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

Iva Guberovic¹, Sarah Hurtado-Bagès¹, Ciro Rivera-Casas², Gunnar Knobloch³,⁴, Roberto Malinverni¹, Vanesa Valero¹, Michelle M. Leger⁵, Jesús García⁶, Jerome Basquin⁷, Marta Gómez de Cedrón⁸, Marta Frigolé-Vivas⁹, Manjinder S. Cheema⁹, Ainhoa Pérez¹, Juan Ausió⁹, Ana Ramírez de Molina⁸, Xavier Salvatella¹,⁶,¹⁰,¹¹, Iñaki Ruiz-Trillo⁵,¹¹, Jose M. Eirin-Lopez², Andreas G. Ladurner³,⁴ and Marcus Buschbeck¹,¹²,¹³

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. In the present study, we investigated the evolution of the macrodomain-containing histone variant macroH2A1.1, an integral chromatin component that limits nuclear NAD⁺ consumption by inhibiting poly(ADP-ribose) polymerase 1 in vertebrate cells. We found that macroH2A originated in premetazoan protists. The crystal structure of the macroH2A macrodomain from the protist Capsaspora owczarzaki 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 lifecycle suggested that the metabolic function of macroH2A was associated with nonproliferative stages. Taken together, we provide insight into the evolution of a chromatin element involved in compartmental NAD regulation, relevant for understanding its metabolism and potential therapeutic applications.

¹Cancer and Leukaemia Epigenetics and Biology Program, Josep Carreras Leukaemia Research Institute, Badalona, Spain. ²Environmental Epigenetics Laboratory, Institute of Environment, Florida International University, Miami, FL, USA. ³Biomedical Center, Department of Physiological Chemistry, LMU Munich, Martinsried Planegg, Germany. ⁴Eisbach Bio GmbH, Martinsried Planegg, Germany. ⁵Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain. ⁶Institute for Research in Biomedicine, The Barcelona Institute of Science and Technology, Barcelona, Spain. ⁷Max Planck Institute of Biochemistry, Department of Structural Cell Biology, Martinsried Planegg, Germany. ⁸Molecular Oncology Group, IMDEA Food Institute, CEI UAM-+CSIC, Madrid, Spain. ⁹Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, Canada. ¹⁰Joint BSC-IRB Research Programme in Computational Biology, Barcelona, Spain. ¹¹ICREA, Barcelona, Spain. ¹²Program for Predictive and Personalized Medicine of Cancer, Germans Trias i Pujol Research Institute, Badalona, Spain. ¹³✉ e-mail: mbuschbeck@carrerasresearch.org

NATURE STRUCTURAL & MOLECULAR BIOLOGY | VOL 28 | DECEMBER 2021 | 1009–1019 | www.nature.com/nsmb
macroH2A histone variants emerged 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 (Fig. 1a). To understand when this role of macroH2A1.1 evolved, we first aimed to determine when the fusion between the histone fold and the met abolite-binding macrodomain occurred. 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 Tables 1 and 2). Previously, we reported the presence of macroH2A in a unicellular filasterean, Capsaspora spp. Together with animals and fungi, filastereans belong to the group of opisthokonts28. In the gene corresponding to human MACROH2A2 appeared after the divergence of the other phyla, and as previously reported16,27, macroH2A was absent in the most divergent macroH2A-containing species in comparison to vertebrates. On the other hand, and as previously reported17, macroH2A was absent from several nonvertebrate species, such as Drosophila melanogaster and Caenorhabditis elegans, as well as tunicates (Fig. 1b), which is indicative of lineage-specific losses. The second macroH2A gene corresponding to human MACROH2A2 appeared after the divergence of the ancestral MACROH2A1 in a common ancestor of jawed vertebrates (Fig. 1b). All three vertebrate macrodomains show a substantial level of conservation that placed them closer to the highly conserved, replication-coupled histone H2B than the fast-evolving histone variant H2A.Bbd (Fig. 1c). Comparison of the amino acid sequences corresponding to exon 5 indicated that nonvertebrate macroH2A is most similar to macroH2A1.1 (Fig. 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 Asp203 and Gly224 (refs. 17,23).

Taken together, these results allowed us to describe the evolutionary order of events that resulted in the generation of the different macroH2A histone variants present in vertebrates (Fig. 1e). Importantly, we found that macroH2A is much older than previously reported and originated in a metazoan ancestor. It was exciting to find that the first macroH2A gene is similar to vertebrate macroH2A1.1. This suggests that the function of the macroH2A histone variant in nuclear NAD+ metabolism might be ancient. The high conservation of the macrodomains of macroH2A1.2 and macroH2A2 suggests that they have acquired new, 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, one of the closest unicellular relatives of animals24. The macrodomain of C. owczarzaki interacted with ADP-ribose in a similar fashion to the murine macroH2A1.1 macrodomain, as suggested by nuclear magnetic resonance (NMR)-based binding spectra (Fig. 2a and Extended Data Fig. 1a,b). However, the ADP-ribose binding by the Capsaspora macrodomain was 8x stronger than by the macrodomain of mouse macroH2A1.1, with equilibrium dissociation constants (Kd) of 1.3 µM and 11.3 µM, respectively (Fig. 2b and Extended Data Fig. 1c–e). The thermodynamic profiles indicated that the two macrodomains bind ADP-ribose using a different binding mode. The Capsaspora macrodomain bound ADP-ribose through favorable enthalpic and entropic contributions, whereas the mouse macroH2A1.1 macrodomain bound it mainly in an enthalpy-driven manner, which was partially compensated by an unfavorable entropic contribution (Fig. 2c). Furthermore, Capsaspora macroH2A macrodomain showed high selectivity for ADP-ribose binding, because its affinity toward ADP was ~50-fold lower, and no interaction was observed with related nucleotides (ATP, AMP, GDP) or ribose (Extended Data Fig. 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 absence of ADP-ribose by protein X-ray crystallography (Fig. 2d,e and Supplementary Table 3). The unliganded macroH2A macrodomain crystallized in the space group P2₁,2 and could be refined to 1.4 Å (0.14 nm) resolution (Protein Data Bank (PDB) accession no. 7NY6), whereas the ADP-ribose-bound protein crystallized in the space group P3,21 and was refined to 2.0 Å resolution (PDB accession no. 7NY7). The obtained globular structures, with seven central β-sheets in the characteristic 1276354 order surrounded by α-helices (Extended Data Fig. 2a,b), showed high structural similarity to the previously described macrodomains24,30. Although the C-terminal α-helix was not resolved in the ADP-ribose-bound form of C. owczarzaki, its root mean square deviation was only 0.5 Å to the human ADP-ribose-bound macroH2A1.1 macrodomain (PDB accession no. 3IID). The electron density of ADP-ribose clearly revealed that the ligand sitsit in the central binding pocket macH2A macrodomain (Fig. 2f). The residues important for ligand binding included Asp203, which established a H bond with the adenine amino group, and Phe352, which stabilized the adenine ring via π-electron stacking of the aromatic rings. Furthermore, the amino group of the side chain of Asn316 established a H bond with the distal ribose of ADP-ribose (Extended Data Fig. 2c and Supplementary Table 4), which can explain the selectivity for ADP-ribose over ADP (Extended Data Fig. 1f). On ADP-ribose binding, the side chains of Glu225 and Asn316 move toward each other and establish a H bond, resulting in a conformation that encloses the central diphosphate moiety of ADP-ribose (Fig. 2g).

In summary, the capacity of the macroH2A macrodomain to bind ADP-ribose is conserved in the protist C. owczarzaki. 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 evolutionarily 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 Fig. 3a). However, the multiple sequence alignment of >300 macroH2A macrodomain sequences delineated two well-conserved regions of the protein, one located toward the N terminus, overlapping with the region encoded by exon 5, and one toward the C terminus of the macrodomain fold (Fig. 3a). It is of interest that, with the exception of Ser275, 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

---

1010 NATURE STRUCTURAL & MOLECULAR BIOLOGY | VOL 28 | DECEMBER 2021 | 1009–1019 | www.nature.com/nsmb
were invariant or structurally related, although the \textit{Capsaspora} macroH2A macrodomain established a lower number of bonds with ADP-ribose (Fig. 3a, Extended Data Fig. 3b and Supplementary Table 4). Mapping the conservation rate on the crystal structure of the apo-form of the \textit{Capsaspora} macroH2A macrodomain indicated a high conservation of the inner part of the ADP-ribose-binding pocket (Fig. 3b), whereas the surface regions were more variable (Fig. 3c, Extended Data Fig. 3c and Supplementary Video 1). The only two residues involved in ADP-ribose binding and with a low degree of conservation were Gln225 and Asn316 in the \textit{Capsaspora} macroH2A, which correspond to Glu225 and Arg315 in human macroH2A1.1. Both pairs of residues close the binding pocket on ADP-ribose binding. However, the polar uncharged side chains of Glu225 and Asn316 establish an H-bond (Fig. 2g), whereas the corresponding residues Glu225 and Arg315 in human macroH2A1.1 establish a salt bridge to close the binding pocket on ADP-ribose binding\cite{1}. Despite the slight difference in the orientation of these side chains in the \textit{Capsaspora} apo form from the human apo form (Fig. 3d), both macrodomains adopt an almost identical conformation on ADP-ribose binding (Fig. 3e).
Taken together, the low deduced conservation at the level of primary sequence contrasts with the highly conserved three-dimensional structure of C. owczarzaki and human macrodomains. The residues forming the inner ADP-ribose-binding pocket are highly conserved between C. owczarzaki and vertebrates. Two nonconserved residues corresponding to Capsaspora Gln225 and Asn316 close the binding pocket of the macroH2A macrodomain after an ADP-ribose-induced fit.

Fig. 2 | Structure and ADP-ribose binding of the protist macroH2A macrodomain. a, STD-NMR spectra indicate the presence of the interaction between purified macrodomains of Capsaspora macroH2A (mH2A) and mouse macroH2A11 (mH2A1.1) with ADP-ribose. b, The ITC of Capsaspora mH2A and mouse mH2A1.1 with ADP-ribose. 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, indicating different contributions of enthalpy ($\Delta H$) and entropy ($-T\Delta S$) to Gibb's free energy ($\Delta G$). Data represent mean ($n \geq 3$) ± s.d. d, Crystal structure of the apo-form of the Capsaspora mH2A macrodomain showing entrance to the ADP-ribose-binding pocket, and rotated by $-90^\circ$ around the x axis. A C-terminal $\alpha$-helix is marked by an asterisk. e, Crystal structure of Capsaspora mH2A macrodomain in complex with ADP-ribose in same orientation as in d, f. Composite 2Fo – Fc electron density map (gray mesh, contoured at 1σ) of ADP-ribose bound to the Capsaspora macroH2A macrodomain. g, Zoom into the binding pocket of superimposed crystal structures of apo- (light-blue) and ADP-ribose-bound (deep-teal) forms of Capsaspora macroH2A macrodomains showing significant reorganization on ligand binding. Arrows indicate the change in position from apo to bound for residues Asp203, Gln225, Asn316 and Phe352. The side chain of Gln225 is not modeled due to ambiguous electron density. One possible conformation is shown as a transparent stick.
Succeeding substitutions of Gln225 and Asn316 reduce affinity. We hypothesized that the specific substitution of amino acids Gln225 and Asn316 drove the evolution toward the decreased ADP-ribose-binding affinity from C. owczarzaki 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 (Fig. 4a, Extended Data Fig. 4 and Supplementary Table 5). The obtained results revealed that the protist-characteristic residue at position 225 (Gln225Glu) occurred early during macroH2A evolution, most likely in the ancestor of bilaterians and cnidarians, and was subsequently maintained in most animals (Fig. 4a). On the other hand, the transition to arginine at position 316 (Asn316Arg) shows more variability, as revealed by its prevalence only at specific protostome lineages, while becoming almost ubiquitous in hemichordates and chordates (Fig. 4a and Extended Data Fig. 5a–c).

To understand the physiological consequence of this evolutionary course, we individually introduced Gln225Glu and Asn316Arg mutations in the Capsaspora macroH2A1 (mH2A1) macrodomain and confirmed that the mutant proteins were folded (Extended Data Fig. 5d). It is interesting that the Gln225Glu substitution resulted in a sevenfold decrease in the binding affinity of mH2A1.1 to ADP-ribose compared with wild-type (WT) Capsaspora protein (Fig. 4b). Thus, the affinity and the thermodynamic properties of ADP-ribose binding by the Gln225Glu Capsaspora mutant were more similar to those of the mouse macroH2A1.1 macrodomain (Extended Data Fig. 5e,f). Strikingly, the Asn316Arg substitution resulted in loss of ADP-ribose binding (Fig. 4b and Extended Data Fig. 5e). It is interesting that we observed alternative splicing of exon 7 affecting position 316 in Mollusca (Fig. 4c and Extended Data Fig. 6a). The alternatively spliced exon 7 was present in >25% of analyzed species (Fig. 4c and Extended Data Fig. 6b) and strikingly determined either a protist- or a vertebrate-characteristic configuration of amino acid 316 (Fig. 4d and Extended Data Fig. 6c).

Taken together, these results shed light on the evolutionary events that affected the affinity of macroH2A1 macrodomains toward ADP-ribose. The exchange of Gln225Glu proved to be a determining factor for the decrease in ADP-ribose affinity over the course of evolution and preceded the epistatic change of Asn316Arg.

Dynamic regulation of Capsaspora metabolism. To understand how the protist macroH2A is related to metabolism, we used C. owczarzaki as a model organism. C. owczarzaki has a dynamic lifecycle composed of three stages: filopodial, aggregative and cystic (Fig. 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 cells, they are proliferative. In less advantageous environments, C. owczarzaki can transition to a spore-like cystic stage, which represents a resistance form with much smaller, nonproliferative cells. Using previously generated, stage-specific RNA-sequencing (RNA-seq) data, we analyzed the expression levels of a curated list...
of >400 metabolic genes, including genes encoded by mitochondria (Supplementary Table 6). 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 (Fig. 5b and Extended Data Fig. 7a). The larger groups contained 103 and 192 genes that were specifically up- or downregulated, respectively, in the cystic stage compared with filopodial and aggregative stages. We refer to these groups as Cys high and Cys low. Of the metabolic genes, 25 and 29 formed the smaller clusters that were differentially expressed between the filopodial and aggregative stages, placing the genes involved in catabolism had a relatively lower representation in the Cys high cluster than in the Cys low one (Fig. 5c), they seemed to be more directed toward efficient mitochondrial ATP generation. For instance, mitochondrial genes encoding components of the respiratory chain were exclusively found in Cys high (Fig. 5d). In terms of anabolism, the relative proportion of anabolism-related genes in the Cys high and Cys low gene clusters was the same (Fig. 5c), but they had a different metabolic implication. NAD+ and NADP biosynthesis pathways were strikingly overrepresented in the Cys high cluster (Fig. 5d). Intrigued by this observation, we examined the dynamic expression of an extended set of genes related to NAD+ metabolism, including effectors such as PARP1 and macroH2A. Unsupervised clustering distinguished four clusters related to their differential expression in the three life stages (Fig. 5c). Clusters 1 and 2 contained genes with increased expression in the two proliferative stages and decreased expression in the non proliferative cystic stage (Fig. 5e). In line with the proliferative characteristics, they included most of the genes encoding
Fig. 5 | A high macroH2A:PARP1 ratio is associated with C. owczarzaki's cystic stage. a, The Capsaspora lifecycle has three life stages: aggregative (Agg), filopodial (Filo) and cystic (Cys). The Cys stage is spore-like and nonproliferative. b, Most metabolic genes are differentially expressed in the three life stages of C. owczarzaki and fall into four major groups: Cys high (103), Cys low (192), Filo high (25) and Agg high (29). Data represent a distribution of z-scores of expressed genes (n = 3 biologically independent samples). Box plot parameters are detailed in 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 showing the relative contribution of different pathways to the anabolic (top) and catabolic (bottom) 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%. CoA, coenzyme A; FAD, flavin adenine dinucleotide; GSH, glutathione. e, Differentially expressed genes related to NAD+ metabolism group into four clusters. Genes were classified in eight groups and color coded as shown: macroH2A, other macrodomain proteins, sirtuins, Nudix proteins, PARPs, PARGs, DNA repair and NAD biosynthesis. f, Western blots of total Capsaspora cell extracts from the three life stages using C. owczarzaki-specific antibodies for macroH2A and PARP1, and histone H3 as a loading control. A representative western blot is shown (n = 3 independent assays). An uncropped blot image is available as Source data. g, Changes in the relative ratio of macroH2A (mH2A) and PARP1 as determined by mass spectrometry. The value in Filo has been set to 1.
Fig. 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 macroH2A11 (mH2A11) in the nucleus is connected with mitochondrial respiration through NAD+ metabolism in vertebrates. NAM, Nicotinamide; NMN, Nicotinamide mononucleotide; OXPHOS, Oxidative phosphorylation. b, In vitro PARP1 auto-PARylation activity (act.) induced by nicked DNA measured by anti-PAR western blotting. Naphthol Blue staining shows the increasing amounts of macrodomains that were titrated into the reaction. A representative western blot is shown (n = 3 independent assays). c, Schematic overview of the constructs that have been introduced into C. owczarzaki unicellular filasterean. Chimera (c.) and the de novo synthesis enzyme quinolinate phosphoribosyl transferase (NAMPT) and the de novo synthesis enzyme quinolinate phosphoribosyl transferase (QFRPT) (Fig. 5c). It is noteworthy that the RNA expression patterns of PARP1 and macroH2A in cluster 2 were highly similar, suggesting coregulation at the transcriptional level (Fig. 5e). At the protein level, PARP1 was readily detected in both proliferative stages by western blotting, but its levels dropped in the cystic stage (Fig. 5e and Extended Data Fig. 7c). For macroH2A, we detected a doublet at the expected size, which collapsed into a single band in the cystic stage (Fig. 5f and Extended Data Fig. 7d). To confirm this change in ratio, we extracted the information on PARP1 and macroH2A from available shotgun mass proteomic data. Although this approach is not quantitative at the absolute level, it demonstrated that the relative ratio of macroH2A and PARP1 was the highest in the cystic stage (Fig. 5g), consistent with our western blotting results. 

Taken together, gene expression data indicated that metabolism is dynamically regulated between the three life stages of the unicellular filasterean C. owczarzaki. It is possible that the cells in the cystic stage use salvaged and de novo synthesized NAD+ for life-sustaining oxidative phosphorylation and ATP production primarily fueled 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 nonproliferative stage.
Together, these metabolic adaptations can explain how *C. owczarzaki* 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 (Fig. 6a). In accordance with its increased affinity toward ADP-ribose, the macrodomain of *Capsaspora* macroH2A had an increased inhibitory capacity toward PARP1 in vitro than the murine macroH2A1.1 macrodomain. This was abolished by mutations in the ADP-ribose-binding pocket, such as Gly224Glu and Gly314Glu (Fig. 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 *C. owczarzaki*. Therefore, we used an orthogonal approach and introduced the *Capsaspora* protein into human HepG2 cells, stably depleted of all macroH2A isoforms\(^{46}\). To avoid confounding influences caused by any differences in histone-fold or linker sequences, we fused the WT and mutant macrodomains of *Capsaspora* macroH2A to the histone-fold and linker region of mouse macroH2A1.1, respectively (Fig. 6c). The expression levels of the WT reached approximately half the level of full-length mouse macroH2A1.1 and were in the range of the endogenous levels of macroH2A proteins in parental HepG2 cells (Fig. 6d and Extended Data Fig. 8a–c). We confirmed that the green fluorescent protein (GFP)-tagged chimeric and murine proteins were fully incorporated into chromatin (Extended Data Fig. 8d) and in contact with PARP1 (Extended Data Fig. 8e). Key metabolic genes were largely unaffected by the expression of the different macroH2A constructs (Fig. 6e). Furthermore, the mitochondrial content was similar across the four cell lines (Fig. 6f). However, when analyzing the oxygen consumption, we found that both the basal and the maximal mitochondrial respiration increased in the presence of mouse macroH2A1.1 (Fig. 6g). Strikingly, despite its lower expression level, this was even more pronounced in the case of the chimeric protein containing the *Capsaspora* WT macrodomain and translated into a significantly increased calculated ATP production (Fig. 6g). The Gly224Glu 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 mitochondrial respiration, when compared with mouse macroH2A1.1.

Taken together, the results of the present 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 on 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 breviate, with the original gene resembling that of macroH2A1.1. The presence of the macroH2A gene was retained and further diversified in vertebrates, whereas it was sporadically lost in some invertebrates with accelerated evolution, such as *Drosophila* spp. It is of interest that the loss of macroH2A in these species correlates with a reduced number of PARP genes\(^{46}\). Gene duplication in a common ancestor of vertebrates resulted in the appearance of macroH2A2, an isoform deficient in ADP-ribose binding. Consecutively, the alternatively spliced exon encoding macroH2A1.2 appeared in a common ancestor of jawed vertebrates, adding to a second example of an NAD signaling-inert macroH2A histone variant that can be incorporated into chromatin. The presence of macroH2A-encoding genes in aprotist 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 ten 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, Gln225 and Asn316. The ancestral sequence reconstruction is consistent with the Gln225Glu 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 Asn316Arg replacement seems to have occurred early during macroH2A evolution, but was sporadically lost in several protostome groups. Asn316Arg is strongly represented among deuterostomes, although the physiological reason remains less clear. It is of interest that a similar course of evolution was recently reported for hemoglobin, where only two historical substitutions in the ancestral protein decrease oxygen affinity, while enabling tetramerization and cooperativity\(^{48}\).

**NAD\(^+\) and macroH2A sustain nonproliferative life stages.** The decreased stringency of macroH2A1.1-dependent regulation was probably selected for along the vertebrate stem lineage. In vertebrates, macroH2A1.1 takes part in crosscompartmental regulation of NAD metabolism by inhibiting PARP1 activity in the nucleus. This function scales with the expression of the macroH2A1.1 isoform and was particularly prominent in nonproliferative myotubes where macroH2A1.1 is expressed at slightly higher levels than PARP1 (refs. 11,14). At the present, it is unclear how the inhibitory effect of macroH2A1.1 is mediated at 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\(^{14}\).

*C. owczarzaki* has three different life stages: two proliferative stages and a nonproliferative, spore-like, cystic stage, to which it
transitions in unfavorable environmental conditions\textsuperscript{9, 10}. 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 nonproliferative stage of 	extit{C. owczarzaki}, similar to what was observed in myotubes. The cystic stage was further characterized by high expression of catabolic, mitochondrially encoded genes encoding components of the respiratory chain. Although anabolic pathways were overall downregulated in the 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\textsuperscript{7, 8}. 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\textsuperscript{11}. It will be interesting to test whether a similar mechanism enables C. owczarzaki to re-enter the proliferative stages of its lifecycle. Taken together, our results indicate that 	extit{C. owczarzaki} uses a combination of both eukaryotic and bacterial mechanisms for survival in nutrient-poor environments. Our data indicate that NAD biosynthesis is channeled to life-sustaining catabolic reactions in the cystic stage. This coincides with a high macroH2A:PARP1 ratio, which has the potential to limit nuclear NAD\textsuperscript{+} consumption by PARP1. The experimental proof is pending the development of genetic tools.

The need for compartmental regulation of NAD metabolism. NAD homeostasis is vital for optimal cell function and, by extension, for organismal health\textsuperscript{12}. The NAD\textsuperscript{+} pools of independent compartments communicate and are connected through the shared NAMPT reaction of the salvage pathway, and transport of NAD\textsuperscript{+} and its precursors\textsuperscript{13}, and thus creating a requirement for communicating and regulating local needs. But why is the compartmental regulation of NAD levels essential? NAD\textsuperscript{+}-dependent enzymes differ in their Michaelis–Menten constant ($K_M$) as much as 100-fold, from $2\mu M$ to $1,000\mu M$\textsuperscript{14}. Thus, their activities are controlled by the local NAD\textsuperscript{+} concentrations, which differ between cellular compartments\textsuperscript{15-17}. This is well illustrated by the example of nuclear NAD\textsuperscript{+}-consuming enzymes PARP1, sirtuin 1 (SirT1) and PARP2, which have decreasing NAD\textsuperscript{+} affinity, respectively. They form a regulatory loop, whereby SirT1 cannot inhibit PARP1, whereas PARP2 regulates the activity of SirT1, depending on the nuclear NAD\textsuperscript{+} levels\textsuperscript{18}. Furthermore, some of the nuclear NAD\textsuperscript{+}-consuming enzymes, such as SirT6, have an even higher NAD\textsuperscript{+} affinity than PARP1. It is interesting that macroH2A1.1 has been shown to interact with ADP-ribosylated SirT7 in an ADP-ribose-binding pocket-dependent manner\textsuperscript{19}. This raises the possibility that macroH2A1.1 may be involved in a more general regulation of nuclear NAD\textsuperscript{+} metabolism and ADP-ribose signaling.

Compartmental NAD\textsuperscript{+} regulation in the context of evolution. The fine-tuned regulation of NAD\textsuperscript{+} consumption and localization is particularly relevant during cell differentiation, when the requirements for NAD\textsuperscript{+} compartmentalization change\textsuperscript{11}. It is conceivable that non-proliferative states require less NAD\textsuperscript{+} 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\textsuperscript{+} regulator was first demonstrated in differentiated myotubes\textsuperscript{21}. However, macroH2A1.1 is expressed in a wide array of tissues\textsuperscript{22}. In addition, its upregulation is also observed during the differentiation of other tissues apart from muscle, such as skin and colon\textsuperscript{23, 24}. 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 C. owczarzaki, one of the closest relatives of animals with a complex lifecycle, shares several mechanisms of spatial cell differentiation with animals\textsuperscript{22, 23, 24, 25}.

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\textsuperscript{26}. 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 a fast response to environmental cues. It has been suggested that changes at the periphery of metabolic networks, possibly encoded by nonessential 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\textsuperscript{26}. 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 multicellular life itself because several animal species have lost macroH2A and macroH2A-deficient mice are viable\textsuperscript{26, 27}.

In the present study, we described the evolution of a histone variant that can act as an inhibitor of nuclear NAD\textsuperscript{+} 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 signaling\textsuperscript{28}. 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\textsuperscript{29}, DNA repair\textsuperscript{30} and transcription\textsuperscript{31}.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-021-00692-5.

Received: 7 April 2021; Accepted: 28 October 2021; Published online: 9 December 2021

References

1. Covarrubias, A. J., Perrone, R., Grozio, A. & Verdin, E. NAD\textsuperscript{+} 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\textsuperscript{+} 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. Antioxid. Redox Signal 28, 251–272 (2018).
4. Palazzo, L., Mikolčević, P., Mikoč, A. & Ahel, I. ADP-ribose signalling and human disease. Open Biol. https://doi.org/10.1098/rsob.190041 (2019).
5. Cambronne, X. A. & Kraus, W. L. Location, location, location: compartmentalization of NAD\textsuperscript{+} synthesis and functions in mammalian cells. Trends Biochem. Sci. 45, 858–873 (2020).
6. Stromoland, 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\textsuperscript{+} 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 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. Pittinen, 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\textsuperscript{+} 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. Rya, K. W. et al. Metabolic regulation of transcription through compartmentalized NAD⁺ biosynthesis. Science 360, eaan5780 (2018).

14. Hurtado-Bagés, S. et al. The histone variant macroH2A1 regulates key genes for myogenic cell fusion in a spliceosome-dependent manner. Cells 9, 1109 (2020).

15. Oláh, G. et al. Differentiation-associated downregulation of poly(AD-ribose) polymerase-1 expression in myoblasts serves to increase their resistance to oxidative stress. PLoS ONE 10, e0134227 (2015).

16. Rack, J. G. M., Perina, D. & Ahel, J. Macromdomains: structure, function, evolution and catalytic activities. Annu. Rev. Biochem. 85, 431–54 (2016).

17. Karras, G. I. et al. The macro domain is an ADP-ribose binding module. EMBO J. 24, 1911–1920 (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 macromdomain-containing histone rearranges chromatin upon sensing PARP1 activation. Nat. Struct. Mol. Biol. 16, 923–929 (2009).

20. Jankevicus, G. et al. A family of macromdomain proteins reverses cellular mono-ADP-ribosylation. Nat. Struct. Mol. Biol. 20, 508–514 (2013).

21. Rosenthal, F. et al. Macromdomain-containing proteins are new mono-ADP-ribosylehydrolases. 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. https://doi.org/10.1038/nrm.2016.166 (2017).

23. Kustatscher, G., Hothorn, M., Pugiez, C., Scheffzek, K. & Ladurner, A. G. Splicing regulates NAD metabolism binding to histone macroH2A. Nat. Struct. Mol. Biol. 12, 624–625 (2005).

24. Kozłowski, M. et al. MacroH2A histone variants limit chromatin plasticity through two distinct mechanisms. EMBO Rep. 19, e44445 (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. Osararhimi, 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., Ausiò, J. & Eirín-López, J. M. The characterization of macroH2A beyond vertebrates reveals differences across tissues and subcellular fractions. EMBO J. 35, 280–290 (2016).

28. Torruella, G. et al. Phyllogenomics reveals convergent evolution of lifestyles in close relatives of animals and fungi. Curr. Biol. 25, 2404–2410 (2015).

29. Sébé-Pedrós, A. et al. Regulated aggregative multicellularity in a close unicellular relative of metazoans. eLife 2013, e01287 (2013).

30. Allen, M. D., Buckle, A. M., Cordell, S. C., Löwe, J. & Borycki, 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. Sébé-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., Corteggo, A., Valente, C., Grimaldi, G. & Palazzo, L. Targeting ADP-ribosylation as an antimicrobial strategy. Biochem. Pharmacol. 167, 13–26 (2019).

34. Chen, W., Sneekens, 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. USA 79, 1163–1165 (1982).

36. Fornasier, E. F. et al. Precisely measured protein lifetimes in the mouse brain reveal differences across tissues and subcellular fractions. Nat. Commun. https://doi.org/10.1038/s41467-018-06519-0 (2018).

37. Mathieson, T. et al. Systematic analysis of protein turnover in primary cells. Nat. Commun. 9, 689 (2018).

38. Pillar, 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, E., 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 352, 1474–1477 (2016).

44. Simonet, N. G. et al. SirT7 auto-ADP-ribosylation regulates glucose starvation response through mH2A1. Sci. Adv. https://doi.org/10.1101/719559 (2020).

45. Crepe, C. et al. MacroH2A1 regulates the balance between self-renewal and differentiation commitment in embryonic and adult stem cells. Mol. Cell. Biol. 32, 1442–1452 (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. Sébé-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. USA 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, 4523–4533 (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. https://doi.org/10.1016/j.celrep.2015.04.022 (2015).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021
Methods

Molecular data mining. MacroH2A sequences were collected from the 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 transcripts for jawless fish (hagfish and lamprey; BioProject accession nos. PRJDB4902 and PRJNA292033, respectively) and the bowfin (BioProject accession no. PRJNA292033) were carried out using Trinity software package on Bioconductor.

Capsaspora data in the described assembly of the G. owczarzaki (BioProject accession nos. PRJDB4902 and PRJNA292033, respectively) and lineage, de novo assembly of transcriptomes for jawless fish (hagfish and lamprey; BioProject accession nos. PRJDB4902 and PRJNA292033, respectively) and vertebrate lineage, de novo assembly of transcriptomes for jawless fish (hagfish and lamprey; BioProject accession nos. PRJDB4902 and PRJNA292033, respectively) and vertebrate of interest were added using the TimeTree database.

Phylogenetic and evolutionary analyses. If not stated otherwise, phylogenetic and molecular evolutionary analyses were conducted using MEGA X v.10.1.7 (ref. 72). MacroH2A phylogeny and ancestral state sequences were inferred by the maximum likelihood method with the LG substitution model, including gamma-distributed variation among sites. Positions with <95% site coverage were eliminated, so 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 a nonparametric bootstrap method (1,000 replicates).

Protein sequence divergence was estimated using uncorrected differences (P-distance), 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.

Isotermal titration calorimetry. Isotermal titration calorimetry (ITC) was performed as previously described (6). Before the experiment, proteins were dialyzed overnight against 50 mM KH2PO4 and 1 mM DTT, pH 7.4, to 280-nm wavelength using calculated molar extinction coefficients. The nucleotides and ribose were prepared in the same buffer in the concentration range 1–1.5 mM. The concentration was determined by absorbance measurements at 280-nm wavelength.

Plasmids. For plasmid construction, we used standard cloning techniques. Capsaspora macroH2A sequence (National Center for Biotechnology Information (NCBI) sequence ID: XM_004347791.1) was synthesized by Life Technologies Inc. (Thermo Fisher Scientific) and cloned into pDNA3.1 backbone with an N-terminal GFP tag. C. owczarzaki (Gly224Glu, Gly225Glu, Asn316Arg) and murine (Gly224Glu) 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 a pETM-11 backbone with an N-terminal 6 His-tag. Capsaspora PPAR1 (NCBI sequence ID: XP_004363957.1) was cloned into N- (pLVX-Puro) and C- (Clones) adding an N-terminal enhanced GFP tag. Mouse–C. owczarzaki chimeras were generated by sequential cloning of fragments. First, histo- and link- and domain sequences from mouse macroH2A1.1 were inserted into the backbone, followed by the insertion of either WT or mutant (Gly224Glu) Capsaspora macroH2A macrodomain. The pLVX-Puro with enhanced GFP alone was cloned and kindly provided by M. Gamble. All sequences were verified by sequencing.

Protein production and purification. Rosetta (DE3), chemically competent Escherichia coli were transformed with bacterial expression vectors and grown in lysogenic broth medium supplemented with 34 μg ml−1 of chloramphenicol and 50 μg ml−1 of kanamycin at 37 °C overnight. The culture was used to inoculate 11 of this Terrific Broth medium (Sigma) and grown at 37 °C and 200 r.p.m. until reaching an absorbance of 0.6–0.6. The protein expression was then induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside for 16 h at 20 °C. The next day, bacteria were pelleted by centrifugation at 10,000 g for 15 min at 4 °C. The bacterial pellet was resuspended in 50 mM NaH2PO4 and 1 mM imidazole and 1 mM DTT, pH 7.4, supplemented with 1 mg ml−1 of lysozyme, 10 μg ml−1 of DNase I and protease inhibitors (Roche cOmplete EDTA free). The lysates were cleared by centrifugation at 30,000 g for 45 min at 4 °C. Subsequently, the clarified lysates were incubated with Ni–NTA beads (QIAGEN) for 1 h at 4 °C and passed over a gravity flow column (BioRad). After washing the beads with 3 column volumes using 50 mM NaH2PO4 and 1 mM imidazole, pH 7.4, the proteins of interest were eluted with 50 mM Tris, 300 mM NaCl, 300 mM imidazole and 1 mM DTT, pH 7.4. The eluted proteins were dialyzed overnight into a phosphate buffer (50 mM KH2PO4 and 1 mM DTT, pH 6.5), unless stated otherwise. The purified proteins were concentrated using a 3-kDa molecular mass cutoff, centrifugal concentrator (Amicon), and then flash-frozen in liquid nitrogen and stored at −80 °C.

STD-NMR. Saturation transfer difference NMR (STD-NMR) experiments were performed as described elsewhere (73). Briefly, 10 mM His-tagged Capsaspora macroH2A macrodomain, or murine macroH2A1.1 WT or mutant macrodomains, was dialyzed and precipitated in a deuterated water buffer containing 16.2 mM NaH2PO4, 3.8 mM Na2HPO4, 1 mM tris(2-carboxyethyl)phosphine, pH 7.4 and 10 μM disuccinimidyl azide. N1,N2-tetramethylpropane-1,2-diamine were obtained from Sigma-Aldrich. The nucleotides and ribose were prepared in the same buffer in the concentration range 1–1.5 mM. The concentration was determined by absorbance measurements at 280-nm wavelength using calculated molar extinction coefficients. The nucleotides and ribose were prepared in the same buffer in the concentration range 1–1.5 mM. The concentration was determined by absorbance measurements at 280-nm wavelength using calculated molar extinction coefficients. The nucleotides and ribose were prepared in the same buffer in the concentration range 1–1.5 mM. The concentration was determined by absorbance measurements at 280-nm wavelength using calculated molar extinction coefficients. The nucleotides and ribose were prepared in the same buffer in the concentration range 1–1.5 mM. The concentration was determined by absorbance measurements at 280-nm wavelength using calculated molar extinction coefficients.
Crystals of *Capsaspora* macroH2A macrodomain in apo-form were obtained in sitting drop vapor diffusion experiments performed at 20°C by mixing 100 nl of 0.1 M Bis-Tris, pH 5.5 and 25% (v/v) poly(ethylene glycol) 3350 (PEG-3350) with 200 nl of a solution containing the protein at 27 mg ml\(^{-1}\). Crystals of *Capsaspora* macroH2A macrodomain in complex with ADP-ribose were obtained in sitting drop vapor diffusion experiments performed at 20°C by mixing 100 nl of 0.2 M ammonium tartarate dibasic and 20% (v/v) PEG-3350 with 200 nl of a solution containing the protein at 21 mg ml\(^{-1}\) and 3.8 M 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 the in-house X-ray source of the Crystalization Facility of the Max Planck Institute of Biochemistry. All datasets were processed using XDS\(^29\).

The structures were solved by molecular replacement using the human macroH2A1.1 macrodomain (PDB accession no. 6M8Q, ref. \(^79\)) as a search model. Model building and refinement were performed in COOT\(^80\) and the structures refined using PHENIX.\(^81\) Model and restraints for ADP-ribose were prepared using Phenix.Elbow\(^82\). A summary of the data collection and refinement statistics is shown in Supplementary Table 3. The Ramachandran statistics for the final refined models were 97.86% favored and 2.14% allowed (apo, PDB accession no. 7NY6), and 96.55% favored and 3.45% allowed (ADP-ribose bound, PDB accession no. 7NY7). UCSF Chimera software\(^83\) and the PyMOL Molecular Graphics System (Schrödinger, LLC) have been used for visualization.

**Antibodies.** We generated a 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 used a mix of three 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 1:1,000 was used for western blotting): anti-histone H3 (RRID: AB_302613, 1:10,000); anti-PAR (RRID: AB_2227987); anti-GRP (RRID: AB_1196614, 1:5,000); anti-macroH2A1.1 (ref. \(^79\), \(^80\), 1:500); anti-macroH2A1.2 (RRID: AB_1950388); anti-His-tag (RRID: AB_2744546, 1:2,000); anti-SAF-A\(^+\) (1:500); anti-PARP-1 (RRID: AB_698884), and fluorophore-conjugated secondary anti-mouse and anti-rabbit (RRID: AB_621842 and RRID: AB_621843 at 1:20,000).

**PARP1 activity assay.** In vitro PARP1 activity was measured using an auto-PARylation assay as previously described\(^5\). Briefly, 0.3X activated DNA (diluted from 10X activated DNA, Trevigen) and 200 µM NAD\(^+\) were added to the buffer containing 10 mM Tris-HCl, pH 7.5, 135 mM NaCl, 0.1% Triton, and 100 µM ADP-ribose, and incubated for 1 h at 37°C. The reaction was stopped by the addition of proteinase K (10 µg ml\(^{-1}\)). The assay was performed in triplicates. Values were normalized to two normalizing genes (*RPLP0* and *GAPDH*) and 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 (10 mM Tris-HCl, pH 8.5, 5 mM EDTA, 0.5% SDS, 200 mM NaCl and 0.1 mg ml\(^{-1}\) of proteinase K) was added directly to the pellets, and the samples were incubated overnight at 37°C while shaking (Thermoblock). Proteinase K was inactivated by incubating the samples for 10 min at 99°C. An equal volume of isopropanol was added to the lysates and left incubating for 20 min at 25°C under constant shaking to precipitate the DNA. The precipitated DNA was pelleted by centrifuging at 10,000 r.p.m. for 10 min at 4°C. To the supernatant, 1 µg ml\(^{-1}\) of RNase A was added and incubated at 37°C for 45 min before 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 qPCR with oligos of mitochondrial (MT-TL1, MT-ND2) and genomic DNA (ACTB, NCOA3). Results were demonstrated as a mitochondrial:genomic DNA ratio. The sequences of all primers used are given in Supplementary Table 7.
Analysis of mitochondrial oxidative phosphorylation. Mitochondrial respiration was monitored with the XFe-96 Cell Bioanalyzer (Seahorse Biosciences). Optimal cell density and drug concentrations had been previously determined. A standard MitoStress assay was performed. Briefly, 20,000 cells were plated in an XFe-96 well-plate, and cells were kept for 6 h in DMEM–10% fetal bovine serum to allow the cells to attach. Then, the medium was changed to 10 mM glucose, 2 mM glutamine and 1 mM pyruvate XFe DMEM (5 mM HEPES, pH 7.4), and cells were incubated for 1 h at 37 °C without CO₂. Three different modulators of mitochondrial respiration were sequentially injected. After determination of the basal oxygen consumption rate, 1.5 mM 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 mM carbonyl cyanide-4-(trifluoromethoxy)phenylhydrazone (FCCP) was injected to free the gradient of H⁺ from the mitochondrial intermembrane space, and thus activate the respiration chain. Finally, 0.75 μM antimycin A and 0.75 μM 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 s.d. In all box plots, the box signifies the upper (75th) and lower (25th) quartiles, the median is represented by a horizontal line within the box and the mean by a rhombus within the box. The upper whisker extends from the upper hinge to the largest value no further than 1.5x the interquartile range (IQR) from the hinge (that is, the IQR is the distance between the first and third quartiles). The lower whisker extends from the hinge to the smallest value, at most 1.5x the IQR of the hinge. The statistical test comparison used to calculate P values as well as P values as set as the 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 using Inkscape (inkscape.org).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The reported protein structures are deposited in the Protein Data Bank with PDB accession nos. 7NV6 (unliganded Capsaspora macroH2A2 domain) and 7NV7 (ADP-ribose bound Capsaspora macroH2A2 domain). Source data are provided with this paper.

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

References
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. B, 127–134 (1995).
85. Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate 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., Dahn, K., Renz, A., Ramspacher, U. & Richter, A. Nuclear-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).

Acknowledgements
We thank D. 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 IGT, the Crystalization Facility of the 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. I.G. was a fellow of the Marie Skłodowska Curie Training network ‘ChroMe’ (H2020-MSCA-ITN-2015-675610, awarded to M.B. and A.G.L.). The project was further supported by national grants (nos. RTI2018-094005-B-I00 and BFU2015-66559-P from FEDER/Ministerio de Ciencia e Innovación—Agencia Estatal de Investigación to M.B.). Research in the participating labs was further supported by the following grants: the Marie Skłodowska Curie ITN ”ChroMe” (H2020-MSCA-ITN-2015-675610, to M.B.), MINECO-ISCIII no. PIE16/0011 (to M.B.); the Deutsche José Carreras Project (ID 21324967—SFB 1064 and Project ID 325871075—SFB 1309 to A.G.L.), Ramón Areces Foundation (to A.R.M.), National Science Foundation (EF-1921402 to I.M.E.L.), 2015 International Doctoral Fellowship La Caixa-Severo Ochoa (to M.F.V.), Marie Skłodowska-Curie
Individual Fellowship (no. 747789 to M.M.L.), Juan de la Cierva-Incorporación (IJC2018-036657-I to M.M.L., ERC-2012-CoG-616960 to I.R.T.), MINECO (BFU2017-90114-P to I.R.T.), AGAUR (2017-SGR-324 to X.S.) and MINECO (BIO2015-70092-R and ERC-2014-CoG-648201 to X.S.). Research at the IJC is supported by the 'La Caixa' Foundation, Fundació Internacional Josep Carreras, Celgene Spain and the CERCA Programme/Generalitat de Catalunya.

Author contributions
I.G., S.H.B. and M.B. conceived the study. I.G., C.R.C., M.M.L. and R.M. curated the data. C.R.C., G.K., J.B., R.M. and M.G.C. performed the formal analysis. M.B. and A.G.L. acquired the funding. I.G., S.H.B., V.V., M.G.C., C.R.C., J.M.E.L., M.M.L. and J.G. performed the investigations. I.G., S.H.B. and M.B. administered the project. I.G., C.R.C., J.M.E.L., M.G.C. and M.F.V. provided the methodology. M.S.C., J.A., X.S., I.R.T. and A.R.M. organized the resources. M.B., A.G.L. and G.K. supervised the study. I.G., S.H.B., V.V., J.G., M.G.C. and A.P. validated the project. I.G., C.R.C., G.K. and R.M. visualized the project. I.G. and M.B. wrote the original draft of the manuscript. C.R.C., G.K., M.M.L., A.G.L., J.M.E.L., I.R.T. and X.S. wrote, reviewed and edited the manuscript.

Competing interests
G.K. is an employee and A.G.L. a cofounder and managing director of Eisbach Bio GmbH, a biotechnology company. The remaining authors declare no competing interests.

Additional information
Extended data Extended data are available for this paper at https://doi.org/10.1038/s41594-021-00692-5.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41594-021-00692-5.

Correspondence and requests for materials should be addressed to Marcus Buschbeck.

Peer review information Nature Structural and Molecular Biology thanks Matthew Gamble and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Beth Moorefield was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. Peer review reports are available.

Reprints and permissions information is available at www.nature.com/reprints.
Extended Data Fig. 1 | See next page for caption.
Extended Data Fig. 1 | The *Capsaspora* macroH2A macrodomain specifically binds ADP-ribose (accompanying Fig. 2). a, Scheme of the saturation transfer difference nuclear magnetic resonance (STD-NMR) experiment, which involves recording two analogous 1D NMR spectra: the reference spectrum of a ligand without saturation of protein signals (irradiating at a position far enough from any signal in the 'H spectrum to avoid saturation), and a spectrum recorded after selective saturation of the protein and bound molecules. The difference spectrum shows only the signals of the binders. b, STD-NMR experiments indicate the presence of interaction between *Capsaspora* macroH2A macrodomain (mH2A) and ADP-ribose (ADPR) comparable to the interaction with mouse macroH2A1.1 macrodomain (mH2A1.1) and ADPR. No interaction is observed between ADPR and the loss-of-function mutant mH2A1.1<sup>G224E</sup> that we included as a control. STD spectrum for ADPR in the absence of protein is shown as a reference. The STD effect (%) is quantified as $I_{STD}=100 \times (I_0 - I_{sat})/I_0$. c, d, e, Representative plots of raw heat evolution in isothermal titration calorimetry (ITC) experiments. ADPR was titrated into a solution with either purified *Capsaspora* mH2A (c), mouse mH2A1.1 (d) macrodomain, or into the buffer alone (e) as a control. The integration of the raw heat evolution results in a curve representing the equilibrium binding isotherm, represented below the raw heat evolution plots (c, d). f, Table showing the results of ITC binding affinity measurements ($K_d$, $N$, $\Delta H$ and $-T\Delta S$) of *Capsaspora* mH2A macrodomain and a range of nucleotides (ATP, ADP, AMP, GDP) and ribose. $K_d$ values for ATP, AMP, GDP and ribose are estimated to be higher than 100 µM, as their ITC binding curves show weak or absence of binding. If no binding occurred or binding was below ITC detection limit, reaction stoichiometries and thermodynamics could not be determined (ND).
Extended Data Fig. 2 | See next page for caption.
Extended Data Fig. 2 | The apo and ADPR-bound-structure of the Capsaspora macrodomain (accompanying Fig. 2). a, Topology diagram of apo-form of Capsaspora mH2A macrodomain. The topology diagram shows β-sheets in a conserved 1276354 order, surrounded by 6 α-helices. b, Topology diagram of Capsaspora mH2A macrodomain in complex with ADP-ribose (ADPR). The topology diagram shows β-sheets in a conserved 1276354 order, surrounded by 5 α-helices. The C-terminal α-helix observed in the apo-structure (a) was not detected in the ADPR-bound structure. This prevented the modelling of the C-terminal 13 amino acids, while the apo structure could be fully refined all the way to the C-terminus of the macrodomain. Whether this is the result of a conformational change upon ligand binding or an artefact from the crystal packing remains to be determined. However, we previously observed that ADP-ribose binding induces a conformational change in the human macroH2A1.1 macrodomain, specifically by a 30° rotation of the C-terminal α-helix away from the globular macrodomain fold (Timinszky et al., 2009, NSMB). c, The schematic representation of ADPR interaction with Capsaspora macroH2A macrodomain. All non-covalent interactions between ADPR and the macrodomain are summarized in the LigPlot diagram. ADPR ligand is represented in thick purple line and the amino acid residues of Capsaspora macroH2A macrodomain in thin orange lines. Hydrogen bonds are represented by the dashed lines between atoms involved, while the circles or semicircles with radiating lines represent atoms or residues involved in hydrophobic contacts between protein and ligand.
Extended Data Fig. 3 | See next page for caption.
Extended Data Fig. 3 | Conservation of the macrodomain and bonds established with ADP-ribose (accompanying Fig. 3). a, Pairwise sequence identity (%) of macrodomains from representative macroH2A sequences of 9 different phyla. For the vertebrates, amino acid sequences corresponding to macroH2A1.1 have been used. b, Table indicating the amino acids establishing H-bond or other bonds with ADP-ribose for *Capsaspora* macroH2A (see also Extended Data Fig. 2c) and human macroH2A1.1 (Timinszky et al., 2009). c, Snapshots of the apo structure of *Capsaspora* macroH2A macrodomain starting from the front perspective of the binding pocket and subsequently rotated by 90°, 180° and 270° around y-axis. The level of conservation is calculated by Consurf server and is projected on the protein surface. See also the movie provided as Supplementary Data File S5. Please note that in the movie conservation degree is color-coded from low (turquoise), intermediate (white) to high (magenta).
Extended Data Fig. 4 | Inferred ancestral amino acid states at sites 225 and 316 (accompanying Fig. 4). Simplified version of the maximum likelihood tree shown in Supplementary Data File S3. The tree shows a set of possible amino acid (states) at each ancestral node based on their inferred likelihood at sites 225 (in bold) and 316 (in italics). Protein numeration is based on the full-length *Capsaspora* macroH2A. Probability cutoff is 0.95 and the ambiguous states are not shown. Collapsed branches in choral represent the vertebrate-characteristic pair E225-R316 (diffused when only one was inferred), in turquoise the *Capsaspora*-characteristic pair Q225-N316 (diffused when only one was inferred), and in brown the species with E225 (vertebrate-like) and N316 (*Capsaspora*-like) residues.
Extended Data Fig. 5 | See next page for caption.
Extended Data Fig. 5 | The identity and biophysical consequences of evolution of residues 225 and 316 closing the binding pocket (accompanying Fig. 4).

a, The pie charts showing the identity of other residues at positions 225 and 316. Residue colors are based on their biochemical properties: positively charged in blue, negatively charged in red, polar in gray, and non-polar in black. 
b, The pie charts showing the identity of the residues at the position 316 which pair with either protist-characteristic Q225, or vertebrate-characteristic E225. The numbering of the residues is based on the Capsaspora mH2A. In particular, the protist-characteristic residue Q225 was paired with varying residues at 316.
c, The pie charts showing the identity of the residues at the position 225 which pair with either protist-characteristic N316, or vertebrate-characteristic R316. The numbering of the residues is based on the Capsaspora mH2A. Strikingly, R316 was paired with E225 or the biochemically similar D225 in more than 95% of the species, while its pairing with Q225 was rare.
d, Thermal shift assay profile of melting temperatures of wild type, Q225E and N316R mutants of Capsaspora macroH2A macrodomains (mH2AWT, mH2AQ225E and mH2AN316R, respectively) show that the folding of the mutant proteins is unaffected. The subtle shift in the melting temperature of Q225E mutant macrodomain to higher temperatures might be indicative of an increase in stability.
e, Representative plots of raw heat evolution in isothermal titration calorimetry (ITC) experiments. ADP-ribose was titrated into a solution with either purified Capsaspora Q225E (left panel) or N316R (right panel) mutant macroH2A macrodomain. The integration of the raw heat evolution resulted in a curve representing the equilibrium binding isotherm, represented in the inset.
f, Table showing the thermodynamics of the binding interaction calculated on the basis of ITC data shown in Fig. 4b.
a Vertebrate MACROH2A1 gene

Mollusca MACROH2A gene

b Mollusca classes with alternative exon 7

C Focused analysis of the residue identity at position 316 in Mollusca with alternative splicing

Extended Data Fig. 6 | See next page for caption.
Extended Data Fig. 6 | Alternative splicing of exon 7 in molluscs (accompanying Fig. 4). a, The organization of the human MACROH2A1 gene and the corresponding gene from the mollusc Crassostrea gigas (Gene ID: 105347164) are represented on an arbitrary scale for comparison purposes. For simplification only the alternative exon 5 defining macroH2A1.1 is represented for human MACROH2A1. Both alternative exons 7 of the Mollusca macroH2A gene are represented. b, Transcriptomic data was available for 50 Mollusca species from three of eight different classes. The number of species per class is indicated (left panel). The percentage of species per Mollusca class containing alternatively spliced macroH2A isoforms are shown (right panel). c, Alignment of C-terminal alternative exons from the mollusc species in which we detected the 2 alternative splice isoforms. Almost all species have both R316 and N316-containing isoforms.
Extended Data Fig. 7 | Metabolic dynamics of *Capsaspora* life stages (accompanying Fig. 5). a, Heatmap of 349 metabolic genes differentially expressed across the three life stages of *Capsaspora*. b, Table showing the ‘other’ genes in the Cys high cluster (Fig. 5c) grouped according to their physiological role. c, Antibody specific for *Capsaspora* PARP1. Rabbits were immunized with three different peptides corresponding to the N-terminal part of PARP1. The antibody specificity was corroborated by immunoblotting bacterial lysates: non-transformed (non-transf.), transformed with a plasmid encoding for N-terminal tagged fragment of PARP1 corresponding to amino acids 1-350 but not induced (non-ind.), and bacteria producing His-tagged PARP1 (His-N-PARP1). Representative immunoblot is shown (n = 3). d, Antibody specific for *Capsaspora* macroH2A. Rabbits were immunized with purified His-tagged macroH2A macrodomain (AA: 182-368). The specificity was corroborated by immunoblotting using stable cell lines expressing GFP-tagged constructs of mouse macroH2A1.1 (mH2A1.1) and *Capsaspora* macroH2A (mH2A) and purified His-tagged macrodomain of *Capsaspora* mH2A. Representative immunoblot is shown (n = 3).
a  GFP expression in established stable cell lines

b  Quantification of GFP expression in established stable cell lines

c  

d  Mixed soluble fraction  Chromatin

Extended Data Fig. 8 | See next page for caption.
Extended Data Fig. 8 | Characterization of stable DKD cell lines (accompanying Fig. 6). a, Fluorescence intensity of stable DKD cell lines stably expression either GFP or GFP-tagged constructs of mouse macroH2A1.1 (mH2A1.1) or chimeric wild type or G224E proteins containing the Capsaspora macrodomain as illustrated in Fig. 6c. Scale represents 50 µm. Cells were checked daily for GFP expression, and the representative images are shown. b, Flow cytometric detection of GFP in the same cells lines as A. Bar plot to the left compares mean fluorescence intensities of all four stable cell lines. Bar plot in the center shows GFP fluorescence in cell lines stably expressing macroH2A constructs. The bar plot to the right shows high percentage (≥ 99%) of GFP positive cells. c, Stable cell lines expressing the GFP-tagged mouse mH2A1.1 were compared to a similar cell line expressing GFP-tagged mouse macroH2A1.2 (mH2A1.2), control DKD cells and the parental HepG2 cell line. Immunoblotting using anti-GFP and anti-mH2A1.2 specific antibodies allows evaluating the expression level of the exogenous mH2A1.1 relative to endogenous mH2A1.2. Comparison to Extended Data Fig. 7b indicates that the chimeric proteins are expressed at a similar range or slightly higher than endogenous macroH2A proteins. Representative immunoblot is shown (n = 3). d, Immunoblot analysis after cell fractionation shows that all macroH2A constructs but not GFP alone are incorporated in chromatin. Representative immunoblot is shown (n = 3). e, The chimeric Capsaspora-mouse macroH2A construct interacts with PARP1 as detected by immunoblotting after co-immunoprecipitation. Representative immunoblot is shown (n = 4).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  *Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

| Data collection | A local DNA sequence database was created from GenBank database by Blast search using human sequences as a query. |
|-----------------|----------------------------------------------------------------------------------------------------------|
| Data analysis   | The following software were used in the study for data analysis: |
|                 | 1. Trinity software (version 2.2.0) for de novo transcriptome assembly, Galaxy web platform (https://usegalaxy.org/) |
|                 | 2. FastQC (2020) for quality control (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) |
|                 | 3. MAFFT (version 7) for MSA (https://mafft.cbrc.jp/alignment/server/) |
|                 | 4. Jalview (version 2) for MSA (https://www.jalview.org/) |
|                 | 5. BioEdit (version 7.2) for potential error editing (https://bioedit.software.informer.com/7.2/) |
|                 | 6. WebLogo3 (version 3) for logo plot generation (http://weblogo.threeplusone.com/) |
|                 | 7. Morpheus (version 07.01.2021) for homology matrix visualization (https://software.broadinstitute.org/morpheus/) |
|                 | 8. MEGA X (version 10.1.7) for phylogenetic and molecular evolutionary analyses (https://www.megasoftware.net/) |
|                 | 9. TimeTree database (version 07.01.2021) for estimating divergence time between pairs of taxa (http://www.timetree.org/) |
|                 | 10. DnaSP (version 6) for calculating the amount of codon usage bias (http://www.cbs.dtu.dk/services/DnaSP/) |
|                 | 11. featureCounts software (v.2.0.1) for quantification of RNAseq reads (http://subread.sourceforge.net/) |
|                 | 12. DESeq2 package (version 1.32.0) for differential gene expression analysis (https://bioconductor.org/packages/release/bioc/html/DESeq2.html) |
|                 | 13. EGGANOG (version 5.0) for functional annotation of genes (http://egganog5.embl.de/#/app/home) |
|                 | 14. GHOSTKOala (version as of 01.03.2021) for functional annotation of genes based on K number (https://www.kegg.jp/ghostkoala/) |
|                 | 15. Orthofinder (version 2.5.4) for identification of gene orthologs (https://github.com/davidemms/OrthoFinder) |
|                 | 16. MicroCal PEAKIT Analysis software (provided by Malvern Panalytical 2018) was used for ITC data analysis. |
|                 | 17. COOT (version 0.9.6), phenoX refine and phenix.elbow (versions as of 01.06.2019) were used for resolving the protein crystal structures (https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/, https://phenix-online.org/documentation/reference/refinement.html) |
phenix-online.org/documentation/reference/elbow.html
19. Consurf tool (version as of 01.03.2021) for projection of evolutionary conservation scores (https://consurf.tau.ac.il/)
20. UCSF Chimera (01.06.2019) and PyMOL (version 2.4) for protein visualization (https://www.cgl.ucsf.edu/chimera/ , https://pymol.org/2/)
21. ImageStudiolite (version 5.x) for the analysis of immunoblots (https://www.licor.com/bio/image-studio-lite/)
22. Inkscape vector graphics on XQuartz (version 2.7.11) editor for creating and editing figures.
23. LIGPLOT (version 4.5.3) for interaction and bond analysis.
24. PRALINE (version as of 01.03.2019) for multiple sequence alignment.
25. STAR (version 2.7.3a) realignment of RNA sequencing reads and inclusion of the transcriptome data and featureCounts software (version 2.0.1) for quantification.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. Github). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files. A data availability statement has been added to the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nrc-reporting-summary-list.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size
We did not perform sample size calculation prior to experimentation but have chosen similar sample sizes to those published in similar experimental studies elsewhere. Details on used samples sizes are given in the Methods section and Figure legends.

Data exclusions
No data was excluded.

Replication
All experiments were replicated in several replicates (n stated in the figure legend of each figure).

Randomization
The design of our study did not require any formal sample randomization.

Blinding
Our study includes biophysical types of analysis that does not allow for user-induced biases. As such we decided to omit blinding. Some experimenters were delegated to technicians that were not aware of sample identity.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ☒ | Antibodies |
| ☒ | Eukaryotic cell lines |
| ☒ | Palaeontology and archaeology |
| ☒ | Animals and other organisms |
| ☒ | Human research participants |
| ☒ | Clinical data |
| ☒ | Dual use research of concern |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ☒ | ChIP-seq |
| ☒ | Flow cytometry |
| ☒ | MRI-based neuroimaging |
Antibodies

Antibodies used
- anti-Capsaspora macroH2A and anti-Capsaspora PARP1 antibodies were generated in this study as described in the methods section and validated by WB as shown in Extended Data Figure 7.
- anti-histone H3 (AbCam; ab-1791) was validated by ChIP, IP and WB according to information provided by the manufacturer.
- anti-PAR (Trevigen; 4336-APC-050) has been validated for ELISA and WB by the manufacturer.
- anti-GFP (Santa Cruz Biotechnology; sc-9996) has been validated by Western blotting and other methods according to the provider's information.
- anti-macroH2A1.1 (Spoorn et al., 2009) was validated by Western blot in the cited paper and by us.
- anti-macroH2A1.2 (Cell Signaling; 4827S) was validated by Western blot by the manufacturer and according to our experience has a weak but tolerable cross-reactivity with macroH2A1.1.
- anti-His-tag (Immunoprecise antibodies; IP-A-2CB1) has been validated for Western blotting, immunofluorescence and other methods by the manufacturer.
- anti-SAF-A has been provided Frank Fackelmayer and has been validated by Western blotting (Kipp et al., 2000, MCB).
- anti-PARP-1 (Abcam, Ab6079; Trevigen, 4338-MC-50) have been validated by Western blotting by both providers.
- fluorophore-conjugated secondary anti-mouse and anti-rabbit (LI-COR Biosciences IODye 680RD and IODye 800CW) have been validated by Western blotting and other methods by the manufacturer.

Validation
The validation of the antibodies generated in this study was done by western blot as shown in the supplementary figure S6C and S6D.

Eukaryotic cell lines

Policy information about cell lines

| Cell line source(s) | HepG2 and HEK293T cell lines was obtained from ATCC. |
|---------------------|------------------------------------------------------|
| Authentication      | Cells were authenticated by STR profiling.          |
| Mycoplasma contamination | Cells are mycoplasma negative. The mycoplasma test was performed by PCR every 2 weeks. |
| Commonly misidentified lines (See ICCLAC register) | Used cell lines are not registered with ICCLAC. |

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

| Laboratory animals | The protist capsaspora owkzarzaki has been grown as described in literature cited in materials and methods. |
|---------------------|---------------------------------------------------------------------------------------------------|
| Wild animals        | N.A                                                                                               |
| Field-collected samples | NA                                                                                              |
| Ethics oversight    | NA                                                                                               |

Note that full information on the approval of the study protocol must also be provided in the manuscript.