *PARP* genes[16]. Gene duplication in a common ancestor of vertebrates resulted in the appearance of macroH2A2, an ADP-ribose binding deficient isoform. Consecutively, the alternatively spliced exon encoding for macroH2A1.2 appeared in a common ancestor of jawed vertebrates, adding to a second example of NAD signalling-inert macroH2A histone variant which can be incorporated into chromatin. The presence of macroH2A-encoding genes in haptist and breviate suggest that the fusion of a macrodomain and a histone-fold occurred before the split between fungi and animals, and thus much earlier than previously thought. Due to the scarcity of the data, we cannot fully rule out the possibility that the appearance of macroH2A in the haptist or breviate was caused by independent fusion events as an example of convergent evolution.

Macrodomains are present in all forms of life. Some viral and bacterial macrodomains bind ADP-ribose[33], while the first extensively characterized ADP-ribose-binding macrodomain was archaeal[17]. Interestingly, histone proteins, including the histone variant macroH2A, have a longer half-life than average cellular proteins[34–37]. It is intriguing to speculate that the fusion of a macrodomain to a histone provided cells with an abundant and strictly nuclear element with an increased half-life for the regulation of NAD-dependent reactions. The ability to influence NAD^+-dependent reactions in a compartmentalized manner might have provided an advantage to eukaryotes, allowing them to adapt to changes in their environment by adopting different states, consistent with our analysis of *Capsaspora* life stages. In addition, early macroH2A might have had functions in ADP-ribose signalling.

Evolution reduced ADP-ribose binding affinity

A major conclusion of our study is that ADP-ribose binding is the most ancient trait of the histone variant macroH2A. The comparative analysis of *Capsaspora* and vertebrate macroH2A macrodomains provides us with a better understanding of how the ADP-ribose and NAD metabolism-related functions of macroH2A were shaped through evolution. The *Capsaspora* macrodomain bound ADP-ribose with almost 10 times higher affinity than its mouse counterpart, and consequently was a much more potent PARP1 inhibitor. We were able to map this functional difference to the

substitution of only two residues that close the binding pocket in the *Capsaspora* macrodomain, glutamine 225 and asparagine 316. The ancestral sequence reconstruction is consistent with the Q225E replacement occurring as early as in the common ancestor of cnidarians and bilaterian metazoans leading to a decreased ADP-ribose affinity, which was maintained in most animal groups. Similarly, the N316R replacement seems to have occurred early during macroH2A evolution, but was sporadically lost in several protostome groups. N316R is strongly represented among deuterostomes, although the physiological reason remains less clear. Interestingly, a similar course of evolution was recently reported for haemoglobin, where only two historical substitutions in the ancestral protein decrease oxygen affinity, while enabling tetramerization and cooperativity[38].

NAD^+ and macroH2A sustain non-proliferative life stages

The decreased stringency of macroH2A-dependent regulation was likely selected for along the vertebrate stem lineage. In vertebrates, macroH2A1.1 takes part in cross-compartmental regulation of NAD metabolism by inhibiting PARP1 activity in the nucleus. This function scales with the expression of the macroH2A1.1 isoform and was particular prominent in non-proliferative myotubes where macroH2A1.1 is expressed at slightly higher levels than PARP1[11,14]. At the present, it is unclear how the inhibitory effect of macroH2A1.1 is mediated on the molecular level. We speculate that the binding of the macrodomain to mono-ADP-ribosylated PARP1 or PARP1 modified with short-chain PARylation could interfere with conformational changes required for PARP1 activity[14].

Capsaspora has three different life stages, two proliferative stages and a non-proliferative, spore-like, cystic stage, to which it transitions in unfavourable environmental conditions[29]. We found that the ratio between macroH2A and PARP1 was the highest in the cystic stage, suggesting that the putative macroH2A-dependent compartmental regulation of NAD metabolism might mostly operate in the non-proliferative stage of *Capsaspora,* similar to what was observed in myotubes. The cystic stage was further characterized by high expression of catabolic, mitochondrial encoded genes encoding components of the respiratory chain. While anabolic pathways were overall downregulated in cystic stage, genes involved in biosynthesis of NAD and precursors were upregulated. Curiously, several reports indicate the importance of ADP-ribose and NAD during sporulation and germination of bacterial spores[39,40]. Bacteria might rely on their NAD-based redox potential for germination, more specifically the accumulation of the reduced forms may have an important role in the initiation of germination[41]. It will be interesting to test whether a similar mechanism enables *Capsaspora* to re-enter the proliferative stages of its life cycle. Taken together, our results indicate that *Capsaspora* uses a combination of both eukaryotic and bacterial mechanisms for survival in nutrient-poor environments. Our data indicates that NAD biosynthesis is channelled to life-sustaining catabolic reactions in the cystic stage. This coincides with a high macroH2A-to-PARP1 ratio, which has the potential to limit nuclear NAD^+ consumption by PARP1. The experimental proof is pending the development of genetic tools.

The need for the compartmental regulation of NAD metabolism

NAD homeostasis is vital for optimal cell function, and by extension, for organismal health[1,2]. The NAD^+ pools of independent compartments communicate and are connected through the shared NAMPT reaction of the salvage pathway, and transport of NAD^+ and its precursors[5], thus creating a requirement for communicating and regulating local needs. But why is the compartmental regulation of NAD levels essential? NAD^+-dependent enzymes differ in their K_m as much as 100-fold, from 2-1000 μM[7,42]. Thus, their activities are controlled by the local NAD^+ concentrations, which differ between cellular compartments[13,43]. This is well illustrated by the example of nuclear NAD^+ consuming enzymes PARP1, sirtuin 1 (SirT1) and PARP2, which have decreasing NAD^+ affinity, respectively. They form a regulatory loop, whereby SirT1 can inhibit PARP1, while PARP2 regulates the activity of SirT1, depending on the nuclear NAD^+ levels[7]. Furthermore, some of the nuclear NAD^+ consuming enzymes, such as SirT6, have an even higher NAD^+ affinity than PARP1. Interestingly, macroH2A1.1 has been shown to interact with ADP-ribosylated SirT7 in an ADP-ribose-binding pocket dependent manner[44]. This raises the possibility that macroH2A1.1 may be involved in a more general regulation of nuclear NAD^+ metabolism and ADP-ribose signalling.

Compartmental NAD^+ regulation in the context of evolution

The fine-tuned regulation of NAD^+ consumption and localization is particularly relevant during the cell differentiation, when the requirements for NAD^+ compartmentalization change[13]. It is conceivable that non-proliferative states require less NAD^+ for nuclear functions, such as replication-associated DNA repair, and thus benefit from prioritizing NAD usage for life-sustaining functions, such as ATP production through mitochondrial respiration. The function of macroH2A1.1 as a nuclear NAD^+ regulator has first been demonstrated in differentiated myotubes[11]. However, macroH2A1.1 is expressed in a wide array of tissues[24]. Additionally, its upregulation is also observed during the differentiation of other tissues apart from muscle, such as skin and colon[45,46]. This suggests that its function is more widespread and generally related to cell differentiation and increased cellular plasticity in animals. In this regard, it is worth noting that *Capsaspora*, one of the closest relatives of animals with a complex life cycle, shares several mechanisms of spatial cell differentiation with animals[31,47].

Furthermore, the consolidation of macroH2A with decreased ADP-ribose affinity in vertebrates coincided with the diversification of NAD biosynthesis pathways, which provided additional elements of regulation[48]. The increased complexity of higher organisms requires intricate fine-tuning of cell processes. This is often achieved by increasing the number of proteins in a regulatory network to allow for efficient sensing of subtle changes in the environment, thus enabling fast response to environmental cues. It has been suggested that changes at the periphery of metabolic networks, possibly encoded by non-essential genes, are more likely to endow the system with the high

probability of gaining beneficial changes than the changes in the rigid core of the pathway encoded by essential genes[49]. Our data suggest that the emergence of macroH2A in protists could be such a peripheral change in the network of NAD metabolism and that it has been selected for during the evolution of metazoans. However, macroH2A is not essential for multi-cellular life *per se* as several animal species have lost macroH2A and macroH2A-deficient mice are viable[50].

Here, we described the evolution of a histone variant that can act as an inhibitor of nuclear NAD^+ consumption, adding a unique mechanism for compartmental metabolic regulation. A better understanding of these regulatory mechanisms will be informative for ongoing development of the therapies targeting NAD metabolism and signalling[1,2]. Future work will have to further elucidate how the metabolic function of macroH2A integrates with its other molecular functions, such as the regulation of higher-order chromatin architecture[32], DNA repair[51] and transcription[52].

Acknowledgements:

We would like to thank David Corujo for help and advice on many occasions and the members of the MSCA Innovative Training Network 'ChroMe', the Buschbeck and Ladurner laboratories for constructive discussions throughout the development of this interdisciplinary project. For technical support, we thank the research service facilities of IJC and IGTP, the Crystallization Facility of Max Planck Institute of Biochemistry, the ICTS NMR facility from the Scientific and Technological Centres of the University of Barcelona and Biophysics Core Facility of BMC-LMU.

Author contributions:

Conceptualization: IG, SHB and MB
Data curation: IG, CRC, MML, RM
Formal analysis: CRC, GK, JB, RM, MGC
Funding acquisition: MB, AGL
Investigation: IG, SHB, VV, MGC, CRC, JMEL, MML, JG
Project administration: IG, SHB and MB
Methodology: IG, CRC, JMEL, MGC, MFV
Resources: MSC, JA, XS, IRT, ARM
Supervision: MB, AGL, GK
Validation: IG, SHB, VV, JG, MGC, AP
Visualization: IG, CRC, GK, RM
Writing – original draft: IG and MB.
Writing – review and editing: CRC, GK, MML, AGL, JMEL, IRT, XS

Funding:

IG was a fellow of the Marie Skłodowska Curie Training network 'ChroMe' H2020-MSCA-ITN-2015-675610 (awarded to MB and AGL). The project was further supported by the national grants RTI2018-094005-B-I00 and BFU2015-66559-P from FEDER / Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación (MB).

Research in the participating labs was further supported by the following grants: the Marie Skłodowska Curie Training network 'INTERCEPT-MDS' H2020-MSCA-ITN-2020-953407 (MB), MINECO-ISCIII PIE16/00011 (MB); the Deutsche José Carreras Leukämie Stiftung DJCLS 14R/2018 (MB), AGAUR 2017-SGR-305 (MB), Fundació La Marató de TV3 257/C/2019 (MB), German Research Foundation Project-ID 213249687 - SFB 1064 and Project-ID 325871075 - SFB 1309 (AGL), the Spanish Ministry of Science PID2019-110183RB-C21 (ARM), Community of Madrid P2018/BAA-4343-ALIBIRD2020-CM (ARM), Ramón Areces Foundation (ARM), National Science Foundation EF-1921402 (JMEL), 2015 International Doctoral Fellowship La Caixa-Severo Ochoa (MFV), Marie Skłodowska-Curie Individual Fellowship #747789 (MML), Juan de la Cierva-

Incorporación IJC2018-036657-I (MML), ERC-2012-CoG-616960 (IRT), MINECO BFU2017-90114-P (IRT), AGAUR 2017 SGR 324 (XS), MINECO BIO2015-70092-R) (XS) and ERC-2014-CoG-648201 (XS).
Research at the IJC is supported by the 'La Caixa' Foundation, the Fundació Internacional Josep Carreras, Celgene Spain and the CERCA Programme/Generalitat de Catalunya.

Competing interest statement.
GK is an employee of and AGL is co-founder and managing director of Eisbach Bio GmbH, a biotechnology company. The remaining authors declare no competing interests.

LEGENDS

Figure 1. Evolutionary history of macroH2A with origin in protists.
a, Schematic representation of the two genes and the alternative exon in mammals leading to three macroH2A (mH2A) proteins of which only the macrodomain of macroH2A1.1 is able to bind ADP-ribose, inhibit PARP1 and thus reduce nuclear NAD^+ consumption.
b, Schematic tree based on the estimated divergence times of selected species. The events of macroH2A origin (arrow), gene duplication and splicing isoform emergence are represented in the tree. Dotted gray lines depict events of macroH2A loss. A full protein phylogeny based on macroH2A macrodomain sequences is showed in Supplementary Data File S3.
c, Estimated rates of evolution for macroH2A proteins in vertebrates. For each macroH2A variant, the evolutionary rate of the macrodomain is depicted. Pairwise amino acid identities are represented based on the species divergence times as defined in the TimeTree database. Evolutionary rates for fast-evolving histone H2A.Bbd, slow-evolving histones H2B were included as references. MYA, million years ago.
d, Logo plots comparing the amino acid sequence encoded by exon 5 of macroH2A from vertebrate macroH2A1.1, macroH2A1.2 and macroH2A2 and invertebrate macroH2A (including protists). Residue colors are based on their biochemical properties: positively charged in blue, negatively charged in red, hydrophobic in black and polar in gray. Residues important for ADP-ribose (ADPR) binding in vertebrate macroH2A1.1 are marked by asterisk.
e, Schematic overview of the evolutionary history of macroH2A. MYA, million years ago.

Figure 2. Structure and ADP-ribose binding of the protist macroH2A macrodomain
a, Saturation transfer difference (STD)-NMR spectra indicate the presence of the interaction between purified macrodomains of *Capsaspora* macroH2A (mH2A) and mouse macroH2A1.1 (mH2A1.1) with ADP-ribose (ADPR).

b, ITC of *Capsaspora* mH2A and mouse mH2A1.1 with ADPR. The calculated K_d values are indicated. A representative graph of four independent experiments is shown.
c, Thermodynamic parameters of (**b**) shown as signature plots indicate different contributions of enthalpy (ΔH) and entropy (-TΔS) to the Gibb's free energy (ΔG). Data represent mean (n ≥ 3) ± SD.
d, Crystal structure of apo-form of *Capsaspora* mH2A macrodomain showing entrance to ADP-ribose binding pocket, and rotated by -90º around x-axis. C-terminal α-helix is marked by asterisk.
e, Crystal structure of *Capsaspora* mH2A macrodomain in complex with ADPR in same orientation as in (**d**).
f, Composite omit 2Fo-Fc electron density map (gray mesh, contoured at 1σ) of ADPR bound to the *Capsaspora* macroH2A macrodomain.
g, Zoom into the binding pocket of superimposed crystal structures of apo- (light blue) and ADPR-bound (deep-teal) forms of *Capsaspora* macroH2A macrodomains shows significant reorganization upon ligand binding. Arrows indicate the change in position from apo to bound for residues D203, Q225, N316 and F352. The side chain of Q225 is not modeled due to ambiguous electron density. One possible conformation is shown as a transparent stick.

Figure 3. Evolutionary change in the binding mode of ADP-ribose
a, Multiple sequence alignment of >300 macroH2A1.1 (mH2A1.1)-like macrodomains indicates two highly conserved regions which highly overlap with ADP-ribose (ADPR)-binding residues. Amino acids encoded by exon 5 and residues establishing bonds with ADPR in *Capsaspora* macroH2A (mH2A) macrodomain (Extended Data Figure 2c and 3b) or human mH2A1.1 are indicated. Arrows point to the two key less-conserved residues involved in ADPR binding, namely Q225 and N316 in *Capsaspora* macroH2A, which correspond to residues E225 and R315 in human mH2A1.1.
b and **c,** Consurf analysis of structural conservation of mH2A1.1-like macrodomain mapped on the apo-form of *Capsaspora* mH2A macrodomain. Horizontal cut shows strong conservation of the interior of the binding pocket (**b**), which is closed by the less conserved amino acids Q225 and N316 (**c**).
d, Overlay of apo-forms of the macrodomains of *Capsaspora* macroH2A (dark-teal) and human macroH2A1.1 (light red, PDB: 2FXK) highlighting residues participating in the binding of ADPR. Indicated residue numbers refer to the amino acids in *Capsaspora* protein.
e, Same as in (**d**) for the ADPR-bound forms (human macroH2A1.1 PDB: 3IID). ADPR is shown in green.

Figure 4. Transition from protist-characteristic to vertebrate-characteristic key residues.
a, Simplified phylogenetic tree of macroH2A representing ancestral sequence reconstruction (ASR) with the focus on the identity of the amino acid residues corresponding to positions 225 and 316 in

Capsaspora. Complete phylogeny in Supplementary Data File S3 and with ASR in Extended Data Figure 4. Circles indicate three key nodes. The vertebrate-like or protist-like identity of residues 225 and 316 is indicated.

b, ITC of *Capsaspora* macroH2A macrodomain mutants (Q225E and N316R) with ADP-ribose (ADPR). Data of wild-type *Capsaspora* macroH2A macrodomain (WT) and the macrodomain of mouse macroH2A1.1 (mH2A1.1) from Figure 1b were included as reference. The calculated K_d values are indicated. A representative graph of four independent experiments is shown.

c, Pie chart showing that almost 25% of analyzed *Mollusca* species have mollusc-specific alternative splicing of macroH2A (AS mH2A). Scheme in the right panel shows the alternatively spliced exon 7, specifying the amino acid residues this splicing affects. Residue 316 is outlined by an arrowhead.

d, For the subset of *Mollusca* species with alternative splicing of exon 7, the identity of the amino acid corresponding to 316 is plotted as percentage of species for both alternative exons. In the majority of cases, isoform 7.1 contains the protist-characteristic residue N316 and the isoform 7.2 has exclusively the vertebrate-characteristic R316.

Figure 5. A high macroH2A to PARP1 ratio is associated with *Capsaspora's* cystic stage.

a, *Capsaspora* life cycle has three life stages: aggregative (Agg), Filopodial (Filo) and cystic (Cys). Cys stage is spore-like and non-proliferative.

b, The majority of metabolic genes are differentially expressed in the three life stages of *Capsaspora* and fall into 4 major groups: Cys high (103), Cys low (192), Filo high (25) and Agg high (29). Data represent a distribution of Z-score of expressed genes (n = 3 biologically independent samples). Box plot parameters are detailed in the Methods section under Statistical Analysis.

c, Metabolic genes were classified as anabolic, catabolic, context-dependent, other and unknown. Pie charts indicate their proportion in Cys high and Cys low clusters identified in (**b**).

d, Column charts show the relative contribution of different pathways to the anabolic (top panel) and catabolic (bottom panel) component of the group of genes in Cys high and Cys low. The total number of anabolic and catabolic genes shown in (**c**) was set to 100%.

e, Differentially expressed genes related to NAD^+ metabolism group into 4 clusters. Genes were classified in 8 groups and color coded as shown in the legend: macroH2A, other macrodomain proteins, sirtuins, Nudix proteins, PARPs, PARG-type hydrolases, DNA repair and NAD biosynthesis.

f, Immunoblot (WB) of total *Capsaspora* cell extracts from the 3 life stages using *Capsaspora* specific antibodies for macroH2A and PARP1 and histone H3 as a loading control. A representative immunoblot is shown (n = 3).

g, Changes in the relative ratio of macroH2A (mH2A) and PARP1 as determined by mass spectrometry (MS). Value in filopodial has been set to 1.

Figure 6. Differential impact of a protist and a vertebrate macroH2A macrodomain on PARP1 activity and cell metabolism.

a, Scheme showing how the inhibition of PARP1 by macroH2A1.1 (mH2A1.1) in the nucleus is connected with mitochondrial respiration through NAD^+ metabolism in vertebrates.

b, *In vitro* PARP1 auto-PARylation activity induced by nicked DNA is measured by anti-PAR immunoblotting. Naphtol blue staining shows the increasing amounts of macrodomains that were titrated into the reaction. A representative immunoblot is shown (n = 3).

c, Schematic overview of the constructs that have been introduced into macroH2A-deficient HepG2 cells (referred to as DKD cells).

d, The exogenous expression of constructs illustrated in **c** is shown by immunoblotting using anti-macroH2A1.1, anti-*Capsaspora* macroH2A, and anti-GFP. Anti-Histone H3 is included as a loading control. A representative immunoblot is shown (n = 3).

e, The expression of genes encoding key components and regulators of indicated metabolic pathways was analyzed by RT-qPCR in the cell lines described in **c** and **d**. Data represent mean (n = 3) ± SD. A two-tailed Student's t-test was used to make the indicated pairwise comparisons, * $p < 0.05$.

f, Mitochondrial content was assessed by measuring the relative amount of unique sequences in mitochondrial and genomic DNA. Mitochondrial DNA (mtDNA) content normalized to nuclear DNA (gDNA) with MT-ND2 and NDUFV1. Data represent mean (n = 3) ± SD. A two-tailed Student's t-test to analyze the presence of statistical differences.

g, The oxygen consumption of stable cell lines described in **c** and **d** was measured in routine culture and after subsequent addition of the ATPase inhibitor oligomycin (O), the uncoupling compound FCCP and the electron transport chain inhibitors rotenone and antimycin A (AA) (left panel). The bar plot shows the resulting ATP production (right panel). Data represent mean (n ≥ 4) ± SD. Ordinary one-way ANOVA was used to make the indicated comparisons between different groups of samples. * $p < 0.05$, **** $p < 0.0001$.

REFERENCES

1. Covarrubias, A. J., Perrone, R., Grozio, A. & Verdin, E. NAD+ metabolism and its roles in cellular processes during ageing. *Nat. Rev. Mol. Cell Biol.* **22**, 119–141 (2021).
2. Rajman, L., Chwalek, K. & Sinclair, D. A. Therapeutic Potential of NAD-Boosting Molecules: The In Vivo Evidence. *Cell Metab.* **27**, 529–547 (2018).
3. Xiao, W., Wang, R. S., Handy, D. E. & Loscalzo, J. NAD(H) and NADP(H) Redox Couples and Cellular Energy Metabolism. *Antioxidants Redox Signal.* **28**, 251–272 (2018).
4. Palazzo, L., Mikolčević, P., Mikoč, A. & Ahel, I. ADP-ribosylation signalling and human disease. *Open Biol.* **9**, (2019).
5. Cambronne, X. A. & Kraus, W. L. Location, Location, Location: Compartmentalization of NAD+ Synthesis and Functions in Mammalian Cells. *Trends Biochem. Sci.* **45**, 858–873 (2020).
6. Strømland, Ø. *et al.* Keeping the balance in NAD metabolism. *Biochem. Soc. Trans.* **47**, 119–130 (2019).

7. Cantó, C., Menzies, K. J. & Auwerx, J. NAD+ Metabolism and the Control of Energy Homeostasis: A Balancing Act between Mitochondria and the Nucleus. *Cell Metab.* **22**, 31–53 (2015).
8. Altmeyer, M. & Hottiger, M. O. Poly(ADP-ribose) polymerase 1 at the crossroad of metabolic stress and inflammation in aging. *Aging (Albany. NY).* **1**, 458–469 (2009).
9. Bai, P. *et al.* PARP-1 inhibition increases mitochondrial metabolism through SIRT1 activation. *Cell Metab.* **13**, 461–468 (2011).
10. Pirinen, E. *et al.* Pharmacological inhibition of poly(ADP-ribose) polymerases improves fitness and mitochondrial function in skeletal muscle. *Cell Metab.* **19**, 1034–1041 (2014).
11. Posavec-Marjanović, M. *et al.* MacroH2A1.1 regulates mitochondrial respiration by limiting nuclear NAD+ consumption. *Nat. Struct. Mol. Biol.* **24**, 902–910 (2017).
12. Luo, X. *et al.* PARP-1 Controls the Adipogenic Transcriptional Program by PARylating C/EBPβ and Modulating Its Transcriptional Activity. *Mol. Cell* **65**, 260–271 (2017).
13. Ryu, K. W. *et al.* Metabolic regulation of transcription through compartmentalized NAD+ biosynthesis. *Science (80-.).* **360**, (2018).
14. Hurtado-Bagès, S. *et al.* The Histone Variant MacroH2A1 Regulates Key Genes for Myogenic Cell Fusion in a Splice-Isoform Dependent Manner. *Cells* **9**, (2020).
15. Oláh, G. *et al.* Differentiation-associated downregulation of poly(ADP-Ribose) polymerase-1 expression in myoblasts serves to increase their resistance to oxidative stress. *PLoS One* **10**, 1–22 (2015).
16. Rack, J. G. M., Perina, D. & Ahel, I. Macrodomains: structure, function, evolution and catalytic activities. *Annu. Rev. Biochem.* **85**, 1–24 (2016).
17. Karras, G. I. *et al.* The macro domain is an ADP-ribose binding module. *EMBO J.* **24**, 1911–20 (2005).
18. Singh, H. R. *et al.* A Poly-ADP-Ribose Trigger Releases the Auto-Inhibition of a Chromatin Remodeling Oncogene. *Mol. Cell* **68**, 860-871.e7 (2017).
19. Timinszky, G. *et al.* A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation. *Nat. Struct. Mol. Biol.* **16**, 923–9 (2009).
20. Jankevicius, G. *et al.* A family of macrodomain proteins reverses cellular mono-ADP-ribosylation. *Nat. Struct. Mol. Biol.* **20**, 508–14 (2013).
21. Rosenthal, F. *et al.* Macrodomain-containing proteins are new mono-ADP-ribosylhydrolases. *Nat. Struct. Mol. Biol.* **20**, 502–507 (2013).
22. Buschbeck, M. & Hake, S. B. Variants of core histones and their roles in development , stem cells and cancer. *Nat. Publ. Gr.* (2017) doi:10.1038/nrm.2016.166.
23. Kustatscher, G., Hothorn, M., Pugieux, C., Scheffzek, K. & Ladurner, A. G. Splicing regulates NAD metabolite binding to histone macroH2A. *Nat. Struct. Mol. Biol.* **12**, 624–5 (2005).
24. Kozlowski, M. *et al.* MacroH2A histone variants limit chromatin plasticity through two distinct mechanisms. *EMBO Rep.* **19**, 1–13 (2018).
25. Chen, H. *et al.* MacroH2A1.1 and PARP-1 cooperate to regulate transcription by promoting CBP-mediated H2B acetylation. *Nat. Struct. Mol. Biol.* **21**, 981–989 (2014).
26. Ouararhni, K. *et al.* The histone variant mH2A1.1 interferes with transcription by down-regulating PARP-1 enzymatic activity. *Genes Dev.* **20**, 3324–3336 (2006).
27. Rivera-Casas, C., Gonzalez-Romero, R., Cheema, M. S., Ausió, J. & Eirín-López, J. M. The characterization of macroH2A beyond vertebrates supports an ancestral origin and conserved role for histone variants in chromatin. *Epigenetics* **11**, 415–425 (2016).
28. Torruella, G. *et al.* Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. *Curr. Biol.* **25**, 2404–2410 (2015).
29. Sebé-Pedrós, A. *et al.* Regulated aggregative multicellularity in a close unicellular relative of metazoa. *Elife* **2013**, 1–20 (2013).
30. Allen, M. D., Buckle, A. M., Cordell, S. C., Löwe, J. & Bycroft, M. The crystal structure of AF1521 a protein from Archaeoglobus fulgidus with homology to the non-histone domain of macroH2A. *J. Mol. Biol.* **330**, 503–511 (2003).
31. Sebé-Pedrós, A. *et al.* High-Throughput Proteomics Reveals the Unicellular Roots of Animal Phosphosignaling and Cell Differentiation. *Dev. Cell* **39**, 186–197 (2016).
32. Douet, J. *et al.* MacroH2A histone variants maintain nuclear organization and heterochromatin architecture. *J. Cell Sci.* **130**, 1570–1582 (2017).

33. Catara, G., Corteggio, A., Valente, C., Grimaldi, G. & Palazzo, L. Targeting ADP-ribosylation as an antimicrobial strategy. *Biochem. Pharmacol.* **167**, 13–26 (2019).
34. Chen, W., Smeekens, J. M. & Wu, R. Systematic study of the dynamics and half-lives of newly synthesized proteins in human cells. *Chem. Sci.* **7**, 1393–1400 (2016).
35. Commerford, S. L., Carsten, A. L. & Cronkite, E. P. Histone turnover within nonproliferating cells. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 1163–1165 (1982).
36. Fornasiero, E. F. *et al.* Precisely measured protein lifetimes in the mouse brain reveal differences across tissues and subcellular fractions. *Nat. Commun.* **9**, (2018).
37. Mathieson, T. *et al.* Systematic analysis of protein turnover in primary cells. *Nat. Commun.* **9**, 1–10 (2018).
38. Pillai, A. S. *et al.* Origin of complexity in haemoglobin evolution. *Nature* **581**, 480–485 (2020).
39. Huh, J. W., Shima, J. & Ochi, K. ADP-ribosylation of proteins in Bacillus subtilis and its possible importance in sporulation. *J. Bacteriol.* **178**, 4935–4941 (1996).
40. Setlow, P. & Kornberg, A. Biochemical studies of bacterial sporulation and germination. *J. Biol. Chem.* **245**, 3637–3644 (1970).
41. Setlow, R. & Setlow, P. Levels of oxidized and reduced pyridine nucleotides in dormant spores and during growth, sporulation, and spore germination of Bacillus megaterium. *J. Bacteriol.* **129**, 857–865 (1977).
42. Berger, F., Lau, C., Dahlmann, M. & Ziegler, M. Subcellular compartmentation and differential catalytic properties of the three human nicotinamide mononucleotide adenylyltransferase isoforms. *J. Biol. Chem.* **280**, 36334–36341 (2005).
43. Cambronne, X. A. *et al.* Biosensor reveals multiple sources for mitochondrial NAD+. *Science (80-.).* **352**, 1474–1477 (2016).
44. Simonet, N. G. *et al.* SirT7 auto-ADP-ribosylation regulates glucose starvation response through mH2A1. *Sci. Adv.* **6**, (2020).
45. Creppe, C. *et al.* MacroH2A1 regulates the balance between self-renewal and differentiation commitment in embryonic and adult stem cells. *Mol. Cell. Biol.* **32**, 1442–52 (2012).
46. Sporn, J. C. & Jung, B. Differential regulation and predictive potential of macroH2A1 isoforms in colon cancer. *Am. J. Pathol.* **180**, 2516–2526 (2012).
47. Sebé-Pedrós, A. *et al.* The Dynamic Regulatory Genome of Capsaspora and the Origin of Animal Multicellularity. *Cell* **165**, 1224–1237 (2016).
48. Bockwoldt, M. *et al.* Identification of evolutionary and kinetic drivers of NAD-dependent signaling. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 15957–15966 (2019).
49. Morowitz, H. J. A theory of biochemical organization, metabolic pathways, and evolution. *Complexity* **4**, 39–53 (1999).
50. Pehrson, J. R., Changolkar, L. N., Costanzi, C. & Leu, N. A. Mice without macroH2A histone variants. *Mol. Cell. Biol.* **34**, 4523–33 (2014).
51. Sebastian, R. *et al.* Epigenetic Regulation of DNA Repair Pathway Choice by MacroH2A1 Splice Variants Ensures Genome Stability. *Mol. Cell* **79**, 836-845.e7 (2020).
52. Lavigne, M. D. *et al.* Composite macroH2A/NRF-1 Nucleosomes Suppress Noise and Generate Robustness in Gene Expression. *Cell Rep.* **11**, (2015).

MATERIAL AND METHODS

Molecular data mining

MacroH2A sequences were collected from GenBank database by Blast searches using human sequences as a query. For better representation of species, especially in the transition to the vertebrate lineage, de novo assembly of transcriptomes for jawless fish (hagfish and lamprey; bioproject acc. numbers: PRJDB4902 and PRJNA292033, respectively) and the bowfin (bioproject acc. number: PRJNA292033) were carried out using Trinity software, version 2.2.0 in the Galaxy web platform[53]. Briefly, paired-end SRA Fastq files were uploaded from the European Nucleotide Archive (ENA) to the Galaxy platform and their quality analyzed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). All left and right reads were concatenated in two separate files and used as input for Trinity with default parameters. Assembled sequences were used to create local nucleotide databases and Blast searches were performed as described above.

Overall, 467 sequences encompassing 327 metazoan (58 vertebrate and 269 non-vertebrate) and 3 non-metazoan species were retrieved. Three macroH2A variants (macroH2A1.1, macroH2A1.2 and macroH2A2) were collected for all vertebrate species except for species displaying only macroH2A1.1. For all macroH2A proteins, only the globular part of the macrodomain (amino acids 182 to the end in human sequences) was used in the analyses, unless stated otherwise. Manual Blast searches were performed for curating the data of underrepresented species, specifically for filasterean and ichthyosporean species using available information[28,54–59].

Sequence alignments

Multiple sequence alignments were performed using MAFFT version 7[60] and Jalview version 2[61], and edited for potential errors in BioEdit (version 7.2). Logo plots were generated based on the aligned sequences using WebLogo3[62]. Multiple sequence alignments of the macroH2A1.1-like macrodomains sequences from 305 species with one macroH2A1.1-like isoform was generated using the alignment tool PRALINE[63]. Conservation score indicates the conservation of amino acid biochemical properties on the scale from 1 to 6 and is represented in height and the color of the bar. For pairwise sequence homology analysis, we used blast homology search and macroH2A macrodomain sequences of representative organisms as a query. The resulting percent homology rate is represented in the homology matrix visualized using Morpheus matrix analysis and visualization software (software.broadinstitute.org/morpheus/).

Phylogenetic and evolutionary analyses

If not stated otherwise, phylogenetic and molecular evolutionary analyses were conducted using MEGA X version 10.1.7[64]. MacroH2A phylogeny and ancestral sequence states were inferred by

using the Maximum Likelihood method with the LG substitution model[65], and including gamma-distributed variation among sites. Positions with less than 95% of site coverage were eliminated, and thus the analysis involved 260 amino acid sequences and a total of 181 positions in the final dataset. The reliability of the reconstructed topology was contrasted by nonparametric bootstrap (BS) method (1000 replicates).

Protein sequence divergence was estimated using uncorrected differences (p-distances, partial deletion 95%) and the rates of evolution were estimated by correlating pairwise protein divergences between pairs of taxa with their corresponding divergence times as defined by the TimeTree database[66].

Metabolism-focused analysis of high-content data from *Capsaspora*

Previously generated raw RNA sequencing data[29] was re-aligned using STAR version 2.7.3a[67]. The function genomeGenerate was used to include the mitochondrial transcriptome data in the described assembly of the *Capsaspora* genome[68] and quantified using featureCounts software ver. 2.0.1[69]. Statistical analysis was performed using DESeq2[70] using Likelihood ratio test selecting genes that show an adjusted p-value equal or less than 0.01. Clusterization of the data was obtained using DEGreport software package on BioconductoR.

Functional annotation was performed using EGGnog 5.0 using precomputed clusters and phylogenies[71]. The curated list of metabolic genes from KEGG[72] was generated by retrieving *Capsaspora* genes using EGGnog 5.0 and GHOSTKoala[73]. This was further complemented by using Orthofinder[74] to identify *Capsaspora* orthologs of the human queries. The absence of orthologues was confirmed by Blast search. The resulting list has further been manually curated and genes involved in multiple pathways have been assigned to their parent metabolic pathway and categorized as anabolic, catabolic, context-dependent (anabolic or catabolic), other and unknown.

The relative abundance of proteins was extracted from previously generated proteomics data using raw measurements and averaging three replicates per condition[31].

Plasmids

For plasmid construction, we used standard cloning techniques. *Capsaspora* macroH2A sequence (NCBI sequence ID: XM_004347479.1) was synthesized Life Technologies Inc. (Thermo Fisher) and cloned into pcDNA3.1 backbone with an N-terminal GFP-tag. Capsaspora (G224E, Q225E, N316R) and murine (G224E) macroH2A binding pocket mutants were generated via Stratagene's site-directed mutagenesis QuickChange protocol. Macrodomain sequences (corresponding to amino acids 182-368 for *Capsaspora* macroH2A and 182-369 for murine macroH2A1.1) were cloned into pETM-11 backbone with an N-terminal 6 His-tag. *Capsaspora* PARP1 (NCBI sequence ID: XP_004363957.1) N-terminal sequence (corresponding to amino acids 1-350) was amplified from *Capsaspora* cDNA and cloned into pETM-11 backbone with an N-terminal 6 His-tag.

For mammalian expression constructs, the full-length sequences of *Capsaspora* macroH2A, or mouse macroH2A1.1 or macroH2A1.2 were cloned into pLVX-Puro lentiviral backbone (Clontech) adding an N-terminal EGFP-tag. Mouse-*Capsaspora* chimeras were generated by sequential cloning of fragments. First, histone fold and linker domain sequences from mouse macroH2A1.1 were inserted into the backbone, followed by the insertion of either wildtype or mutant (G224E) *Capsaspora* macroH2A macrodomain. pLVX-Puro with EGFP alone was cloned and kindly provided by Matthew Gamble[75]. All sequences were verified by sequencing.

Protein production and purification

Rosetta (DE3) chemically competent *E. coli* were transformed with bacterial expression vectors and grown in LB medium supplemented with 34 µg/mL chloramphenicol and 50µg/mL kanamycin at 37ºC overnight. The culture was used to inoculate 1L terrific broth medium and grown at 37ºC and 200rpm until reaching an optical density at 600 nm of 0.4-0.6. The protein expression was then induced with 0.5mM isopropyl-β-d-1-thiogalactopyranoside (IPTG) for 16h at 20ºC. The next day, bacteria were pelleted by centrifugation at 10,000g for 15min at 4ºC. The bacterial pellet was lysed in 50mM Tris, 300mM NaCl, 10mM imidazole, 1mM DTT (pH 7.4) supplemented with 1mg/mL lysozyme, 10 µg/mL DNase I and protease inhibitors (Roche cOMPLETE EDTA free). The lysates were cleared by centrifugation at 30,000g for 45min at 4ºC. Subsequently, the cleared lysates were incubated with Ni-NTA beads (Quiagen) for 1 hour at 4ºC and passed over a gravity flow column 3 times. After washing the beads with 3 column volumes using 50mM Tris, 1M NaCl, 10mM imidazole, 1mM DTT (pH 7.4), the proteins of interest were eluted with 50mM Tris, 300mM NaCl, 300mM imidazole, 1mM DTT (pH 7.4). The eluted proteins were dialysed overnight into a phosphate buffer (50mM KH_2PO_4, 1mM DTT, pH 6.5), unless stated otherwise. The purified proteins were concentrated using 3kDa MWCO centrifugal concentrator (Amicon), and then flash frozen in liquid nitrogen and stored at -80ºC.

Saturation Transfer Difference Nuclear Magnetic Resonance (STD-NMR)

Saturation Transfer Difference Nuclear Magnetic Resonance (STD-NMR) experiments were performed as described elsewhere[76,77]. Briefly, 10µM His-tagged *Capsaspora* macroH2A macrodomain, or murine macroH2A1.1 wild-type or mutant macrodomains were dialysed and prepared in a deuterated water buffer (20mM NaP, 1mM TCEP, pH 7.4 and 10µM DSS). STD spectra of 1mM ADP-ribose in the presence of 10µM macroH2A macrodomains were obtained at 25ºC (298 K) on a Bruker Avance 600 MHz spectrometer equipped with a cryoprobe. A pseudo-2D version of the STD NMR sequence was used for the interleaved acquisition of on-resonance and off-resonance spectra. The on-resonance frequency was set to 0.0ppm and the saturation time was 3s. The STD effect (%) was quantified based on the following equation: $I_{STD} = 100*(I_0 - I_{SAT})/I_0$, where I_{SAT} and I_0 are the intensities of a given signal in the on-resonance and the off-resonance spectra, respectively.

Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) was performed as previously described[17]. Before the experiment, proteins were dialyzed overnight against 50mM KH_2PO_4, 1mM DTT (pH 6.5) at 4°C. The dialyzed proteins were then centrifuged for 20min at 20000 g at 4°C and the protein concentration was determined by absorbance measurements at 280nm wavelength using calculated molar extinction coefficients. The nucleotides and ribose were prepared in the same buffer in the concentration range 1-1.5mM. The concentration of ADP-ribose was additionally confirmed by absorbance measurement at 260nm, using a molar extinction coefficient of 13,500 M^{-1} cm^{-1}. Assays were conducted on the PEAQ-ITC instrument (MicroCal) at 25°C, and experimental data analysis was performed with MicroCal PEAQ-ITC Analysis Software.

Thermal shift assays

Fluorescence-based thermal shift assays were performed as previously described[18]. Briefly, 5µM protein solutions supplemented with 8 x SYPRO Orange were heated in 50mM KH_2PO_4 (pH 6.5) and 1mM DTT from 5°C to 95°C at a ramp rate of 1%. The assays were conducted in MicroAmp Fast 96-well reaction plates sealed with MicroAmp Optical Adhesive Film (Applied Biosystems). The fluorescence measurements at 554nm were normalized to the lowest value before the transition and the maximum fluorescence.

Crystallization, data collection and processing

All crystallization experiments were conducted at the Crystallization Facility of the Max-Planck-Institute for Biochemistry (Martinsried, Germany). Before setting up the crystallization, the proteins were dialyzed overnight against 20mM Bis-Tris (pH 7.0) at 4°C. The dialyzed proteins were then centrifuged for 20min at 20000 g at 4°C and the protein concentration was determined by absorbance measurements at 280nm wavelength using calculated molar extinction coefficients.

Crystals of *Capsaspora* macroH2A macrodomain in apo-form were obtained in sitting drops vapor diffusion experiments performed at 20ºC by mixing 100nL 0.1M Bis-Tris (pH 5.5) and 25% (w/v) PEG3350 with 200nL solution containing the protein at 27mg/mL. Crystals of *Capsaspora* macroH2A macrodomain in complex with ADP-ribose were obtained in sitting drops vapor diffusion experiments performed at 20ºC by mixing 100nL of 0.2M ammonium tartarate dibasic and 20% (w/v) PEG3350 with 200nL solution containing the protein at 21mg/mL and 3.8mM ADP-ribose. Crystals were cryoprotected by soaking in mother liquor supplemented with 30% ethylene glycol and flash cooled in liquid nitrogen. Diffraction data of proteins were collected on the Swiss Light Source or on the in-house X-ray source of the Crystallization Facility of the Max-Planck-Institute for Biochemistry (Martinsried, Germany). All datasets were processed using XDS[78].

The structures were solved by molecular replacement using the human macroH2A1.1 macrodomain (PDB 3IID[19]) as a search model. Model building and real space refinement were performed in COOT[79,80] and the structures refined using PHENIX REFINE[81]. Model and restraints for ADP-ribose was prepared using Phenix.Elbow[82]. A summary of the data collection and refinement statistics is shown in Supplementary Data File S4. The Ramachandran statistics for the final refined models were 97.86 % favored and 2.14 % allowed (apo, PDB: 7NY6), and 96.55 % favored and 3.45 % allowed (ADPr bound, PDB: 7NY7). UCSF Chimera software[83] and the PyMOL Molecular Graphics System (Schrödinger, LLC) have been used for visualization.

The LigPlot diagram for the crystal structure of *Capsaspora* macroH2A macrodomain in complex with ADP-ribose was generated using an online platform LIGPLOT v.4.5.3[84]. Consurf bioinformatic tool[85] was used for the projection of evolutionary conservation scores. Briefly, the conservation analysis of amino acid positions was calculated based on the phylogenetic relations between sequences of 305 species with 1 macroH2A1.1-like isoform and using *Capsaspora* apo macroH2A macrodomain structure as a query. UCSF Chimera software has been used for visualization[83].

Antibodies

We generated specific antibody against *Capsaspora* proteins by immunizing rabbits with purified His-tagged *Capsaspora* macroH2A macrodomain or carrier protein-coupled peptides of Capsaspora PARP1. Specifically, we have used a mix of 3 different peptides corresponding to amino acids 103-114, 132-142 and 299-310 of *Capsaspora* PARP1 protein. Sera were collected from terminal bleeds after three to four rounds of inoculation. The obtained antibody sera were used at 1:150 dilutions. The animal procedures were carried out by the CID-CSIC Antibody Generation Service (Spain) and UVic-animal care facility (Canada).

The following additional antibodies were used (if not stated otherwise a dilution of 1:1000 was used for Western blotting): anti-histone H3 (RRID:AB_302613, 1:10000); anti-PAR (RRID:AB_2272987); anti-GFP (RRID:AB_1196614, 1:5000); anti-macroH2A1.1[86] (1:500); anti-macroH2A1.2 (RRID:AB_1950388); anti-His-tag (RRID:AB_2744546, 1:2000); anti-SAF-A[87] (1:500); anti-PARP-1 (RRID:AB_659884); and fluorophore-conjugated secondary anti-mouse and anti-rabbit (RRID:AB_621842 and RRID:AB_621843 at 1:20000).

PARP1 activity assay

In vitro PARP-1 activity was measured using auto-PARylation assay as previously described[11]. Briefly, 0.3x activated DNA (diluted from 10x activated DNA, Trevigen) and 200 μM NAD+ were added to the buffer containing 50 mM Tris-HCl (pH 8), 50 mM NaCl and 1 mM MgCl2. Subsequently, purified macrodomains were added in the reaction mix in the defined concentrations (10 and 50μM for mouse, and 10, 25 and 50μM for *Capsaspora* macroH2A macrodomains). Finally, 0.2 units/μL PARP-1 HSA enzyme (Trevigen) was added to all reactions and the reaction mix was incubated at

25ºC for 20min. Reactions were stopped by the addition of Laemmli's sample buffer and boiling on 95ºC, separated on SDS-PAGE and analyzed by immunoblotting.

Culture of human and *Capsaspora* cells

Unless stated otherwise, the following conditions were used for cell culture. MacroH2A-depleted HepG2 (DKD) cells[32] and HEK293T (ATCC, CRL-3216) were routinely cultured in Dulbecco's modified Eagle medium (DMEM) containing 4.5 g/L glucose (Gibco) supplemented with 10% v/v fetal bovine serum (FBS) (Gibco), 2 mM L-glutamine (Gibco), 50 U/ml penicillin (Gibco) and 50 mg/ml streptomycin (Gibco). Cells were authenticated, incubated at 37°C in 5% CO2 and periodically checked for mycoplasma contamination. Cells were collected by scraping, washed with PBS and pelleted. *Capsaspora* was cultured, and cells from all three different stages were harvested as described[29]. Before analysis, *Capsaspora* cells were washed with PBS and flash-frozen in liquid nitrogen.

Gene transduction and establishment of stable cell lines

HEK293T cells were used as packaging cells to produce viral particles for lentiviral infections. Four million HEK293T cells were seeded in P10 plates and cultured to 60-70% confluency. At that point the cells were transfected with 10 µg of the lentiviral plasmid of interest and 3 µg of the pCMV-VSV-G plasmid and pCMV-dR8.91. Plasmid DNA was mixed in 1x HBS solution (2x HBS: 272 mM NaCl, 2.8 mM Na_2HPO_4, 55mM HEPES, pH 7) containing 125mM $CaCl_2$ in a total volume of 800 µL. The mix was added onto the cell culture media dropwise and left overnight. Transfection efficiency was controlled using a GFP expression vector. The supernatant containing viral particles produced by HEK293T cells was collected for 24 hours at 48 hours and 72 hours after transfection, filtered using a 0.45µm filter and supplemented with 8µg/mL of polybrene (Sigma-Aldrich). The fresh viral supernatant was added to target cells at 60-70% confluency in six-well plates that were centrifuged 45 min at 1200rpm at 37ºC, incubated at 37ºC for 45min and then cultured overnight in fresh media. The same process was repeated 24h after the first infection. The transduced cells were selected with 2µg/ml puromycin. The necessary selection time was determined by using a negative control plasmid without resistance. This procedure was used to generate all HepG2 stable cell lines. The efficiency of the cell infection was validated by live cell fluorescence and flow cytometry.

Cell fractionation, immunoprecipitation and immunoblotting

For immunoblotting of total cell material, cell pellets were directly resuspended with Laemmli's sample buffer and sonicated using Bioruptor Plus (Diagenode) and incubated at 95ºC for 10min before loading the samples on polyacrylamide gels.

For chromatin lysis, nuclei from mammalian cells were isolated and sonicated to solubilize chromatin, as previously described[11]. Briefly, cells were collected by scraping and lysed in sucrose

buffer (0.32 M sucrose, 10 mM Tris-HCl, pH 8.0, 3 mM $CaCl_2$, 2 mM MgOAc, 0.1% Triton buffer, 1:100 PMSF and 1:200 leupeptin). Samples were passed through a syringe to facilitate membrane disruption and then further incubated for 8 min on ice. Intact nuclei were collected by centrifugation for 3 min at 1,000g and washed once more in sucrose buffer. Washed nuclei were resuspended in lysis buffer (50 mM Tris-HCl, pH 7–8, 135 mM NaCl, 0.1% Triton, 1 mM EDTA, 1 mM DTT, 1:100 PMSF and 1:200 leupeptin), and chromatin was solubilized by progressive sonication with a Bioruptor Plus (Diagenode).

For the co-immunoprecipitation of PARP1, the lysis buffer was complemented with PARG and PARP inhibitors (1μM ADP-HPD from CalBioChem and 1μM olaparib from SelleckChem, respectively). Insoluble material was removed by centrifugation, and lysates were pre-cleared with sepharose-beads. At this step, 5% of the total lysate was kept as input material, and the rest of the lysate was incubated for 3 h with anti-GFP nanobodies coupled to magnetic beads (ChromoTek) which were previously blocked with 1%BSA in lysis buffer. Precipitates were washed three times with lysis buffer containing 1% Triton X-100. For SDS-PAGE and western blotting analysis, typically 1% input and 20% immunoprecipitated material was loaded.

For cell fractionations, nuclei were prepared as described above and the supernatant was kept as the cytosolic fraction. Nuclei were then incubated with high salt buffer buffer (20 mM HEPES, pH 7.9, 410 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 25% glycerol, 0.5% NP-40) for 30 min. Ultracentrifugation at 50,000g was used to separate the chromatin (pellet) and nucleosol (supernatant).

Following the transfer of proteins to the nitrocellulose membrane (GE Healthcare), the membranes were blocked with 5% low fat milk (Nestle) and incubated with primary antibodies overnight at 4ºC. The next day, membrane were washed with TBST and incubated with fluorophore-conjugated secondary antibodies for 1 hour on 25ºC in dark. After washing with TBST, the dried membranes were scanned with Odyssey® CLx Imager and analysed using ImageStudioLite quantification software (LI-COR Biosciences).

RNA and DNA analysis

Total RNA from mammalian cells was isolated by the Maxwell RSC simplyRNA Cells Kit (Promega) using the Maxwell RSC Instrument (Promega) according to the manufacturer's instructions. 1μg of total RNA was used for cDNA synthesis using the First Strand cDNA synthesis kit (Thermo Scientific) using oligo(dT) primers and according to the manufacturer's instructions. Relative cDNA levels were quantified by RT-qPCR (LightCycler 480 II instrument, Roche). All samples were analyzed in technical triplicates. Values were normalized to two reference housekeeping genes (*RPLP0* and *GAPDH*) and are plotted relative to a reference sample set to 1. To measure mitochondrial and genomic DNA, we extracted the total DNA from all cell lines of interest. Briefly, cells were pelleted and the DNA isolation buffer (10mM Tris HCl (pH 8.5), 5mM EDTA, 0,5% SDS, 200mM NaCl,

0.1mg/ml Proteinase K) was added directly on the pellets and the samples were incubated overnight at 37ºC while shaking (Thermoblock). Proteinase K was inactivated by incubating the samples for 10min at 99ºC. Equal volume of isopropanol was added to the lysates and left incubating for 20min at 25ºC under constant shaking to precipitate the DNA. The precipitated DNA was pelleted by centrifuging at 10,000 rpm for 10min at 4ºC. Supernatant was removed and precipitate was washed with ice-cold 70% ethanol and centrifugated at 10,000 rpm for 10min at 4ºC. After removing the ethanol, the pellet was air-dried and resuspended in an appropriate volume of DNase-free water. The obtained DNA was used to perform quantitative PCR with oligos of mitochondrial (MT-TL1, MT-ND2) and genomic DNA (ACTB, NCOA3). Results were demonstrated as mitochondrial to genomic DNA ratio. The sequences of all primers used are given in the Supplementary Data File S8.

Analysis of mitochondrial oxidative phosphorylation

Mitochondrial respiration was monitored with the XFe96 Cell Bionalyzer (Seahorse Biosciences, XFe96). Optimal cell density and drug concentrations have been previously determined[11]. A standard MitoStress assay was performed. Briefly, 20000 cells were plated in an XFe-96 well-plate, and cells were kept for 6 h in DMEM 10% FBS to allow the cells to attach. Then, the medium was changed to 10mM glucose, 2mM glutamine, and 1mM pyruvate XFe DMEM (5mM Hepes, pH7.4), and cells were incubated for 1 h at 37ºC without CO_2. Three different modulators of the mitochondrial respiration were sequentially injected. After basal oxygen consumption rate determination, 1.5µM oligomycin, which inhibits ATPase, was injected to determine the amount of oxygen dedicated to ATP production by mitochondria. To determine the maximal respiration rate or spare respiratory capacity, 1.5 µM carbonyl cyanide-4-(trifluoromethoxy)phenylhydrazone (FCCP) was injected to free the gradient of H^+ from the mitochondrial intermembrane space and thus to activate maximal respiration. Finally, 0.75µM antimycin A and 0.75µM rotenone were added to completely inhibit the mitochondrial respiration.

Statistical analysis and figure editing

In all bar plots, the height of the bar corresponds to the mean value and the bars indicate the standard deviation. In all box plots, the box signifies the upper (75th) and lower quartiles (25th), the median is represented by a horizontal line within the box and the mean is represented by a rhombus within the box. The upper whisker extends from the upper hinge to the largest value no further than 1.5 * IQR from the hinge (where IQR is the inter-quartile range, or distance between the first and third quartiles). The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge. The statistical test and comparison used to calculate p-values as well as p-values set as significance level are reported in each figure and/or figure legend. If not indicated otherwise, a two-tailed Student's t-test was used to assess statistical significance. The number of technical replicates

or independent cell culture experiments is indicated in the relevant figure legend(s). Figures were edited with Inkscape (inkscape.org).

DATA AVAILABILITY

The reported protein structures are deposited in the Protein Data Bank with PDB accession codes with accession codes 7NY6 (unliganded *Capsaspora* macroH2A macrodomain) and 7NY7 (ADP-ribose bound *Capsaspora* macroH2A macrodomain). Source data sets are provided with this paper.

CODE AVAILABILITY

We have exclusively used publicly available packages for bioinformatic analysis and provide their references in the Methods section. If not stated otherwise, we have used default parameters. Specific scripts are available on request.

REFERENCES:

53. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* **44**, W3–W10 (2016).
54. Brown, M. W. *et al.* Phylogenomics Places Orphan Protistan Lineages in a Novel Eukaryotic Super-Group. *Genome Biol. Evol.* **10**, 427–433 (2018).
55. de Mendoza, A., Suga, H., Permanyer, J., Irimia, M. & Ruiz-Trillo, I. Complex transcriptional regulation and independent evolution of fungal-like traits in a relative of animals. *Elife* **4**, (2015).
56. Dudin, O. *et al.* A unicellular relative of animals generates a layer of polarized cells by actomyosin-dependent cellularization. *Elife* **8**, (2019).
57. Grau-Bové, X. *et al.* Dynamics of genomic innovation in the unicellular ancestry of animals. *Elife* **6**, 1–35 (2017).
58. Hehenberger, E. *et al.* Novel Predators Reshape Holozoan Phylogeny and Reveal the Presence of a Two-Component Signaling System in the Ancestor of Animals. *Curr. Biol.* **27**, 2043-2050.e6 (2017).
59. Richter, D., Berney, C., Strassert, J., Burki, F. & de Vargas, C. EukProt: a database of genome-scale predicted proteins across the diversity of eukaryotic life. *bioRxiv* 2020.06.30.180687 (2020) doi:10.1101/2020.06.30.180687.
60. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
61. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
62. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
63. Bawono, P. & Heringa, J. PRALINE: A versatile multiple sequence alignment toolkit. *Methods Mol. Biol.* **1079**, 245–262 (2014).
64. S, K., G, S., M, L., C, K. & K, T. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
65. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307–1320 (2008).
66. Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: A public knowledge-base of divergence times among organisms. *Bioinformatics* **22**, 2971–2972 (2006).
67. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

68. Suga, H. *et al.* The Capsaspora genome reveals a complex unicellular prehistory of animals. *Nat. Commun.* **4**, 1–9 (2013).
69. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
70. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
71. Huerta-Cepas, J. *et al.* EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
72. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* vol. 28 27–30 (2000).
73. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J. Mol. Biol.* **428**, 726–731 (2016).
74. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, (2019).
75. Ruiz, P. D. & Gamble, M. J. MacroH2A1 chromatin specification requires its docking domain and acetylation of H2B lysine 20. *Nat. Commun.* **9**, (2018).
76. Mayer, M. & Meyer, B. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew. Chemie - Int. Ed.* **38**, 1784–1788 (1999).
77. Aretz, J. *et al.* Allosteric Inhibition of a Mammalian Lectin. *J. Am. Chem. Soc.* **140**, 14915–14925 (2018).
78. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).
79. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
80. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
81. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 352–367 (2012).
82. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. Electronic ligand builder and optimization workbench (eLBOW): A tool for ligand coordinate and restraint generation. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **65**, 1074–1080 (2009).
83. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
84. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. Ligplot: A program to generate schematic diagrams of protein-ligand interactions. *Protein Eng. Des. Sel.* **8**, 127–134 (1995).
85. Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).
86. Sporn, J. C. *et al.* Histone macroH2A isoforms predict the risk of lung cancer recurrence. *Oncogene* **28**, 3423–3428 (2009).
87. Fackelmayer, F., Dahm, K., Renz, A., Ramsperger, U. & Richter, A. Nucleic-acid-binding properties of hnRNP-U/SAF-A, a nuclear-matrix protein which binds DNA and RNA in vivo and in vitro. *Eur. J. Biochem.* **221**, 749–757 (1994).

Figure 1.

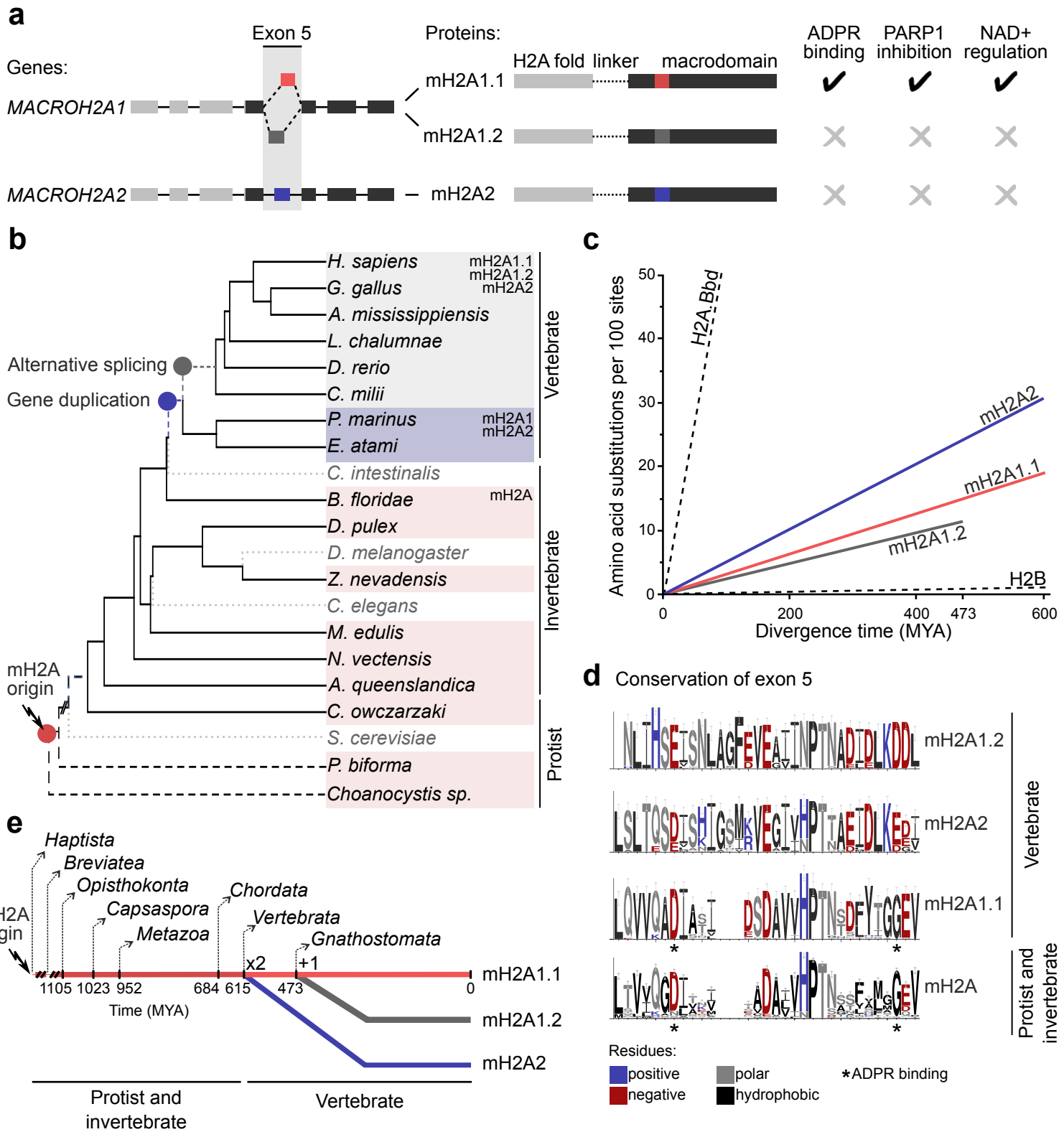
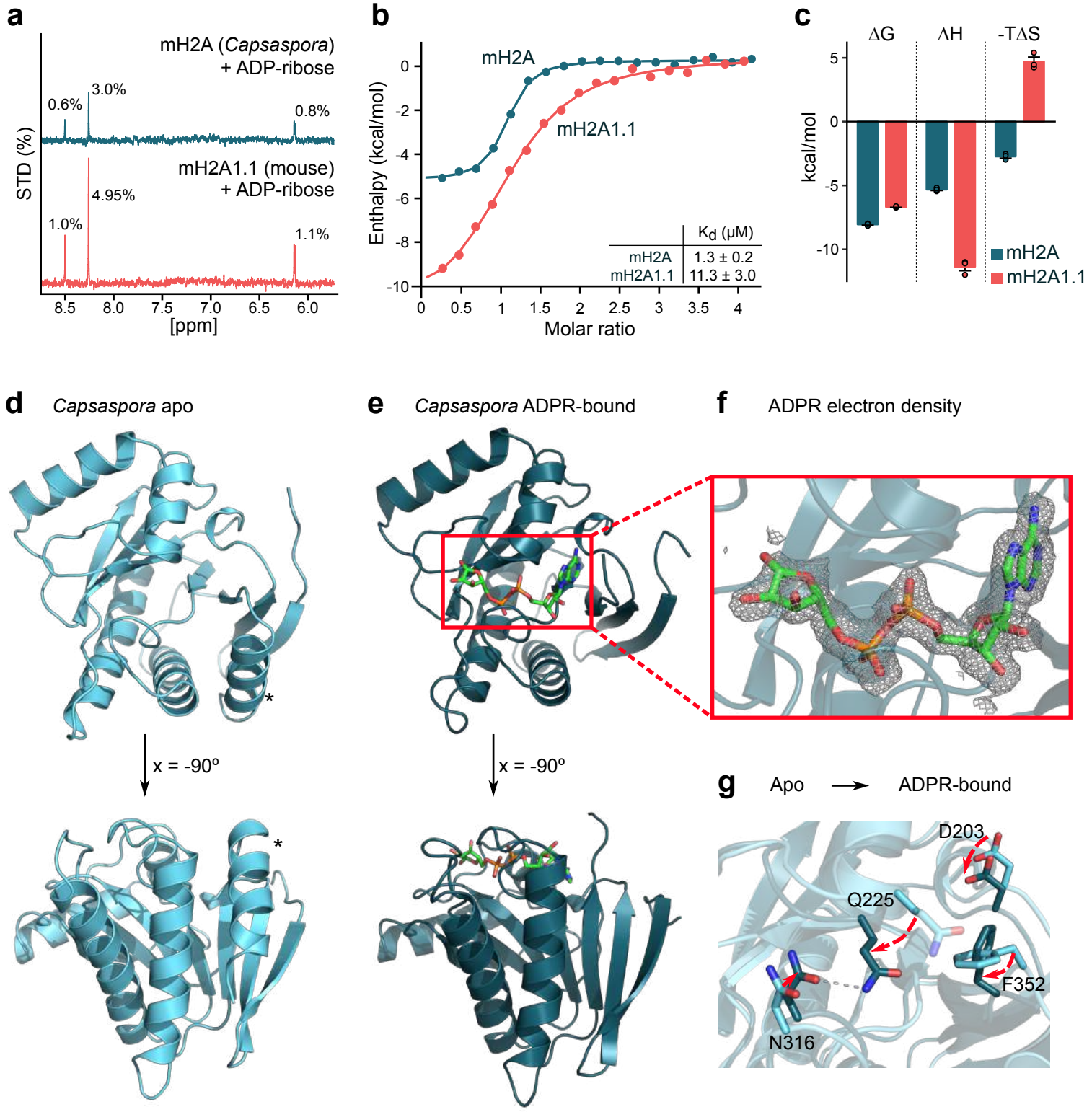

a
Genes:
MACROH2A1
MACROH2A2
Exon 5
Proteins:
H2A fold
linker
macrodomain
mH2A1.1
mH2A1.2
mH2A2
ADPR binding
PARP1 inhibition
NAD+ regulation
b
H. sapiens
G. gallus
A. mississippiensis
L. chalumnae
D. rerio
C. milii
P. marinus
E. atami
C. intestinalis
B. floridae
D. pulex
D. melanogaster
Z. nevadensis
C. elegans
M. edulis
N. vectensis
A. queenslandica
C. owczarzaki
S. cerevisiae
P. biforma
Choanocystis sp.
mH2A1.1 mH2A1.2 mH2A2
mH2A1 mH2A2
mH2A
Vertebrate
Invertebrate
Protist
Alternative splicing
Gene duplication
mH2A origin
c
Amino acid substitutions per 100 sites
H2A.Bbd
mH2A2
mH2A1.1
mH2A1.2
H2B
Divergence time (MYA)
d
Conservation of exon 5
mH2A1.2
mH2A2
mH2A1.1
mH2A
Vertebrate
Protist and invertebrate
Residues:
positive
negative
polar
hydrophobic
*ADPR binding
e
Haptista
Breviatea
Opisthokonta
Capsaspora
Metazoa
Chordata
Vertebrata
Gnathostomata
x2
+1
mH2A origin
1105 1023 952 684 615 473 0
Time (MYA)
mH2A1.1
mH2A1.2
mH2A2
Protist and invertebrate
Vertebrate

Figure 2.


a
mH2A (Capsaspora)
+ ADP-ribose
0.6%
3.0%
0.8%
mH2A1.1 (mouse)
+ ADP-ribose
1.0%
4.95%
1.1%
STD (%)
8.5 8.0 7.5 7.0 6.5 6.0
[ppm]
b
mH2A
mH2A1.1
Enthalpy (kcal/mol)
0 -2 -4 -6 -8 -10
0 0.5 1 1.5 2 2.5 3 3.5 4
Molar ratio
Kd (μM)
mH2A 1.3 ± 0.2
mH2A1.1 11.3 ± 3.0
c
ΔG
ΔH
-TΔS
kcal/mol
5 0 -5 -10
mH2A
mH2A1.1
d
Capsaspora apo
x = -90°
e
Capsaspora ADPR-bound
x = -90°
f
ADPR electron density
g
Apo → ADPR-bound
D203
Q225
F352
N316

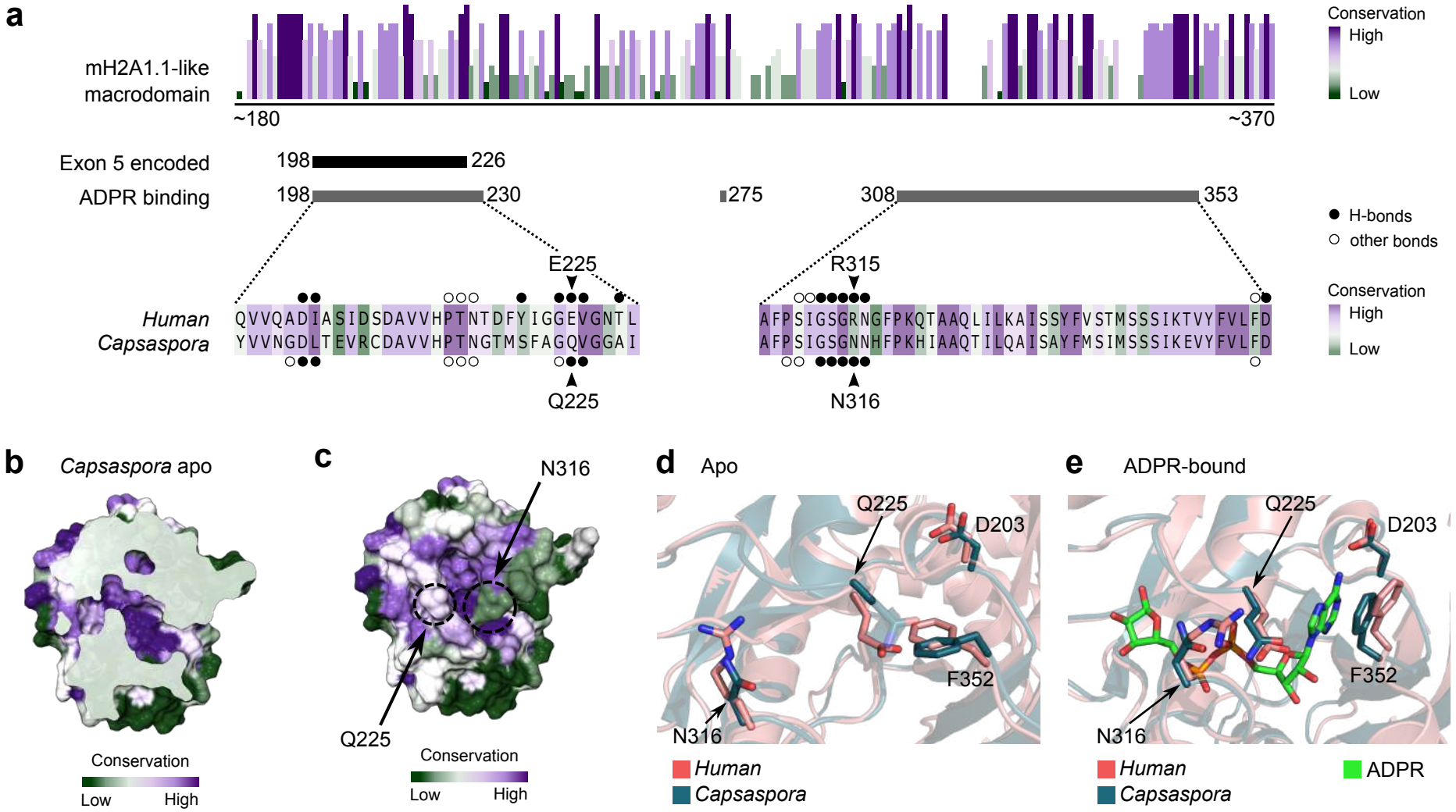
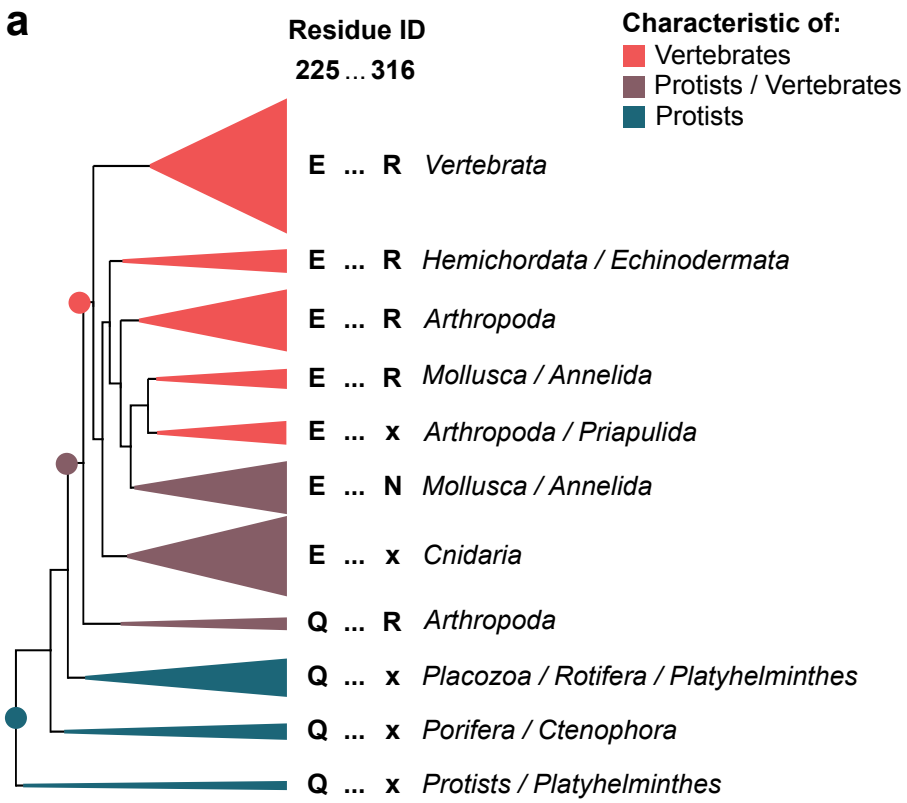
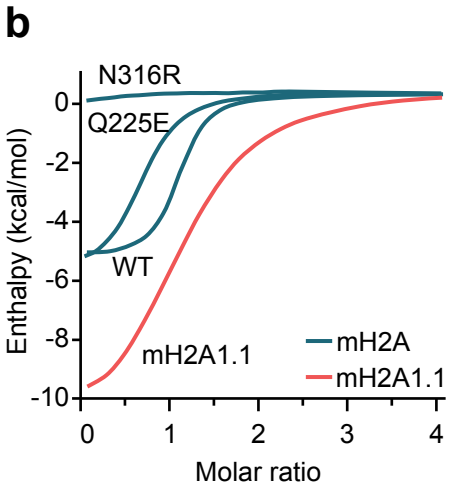
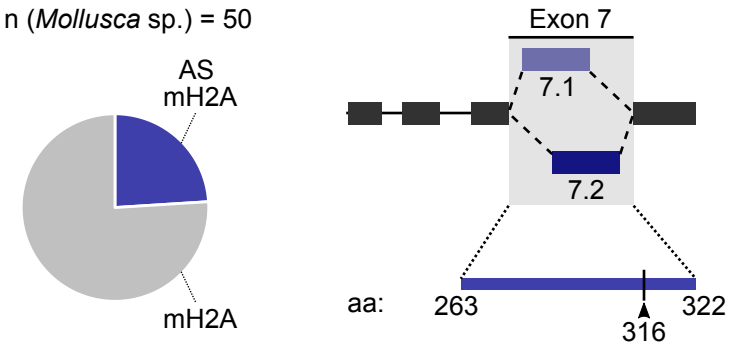

a
mH2A1.1-like
macrodomain
~180
~370
Conservation
High
Low
Exon 5 encoded
198
226
ADPR binding
198
230
275
308
353
H-bonds
other bonds
E225
R315
Human
Capsaspora
QVVQADIASIDSDAVVHPTNTDFYIGGEVGNTL
YVVNGDLTEVRCDAVVHPTNGTMSFAGQVGGAI
AFPSIGSGRNGFPKQTAAQLILKAISSYFVSTMSSSIKTVYFVLFD
AFPSIGSGNNHFPKHIAAQTILQAISAYFMSIMSSSIKEVYFVLFD
Q225
N316
b
Capsaspora apo
Conservation
Low
High
c
N316
Q225
Conservation
Low
High
d
Apo
Q225
D203
F352
N316
Human
Capsaspora
e
ADPR-bound
Q225
D203
F352
N316
Human
Capsaspora
ADPR

Figure 4.





	K_d (μM)
mH2A WT	1.2 ± 0.2
mH2A Q225E	8.6 ± 0.8
mH2A N316R	>100
mH2A1.1 WT	11.3 ± 3.0



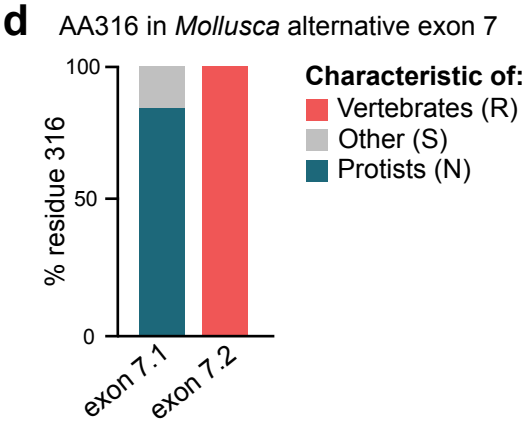


Figure 5.

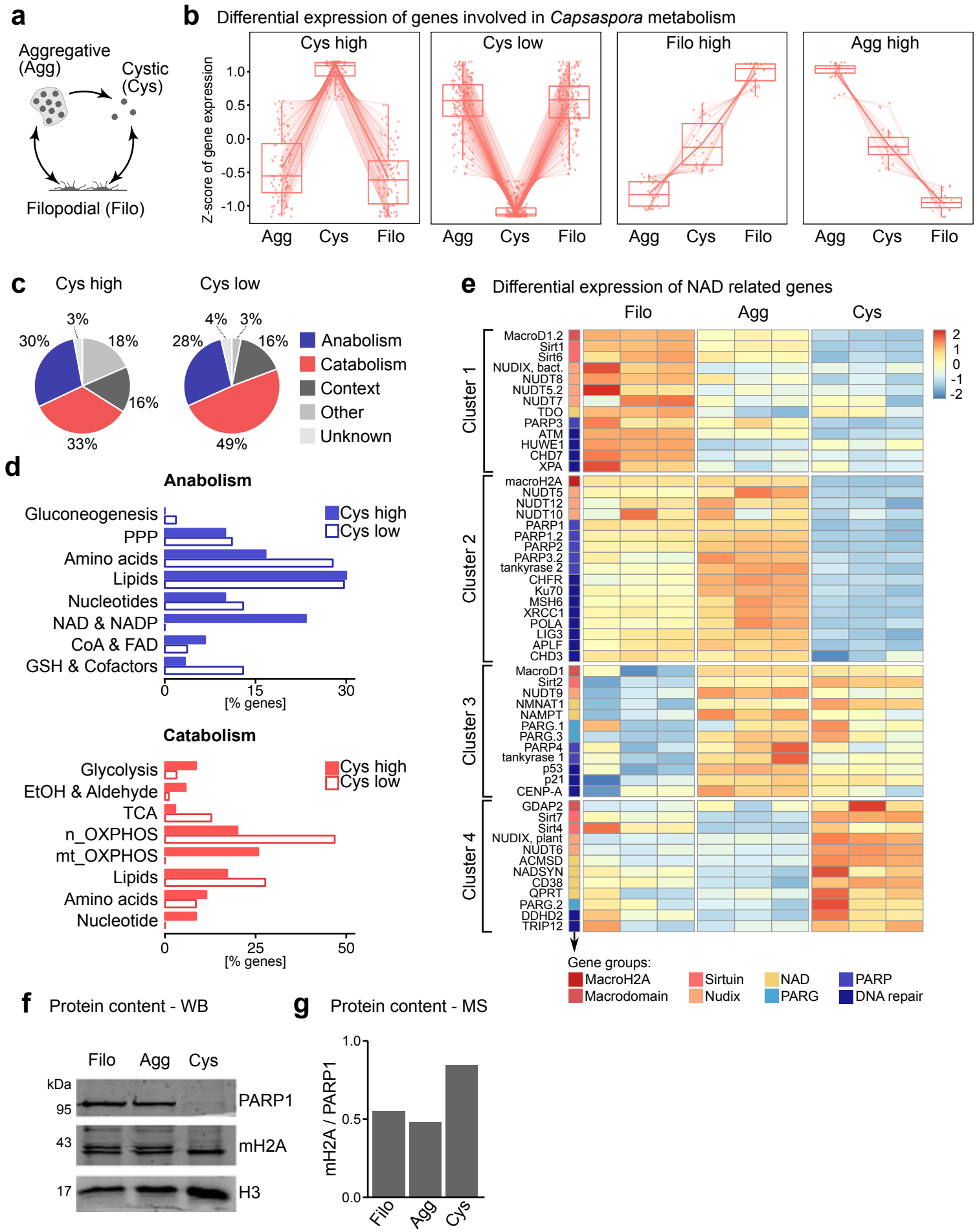
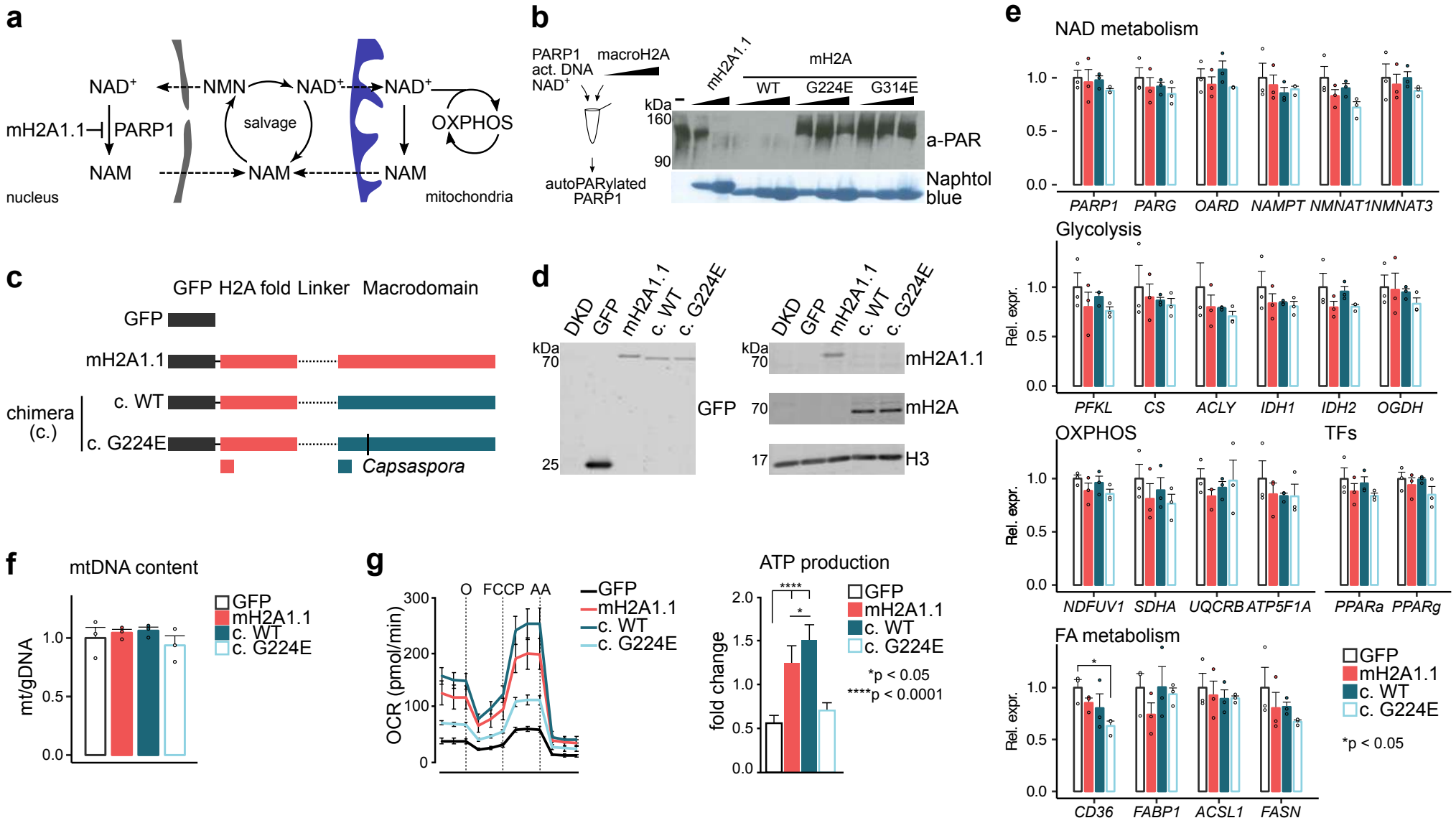

a
Aggregative (Agg)
Cystic (Cys)
Filopodial (Filo)
b
Differential expression of genes involved in Capsaspora metabolism
Cys high
Cys low
Filo high
Agg high
Z-score of gene expression
Agg
Cys
Filo
c
Cys high
Cys low
3%
30%
18%
16%
33%
4%
3%
28%
16%
49%
Anabolism
Catabolism
Context
Other
Unknown
d
Anabolism
Gluconeogenesis
PPP
Amino acids
Lipids
Nucleotides
NAD & NADP
CoA & FAD
GSH & Cofactors
Cys high
Cys low
[% genes]
Catabolism
Glycolysis
EtOH & Aldehyde
TCA
n_OXPHOS
mt_OXPHOS
Lipids
Amino acids
Nucleotide
e
Differential expression of NAD related genes
Filo
Agg
Cys
Cluster 1
MacroD1.2
Sirt1
Sirt6
NUDIX, bact.
NUDT8
NUDT5.2
NUDT7
TDO
PARP3
ATM
HUWE1
CHD7
XPA
Cluster 2
macroH2A
NUDT5
NUDT12
NUDT10
PARP1
PARP1.2
PARP2
PARP3.2
tankyrase 2
CHFR
Ku70
MSH6
XRCC1
POLA
LIG3
APLF
CHD3
Cluster 3
MacroD1
Sirt2
NUDT9
NMNAT1
NAMPT
PARG.1
PARG.3
PARP4
tankyrase 1
p53
p21
CENP-A
Cluster 4
GDAP2
Sirt7
Sirt4
NUDIX, plant
NUDT6
ACMSD
NADSYN
CD38
QPRT
PARG.2
DDHD2
TRIP12
Gene groups:
MacroH2A
Macrodomain
Sirtuin
Nudix
NAD
PARG
PARP
DNA repair
f
Protein content - WB
Filo
Agg
Cys
kDa
95
43
17
PARP1
mH2A
H3
g
Protein content - MS
mH2A / PARP1
Filo
Agg
Cys

Figure 6.


a
NAD+
NMN
NAD+
NAD+
mH2A1.1
PARP1
salvage
OXPHOS
NAM
NAM
NAM
nucleus
mitochondria
b
PARP1 act. DNA NAD+
macroH2A
mH2A1.1
mH2A
WT
G224E
G314E
kDa
160
90
a-PAR
autoPARylated PARP1
Naphtol blue
c
GFP H2A fold Linker Macrodomain
GFP
mH2A1.1
chimera (c.)
c. WT
c. G224E
Capsaspora
d
DKD
GFP
mH2A1.1
c. WT
c. G224E
kDa
70
25
GFP
mH2A1.1
mH2A
H3
e
NAD metabolism
PARP1
PARG
OARD
NAMPT
NMNAT1
NMNAT3
Glycolysis
Rel. expr.
PFKL
CS
ACLY
IDH1
IDH2
OGDH
OXPHOS
TFs
NDUFV1
SDHA
UQCRB
ATP5F1A
PPARa
PPARg
FA metabolism
CD36
FABP1
ACSL1
FASN
GFP
mH2A1.1
c. WT
c. G224E
*p < 0.05
f
mtDNA content
mt/gDNA
g
O
FCCP
AA
OCR (pmol/min)
ATP production
fold change
****
*
*p < 0.05
****p < 0.0001