







# Adenine DNA methylation associated with transcriptionally permissive chromatin is widespread across eukaryotes

Received: 15 April 2024

Accepted: 14 October 2025

Published online: 18 November 2025

 Check for updates

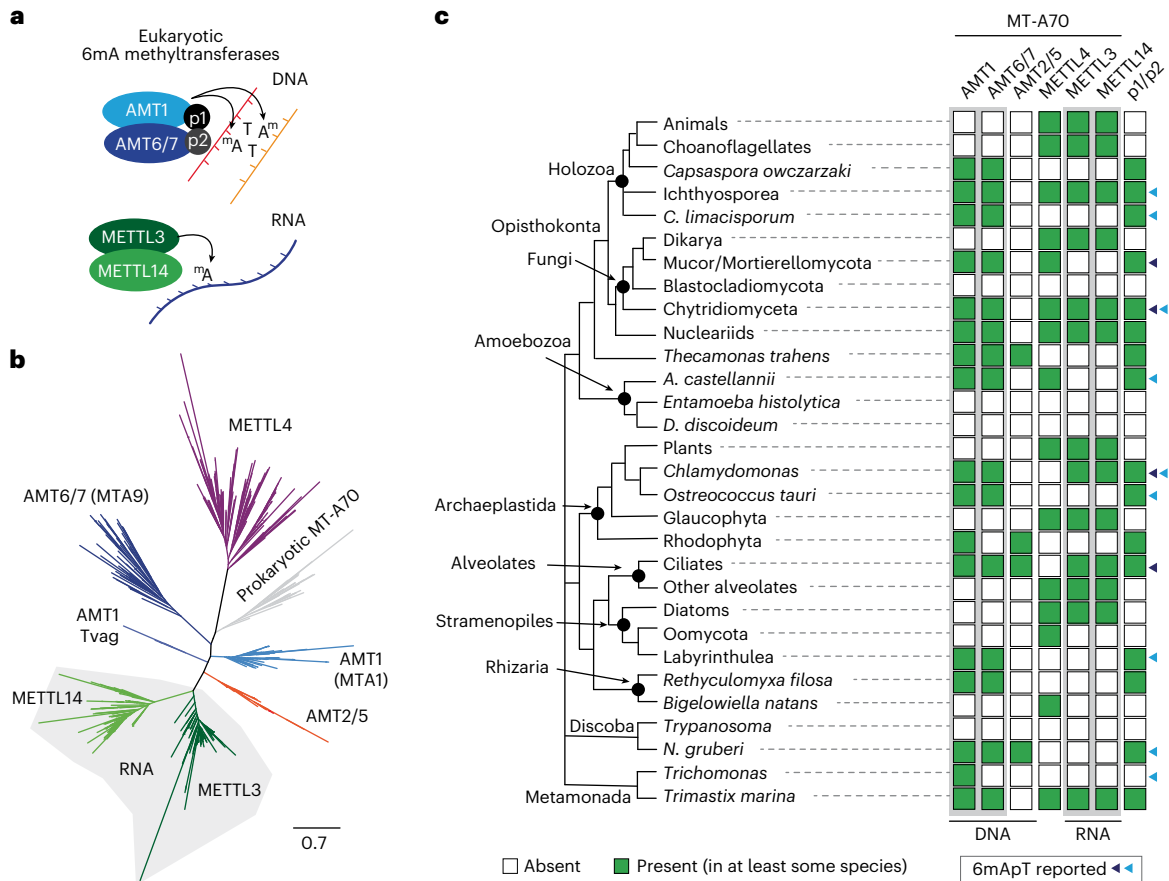
Pedro Romero Charria <sup>1,2</sup>, Cristina Navarrete<sup>3,4</sup>, Vladimir Ovchinnikov<sup>1</sup>, Lan Xu <sup>1,2</sup>, Luke A. Sarre<sup>1,2</sup>, Victoria Shabardina <sup>5</sup>, Ewa Ksiezopolska<sup>3,4</sup>, Elena Casacuberta<sup>5</sup>, David Lara-Astiaso <sup>6,7</sup>, Arnau Seb -Pedr s <sup>3,4,8,9</sup> & Alex de Mendoza <sup>1,2</sup> ✉

DNA methylation is a key regulator of eukaryotic genomes, most commonly through 5-methylcytosine (5mC). In contrast, the existence and function of *N*<sup>6</sup>-methyladenine (6mA) in eukaryotes have been controversial, with conflicting reports resulting from methodological artifacts. Nevertheless, some unicellular lineages, including ciliates, early-branching fungi and the alga *Chlamydomonas*, show robust 6mA signals, raising questions about their origin and evolutionary role. Here we apply Oxford Nanopore sequencing to profile 6mA at base-pair resolution across 18 unicellular eukaryotes representing all major supergroups. We find that robust 6mA patterns occur only in species that encode the adenine methyltransferase AMT1. Notably, 6mA consistently accumulates downstream of transcriptional start sites, positioned between H3K4me3-marked nucleosomes, indicating a conserved association with transcriptional activation. Our results support the idea that the last eukaryotic common ancestor had a dual methylation system, with transcription-linked 6mA and repressive 5mC, which has been repeatedly simplified in both multicellular and unicellular lineages through the loss of the AMT1 pathway.

Modified bases can change the way DNA is interpreted. The best studied modification is 5-methylcytosine (5mC), a widespread eukaryotic mark deposited by DNA methyltransferases (DNMTs)<sup>1–3</sup>, with functions ranging from transposable element (TE) control to gene regulation<sup>4–6</sup>. Other types of base modifications, like hydroxymethyluracil in dinoflagellates or the ‘base J’ in kinetoplastids, are lineage-restricted but abundant in some species<sup>7,8</sup>. In contrast, *N*<sup>6</sup>-methyladenine (6mA) has been reported across various lineages, including animals<sup>9–12</sup>, plants<sup>13</sup>, early-diverging fungi<sup>14,15</sup>, algae<sup>16</sup> and ciliates<sup>17,18</sup>. However,

global 6mA levels are often close to detection limits, with reports questioning its presence in animals, plants or yeast<sup>19–23</sup>. Challenges in 6mA quantification include antibody pull-down specificity issues, bacterial DNA contamination, RNA contamination and sequencing data artifacts<sup>24,25</sup>. Despite this, ciliates, early-diverging fungi and the alga *Chlamydomonas reinhardtii* exhibit high and orthogonally confirmed 6mA levels<sup>14,19</sup>. Thus, inadequate taxon sampling and conflicting results have limited our understanding of 6mA evolution.

<sup>1</sup>School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK. <sup>2</sup>Centre for Epigenetics, Queen Mary University of London, London, UK. <sup>3</sup>Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. <sup>4</sup>Universitat Pompeu Fabra (UPF), Barcelona, Spain. <sup>5</sup>Institut de Biologia Evolutiva (CSIC-UPF), Barcelona, Spain. <sup>6</sup>Department of Haematology, University of Cambridge, Cambridge, UK. <sup>7</sup>Wellcome Trust—Medical Research Council Cambridge Stem Cell Institute, Cambridge, UK. <sup>8</sup>ICREA, Barcelona, Spain. <sup>9</sup>Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK. ✉ e-mail: [a.demendozasoler@qmul.ac.uk](mailto:a.demendozasoler@qmul.ac.uk)



**Fig. 1 | MT-A70 6-methyl-adenine methyltransferases origin and recurrent loss.** **a**, Diagram of the known heterodimers formed by MT-A70 6mA methyltransferases. AMT1 and METTL3 are the active enzymes, whereas their partners are required but do not catalyze the reaction. p1 and p2 are described proteins from the ciliate *O. trifallax* that are required for AMT1 function<sup>18</sup>. **b**, Maximum-likelihood phylogenetic tree of MT-A70 6mA methyltransferases, including eukaryotic and prokaryotic sequences. Each clade is named according to previous references. *T. vaginalis* AMT1 paralogs highlighted as Tvag in the

tree, yet their clustering as AMT1 sequences is shown in Extended Data Fig. 1a. **c**, Distribution of MT-A70 6mA methyltransferase family members across eukaryotes. For some lineages, the presence suggests that at least some species within the clade contain them. Dark triangles depict species for which 6mApT levels were reported previously, and pale blue triangles are species covered in this study. Phylogenetic relationships across these species are based on consensus in the field<sup>35,82</sup>.

A mechanistic understanding of genomic 6mA in eukaryotes comes from research in ciliates. In ciliates, the 6mA methyltransferases belong to the MT-A70 family, unrelated to DNMTs but sharing structural homology with the m<sup>6</sup>A RNA methyltransferases, METTL3 and METTL4 (refs. 26–28). AMT1 (known as MTA1 in *Oxytricha trifallax*) is the main 6mA writer in ciliates, while accessory enzymes from families AMT6/7 (MTA9 in *O. trifallax*) are dispensable and lack catalytic activity<sup>18,29</sup>. However, AMT6/7 forms a heterodimer with AMT1 during DNA engagement, akin to the METTL3 and METTL14 heterodimer (Fig. 1a)<sup>30</sup>. Additionally, DNA-binding proteins p1 and p2 have a role in targeting genomic DNA by forming a multimeric complex with AMT1 (refs. 18,29). In ciliates, early-diverging fungi and the alga *C. reinhardtii*, 6mA is confined to ApT dinucleotides<sup>14,16–18,28</sup>. It is suggested that AMT1 binds hemimethylated ApT sites, as does DNMT1, thereby maintaining 5mC on CpG dinucleotides<sup>31</sup>. In ciliates and *C. reinhardtii*, 6mA is found in internucleosomal linker DNA<sup>16–18,32</sup>, particularly enriched downstream of the transcriptional start site (TSS) of active genes. This association with transcription contrasts the proposed role of 6mA in silencing TE in animals<sup>11,12,22</sup>. Here 6mA may be evolutionarily flexible, like 5mC, which in different eukaryotes is restricted to transposon silencing or gene bodies<sup>4–6</sup>. However, the lack of data from many eukaryotic lineages limits our ability to reconstruct 6mA evolution and function.

To clarify these uncertainties, in this work, we study the evolutionary history of MT-A70 methyltransferases and we find that the AMT1 pathway can be traced back to the last eukaryotic common ancestor (LECA). Then, using Oxford Nanopore modified base calling, previously validated for identifying this base modification in eukaryotes<sup>33</sup>, we detect consistent 6mA patterns in representatives from the major eukaryotic supergroups.

## Results

### AMT1 is ancestral but has been repeatedly lost in eukaryotes

To clarify the origins and evolution of MT-A70 methyltransferases, we examined 231 eukaryotic genomes and transcriptomes (Supplementary Table 1). Additionally, we searched for noneukaryotic MT-A70 homologs, whose origins remain unclear<sup>1,26</sup>. The resulting MT-A70 phylogenetic tree mirrors the following six established eukaryotic families: AMT1, AMT6/7, AMT2/5, METTL3, METTL4 and METTL4 (Fig. 1b)<sup>26,28</sup>. Notably, prokaryotic sequences form a distinct clade, suggesting a bacterial origin for MT-A70 inherited early in eukaryotic evolution (Fig. 1b and Extended Data Fig. 1). MT-A70 presence in prokaryotes is confined to select eubacterial phyla, with few archaean genomes harboring it, typically in a single copy per species (Extended Data Fig. 1a,b). Asgardarchaeota sequences cluster within prokaryotes rather than as sister to eukaryotes, arguing against

an archaeal donor. Only one Asgardarchaeota of the 302 genomes searched contains a METTL3-like sequence, likely from metagenomic assembly contamination (Extended Data Fig. 1a). Unlike the more widespread bacterial-type Dam 6mA methyltransferase, Mui1-like 6mA MT-A70 in bacteria is less diverse<sup>1</sup>. The monophyletic clustering of bacterial sequences supports a single origin of eukaryotic MT-A70 from an unidentified bacterial donor, or, alternatively, an eukaryotic origin followed by a single lateral gene transfer to prokaryotes, which is less parsimonious. This rules out recurrent transfers from prokaryotes to eukaryotes.

We found that all six MT-A70 families are present in most eukaryotic supergroups (Fig. 1c), suggesting that they duplicated before LECA. The RNA-associated METTL3 and METTL4 co-occur across the species phylogeny, in a pattern similar to the DNA-associated AMT1 and AMT6/7, suggesting that both enzymatic pairs have conserved heterodimer associations across eukaryotes (Fig. 1c and Extended Data Fig. 1c). AMT2/5 is less common, often co-occurring with AMT1, hinting at a potential partnership but one in which AMT2/5 is dispensable. Notably, AMT2/5 is exceptional as it possesses C-terminal ZZ zinc finger domains of unknown function (Extended Data Fig. 1d)<sup>26,28</sup>. In contrast, METTL4 does not co-occur with any MT-A70 families, suggesting that it does not form heterodimers (Fig. 1c and Extended Data Fig. 1c). METTL4 is believed to methylate mitochondrial DNA or small RNAs, but its role is currently not well understood<sup>22,34</sup>. The nematode *Caenorhabditis elegans* METTL4 ortholog (known as DAMT-1) has been suggested as a genomic DNMT<sup>10</sup>, yet this is contested<sup>20</sup>. Beyond the MT-A70 methyltransferases, the proteins p1 and p2 exclusively co-occur with AMT1 (Fig. 1c), indicating an ancestral link with the AMT1 pathway<sup>18</sup>.

Hence, the enzymes responsible for RNA and DNA 6mA exhibit ancient origins, yet their evolution appears largely decoupled to each other in eukaryotes. Notably, AMT1 is consistently found in species with documented robust 6mA levels, such as ciliates, *C. reinhardtii* and early fungi<sup>14,16–18</sup>. Although exceptions exist, AMT1 co-occurs with AMT6/7 and p1/p2 partners (for example, *Trichomonas vaginalis*).

### 6mA correlates with transcription in AMT1-encoding species

Based on the AMT1 distribution, we searched for eukaryotic lineages with detectable genomic 6mA. To avoid bacterial contamination, we focused on species grown axenically. Our selection included 15 species encoding AMT1 and three species lacking AMT1 orthologs as negative controls. Our sampling of ichthyosporeans, *Corallochytrium limacisporum*, a chytrid fungus, amoebozoans, chlorophytes, a glaucophyte, stramenopiles, heteroloboseans and a metamonada spans major eukaryotic branching groups<sup>35</sup>. Genomic DNA was sequenced using Oxford Nanopore R9.4.1 technology for 15 species, while equivalent public data were used for three more<sup>36–38</sup>. Using the Guppy software with the ‘all-context’ Rerio model, we obtained maps for 6mA and 5mC in all species. Global methylation levels varied substantially across species, with the lowest levels observed in species lacking AMT1 (Fig. 2a).

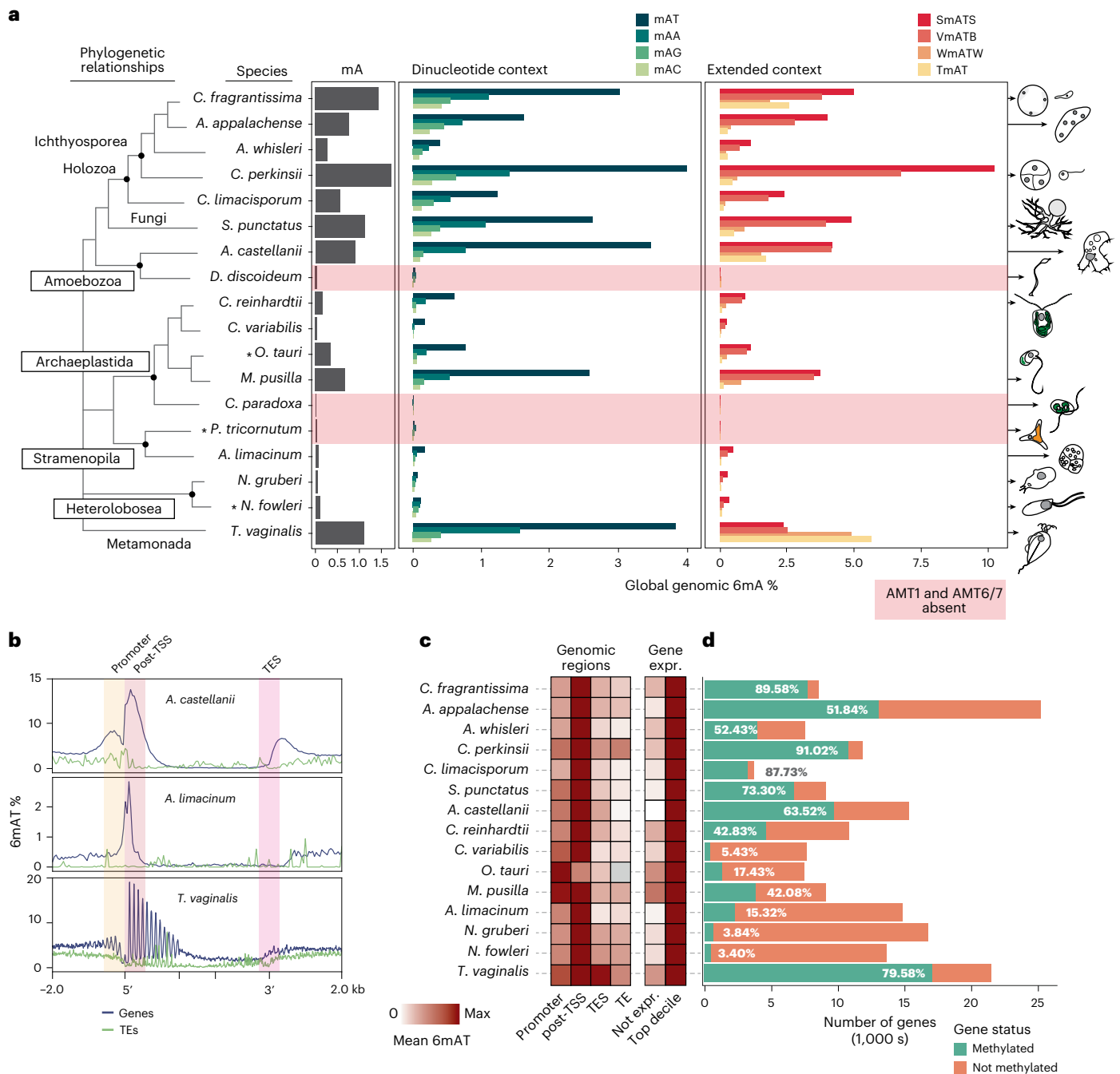
Because AMT1 preferentially methylates ApT dinucleotides in ciliates<sup>14,16,17</sup>, we assessed global 6mA levels across various sequence contexts. In all AMT1-encoding species, AT was the main methylated context, although AA also showed appreciable levels (Fig. 2a). ApT sites displayed symmetric methylation across all AMT1-encoding species (Extended Data Fig. 2a). In contrast, AMT1-lacking species had negligible difference among dinucleotide contexts, with the AA dinucleotide being the most highly enriched (Fig. 2a and Supplementary Fig. 1). Notably, *C. reinhardtii* and early-diverging fungi exhibit higher methylation in the VATB context (V = C/G/A, B = C/G/T)<sup>14,19</sup>, whereas in fungi the TAT trinucleotide showed the lowest methylation<sup>14</sup>. Thus, we calculated global levels for three 4-mer combinations and the TAT trinucleotide to discern AT flanking specificity. Across all AMT1-containing species SATS (S = C or G) exhibited the highest methylation, while TAT and WATW (W = A or T) showed the lowest (Fig. 2a). The VATB context

displayed intermediate levels, suggesting that cytosine and guanine flanking dictate methyltransferase specificity (Fig. 2a). An exception was the metamonad *T. vaginalis*, which encodes four AMT1 paralogues but lacks AMT6/7, p1 or p2 (Fig. 1c). In *Trichomonas*, WATW and TAT were more enriched than SATS (Fig. 2a), indicating the signal reflects distinct methyltransferase preferences and showing that these patterns are not due to Oxford Nanopore base-calling bias.

We then examined the genomic distribution of 6mA. In species without AMT1, 6mA appeared uniformly across genes and TEs (Extended Data Fig. 2b), consistent with the background noise. Conversely, in AMT1-encoding species, a peak of AT methylation emerged proximal to the TSS (Fig. 2b,c). AA methylation mirrored these patterns, suggesting that AA methylation might be an off-target substrate of AMT1 (Extended Data Fig. 2c). Some species displayed a subtle 6mA enrichment in promoter regions (upstream of the TSS; Fig. 2c)<sup>16</sup>, which proved to be an artifact arising from the head-to-head orientation of upstream genes (Extended Data Fig. 2d). Neither transcriptional end sites nor TEs exhibited enrichment in 6mA (Fig. 2c). Genic 6mA correlated with gene transcriptional activity (Fig. 2c). Across all species, silent genes presented lower 6mA levels compared to highly expressed genes, which also harbored more methylated ApT sites (Fig. 2c and Extended Data Figs. 3 and 4). The association with transcription was quite variable; for instance, *Acanthamoeba castellanii* exhibited a gradual link with transcriptional levels, while in ichthyosporeans or prasinophytes all genes above a specific transcriptional threshold presented uniform 6mA levels (Extended Data Figs. 3 and 4). This diversity was also evident in the total number of genes displaying methylation in each genome (Fig. 2d and Supplementary Fig. 2). The ichthyosporean *Chromosphaera perkinsii* exhibited 92% of genes with 6mA, whereas both *Naegleria* species displayed detectable methylation in only 3% genes (Fig. 2d). Notably, most amino acid positions on the AMT1 sequences were under strong purifying selection (Supplementary Table 3), suggesting that the variation in targeting mechanisms rather than in the AMT1 protein itself is responsible for its differential usage, perhaps underlying distinct promoter architectures<sup>39</sup>.

### Stage-specific transcription occurs independent of 6mA

We then tested how variable 6mA is upon transcriptional changes. We chose the amoebozoan *A. castellanii*, the fungus *Spizellomyces punctatus*, the ichthyosporeans *Creolimax fragrantissima* and *C. perkinsii* because these have different cell stages that can be isolated in culture<sup>40–43</sup> (Fig. 3a), high 6mA levels and distinct associations between 6mA and transcriptional level in the steady state (Extended Data Fig. 3). We also collected matched transcriptomic and DNA samples for *Naegleria gruberi* grown at 23 °C and 30 °C. All species presented static 6mA methylomes at ApT and gene-level resolution, where most genes displayed minimal 6mA differences (Fig. 3b,c and Extended Data Fig. 5a,c). In *N. gruberi*, the 644 genes called methylated using R9 nanopore chemistry, showed comparable 6mA levels at 23 °C and 30 °C, despite using R10 chemistry and a distinct analytical pipeline (Extended Data Fig. 5c). When we computed the total number of differentially expressed genes across stages and overlapped these with genes exhibiting >1% change in 6mA levels along the gene body, we observed that thousands of genes change expression without any detectable 6mA alterations (Fig. 3d). In *A. castellanii*, 24% of the trophozoite upregulated genes were also hypermethylated in this stage, following a positive correlation with transcriptional change, yet the rest of trophozoite upregulated genes did not show methylation changes and 5.57% were hypomethylated, implying a limited link between 6mA and transcriptional change. In the *S. punctatus* zoospore, some genes were almost completely silenced in one stage yet retained high 6mA levels as in the transcribed stage (Fig. 3e). Further analysis of genes with variable 6mA (>1.5%) showed no correlation with transcriptional changes (Extended Data Fig. 5b,d). Although 6mA is associated with steady-state transcription in all AMT1-encoding eukaryotes, its link to



**Fig. 2 | Nanopore sequencing reveals widespread 6mA in AMT1-encoding eukaryotic genomes.** **a**, Global methylation levels of 6mA in different sequence contexts across 18 eukaryotes representing deep branching points of eukaryotic diversity. S stands for C or G, W stands for A or T, V stands for A,C,G and B stands for T,C,G. Species with red shade do not encode for AMT1 or its partner AMT6/7. Asterisks indicate Oxford Nanopore datasets obtained from NCBI/ENA and silhouettes represent the shapes of the species. **b**, Average 6mA levels in the AT context over three example species, including genes (blue) and TEs (green; >1,000 bp). Thus, 5' and 3' ends of the features are marked with dashed lines, the

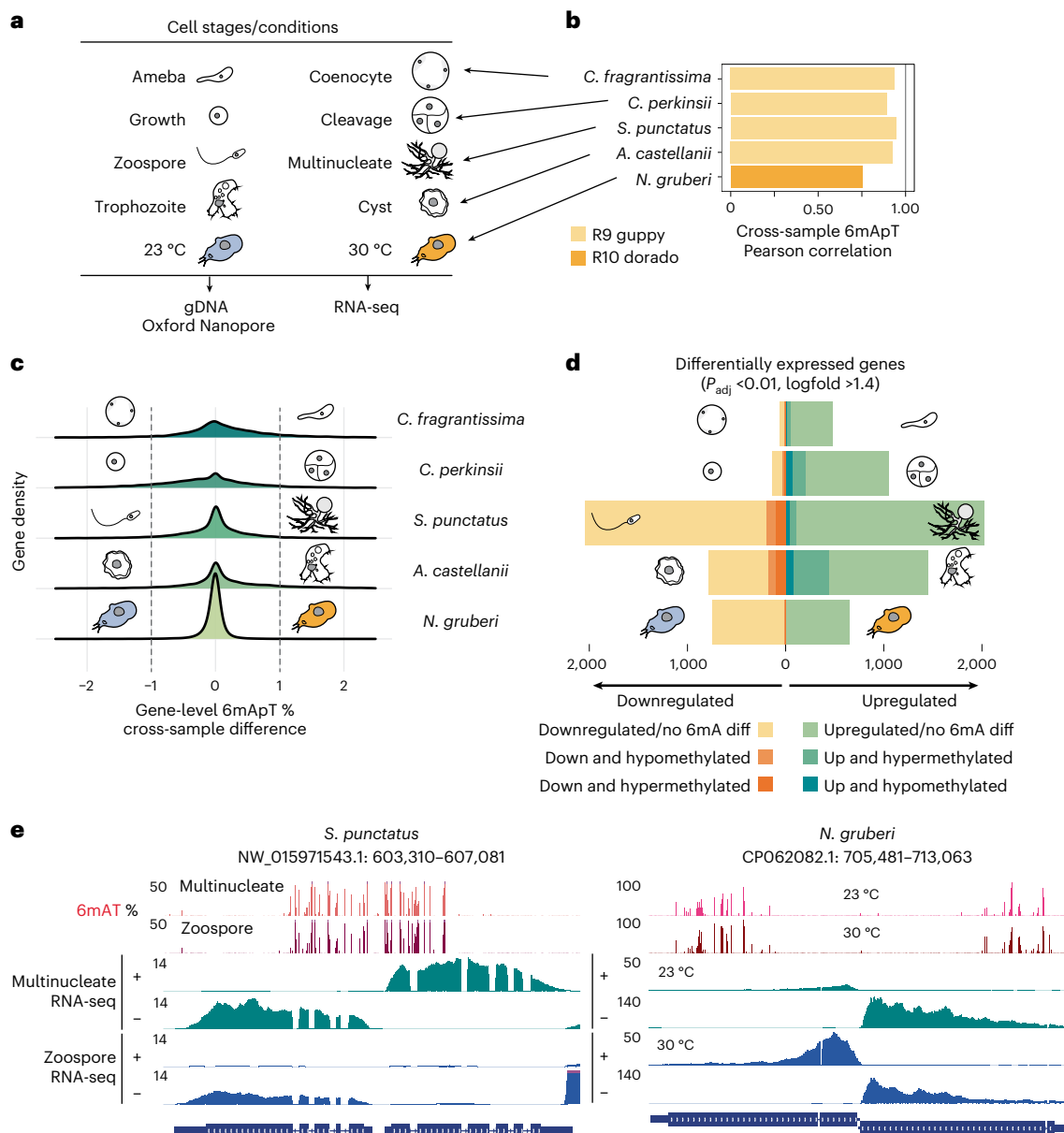
first and last 1,500 bp are unscaled, and the shades indicate the regions used in **c**. **c**, Heatmap displaying average 6mA levels across various genomic regions, including the 500-bp downstream of the post-TSS, the 500 bp centered around the TES and TEs. Also, for gene-body 6mA levels for nonexpressed genes and the top decile of highly expressed genes, see full values in Extended Data Fig. 3. Gray cell indicates missing data, as *O. tauri* has too few repeats. **d**, Number of genes methylated in each genome (see Methods for thresholds). TES, transcriptional end site.

dynamic transcription is weak, suggesting that 6mA reprogramming is dispensable for transcriptional change.

**5mC and 6mA mark different functions in eukaryotes**

As 5mC is the other widespread form of DNA methylation in eukaryotes, we examined its relationship with 6mA within our species set (Fig. 4a).

We found diverse scenarios—species exhibited nondetectable 5mC levels alongside high 6mA (for example, *C. limacisporum*), while others exclusively harbored 5mC (*Phaeodactylum tricoratum*). Some species presented both modifications, and *Dictyostelium discoideum* displayed neither (Fig. 4a). Negligible 5mC levels and the absence of CpG symmetry detected with nanopore matched previous results from



**Fig. 3 | Stable 6mA patterns correlate weakly with stage-specific transcriptional changes.** **a**, Collection of species and stages/treatments profiled in this study. **b**, Global 6mA correlation levels calculated for all ApT sites in each stage. All species were profiled using R9.4.1 ONT chemistry and processed with Guppy, except for *N. gruberi*, which was profiled using R10.4.1 ONT chemistry and processed with Dorado. **c**, Distribution of gene-body 6mA percentage differences between two stages across all five species comparisons. A threshold of  $\geq 1\%$

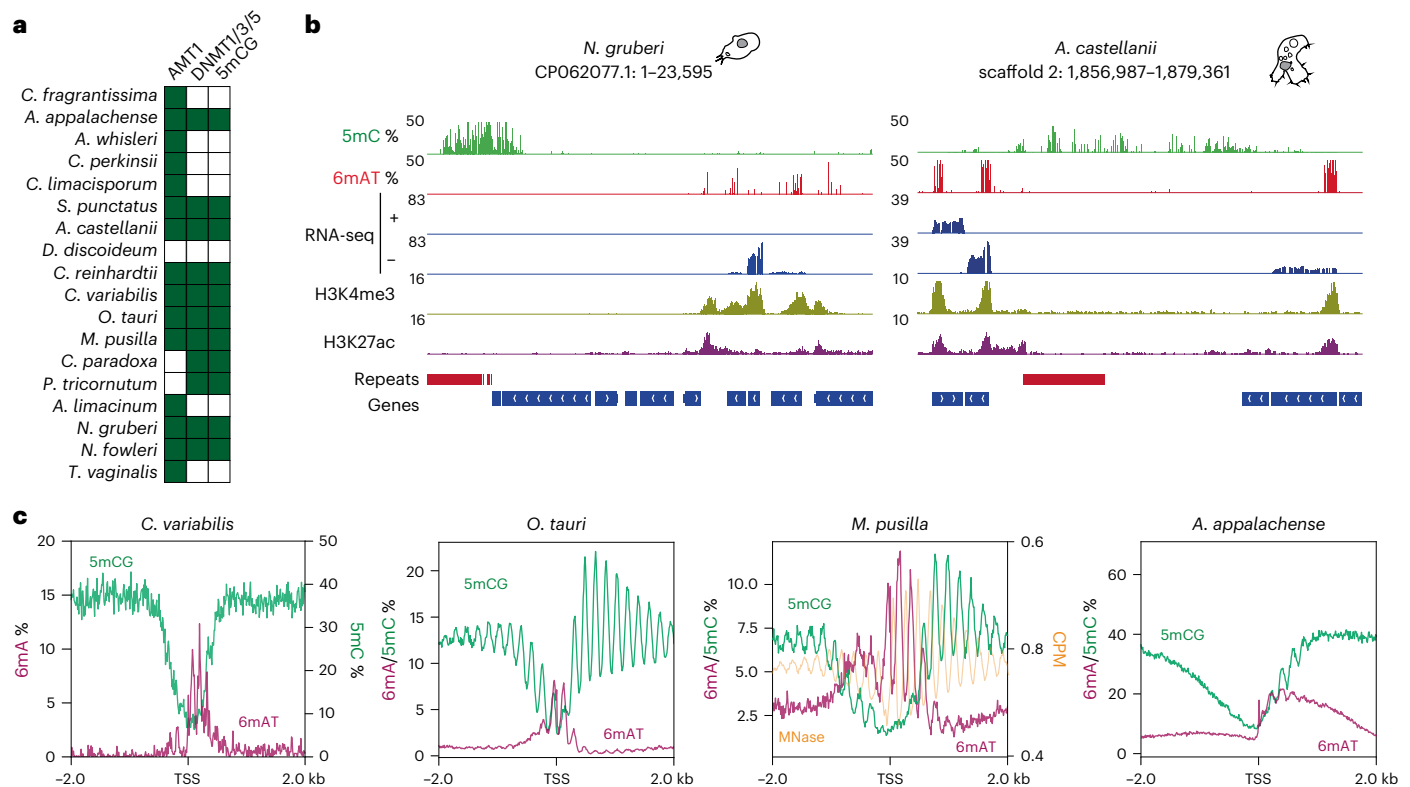
6mA delta was used to classify a gene as hypermethylated or hypomethylated. **d**, Number of differentially expressed genes across stages, divided by 6mA methylation change categories (no change, hypermethylated, hypomethylated). **e**, Genome browser snapshots showing differentially transcribed genes alongside stable 6mA levels in *S. punctatus* and *N. gruberi*. RNA-seq values are normalized to CPM and displayed separately for each strand. ONT, Oxford Nanopore Technologies; CPM, counts per million.

enzymatic methyl-sequencing and whole-genome bisulfite sequencing in these species<sup>44,45</sup> (Extended Data Fig. 6a). This variation suggests that 6mA is not evolutionarily linked to 5mC, corroborating observations in early-diverging fungal lineages<sup>14</sup>.

The recurrent loss of 5mC in eukaryotes is partly attributed to its mutagenic effects<sup>46</sup>. Species with elevated 5mC levels, like vertebrates, often have fewer CpG dinucleotides than expected by chance, because deamination of 5mC into thymine depletes CpGs over evolution<sup>47</sup>. In contrast, animals with low or no 5mC lack this compositional bias<sup>48</sup>. Using a similar approach, we calculated observed-to-expected ApT ratios across our species. Global 6MAT levels did not correlate with ApT dinucleotide depletion, with a relatively constant ratio across species (Extended Data Fig. 6b). Notably, CpG depletion was observed

in species with very low 5mCG levels (for example, *Naegleria*), while prasinophytes exhibited the opposite trend, with high 5mCG levels and enriched CpG composition (Extended Data Fig. 6b) as previously reported<sup>3</sup>. These findings suggest that neither 5mC nor 6mA substantially shapes genomic base composition across the vast diversity of eukaryotes, in contrast with prior observations for 5mC in animals and plants.

One possibility is that 6mA mutagenic effects are most evident in regions where it is enriched, particularly post-TSS. Similarly, in invertebrates with 5mC in gene bodies, methylated genes show lower CpG ratios than unmethylated ones<sup>49,50</sup>. We grouped genes by the presence or absence of methylated ApTs and compared local ApT density. In most species, 6mA showed no clear association with ApT composition



**Fig. 4 | 6mA and 5mC are compartmentalized in eukaryotic genomes.**

**a**, Distribution of DNMTs and the presence of 5mC in the CG context across the species in our dataset. Dark green indicates presence, while white indicates absence. **b**, Genome browser snapshots of *N. gruberi* and *A. castellanii* showing 6mA enrichment downstream of expressed genes and 5mC demarcating transcriptionally silent repetitive regions. RNA-seq (strand-specific) and

ChIP-seq data are scaled using CPM. **c**, Average methylation levels centered at the TSS in species that have 5mC gene-body methylation. For *C. variabilis*, 6mA and 5mC axes are shown separately. For *M. pusilla*, MNase data are shown to highlight the anticorrelation with both 5mC and 6mA peaks, as CPM. MNase, micrococcal nuclease.

(Extended Data Fig. 7). In some species, methylated genes had higher ApT density than unmethylated ones, opposite to patterns seen with 5mC. In early-diverging fungi, 6mA-methylated genes are depleted for TAT motifs. We confirm this in the fungus *S. punctatus*, but not across other eukaryotes (Supplementary Fig. 3). Thus, the recurrent loss of 6mA in eukaryotes is unlikely to be driven by mutagenic effects, which appear limited.

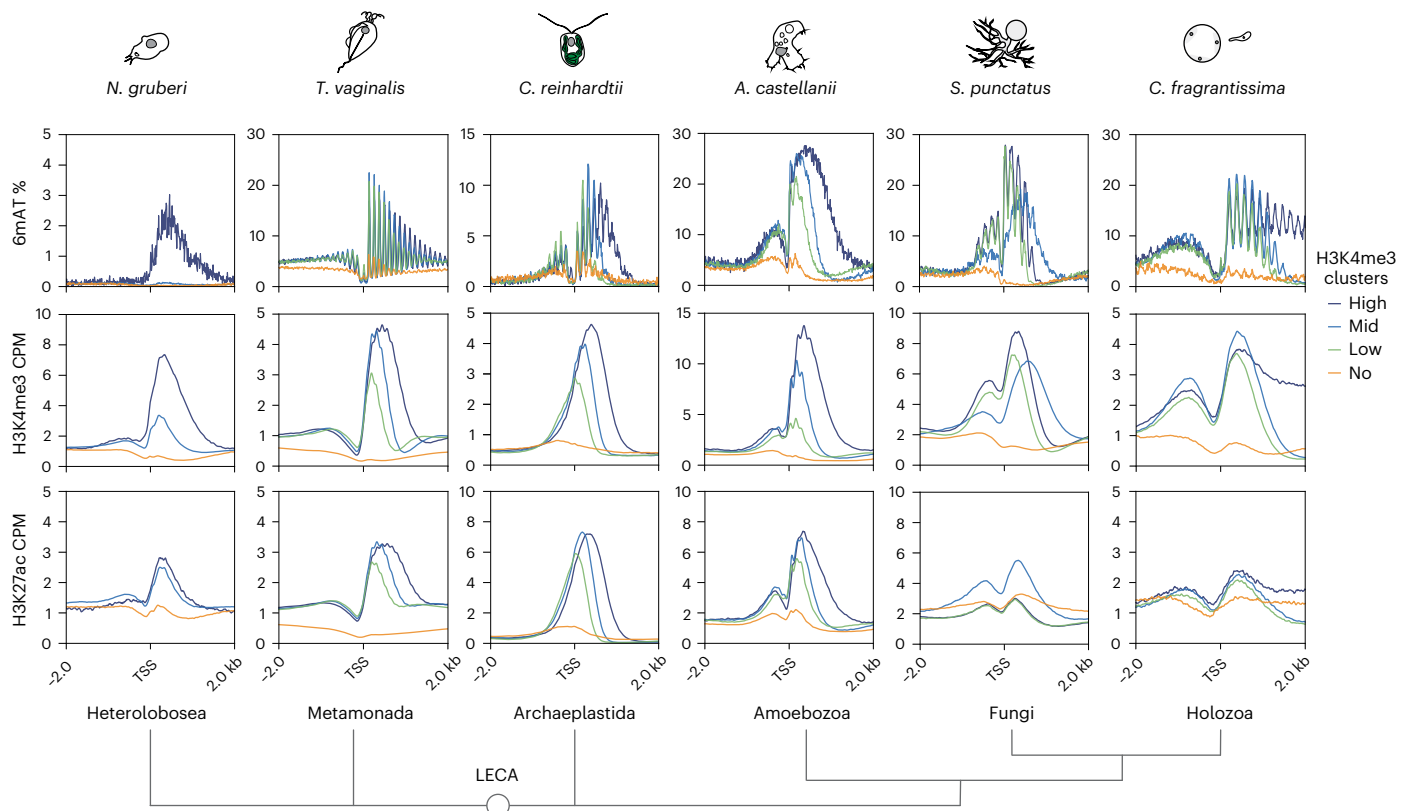
In species with both 6mA and 5mC, we observe two main patterns. In *Acanthamoeba*, *Naegleria*, *Spizellomyces* or *Chlamydomonas*, as in the previously reported early-branching fungi<sup>14,51</sup>, 5mC is confined to transposons while 6mA marks expressed genes (Fig. 4b), suggesting opposing links to transcription. Nevertheless, in several eukaryotes, 5mC is also found within gene bodies, including animals and plants<sup>52,53</sup>, but also the chlorophytes and the ichthyosporan *Amoebidium appalachense* in our dataset<sup>3,45</sup>. In this second pattern, 6mA predominantly occupies the five prime ends of the gene body, while 5mC peaks when 6mA diminishes (Fig. 4c). In all four species, there is a distinct periodic 6mA pattern suggestive of internucleosomal positioning (Fig. 4c). We used micrococcal nuclease sequencing data for *Micromonas pusilla* to confirm that 6mA peaks are anticorrelated with nucleosome positioning (Fig. 4c and Extended Data Fig. 8)<sup>3</sup>. In *Amoebidium*, both modifications coexist in the first three or four nucleosome territories downstream of the TSS (Fig. 3c and Extended Data Fig. 8). In the case of *Micromonas* and *Ostreococcus*, nanopore data align with the reported pattern of internucleosomal 5mC distribution along the gene bodies of prasinophytes<sup>3</sup>. The original study noted that 5mC is weakly linked to transcription in prasinophytes, suggesting a compaction role in these microalgae with very small genomes (for example, *Ostreococcus tauri*, 12.6 Mb)<sup>3</sup>. Notably, 6mA replaces 5mC in the first nucleosomes

downstream of the TSS of active genes, suggesting a role in maintaining transcriptional permissiveness, as inactive genes show less 6mA (Extended Data Figs. 3 and 4).

### Nucleosomes with H3K4me3 colocalize with 6mA

Our findings align with prior reports indicating an enrichment of 6mA in nucleosomal linker DNA<sup>16–18,32</sup>. In ciliates, the deletion of AMT1 disrupts 6mA patterns, resulting in more diffuse nucleosome positioning around the TSS<sup>18,28</sup>. Most species in our dataset also exhibit a periodic 6mA pattern. However, *Naegleria*, *Acanthamoeba* or *Chromosphaera* lack a discernible periodic 6mA pattern (Extended Data Fig. 7). The absence of periodicity might be attributed to irregular nucleosome positioning in these species or potential issues with TSS annotations. However, we could not detect periodic patterns over individual genes, unlike in species that display periodicity (Fig. 4b). Moreover, there is a considerable variation in the average number of nucleosomes covered in 6mA territories across species. For instance, the stramenopile *Aurantiochytrium limacinum* displays only three 6mA peaks, while the metamonad *T. vaginalis* exhibits periodic 6mA peaks across entire gene bodies, albeit with decreasing intensity (Fig. 5 and Extended Data Fig. 7). Thus, the 6mA–nucleosome linker association is consistent across AMT1-encoding eukaryotes, although with lineage-specific variations.

To probe what shapes 6mA patterns and their link to nucleosomes, we performed chromatin immunoprecipitation sequencing (ChIP-seq) for two TSS-associated and transcriptionally active histone marks—histone 3 lysine 4 trimethylation (H3K4me3) and lysine 27 acetylation (H3K27ac)<sup>28,54</sup>. We generated data for four species and reanalyzed datasets for two additional species<sup>55</sup>, covering most divergent eukaryotic groups with comparable ChIP-seq quality (Extended Data Fig. 9a



**Fig. 5 | Nucleosomes with H3K4me3 modification recapitulate 6mA patterns.** Average 6mA (on ApT context), H3K4me3 and H3K27ac across five eukaryotic lineages. For each species, genes have been clustered using the H3K4me3 signal, forming the following four clusters: high, middle, low and no signal. For

*N. gruberi*, because the amount of methylated genes is small, the classification was done independently, dark blue corresponds to 6mA-methylated genes, pale blue corresponds to genes with H3K4me3 with no 6mA and orange for genes that lack both H3K4me3 and 6mA. H3K4me3 and H3K27ac are shown as CPM.

and Supplementary Fig. 4). Both histone marks showed expected enrichments post-TSS with peak widths under 1,000 bp (Fig. 5 and Extended Data Fig. 9b,c)<sup>28,56</sup>. We then grouped genes by H3K4me3 signal and found that 6mA closely overlaps with this histone mark (Fig. 5 and Extended Data Fig. 9d). In contrast, genomic regions beyond H3K4me3 peaks have substantially lower 6mA levels across all species (Extended Data Fig. 9e). Notably, 6mA aligns with H3K4me3 not only in intensity but also in deposition width.

In the ichthyosporean *C. fragrantissima*, some genes were fully covered by H3K4me3, extending beyond typical post-TSS patterns (Fig. 5 and Extended Data Fig. 9d). These broad H3K4me3 peaks cannot be explained solely by head-to-head gene arrangements (Extended Data Fig. 9c,d), and have been associated with low transcriptional levels<sup>57</sup>. Within these H3K4me3 domains, 6mA is present throughout (Extended Data Fig. 10a), supporting a tight link between their deposition in *Creolimax*. Interestingly, in ichthyosporeans, the DNA-binding partner of AMT1—the p1 subunit—contains a PHD domain, which is known to read histone tail methylations<sup>58</sup> (Supplementary Fig. 5).

The other two species displaying divergent 6mA and H3K4me3 patterns are the heterolobosean *N. gruberi* and the metamonad *T. vaginalis*. In *Naegleria*, only 644 genes show 6mA, so we categorized genes based on the presence or absence of 6mA and/or H3K4me3. Genes with 6mA showed the highest H3K4me3 signal (Figs. 4b and 5). However, most genes displayed H3K4me3 without 6mA, suggesting independent deposition (Fig. 5 and Extended Data Fig. 9d). In contrast, H3K27ac was comparable between genes with 6mA and those with only H3K4me3. In *Naegleria*, as in most eukaryotes, the p1 subunit lacks a PHD domain, so colocalization with H3K4me3 would require intermediate proteins.

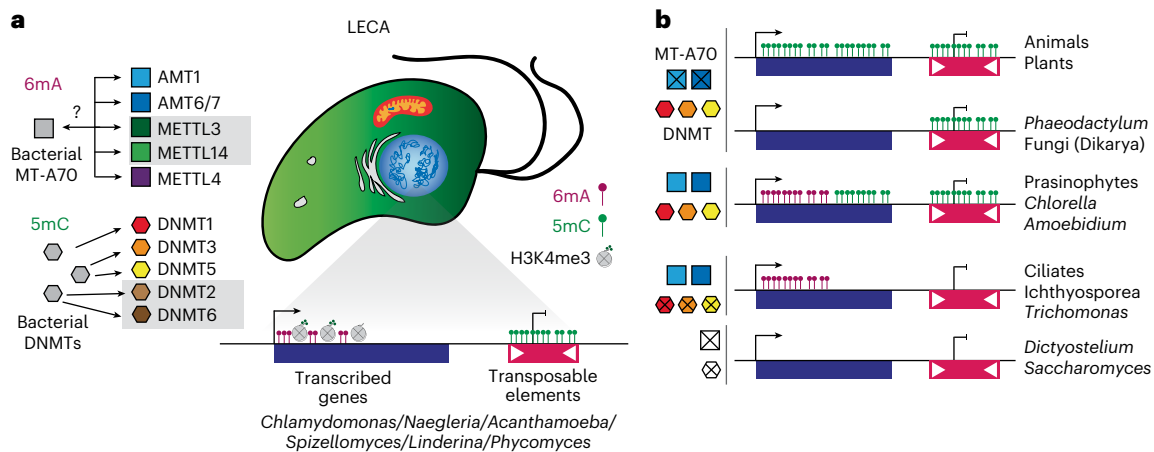
*T. vaginalis* exhibits an opposite pattern; H3K4me3 and H3K27ac correlate with 6mA intensity, but 6mA extends beyond the modified nucleosomes (Fig. 5 and Extended Data Figs. 9d and 10a). In *Trichomonas*, both 6mA and H3K4me3 may depend on transcriptional initiation, but 6mA extends further, spanning broader domains. *Trichomonas* is exceptional as it lacks p1, p2 and AMT6/7 orthologs that could explain this divergent pattern (Fig. 1c).

To directly test whether the H3K4me3 deposition influences 6mA presence, we treated *A. castellanii* with H3K4me3 inhibitors. We used OICR-9429 and Piribedil, which block WDR5 interactions with histone methyltransferases<sup>59,60</sup>. Both inhibitors resulted in a dose-dependent reduction in H3K4me3 levels while leaving overall H3 levels unaffected (Extended Data Fig. 10b). At the highest inhibitor concentrations, we profiled the 6mA methylome in treated and DMSO control samples. Furthermore, 6mA levels showed minimal differences between inhibitor and control groups, both globally and at H3K4me3 peaks (Extended Data Fig. 10c,d), suggesting that H3K4me3 does not directly determine 6mA deposition in these species.

Overall, our data suggest that AMT1-dependent 6mA is tied to nucleosome positioning across eukaryotes, especially with transcription-permissive marks. The hierarchy between these components appears lineage-specific and may be simplified in some lineages.

## Discussion

We show that 6mA is widespread across eukaryotes and trace its origin to the LECA. Our broad sampling supports an ancestral AMT1 pathway, irrespective of the debated root positions of the eukaryotic tree of life<sup>61–63</sup>. Unlike 5mC DNMTs acquired by LECA through independent bacterial transfers<sup>64</sup>, the 6mA MT-A70 methyltransferases were likely inherited in a single event from a bacterial donor (Fig. 6a). Pre-LECA,



**Fig. 6 | Reconstruction of the LECA epigenome.** **a**, Diagram showcasing the distinct evolutionary origins of 6mA (MT-A70) and 5mC (DNMTs) enzymes in the advent of the LECA, and the inferred ancestral epigenome based on distribution and proposed roots of the eukaryotic taxa. Species genus represents examples of extant eukaryotes that still exhibit the inferred ancestral patterns. The question mark symbol indicates that potentially prokaryotes might have acquired

MT-A70 sequences from a pre-eukaryotic lineage. **b**, Derived epigenome patterns in modern eukaryotes associated with simplification in methyltransferase repertoires, with example species/lineages for each pattern. Crossed rectangles or hexagons indicate enzyme losses, which are color coded and shaped as in **a**. The presence of DNMTs does not imply co-occurrence of all three enzymes, but any of them.

MT-A70 underwent duplications into six families, undergoing subfunctionalisation in heterodimeric partnerships and RNA or DNA substrates. Postdiversification, MT-A70 families remained relatively static in eukaryotic evolution, contrasting with DNMT families, which underwent lineage-specific duplications and expansions (Dim-2/RID in fungi, CMT and DRM in plants)<sup>3,4,65</sup>. Concurrently, 6mA patterns remained relatively constant across AMT1-encoding eukaryotes, consistently enriching ApT dinucleotides at TSS. Conversely, 5mC is variable, ranging from gene-body methylation to TE silencing, and methylates sequence contexts beyond CpGs<sup>4-6</sup>. As 5mC is presumed to ancestrally silence TEs<sup>4,6,53</sup>, its faster evolutionary dynamics may result from adaptations to track recurrent transposon invasions, in contrast to the slower concerted co-evolution of the 6mA-AMT1 pathway targeting endogenous genes. Similarly, the targeting of silencing histone modifications appears to evolve more dynamically than those of transcriptionally associated modifications<sup>57</sup>. Our survey suggests that LECA had a dual methylation system, with transcriptionally associated 6mA and repressively associated 5mC (Fig. 6a).

Our study confirms Oxford Nanopore as a reliable platform for 6mA detection<sup>33,66</sup>, showcasing robust and reproducible patterns across species with diverse genome compositions. Long-read base-pair resolution maps avoid pitfalls of other methods<sup>24</sup>. As an example, in *T. vaginalis*, antibody-based pull-downs suggested 6mA enrichment in TE<sup>67</sup>, but this genome is dominated by recently duplicated Maverick retrotransposons<sup>68</sup>, where short-read mapping artifacts are common<sup>25,69</sup>. By contrast, our nanopore data show a strong periodic 6mA pattern on transcribed genes in *Trichomonas*, without enrichment in TEs, consistent with patterns across other eukaryotes<sup>16-18</sup>. Nanopore can provide independent confirmation of conflicting PacBio CLR results, as in *C. elegans*, where 6mA presence remains debated<sup>10,20,70,71</sup>. In *C. elegans*, the proposed 6mA methyltransferase METL-9 belongs to a different family than MT-A70, targeting GGAG rather than ApT<sup>70,71</sup>. Alternative 6mA DNA pathways may exist in eukaryotes, such as the fungal MetB asymmetric methyltransferase<sup>15</sup>, acquired through lateral gene transfer from bacteria. However, in nematodes, the few thousand 6mA sites detected are poorly reproducible across samples and lack the clustered patterns of AMT1-encoding species, likely reflecting the absence of a maintenance methyltransferase<sup>71</sup>. Alternatively, nucleotide salvage pathways incorporating m6A from RNA and mitochondrial DNA into nuclear DNA, or RNA:DNA hybrids, might explain contentious previous reports<sup>22-24,72,73</sup>. By contrast, robust and reproducible ApT

methylation in AMT1-encoding eukaryotes suggests a stable, functionally important chromatin component.

Our finding of recurrent AMT1 pathway loss challenges assumptions about the evolution of DNA methylation. Similar to DNMTs and the RNA m6A METTL3/METTL14 pathway, the 6mA-AMT1 pathway is frequently lost across eukaryotes<sup>4,45</sup>. As 6mA shows no obvious mutagenic effect, these losses may reflect evolutionary contingency, with 6mA role compensated by alternative mechanisms. This has produced diverse combinations of 6mA and 5mC across extant eukaryotes (Fig. 6b).

Thus, 6mA strongly correlates with transcription, is enriched with highly expressed genes, yet its dynamics appear limited. Across five species in our dataset, 6mA shows little change under transcriptional perturbations. In another recently published example, the fungi *Phycomyces blakesleeanus* and *Mucor lusitanicus* also displayed modest 6mA dynamics despite treatment-specific transcriptional changes<sup>15</sup>. Similarly, AMT1 knockouts in ciliates and *M. lusitanicus* result in limited transcriptional changes<sup>18,74</sup>. This resembles gene-body 5mC in plants and invertebrates, stably expressed genes are usually methylated, yet developmental transcriptional changes rarely alter 5mC patterns<sup>4,75,76</sup>. Likewise, while H3K4me3 is closely associated with transcription, its direct role in initiating transcription remains debated<sup>77</sup>. For instance, depleting H3K4me3 can increase RNA polymerase II pausing without affecting transcriptional initiation<sup>78</sup>. Overall, chromatin features such as 6mA, gene-body 5mC and H3K4me3 mark transcriptionally active regions but likely act in permissive or downstream roles rather than directly regulating transcription.

The function of 6mA in nucleosome positioning remains unclear, as ciliates with mutated AMT1 exhibit disordered nucleosomes yet remain generally viable<sup>18,28</sup>, primarily only lethally affected in sexual reproduction. In ciliates, the effect on nucleosome positioning is also stronger in vitro than in vivo<sup>18,28,32</sup>. Some of our surveyed species lack a periodic 6mA pattern, questioning the importance of 6mA in nucleosome positioning. In contrast, our data suggest an ancient link between 6mA and H3K4me3, perhaps also with H2A.Z<sup>28,74</sup>. Given the widespread presence of H3K4me3 as a post-TSS marker<sup>54</sup>, even in species lacking 6mA, the latter may be nonessential or readily substituted in this region. Notably, H3K4me3 and H2A.Z usually anticorrelate with 5mC across eukaryotes<sup>53,79,80</sup>, so the H3K4me3-6mA link may further contribute to chromatin compartmentalization. Our inhibitor experiments in *A. castellanii* show that H3K4me3 is not required for 6mA deposition.

In ichthyosporeans, the p1 PHD domain suggests possible H3K4me3 recognition. Thus, the hierarchical relationship between H3K4me3–6mA may have evolved independently, or their co-occurrence may simply reflect a downstream consequence of transcription rather than a direct causal link. Biophysically, 6mA is believed to decrease double-strand DNA stability, an effect potentially advantageous for transcription at the start of highly expressed genes<sup>22,27</sup>. Yet to understand the 6mA role as an epigenetic mark, identifying potential readers is key. Similarly, although 5mC stiffens DNA<sup>22</sup>, its function across eukaryotes depends more on associated readers or inhibiting transcription factor binding than on intrinsic biophysical properties<sup>81</sup>.

Intriguingly, 5mC is predominantly retained in major multicellular eukaryotic lineages (plants, animals and Dikarya fungi), whereas the AMT1–6mA pathway was lost in these lineages. In plants and animals, 5mC evolved to methylate gene bodies (Fig. 6b)<sup>4,5,53</sup>, replacing 6mA in the post-TSS region, yet this substitution is unlikely to be the sole cause of 6mA loss. The unicellular relatives of plants and animals have gene-body 5mC coexisting with 6mA, suggesting 6mA and 5mC have distinct roles even if found in the same regions (Fig. 6b). In conclusion, our data reshape the understanding of 6mA function and evolution, highlighting that both 6mA and 5mC were integral to the original eukaryotic chromatin toolkit. Although ciliates have spearheaded our understanding of 6mA in eukaryotes, their drastically unusual genome organization, with a micronucleus and a macronucleus, and a lineage-specific expansion of MT-A70s<sup>18,28</sup>, will benefit from complementary insights from alternative lineages. By expanding taxon sampling to tractable species with canonical genomes, our work sets the stage to clarify 6mA functions and to understand why this pathway was lost in major multicellular lineages.

## Online content

Any methods, additional references, Nature Portfolio 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/s41588-025-02409-6>.

## References

- Iyer, L. M., Abhiman, S. & Aravind, L. *Natural History of Eukaryotic DNA Methylation Systems*, Vol. 101 (Elsevier, 2011).
- Lyko, F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat. Rev. Genet.* **19**, 81–92 (2018).
- Huff, J. T. & Zilberman, D. Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. *Cell* **156**, 1286–1297 (2014).
- De Mendoza, A., Lister, R. & Bogdanovic, O. Evolution of DNA methylome diversity in eukaryotes. *J. Mol. Biol.* **432**, 1687–1705 (2020).
- Schmitz, R. J., Lewis, Z. A. & Goll, M. G. DNA methylation: shared and divergent features across eukaryotes. *Trends Genet.* **35**, 818–827 (2019).
- Sarkies, P. Encyclopaedia of eukaryotic DNA methylation: from patterns to mechanisms and functions. *Biochem. Soc. Trans.* **50**, 1179–1190 (2022).
- Rae, P. M. M. & Steele, R. E. Modified bases in the DNAs of unicellular eukaryotes: an examination of distributions and possible roles, with emphasis on hydroxymethyluracil in dinoflagellates. *Biosystems* **10**, 37–53 (1978).
- Borst, P. & Sabatini, R. Base J: discovery, biosynthesis, and possible functions. *Annu. Rev. Microbiol.* **62**, 235–251 (2008).
- Kozioł, M. J. et al. Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat. Struct. Mol. Biol.* **23**, 24–30 (2016).
- Greer, E. L. et al. DNA methylation on N6-adenine in *C. elegans*. *Cell* **161**, 868–878 (2015).
- Wu, T. P. et al. DNA methylation on N(6)-adenine in mammalian embryonic stem cells. *Nature* **532**, 329–333 (2016).
- Zhang, G. et al. N6-methyladenine DNA modification in *Drosophila*. *Cell* **161**, 893–906 (2015).
- Liang, Z. et al. DNA N6-adenine methylation in *Arabidopsis thaliana*. *Dev. Cell* **45**, 406–416 (2018).
- Mondo, S. J. et al. Widespread adenine N6-methylation of active genes in fungi. *Nat. Genet.* **49**, 964–968 (2017).
- Lax, C. et al. Symmetric and asymmetric DNA N6-adenine methylation regulates different biological responses in Mucorales. *Nat. Commun.* **15**, 6066 (2024).
- Fu, Y. et al. N6-methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*. *Cell* **161**, 879–892 (2015).
- Wang, Y., Chen, X., Sheng, Y., Liu, Y. & Gao, S. N6-adenine DNA methylation is associated with the linker DNA of H2A.Z-containing well-positioned nucleosomes in *Pol II*-transcribed genes in *Tetrahymena*. *Nucleic Acids Res.* **45**, 11594–11606 (2017).
- Beh, L. Y. et al. Identification of a DNA N6-adenine methyltransferase complex and its impact on chromatin organization. *Cell* **177**, 1781–1796 (2019).
- Kong, Y. et al. Critical assessment of DNA adenine methylation in eukaryotes using quantitative deconvolution. *Science* **375**, 515–522 (2022).
- O’Brown, Z. K. et al. Sources of artifact in measurements of 6mA and 4mC abundance in eukaryotic genomic DNA. *BMC Genomics* **20**, 445 (2019).
- Douvlataniotis, K., Bensberg, M., Lentini, A., Gylemo, B. & Nestor, C. E. No evidence for DNA N<sup>6</sup>-methyladenine in mammals. *Sci. Adv.* **6**, eaay3335 (2020).
- Boulias, K. & Greer, E. L. Means, mechanisms and consequences of adenine methylation in DNA. *Nat. Rev. Genet.* **23**, 411–428 (2022).
- Musheev, M. U., Baumgärtner, A., Krebs, L. & Niehrs, C. The origin of genomic N<sup>6</sup>-methyl-deoxyadenosine in mammalian cells. *Nat. Chem. Biol.* **16**, 630–634 (2020).
- Kong, Y., Mead, E. A. & Fang, G. Navigating the pitfalls of mapping DNA and RNA modifications. *Nat. Rev. Genet.* **24**, 363–381 (2023).
- Debo, B. M., Mallory, B. J. & Stergachis, A. B. Evaluation of N<sup>6</sup>-methyldeoxyadenosine antibody-based genomic profiling in eukaryotes. *Genome Res.* **33**, 427–434 (2023).
- Iyer, L. M., Zhang, D. & Aravind, L. Adenine methylation in eukaryotes: apprehending the complex evolutionary history and functional potential of an epigenetic modification. *Bioessays* **38**, 27–40 (2016).
- Bochtler, M. & Fernandes, H. DNA adenine methylation in eukaryotes: enzymatic mark or a form of DNA damage? *Bioessays* **43**, e2000243 (2021).
- Wang, Y. et al. A distinct class of eukaryotic MT-A70 methyltransferases maintain symmetric DNA N6-adenine methylation at the ApT dinucleotides as an epigenetic mark associated with transcription. *Nucleic Acids Res.* **47**, 11771–11789 (2019).
- Chen, J. et al. Structural basis for MTA1c-mediated DNA N6-adenine methylation. *Nat. Commun.* **13**, 3257 (2022).
- Liu, J. et al. A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* **10**, 93–95 (2014).
- Sheng, Y. et al. Semi-conservative transmission of DNA N<sup>6</sup>-adenine methylation in a unicellular eukaryote. *Genome Res.* **34**, 740–756 (2024).
- Luo, G.-Z. et al. N<sup>6</sup>-methyldeoxyadenosine directs nucleosome positioning in *Tetrahymena* DNA. *Genome Biol.* **19**, 200 (2018).
- Liu, Q. et al. Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data. *Nat. Commun.* **10**, 2449 (2019).

34. Hao, Z. et al. N<sup>6</sup>-deoxyadenosine methylation in mammalian mitochondrial DNA. *Mol. Cell* **78**, 382–395 (2020).
35. Burki, F., Roger, A. J., Brown, M. W. & Simpson, A. G. B. The new tree of eukaryotes. *Trends Ecol. Evol.* **35**, 43–55 (2020).
36. Thomy, J. et al. Combining nanopore and Illumina sequencing permits detailed analysis of insertion mutations and structural variations produced by PEG-mediated transformation in *Ostreococcus tauri*. *Cells* **10**, 664 (2021).
37. Liechti, N., Schürch, N., Bruggmann, R. & Wittwer, M. Nanopore sequencing improves the draft genome of the human pathogenic amoeba *Naegleria fowleri*. *Sci. Rep.* **9**, 16040 (2019).
38. Giguere, D. J. et al. Telomere-to-telomere genome assembly of *Phaeodactylum tricornutum*. *PeerJ* **10**, e13607 (2022).
39. Lenhard, B., Sandelin, A. & Carninci, P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.* **13**, 233–245 (2012).
40. Bernard, C. et al. A time-resolved multi-omics atlas of *Acanthamoeba castellanii* encystment. *Nat. Commun.* **13**, 4104 (2022).
41. 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**, e08904 (2015).
42. Olivetta, M., Bhickta, C., Chiaruttini, N., Burns, J. & Dudin, O. A multicellular developmental program in a close animal relative. *Nature* **635**, 382–389 (2024).
43. Medina, E. M. et al. Genetic transformation of *Spizellomyces punctatus*, a resource for studying chytrid biology and evolutionary cell biology. *eLife* **9**, e52741 (2020).
44. Drewell, R. A. et al. The *Dictyostelium discoideum* genome lacks significant DNA methylation and uncovers palindromic sequences as a source of false positives in bisulfite sequencing. *NAR Genom. Bioinform* **5**, lqadQ35 (2023).
45. Sarre, L. A. et al. DNA methylation enables recurrent endogenization of giant viruses in an animal relative. *Sci. Adv.* **10**, eado6406 (2024).
46. Pfeifer, G. P. Mutagenesis at methylated CpG sequences. *Curr. Top. Microbiol. Immunol.* **301**, 259–281 (2006).
47. Shen, J. C., Rideout, W. M. 3rd & Jones, P. A. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res.* **22**, 972–976 (1994).
48. Bird, A. P. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* **8**, 1499–1504 (1980).
49. Suzuki, M. M., Kerr, A. R. W., de Sousa, D. & Bird, A. CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res.* **17**, 625–631 (2007).
50. Provataris, P., Meusemann, K., Niehuis, O., Grath, S. & Misof, B. Signatures of DNA methylation across insects suggest reduced DNA methylation levels in Holometabola. *Genome Biol. Evol.* **10**, 1185–1197 (2018).
51. Bewick, A. J. et al. Diversity of cytosine methylation across the fungal tree of life. *Nat. Ecol. Evol.* **3**, 479–490 (2019).
52. Bewick, A. J. & Schmitz, R. J. Gene body DNA methylation in plants. *Curr. Opin. Plant Biol.* **36**, 103–110 (2017).
53. Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**, 916–919 (2010).
54. Grau-Bové, X. et al. A phylogenetic and proteomic reconstruction of eukaryotic chromatin evolution. *Nat. Ecol. Evol.* **6**, 1007–1023 (2022).
55. Ngan, C. Y. et al. Lineage-specific chromatin signatures reveal a regulator of lipid metabolism in microalgae. *Nat. Plants* **1**, 15107 (2015).
56. Sebé-Pedrós, A. et al. The dynamic regulatory genome of *Capsaspora* and the origin of animal multicellularity. *Cell* **165**, 1224–1237 (2016).
57. Navarrete, C., Montgomery, S. A., Mendieta, J., Lara-Astiaso, D. & Sebé-Pedrós, A. Diversity and evolution of chromatin regulatory states across eukaryotes. Preprint at *bioRxiv* <https://doi.org/10.1101/2025.03.17.643675> (2025).
58. Sanchez, R. & Zhou, M.-M. The PHD finger: a versatile epigenome reader. *Trends Biochem. Sci.* **36**, 364–372 (2011).
59. Grebien, F. et al. Pharmacological targeting of the Wdr5-MLL interaction in C/EBP $\alpha$  N-terminal leukemia. *Nat. Chem. Biol.* **11**, 571–578 (2015).
60. Zhang, X. et al. Piribedil disrupts the MLL1-WDR5 interaction and sensitizes MLL-rearranged acute myeloid leukemia (AML) to doxorubicin-induced apoptosis. *Cancer Lett.* **431**, 150–160 (2018).
61. Al Jewari, C. & Baldauf, S. L. An excavate root for the eukaryote tree of life. *Sci. Adv.* **9**, eade4973 (2023).
62. Cerón-Romero, M. A., Fonseca, M. M., de Oliveira Martins, L., Posada, D. & Katz, L. A. Phylogenomic analyses of 2,786 genes in 158 lineages support a root of the eukaryotic tree of life between opisthokonts and all other lineages. *Genome Biol. Evol.* **14**, evac119 (2022).
63. Williamson, K. et al. A robustly rooted tree of eukaryotes reveals their excavate ancestry. *Nature* **640**, 974–981 (2025).
64. Jurkowski, T. P. & Jeltsch, A. On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. *PLoS ONE* **6**, e28104 (2011).
65. Ponger, L. & Li, W.-H. Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Mol. Biol. Evol.* **22**, 1119–1128 (2005).
66. Tourancheau, A., Mead, E. A., Zhang, X.-S. & Fang, G. Discovering multiple types of DNA methylation from bacteria and microbiome using nanopore sequencing. *Nat. Methods* **18**, 491–498 (2021).
67. Lizarraga, A. et al. Adenine DNA methylation, 3D genome organization and gene expression in the parasite *Trichomonas vaginalis*. *Proc. Natl Acad. Sci. USA* **117**, 13033–13043 (2020).
68. Pritham, E. J., Putliwala, T. & Feschotte, C. Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene* **390**, 3–17 (2007).
69. Lentini, A. et al. A reassessment of DNA-immunoprecipitation-based genomic profiling. *Nat. Methods* **15**, 499–504 (2018).
70. Ma, C. et al. A novel N<sup>6</sup>-deoxyadenine methyltransferase METL-9 modulates *C. elegans* immunity via dichotomous mechanisms. *Cell Res.* **33**, 628–639 (2023).
71. Zhang, J. et al. 6mA-Sniper: quantifying 6mA sites in eukaryotes at single-nucleotide resolution. *Sci. Adv.* **9**, eadh7912 (2023).
72. Abakir, A. et al. N<sup>6</sup>-methyladenosine regulates the stability of RNA:DNA hybrids in human cells. *Nat. Genet.* **52**, 48–55 (2020).
73. Febrimarsa et al. Randomly incorporated genomic N<sup>6</sup>-methyldeoxyadenosine delays zygotic transcription initiation in a cnidarian. *EMBO J.* **42**, e112934 (2023).
74. Duan, L. et al. Methyl-dependent auto-regulation of the DNA N<sup>6</sup>-adenine methyltransferase AMT1 in the unicellular eukaryote *Tetrahymena thermophila*. *Nucleic Acids Res.* **53**, gkaf022 (2025).
75. Zilberman, D. An evolutionary case for functional gene body methylation in plants and animals. *Genome Biol.* **18**, 87 (2017).
76. Dixon, G. & Matz, M. Changes in gene body methylation do not correlate with changes in gene expression in Anthozoa or Hexapoda. *BMC Genomics* **23**, 234 (2022).
77. Howe, F. S., Fischl, H., Murray, S. C. & Mellor, J. Is H3K4me3 instructive for transcription activation?. *Bioessays* **39**, 1–12 (2017).
78. Wang, H. et al. H3K4me3 regulates RNA polymerase II promoter-proximal pause-release. *Nature* **615**, 339–348 (2023).

79. De Mendoza, A. et al. Convergent evolution of a vertebrate-like methylome in a marine sponge. *Nat. Ecol. Evol.* **3**, 1464–1473 (2019).
80. Schwaiger, M. et al. Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Res.* **24**, 639–650 (2014).
81. Kaluscha, S. et al. Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. *Nat. Genet.* **54**, 1895–1906 (2022).
82. Torruella, G. et al. Phylogenomics reveals convergent evolution of lifestyles in close relatives of animals and fungi. *Curr. Biol.* **25**, 2404–2410 (2015).

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

## Methods

Experiments within this study did not require ethics board approval.

### Sequence search and phylogenetic analysis

A collection of eukaryotic proteomes spanning the broadest diversity of lineages was scanned using HMMER3 (v3.3.1) with the PFAM domain for MT-A70 (ref. 83), using an *e* value of 0.0001 as a threshold. Additionally, AMT6/7 sequence from *Tetrahymena thermophila* was used as a query against the same database using BLASTP (2.9.0+), to obtain divergent orthologs that were filtered out with the hmsearch approach. In parallel, *T. thermophila* MT-A70 AMT1 and AMT7 sequences were used to BLASTP against NCBI 'nr' databases, excluding eukaryotes as taxonomic hits. The resulting proteins were merged into a multisequence alignment using MAFFT L-INS-i mode<sup>84</sup> (v7.475). The alignment was trimmed with trimAl -gappyout parameter<sup>85</sup> (v1.4.rev15), and a maximum-likelihood phylogeny was obtained using IQ-TREE<sup>86</sup> (v2.1.2), with automatic model selection, computing 1,000 ultrafast bootstraps and ALRT replicates as nodal supports. The distribution of MT-A70 hits across our dataset can be found in Supplementary Table 1. To determine the distribution of MT-A70 in prokaryotes, the same BLASTP approach was used, allowing 250 hits against the NCBI 'refseq\_select' database, specifying bacteria and archaea (separately). The 302 Asgardarchaeota genomes were downloaded from NCBI (January 2025) and searched using HMMER3, as for eukaryotes. Another phylogeny was constructed from a subset of the original tree, selecting a few representative eukaryotic species and all prokaryotic outgroups, using the same methods as for the larger tree. To determine the presence of p1 and p2 orthologs, we used BLASTP against the custom eukaryotic proteome database, requiring an *e* value of <0.0001.

### Protist cultures

*C. fragrantissima*, *Abeoforma whisleri* and *C. limacisporum* cells were grown axenically in marine broth liquid medium (Difco, 2216) at 17 °C, 17 °C and 23 °C, respectively. *C. perkinsii* cells were grown axenically at 23 °C in liquid medium (containing 3 g of yeast extract, 3 g of malt extract, 5 g of peptone, 10 g of glucose and 20 g on NaCl per liter of distilled water). *A. castellanii* Neff cells were grown axenically at 23 °C in ATCC medium 712. *N. gruberi* cells were grown axenically at 23 °C and 30 °C in ATCC medium 1034. *A. limacinum* cells were grown axenically at 19 °C in ATCC medium 790. *S. punctatus* cells were grown axenically at 17 °C in liquid medium (containing 2.5 g of yeast extract, 0.5 g of K<sub>2</sub>HPO<sub>4</sub>, 2.5 ml of ethanol, 15 ml of glycerol and 485 ml of Milli-Q water). *D. discoideum* was grown axenically at 23 °C in liquid HL5-C medium (Formedium). *C. reinhardtii* CC-503 was grown on TAP growth media (Thermo Fisher Scientific, A1379801) at room temperature.

We performed DNA extractions for *N. gruberi*, *D. discoideum*, *C. reinhardtii*, *C. perkinsii*, *A. whisleri* and *S. punctatus* from confluent cultures grown into 25-ml flasks for 7 days, and 4 days for *C. limacisporum*. For *A. limacinum*, cells were grown for 7 days and then 1 ml was passed into two 25-ml flasks with fresh medium to enrich for zoospores and then collected 2 days later. In all cases, cells were centrifuged at 3,000g for 5 min, and the supernatant was discarded before DNA extraction.

### Stage-specific sampling

For *A. castellanii* trophozoite DNA, we obtained cells from a confluent 7-day culture. For the cystic stage, cells were grown for 5 days in 25-ml flasks and then media was removed and replaced by 5 ml of encystment medium (containing 3.728 g of KCl, 1.68-g NaHCO<sub>3</sub>, 0.986 g of MgSO<sub>4</sub> × 7H<sub>2</sub>O, 0.03 g of CaCl<sub>2</sub> × 2H<sub>2</sub>O and 0.017 g of 2-amino-2-methyl-1,3-propanediol per 500 ml of distilled water)<sup>87</sup>; after 3 days of incubation at 23 °C, *A. castellanii* cysts were collected for DNA extraction.

*C. fragrantissima* cells were grown until confluency for 5 days in 25-ml flask, and then cells were scratched and passed into 250-ml flask with 25 ml of fresh medium. These new flasks were grown for 48 h under gentle agitation at 17 °C and then were filtered using a 20-µm

cell strainer (pluriSelect) and collected into a 50-ml Falcon tube to separate the amebas from the mature coenocytes, as described previously<sup>41</sup>. Both amebas and coenocytes were collected for independent DNA extraction.

For *C. perkinsii* synchronization, we filtered a 6-day-old culture using a 5-µm filter (puriSelect) to obtain a small-celled population that was then diluted 1:100 in liquid media to grow at 23 °C, as described before<sup>42</sup>. Cells were visually inspected and collected after 54 h and 96 h, and then snap-frozen in liquid N<sub>2</sub> for later DNA extractions.

For *S. punctatus*, zoospore cells were grown until confluency for 7 days in 25-ml flask, and then 1 ml of the culture was spread onto *Spicellomyces*-agar plates (*S. punctatus* liquid medium + 15 g l<sup>-1</sup> of agar) and incubated for 48 h at 30 °C. After that, the active *Spizellomyces* plates were flooded with 1 ml of dilute salt (DS) solution and incubated at room temperature for 1 h to induce the release of zoospores. After the release, the DS solution containing the zoospores was retrieved and filtered using a 10-µm syringe strainer (pluriSelect) as described previously<sup>43</sup>. At that moment, zoospores were collected and frozen in liquid N<sub>2</sub> for later DNA and RNA extractions. For the multinucleated colonies, zoospores obtained as described before grown in liquid media for 24 h at 30 °C, multinucleated cells were collected by scraping the flask and snap-frozen in liquid N<sub>2</sub> for later DNA and RNA extractions.

### Nucleic acid isolation

Before DNA extraction, to break the cell wall of *S. punctatus*, *C. fragrantissima* and synchronized *C. perkinsii* samples, heat shock was applied, the cells were frozen with liquid N<sub>2</sub>, and then thawed at 60 °C, repeating this thrice. For *C. perkinsii* unsynchronized samples and *A. castellanii* cysts, we immersed the pellets in liquid N<sub>2</sub> and then ground them with a pestle and mortar.

The resulting cell pellets for *N. gruberi*, *A. castellanii* amebas, cysts and Piribedil-treated samples, *D. discoideum*, *C. perkinsii* unsynchronized samples, *C. limacisporum* and *A. limacinum* were used for DNA extraction using the Qiagen MagAttract HMW DNA kit following the manufacturer's whole blood protocol. For *S. punctatus*, *C. fragrantissima*, *C. perkinsii* synchronized samples, *A. castellanii* OICR-9429 treated samples and *C. reinhardtii* NEB Monarch Genomic DNA Purification Kit were used following the animal tissue protocol. For *A. whisleri*, we used a phenol/chloroform genomic DNA extraction method.

Total RNA was extracted in biological duplicates from *N. gruberi* cultures grown at 23 °C and 30 °C, as well as from *S. punctatus* zoospores and multinucleate colonies, following NEB Monarch Total RNA Miniprep Kit tough-to-lyse samples protocol after snap-freezing the samples in liquid N<sub>2</sub>.

The genomic DNA of *T. vaginalis* strain G3 was obtained from ATCC, and the DNA of *M. pusilla* Culture Collection of Algae and Protozoa (CCAP) 1965/4, *Chlorella variabilis* NC64A and *Cyanophora paradoxa* CCAP 981/1 were obtained from CCAP (Oban).

### Nanopore sequencing, base calling and methylation analysis

We quantified genomic DNA using a Qubit 3 Fluorometer with the dsDNA BR Assay Kit and assessed DNA size with a TapeStation 2200 using the Genomic DNA ScreenTape Assay. We then started with 1–1.5 µg of high molecular weight genomic DNA that was ligated with nanopore SQK-LSK110 ligation kit following the manufacturer's instructions and sequenced later in MinION R9.4.1 flow cells (see Supplementary Table 3 for details).

The resulting fast5 files were then used as input for Guppy (v6.5.7), and the Rerio modified base-calling model 'res\_dna\_r941\_min\_modbases-all-context\_v001' was specified, aligning the reads to the reference genome (-align\_ref) parameter. The list of reference genomes is provided in Supplementary Table 4 (refs. 41,45,88–91). The resulting BAM files were then merged and sorted using samtools, and then the Oxford Nanopore 'modbam2bed' (v0.9.1) was used to obtain either 6mA (-m 6mA) or 5mC (-m 5mC) basecalls. The resulting

bedMethyl files were then used to extract the neighboring bases for each A position, gathering one base upstream and two downstreams using BEDTools<sup>92</sup>, and the stranded (5' to 3') four-nucleotide context was included for each position. Dinucleotide contexts were divided into AT, AA, AC and AG, and bigwig files were generated using the UCSC bedGraphToBigWig tool (v4). Visualization of average methylation levels on genes, TSS and TEs was obtained using DeepTools2 (v3.5.0)<sup>93</sup>.

For *N. gruberi* temperature-treated DNA samples, we generated libraries using the SQK-LSK114 ligation kit and sequenced them using the PromethION R10.4.1 flow cells. Because there is no compatible R10.4.1 model for Guppy, we switched to Dorado (v0.7.2) for 6mA base calling, using 6mA in sup mode. The 6mA calls were obtained using the Oxford Nanopore Modkit (v0.4.2) pileup function, testing several thresholds for calling a modified base, inspecting them in the browser and finding that 0.995 showed the lowest false positive rate when compared to R9.4.1 for *N. gruberi* (-motif AN O --mod-thresholds a:0.995).

Global methylation levels were computed in R, and the regional average levels of 6mA were computed using the Bioconductor bsseq package<sup>94</sup>, treating the ApT dinucleotides as if they were CpGs. To call an ApT site methylated, we required at least 10× coverage and >10% 6mA, with lower coverage thresholds in *M. pusilla* and *A. whisleri* (5×) and in *C. limacisporum* and *C. variabilis* (3×). To call methylated genes, we required that each gene contain at least three methylated ApT sites. To compute stage-specific methylation differences in gene bodies, we computed the weighted average of 6mApT methylation across the whole gene bodies and subtracted those values. Genes with a difference of >1.5% 6mApT in both directions were selected as the most highly distinct.

### Selection analysis

To assess selective pressure acting on AMT1 gene sequences from the species sequenced in this study, a maximum-likelihood phylogenetic tree was inferred using IQ-TREE, as described in the phylogenetic methods section. A codon-based alignment was generated by combining the protein sequence alignment obtained with MAFFT with the corresponding nucleotide sequences using PAL2NAL (v14). Tests for positive selection were performed using the CodeML implemented in EasyCodeML software (v1.41)<sup>95</sup>, applying the preset Site Model to detect the variation in selective pressure across codon sites.

### RNA-seq preparation and analysis

We used 0.8–1 µg of total RNA from *N. gruberi* cultures grown at 23 °C and 30 °C and *S. punctatus* cell stages to select poly-A tails with the NEBNext Poly(A) mRNA Magnetic Isolation Module kit; then libraries were prepared using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina in duplicate, following the manufacturer's instructions. The resulting libraries were sequenced on a NovaSeq 2000 (NovoGene), yielding 20–30 million paired-end reads per replicate.

Furthermore, publicly available RNA-seq datasets for all species were downloaded from ENA from various studies (Supplementary Table 5)<sup>40–42</sup>. The data were mapped to the reference genomes using HISAT2 (v2.2.1)<sup>96</sup>, restricting intron size to 40,000 bp. To calculate steady-state transcription levels across all species, StringTie was used to obtain the Transcripts per Million measure for all gene models, and bigwig files were generated using DeepTools2, specifying the stranded information when the original libraries had strand-specific information<sup>97</sup>. For the differential expression analysis, Kallisto (0.46) was used to obtain the counts (est\_counts) by aligning reads to reference transcriptomes<sup>98</sup>. The estimated counts were read into R and DESeq2 was used to compute the differential expression, using a *P*-adjusted value of <0.01 as a significant threshold.

### ChIP-seq library preparation and analysis

ChIP-seq was performed as previously described<sup>99</sup> with modifications. Briefly, 100-ng chromatin per species was pooled per ChIP.

Pooled chromatin was incubated at 4 °C for 12–14 h with rotation with 2.5-µg anti-H3K27ac (Abcam, ab4729) or 2.5-µg anti-H3K4me3 (Sigma-Aldrich, 07-473). Immunoprecipitated complexes were captured using a mix of Protein A (Sigma-Aldrich, 16-661) and Protein G magnetic beads (Sigma-Aldrich, 16-662), washed and reverse crosslinked for 30 min at 55 °C followed by an hour at 68 °C. Immunoprecipitated DNA was purified using solid-phase reversible immobilization beads (Beckman Coulter, A63881). The NEBNext Ultra II DNA Library Prep Kit (New England BioLabs) was used to produce the resulting ChIP libraries according to the manufacturer's protocol.

The data for *C. reinhardtii* and *T. vaginalis* were downloaded from ENA, belonging to previous studies<sup>55,100</sup> (Supplementary Table 6, SRR numbers).

All ChIP-seq data were analyzed using fastp to trim reads, and the reads were mapped to the genomes using bowtie2 (v1.2), allowing a maximum insert size of 2,000 bp (-I 2000). Duplicate reads were removed using Sambamba (v0.6.6). DeepTools2 was used to generate bigwig files, compute insert size distributions and visualize epigenomic data, as well as the Integrative Genome Viewer. MACS3 (v3.0.0a7) was used to call peaks<sup>101</sup>, using a *q* value of 0.01 and adjusting the genome size to each species (for example, -q 0.01 -g 4e7), as well as activating the broad parameter for broad peak calling. Broad peaks were defined as those having >1,200 bp using their width distribution as a reference (Extended Data Fig. 9c). The H3K4me3 signal was clustered using DeepTools2 plotHeatmap function, allowing four *k*-means clusters (-k-means 4), providing the four categories shown in Fig. 5.

### Inhibitors and western blot

Histone H3K4me3 suppression drugs OICR-9424 (MedChemExpress, HY-16993) and Piribedil (Cayman Chemical, 3605-01-4) were diluted in DMSO at 50 mM and 100 mM, respectively, as stock solutions. For drug treatment experiments, 2.1 million *A. castellanii* cells were cultured in 42 ml of liquid media at 23 °C for 24 h. After incubation, the cells were treated for 30 h with either 60-µM OICR-9424 or 300-µM OICR-9424, or with 5-µM Piribedil or 250-µM Piribedil. Two control samples were prepared by treating cells with 0.6% DMSO (for OICR-9424) and 0.2% DMSO (for Piribedil), adjusting the DMSO concentration to the highest inhibitor concentrations. Cells were collected by scraping the flasks and centrifuged at 3,000g for 5 min, the supernatant was discarded and the pellets were snap-frozen in liquid N<sub>2</sub> for subsequent histone and DNA extraction.

Histones were extracted using an acid extraction protocol optimized for western blot analysis. Briefly, treated or control *A. castellanii* pellets were washed twice with ice-cold PBS. Cells were resuspended in Triton Extraction Buffer (TEB; PBS containing 0.5% (vol/vol) Triton X-100, 5 mM sodium butyrate and one Complete Mini EDTA-free protease inhibitor tablet per 10 ml) at a density of 3 × 10<sup>7</sup> cells per ml. Cell lysis was performed on ice for 10 min with gentle agitation. Nuclei were pelleted by centrifugation at 650g for 10 min at 4 °C. The nuclei were washed once with half the volume of TEB and centrifuged again under the same conditions. The resulting nuclear pellet was resuspended in 0.2 N HCl at a density of 12 × 10<sup>7</sup> nuclei per ml and incubated overnight at 4 °C to acid-extract histones. Samples were then centrifuged at 650g for 10 min at 4 °C to remove debris. The supernatant containing soluble histones was collected and neutralized with 2 M NaOH at one-tenth of the volume of the supernatant. Histone extracts were aliquoted and stored at -20 °C until further use.

Protein samples (5 µg for H3 and 12 µg for H3K4me3 detection) were mixed with 2× Laemmli sample buffer containing 50-mM dithiothreitol and denatured by boiling at 100 °C for 10 min. Proteins were separated by SDS-PAGE on 15% acrylamide gels and transferred to nitrocellulose membranes. Membranes were blocked for 1 h at room temperature in TBST buffer (50 mM Tris-HCl, pH 7.5; 150 mM NaCl; 0.05% (vol/vol) Tween-20) containing 5% (wt/vol) nonfat dry milk.

Primary antibodies were incubated for 2 h at room temperature in TBST containing 1% (wt/vol) nonfat dry milk. After washing, membranes were incubated with horseradish peroxidase-conjugated secondary antibodies diluted 1:10,000 in TBST with 1% (wt/vol) nonfat dry milk.

Signal of the secondary antibody was detected with a FUSION Solo S (Vilber Lourmat).

The primary antibodies used were antihistone H3 (1:1,000; Abcam, ab1791), histone H3K4me3 (1:500; Active Motif, AB\_2615077) and the secondary antibody goat antirabbit IgG (H + L; Invitrogen, 31462).

### Genome assembly and re-annotation

The genomes of *N. gruberi* and *C. perkinsii* were re-assembled for this study. We used Oxford Nanopore reads and performed conventional base calling with Guppy using the 'sup' model and splitting chimeric reads. The resulting reads were then used as input for Flye (v2.9-b1768) using `-nanopore_hq` option, and allowing two steps of polishing<sup>102</sup>. Publicly available Illumina reads<sup>90</sup> were used for polishing using HyPo<sup>103</sup> (v1.0.3). For *N. gruberi*, the resulting contigs were scaffolded using *Naegleria lovaniensis* chromosome-scale genome<sup>104</sup> with RagTag<sup>105</sup> (v2.1.0). Publicly available RNA-seq for each species<sup>90,106</sup> was aligned to the genomes using HISAT2 with the `-dta` parameter with a maximum intron size of 10 kb, and Stringtie (v2.1.2) was then used to predict transcript models<sup>96,97</sup>. In parallel, Trinity (v2.8.5) was used to build de novo transcriptome assemblies<sup>107</sup>, which were mapped to the genome using GMAP (14 October 2020). The best transcripts were selected using Mikado<sup>108</sup> (v2.3.0), which were then used to generate exon hints for Augustus<sup>109</sup> (v3.4.0). LiftOff<sup>110</sup> (v1.6.3) was used to transfer annotations from the previous genome assemblies to the new assemblies<sup>90,111</sup>, which were then used as coding sequence hints. Then Augustus was trained on the best Mikado gene models for each species, and those species models were used to predict the genes with both exon and coding-sequence hints. In the last step, transcript UTRs were added to the Augustus gene models using PASA (v2.0.1) with the best Mikado transcripts as input<sup>112</sup>.

For *C. limacisporum*, *C. variabilis*, *C. fragrantissima* and *A. whisleri*, the annotations were also updated with available RNA-seq data to include UTRs and improve TSS annotation<sup>41,82,113</sup>. We used a combination of Trinity and Stringtie as described above to generate a non-redundant transcript set with Mikado, which was then used to update gene structures with PASA.

Repeat annotations were generated for species that lacked them, using RepeatModeler2 (v2.0.2a)<sup>114</sup>, including the LTR finding module, and mapping the consensus repeat library against the genome using RepeatMasker (v4.1.2-p1).

### Statistics and reproducibility

No statistical method was used to predetermine sample size. Bulk RNA-seq experiments were performed in two replicates for *N. gruberi* and *S. punctatus*, while all available replicates from public datasets were used for other samples. ChIP-seq experiments were performed with at least two replicates, and Oxford Nanopore sequencing was performed with one replicate per stage or species. Drug treatment experiments and western blot were performed twice with similar results. No data were excluded from the analyses. The experiments were not randomized. The investigators were not blinded to allocation during the experiments or during outcome assessment.

### Reporting summary

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

### Data availability

Nanopore raw sequencing data have been uploaded to BioStudies (<https://www.ebi.ac.uk/biostudies/>) under the accession S-BSST1-363 (ref. 115), and the ChIP-seq data can be found in the GEO submission GSE261870. The ChIP-seq for *C. reinhardtii* was obtained from

GSE59629, and *T. vaginalis* from GSE89662. Publicly available RNA-seq was obtained from PRJNA794325 (*A. castellanii*), PRJNA360056 (*A. whisleri*), SAMD00394225 (*A. limacinum*), PRJNA1091032 (*C. perkinsii*), PRJNA285347 (*C. reinhardtii*), PRJNA849385 (*T. vaginalis*), PRJNA262632 (*C. limacisporum*), PRJNA210187 (*C. variabilis*), GSE46692 (*M. pusilla*), GSE155535 (*O. tauri*), PRJNA642022 (*N. fowleri*), GSE249241 (*A. appalachense*) and GSE68616 (*C. fragrantissima*). Annotation and other analysis files associated with this article can be found at: <https://doi.org/10.5281/zenodo.17174913> (ref. 116). IGV genome browser session can be accessed for *C. fragrantissima* (<https://tinyurl.com/22zs884x>), *S. punctatus* (<https://tinyurl.com/2d7jgzpz>), *T. vaginalis* (<https://tinyurl.com/2ar17o56>), *A. castellanii* (<https://tinyurl.com/22wj72j>), *N. gruberi* (<https://tinyurl.com/mr44r47h>) and *C. reinhardtii* (<https://tinyurl.com/25aww7qk>). Source data are provided with this paper.

### Code availability

All the custom code used for the analysis of this paper can be found on GitHub at [https://github.com/AlexdeMendoza/6mA\\_evolution](https://github.com/AlexdeMendoza/6mA_evolution) and Zenodo (ref. 116).

### References

- Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
- Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
- Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
- Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
- Prescott, D. M. (ed.). *Methods in Cell Biology* Vol. 1, 55–83 (Academic, 1964).
- Collier, J. L. et al. The protist *Aurantiochytrium* has universal subtelomeric rDNAs and is a host for mirusviruses. *Curr. Biol.* **33**, 5199–5207 (2023).
- Matthey-Doret, C. et al. Chromosome-scale assemblies of *Acanthamoeba castellanii* genomes provide insights into *Legionella pneumophila* infection-related chromatin reorganization. *Genome Res.* **32**, 1698–1710 (2022).
- Grau-Bové, X. et al. Dynamics of genomic innovation in the unicellular ancestry of animals. *eLife* **6**, e26036 (2017).
- Price, D. C. et al. Analysis of an improved *Cyanophora paradoxa* genome assembly. *DNA Res.* **26**, 287–299 (2019).
- Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
- Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
- Hansen, K. D., Langmead, B. & Irizarry, R. A. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol.* **13**, R83 (2012).
- Gao, F. et al. EasyCodeML: a visual tool for analysis of selection using CodeML. *Ecol. Evol.* **9**, 3891–3898 (2019).
- Sirén, J., Välimäki, N. & Mäkinen, V. HISAT2—fast and sensitive alignment against general human population. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **11**, 375–388 (2014).
- Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
- Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

99. Lara-Astiaso, D. et al. Immunogenetics. Chromatin state dynamics during blood formation. *Science* **345**, 943–949 (2014).
100. Song, M.-J. et al. Epigenome mapping highlights chromatin-mediated gene regulation in the protozoan parasite *Trichomonas vaginalis*. *Sci. Rep.* **7**, 45365 (2017).
101. Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
102. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
103. Kundu, R., Casey, J. & Sung, W.-K. HyPo: super fast & accurate polisher for long read genome assemblies. Preprint at *bioRxiv* <https://doi.org/10.1101/2019.12.19.882506> (2019).
104. Joseph, S. J. et al. Comparative genomic and transcriptomic analysis of *Naegleria fowleri* clinical and environmental isolates. *mSphere* **6**, e0063721 (2021).
105. Alonge, M. et al. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* **23**, 258 (2022).
106. Grau-Bové, X., Ruiz-Trillo, I. & Irimia, M. Origin of exon skipping-rich transcriptomes in animals driven by evolution of gene architecture. *Genome Biol.* **19**, 135 (2018).
107. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
108. Venturini, L., Caim, S., Kaithakottil, G. G., Mapleson, D. L. & Swarbreck, D. Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *Gigascience* **7**, giy093 (2018).
109. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).
110. Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639–1643 (2021).
111. Fritz-Laylin, L. K. et al. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell* **140**, 631–642 (2010).
112. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
113. Blanc, G. et al. Deep RNA sequencing reveals hidden features and dynamics of early gene transcription in *Paramecium bursaria* chlorella virus 1. *PLoS ONE* **9**, e90989 (2014).
114. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA* **117**, 9451–9457 (2020).
115. De Mendoza, A., Romero Charria, P., Shabardina, V. & Casacuberta, E. Adenine DNA methylation associated with transcriptionally permissive chromatin is widespread across eukaryotes. *BioStudies* <https://doi.org/10.6019/S-BSS1363> (2025).
116. De Mendoza, A. & Romero Charria, P. AlexdeMendoza/6mA\_Evolution: 6mA\_Evolution\_publicationRelease (v1.0). *Zenodo* <https://doi.org/10.5281/zenodo.17174913> (2025).

## Acknowledgements

We would like to thank I. Ruiz-Trillo and M. Antó from the Institute of Evolutionary Biology in Barcelona for sharing some of the initial cultures used in this study; P. Hurd, Z. O'Neill and N. Forrester at QMUL for helping with histone extraction and western blot and B. Mahmoud for helping with cultures and DNA extractions. The authors would also like to thank O. Dudin and O. Bogdanovic for their critical review of the manuscript. This work used computing resources from Queen Mary University of London's Apocrita HPC facilities. This work was funded by the Horizon 2020 Framework Programme (European Research Council Starting Grant action 950230 to A.d.M.). P.R.C. was funded by a QMUL PhD fellowship. C.N. was funded by FPI PhD fellowship from the Spanish Ministry of Science and Innovation, the research in ASP group was supported by the European Research Council (ERC-StG, 851647) and the Spanish Ministry of Science and Innovation (grant PID2021-124757NB-I00), as well as the Spanish Ministry's support to the EMBL partnership, the Centro de Excelencia Severo Ochoa and the CERCA Programme (Generalitat de Catalunya). V.S. and E.C. were supported by grant PID2020-120609GB-I00 from MCIN/AEI/10.13039/501100011033 and grant 2021SGR00751 from the Departament de Recerca i Universitats de la Generalitat de Catalunya.

## Author contributions

A.d.M. conceived and designed the study. P.R.C. obtained nanopore sequencing for all but two species. P.R.C., V.O., L.A.S. and A.d.M. performed bioinformatic analysis. P.R.C., L.X. and E.K. performed protein extraction experiments. C.N., D.L.-A. and A.S.-P. performed the ChIP-seq experiments. V.S. and E.C. provided nanopore data for *A. whisleri* and shared cultures. A.d.M. and P.R.C. drafted the manuscript, and all authors critically read and commented on the manuscript.

## Competing interests

All authors declare no competing interests.

## Additional information

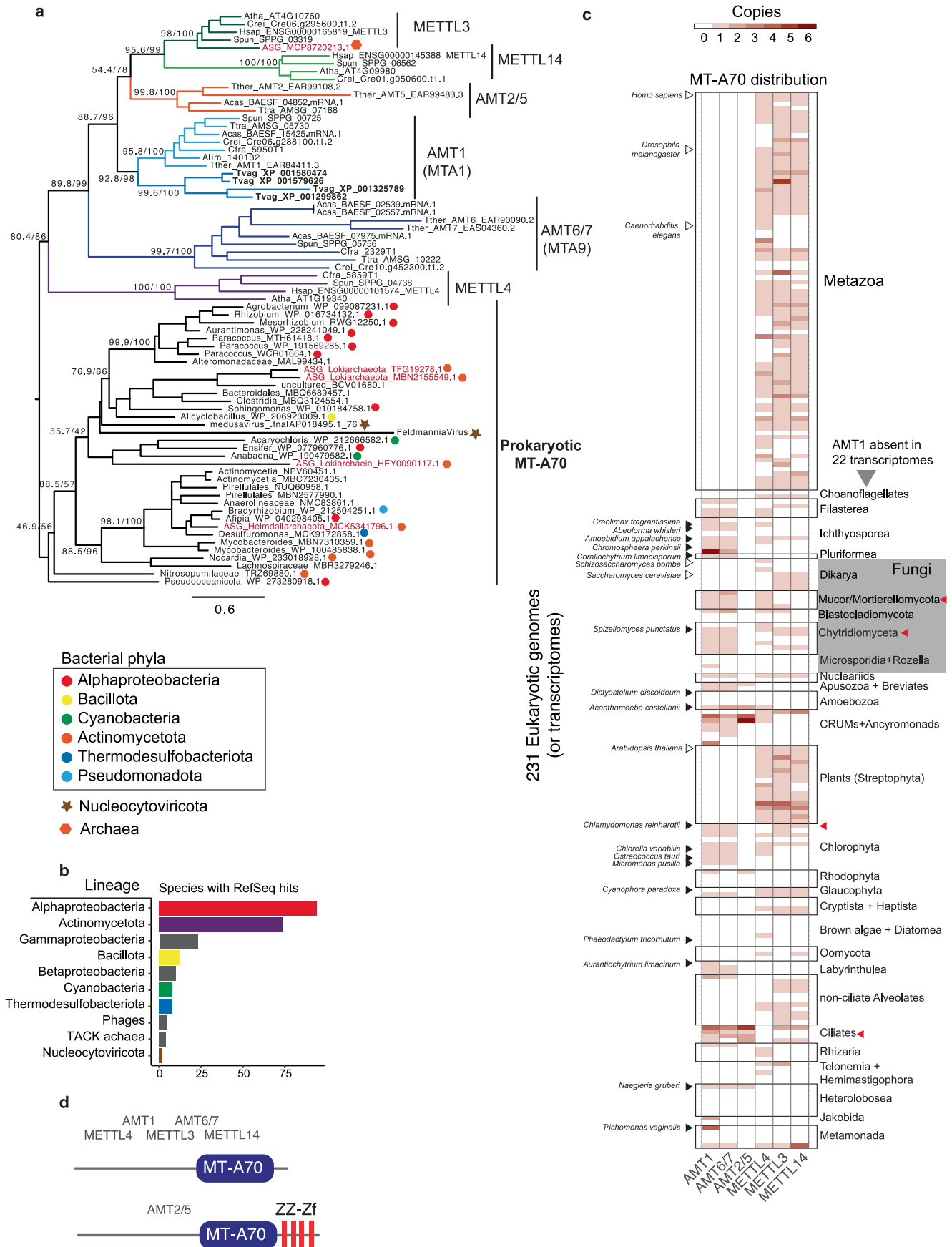
**Extended data** is available for this paper at <https://doi.org/10.1038/s41588-025-02409-6>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41588-025-02409-6>.

**Correspondence and requests for materials** should be addressed to Alex de Mendoza.

**Peer review information** *Nature Genetics* thanks Stephen Mondo and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

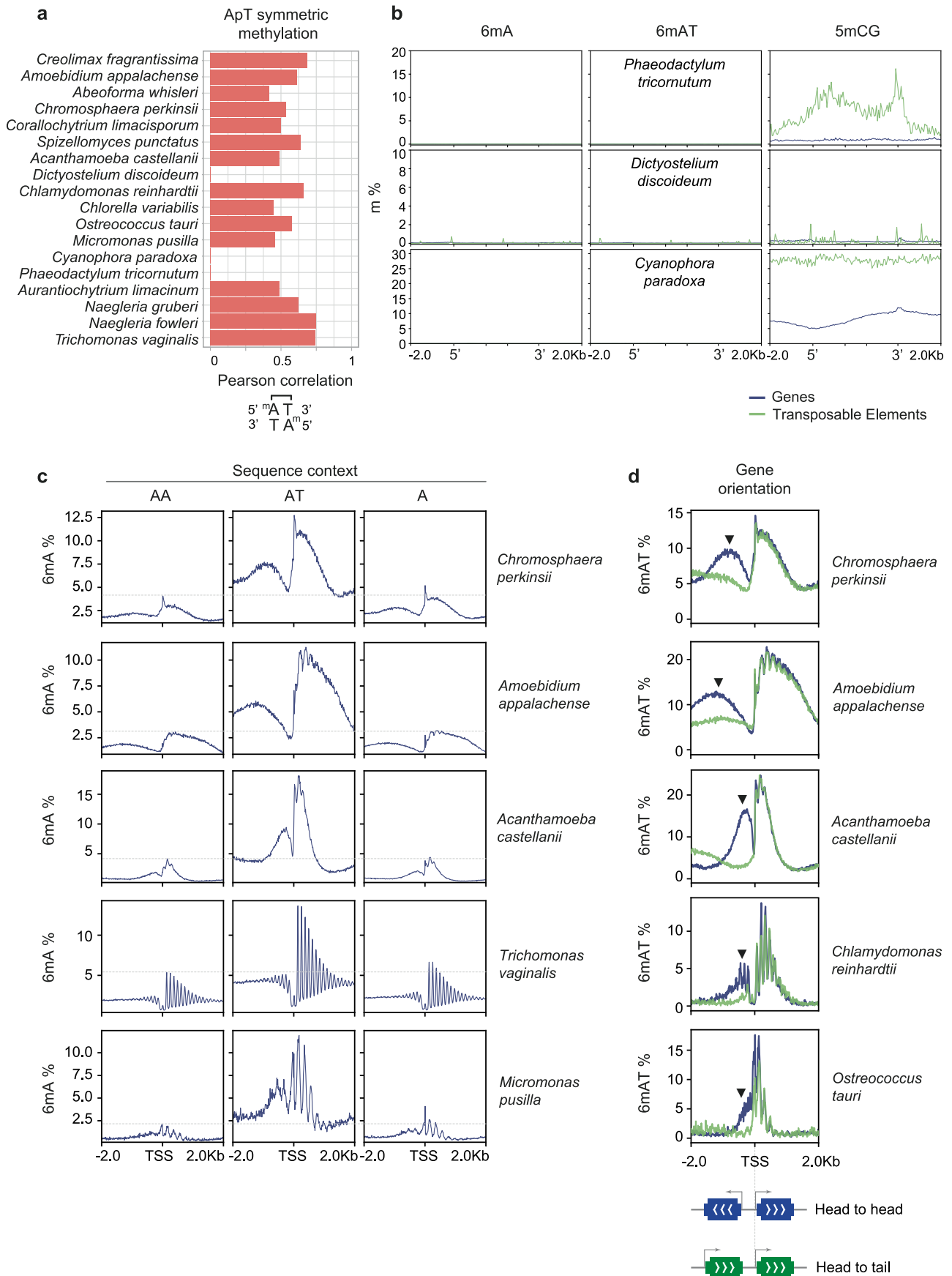
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



Extended Data Fig. 1 | See next page for caption.

**Extended Data Fig. 1 | Bacterial outgroups, structure and distribution of MT-A70. a**, Maximum likelihood phylogenetic tree with selected species from Fig. 1b highlighting the prokaryotic outgroups, color coded by bacterial lineage (for those bacteria with genus/species defined). *Trichomonas vaginalis* AMT1 sequences are highlighted in bold, whereas Asgardarchaeota are highlighted in red. Nodal supports represent SH-aLRT and ultrafast bootstrap values. **b**, Taxonomy count of MT-A70 BlastP searches against NCBI RefSeq\_select database (January 2024). **c**, Heatmap showing the number of MT-A70 family

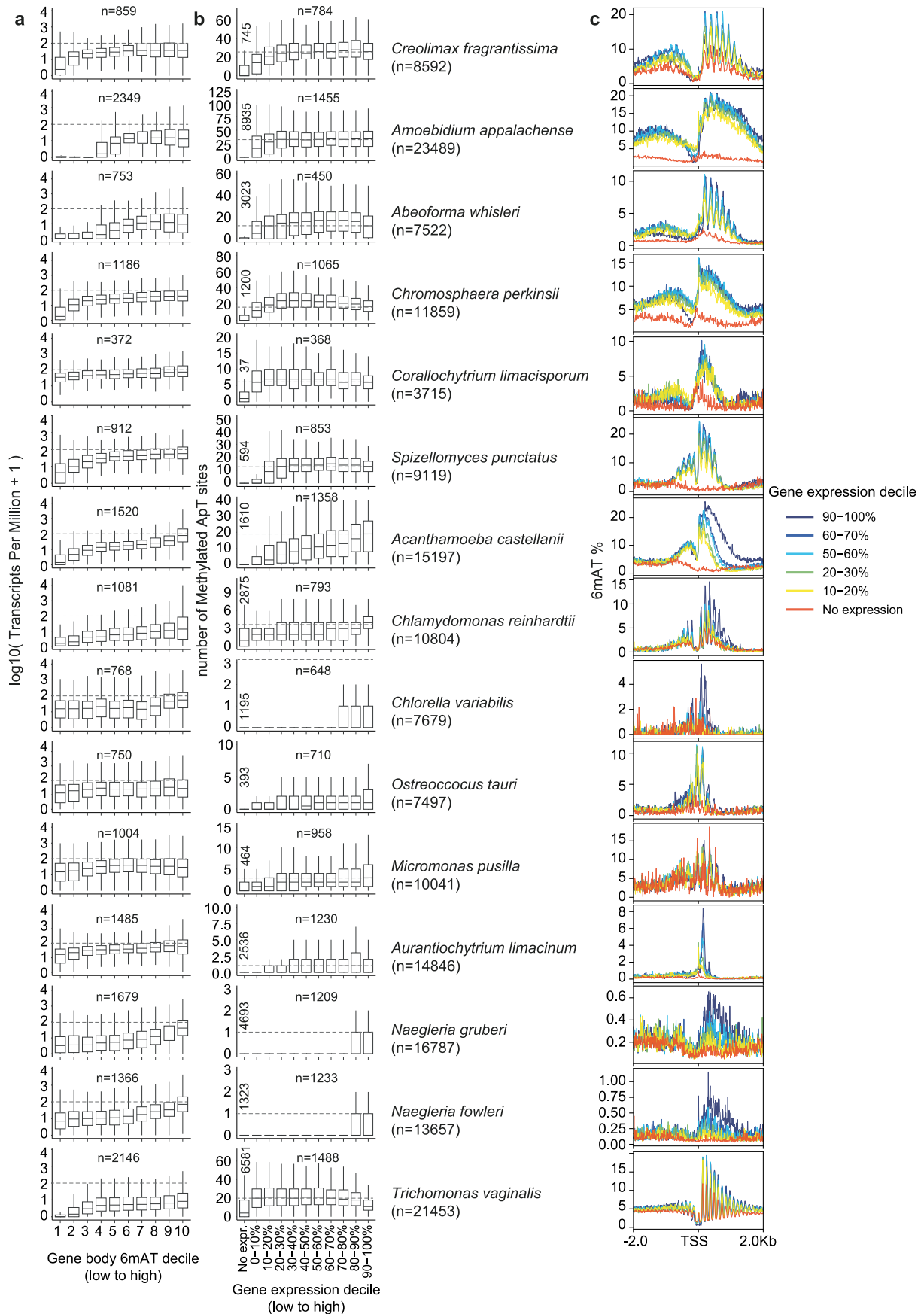
members across a collection of 231 eukaryotic species covering a wide scope of known diversity. Classification obtained from a phylogenetic tree in Fig. 1b. Black triangles are the species sequenced in this study, and red arrows indicate lineages/species for which 6mA has been previously confirmed at robust high levels, white triangles indicate species for which 6mA is contested. **d**, Domain architectures of eukaryotic MT-A70 families, defined with Pfam domains MT-A70 (PF05063) and ZZ (PF00569).



Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2 | Patterns of 6mA distribution in eukaryotes. a,** Global correlation of 6mA levels in ApT dinucleotides. Pearson's correlation was calculated for ApT positions with at least 10x coverage on each strand. Please note the lack of correlation in AMT1-lacking species. **b,** Average methylation levels in 6mA and 5mC for three species that lack AMT1 or AMT6/7. *Dictyostelium discoideum* also lacks DNMTs and has previously been reported to lack 5mC; thus, the signal is likely noise (also driven by the low number of annotated TEs—88). Borders of genes and TEs are represented by the 5' and 3' sites. The first 1,500 bp post-TSS and before the TES are unscaled. **c,** Average 6mA levels around the

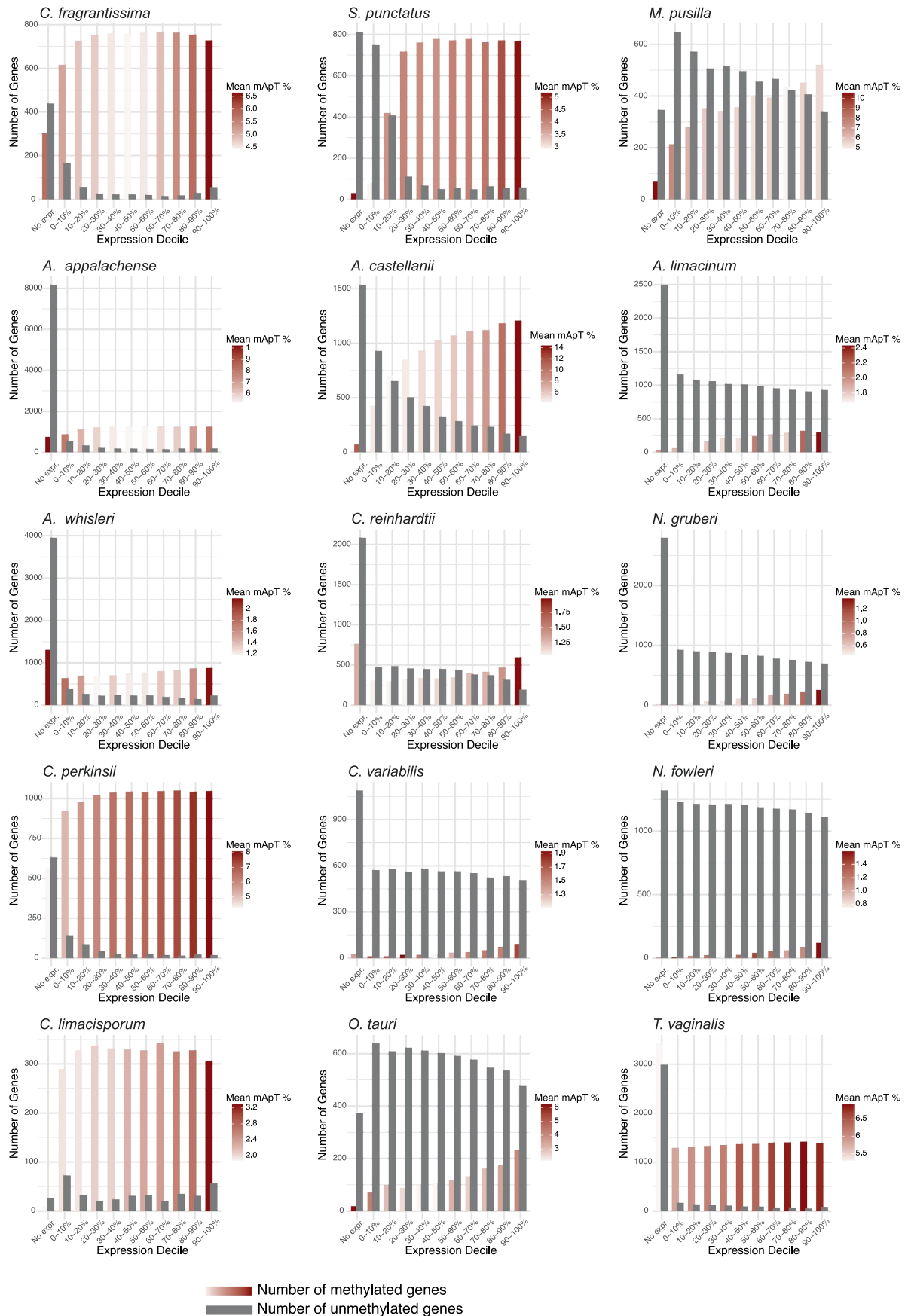
transcriptional start site (TSS) in five species with high 6mA levels (Fig. 2a), divided by sequence context. AT dinucleotides are the preferentially methylated context, but the AA context also gets residual methylation on the same regions, yet on average less than the total 6mA methylation. The dashed line highlights the max reached by AA methylation. **d,** Average 6mA level plot around the TSS in species with marked 'promoter' methylation patterns. Genes with a head-to-head orientation are displayed in blue and genes with a head-to-tail orientation are displayed in green. The black arrow shows the artifactual 'promoter' peak driven by head-to-head-oriented genes.



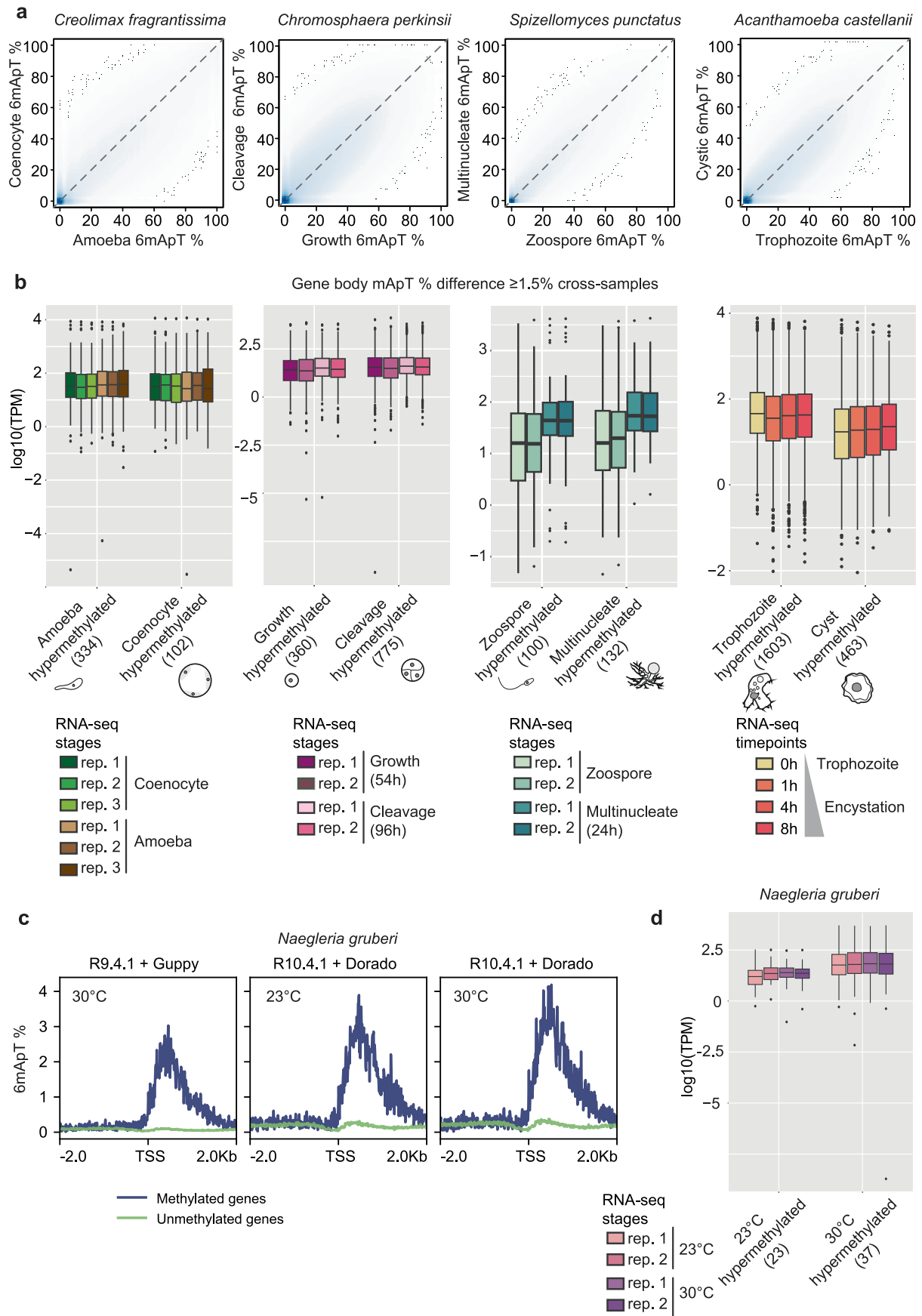
Extended Data Fig. 3 | See next page for caption.

**Extended Data Fig. 3 | 6mA correlates with higher transcription across eukaryotes, encoding AMT1.** **a**, Distribution of transcription levels of genes divided into ten deciles sorted by the gene body 6mA levels in the AT context. Number of genes in deciles shown on top. Boxplots show the median (centerline), interquartile range (box), and the smallest and largest values within 1.5× IQR (whiskers). **b**, Distribution of methylated AT sites (>10% 6mA) in gene bodies sorted in ten deciles by transcriptional level (numbers per deciles shown on top),

where not expressed genes (TPM < 1) are all shown separately (total number on no. expr. category). 'n = ' in parenthesis indicates the total amount of genes shown for each species. Boxplots show the median (centerline), interquartile range (box), and the smallest and largest values within 1.5× IQR (whiskers). **c**, Average 6mA levels in the AT context around the TSS for genes sorted by transcriptional levels as in **b**. Decile of expression color code shown in legend. TPM, Transcripts per Million.



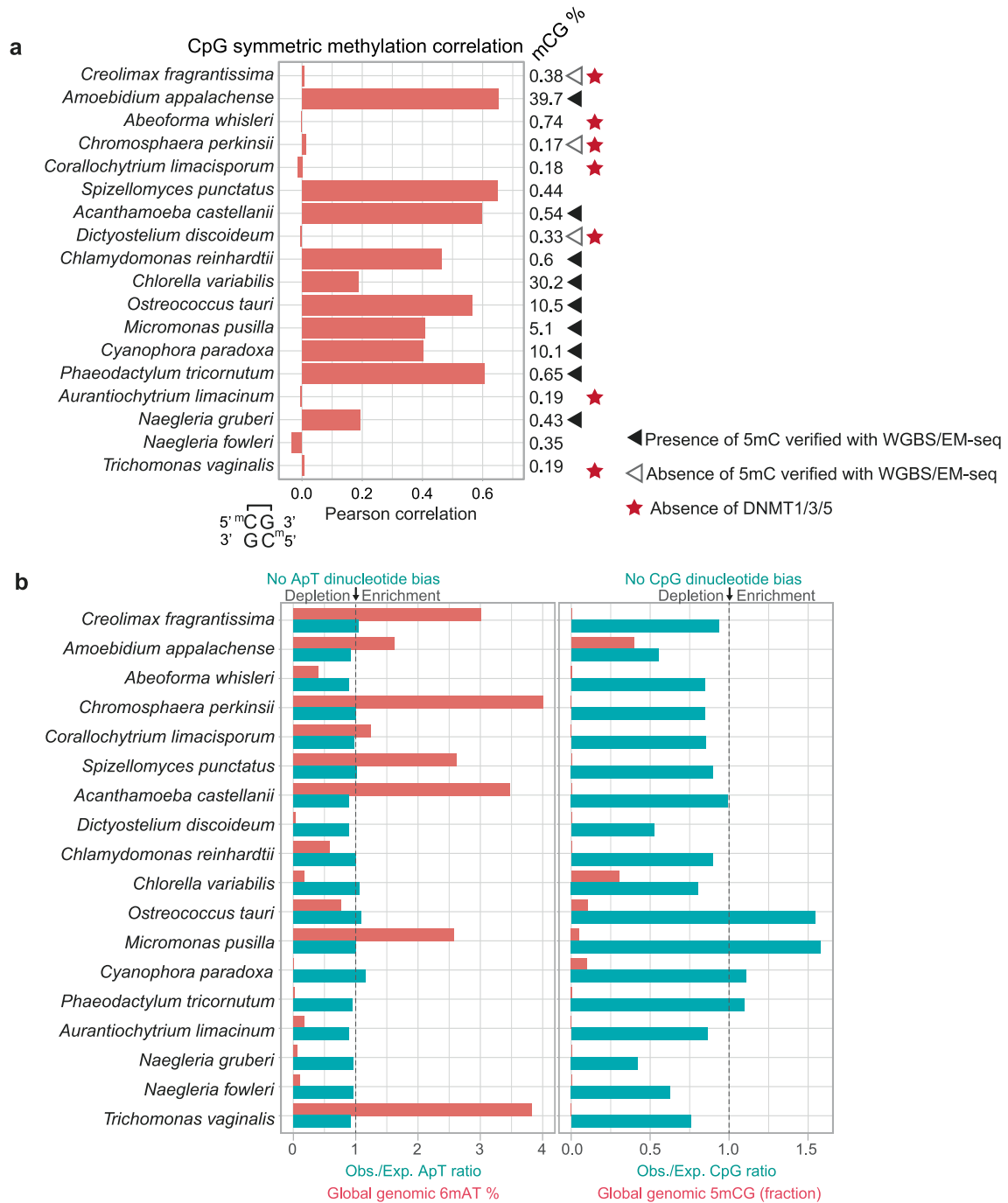
**Extended Data Fig. 4 | Correlation of 6mA with transcription varies among eukaryotes encoding AMT1.** The figure displays the number of 6mA-methylated versus unmethylated genes across transcriptional deciles. For methylated genes, bars are colour-coded to reflect the average 6mA at ApT percentage, while unmethylated genes are represented by grey bars.



Extended Data Fig. 5 | See next page for caption.

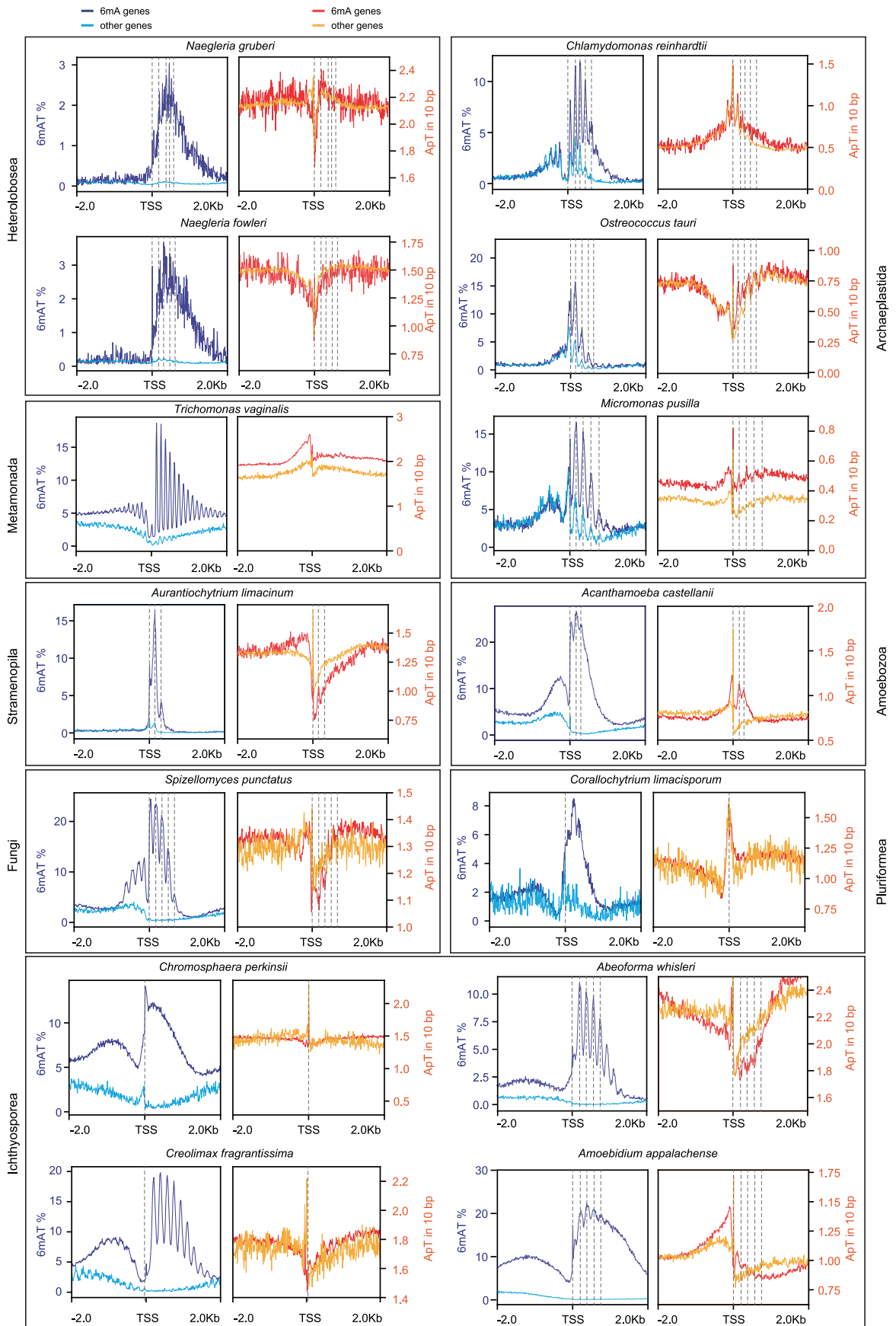
**Extended Data Fig. 5 | 6mA methylation is modestly responsive to transcriptional changes.** **a.** Comparison of 6mA levels in ApT dinucleotides across two cell stages in four species. Only ApT sites with at least 10x coverage in both samples are shown. **b.** Boxplot displaying the transcriptional levels of genes with the variable 6mA levels ( $\geq 1.5\%$  6mApT gene difference) across cell stages. Total number of genes in each category shown in the x-axis labels. Each RNA-seq sample is color coded according to each species legend. Boxplots show the median (centerline), interquartile range (box), and the smallest and largest values within  $1.5 \times$  IQR (whiskers); points outside this range are plotted as

outliers. **c.** Average methylation levels on methylated and unmethylated genes in *N. gruberi*. Methylated genes were defined using R9.4.1 and Guppy as base modification algorithm, whereas 23 °C and 30 °C samples were sequenced using R10.4.1 pore chemistry and 6mA calls done using Dorado. **d.** Boxplot displaying the transcriptional levels of genes with the variable 6mA levels ( $> 1.5\%$  6mApT gene difference) across *N. gruberi* temperature treatments. Boxplots show the median (centerline), interquartile range (box), and the smallest and largest values within  $1.5 \times$  IQR (whiskers); points outside this range are plotted as outliers.



**Extended Data Fig. 6 | 6mA methylation is not associated with a depletion of AT dinucleotides.** **a**, Global correlation of 5mC levels at CpG dinucleotides. Pearson's correlation was calculated for CpG positions with at least 10 $\times$  coverage on each strand, noting the lack of correlation in species lacking DNMT1/3/5. Global 5mCG levels were estimated using Oxford Nanopore data. Black triangles indicate species for which WGBS/EM-seq data have confirmed the presence of this modification, while white triangles indicate confirmed absence. **b**, The global genomic 6mA levels in AT context (red) are shown next to the observed versus

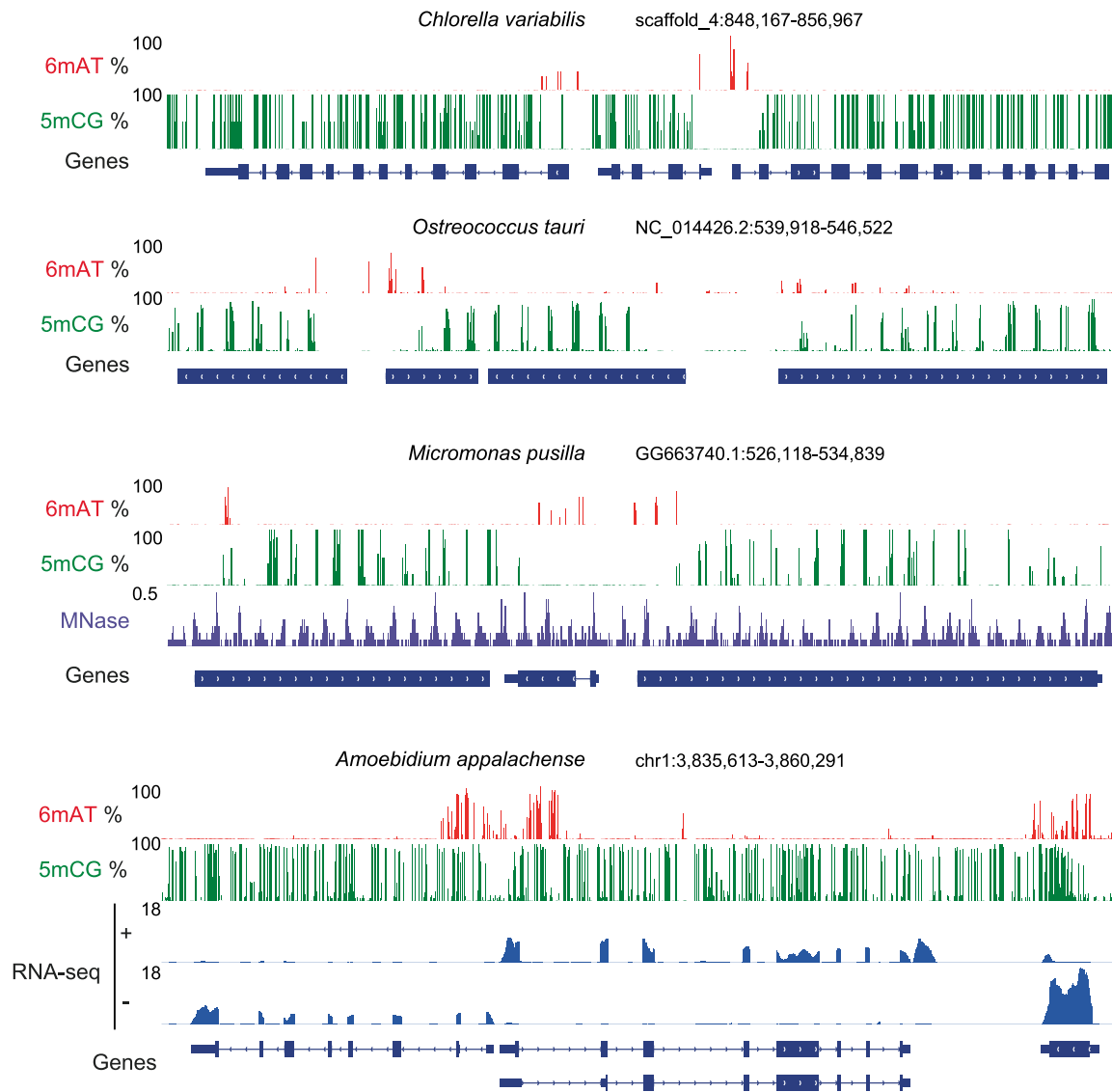
expected ApT dinucleotide ratios (turquoise) in each of sampled genomes. An observed-to-expected ratio of 1 indicates that ApT dinucleotides are present at the expected ratio given the genomic AT%, whereas <1 levels indicate depletion and >1 levels indicate enrichment. For comparison, global genomic 5mC levels (shown as a fraction with maximum being 1.0 in red) in the CG context are shown next to the observed versus expressed CG dinucleotide context ratio (turquoise) for the same set of species.



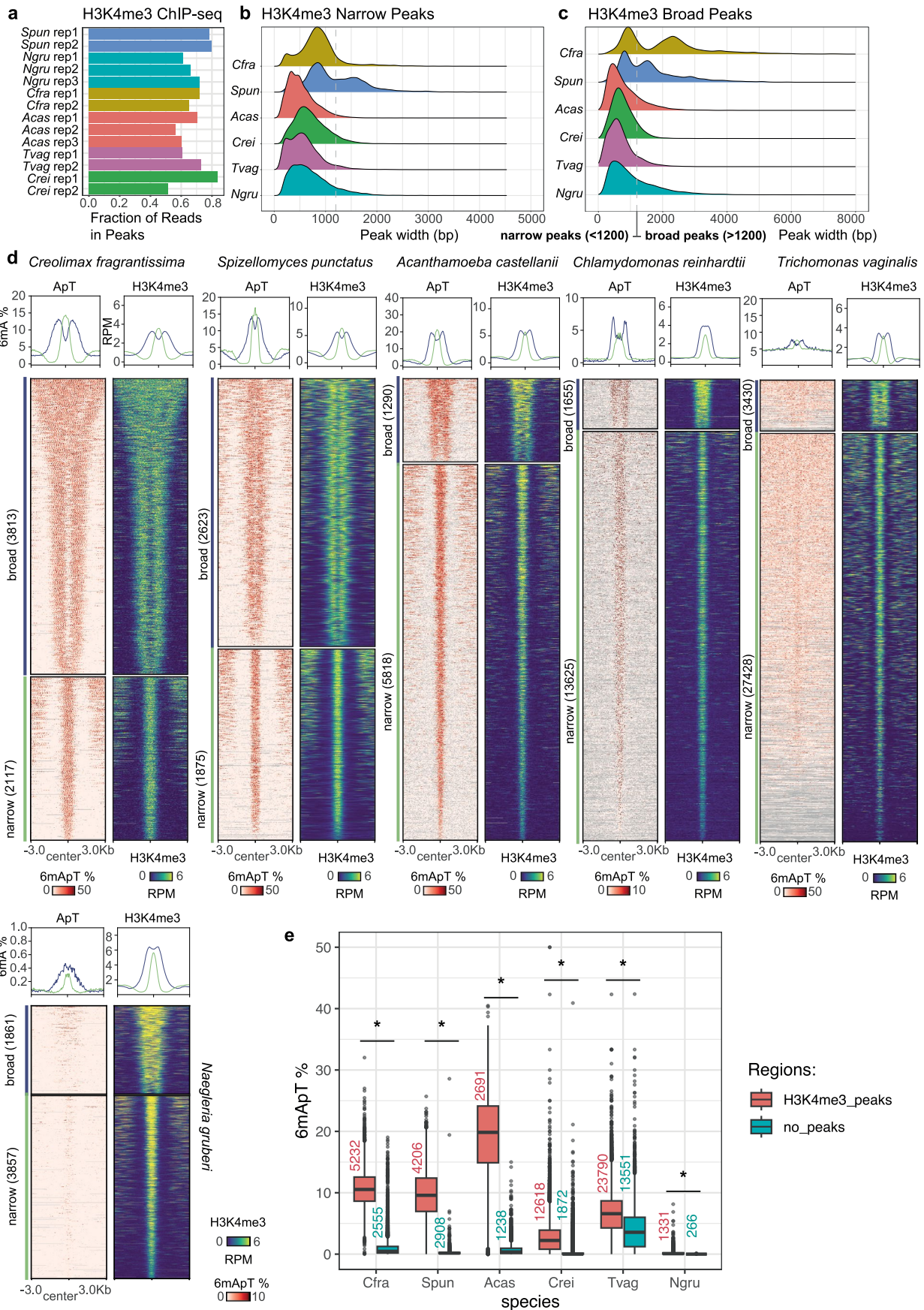
Extended Data Fig. 7 | See next page for caption.

**Extended Data Fig. 7 | 6mA rarely recapitulates the ApT dinucleotide composition downstream of the TSS.** Average 6mA levels in the AT context shown side by side to average ApT content in genes separated by their methylation status. Methylated genes are shown in dark blue (6mA plot) and dark red (ApT content plot), whereas unmethylated genes are shown in pale

blue (6mA plot) and orange (ApT content plot). Windows for averaging are 10 bp long. Enrichment of ApT in the TSS is likely driven by lack or inaccurate UTR annotations in many gene models, with 5' end of genes being the start codon 'ATG'. Gray dashed lines indicate the TSS position and the 6mA/T peaks in each species.



**Extended Data Fig. 8 | 5mC in gene bodies coexists with post-TSS 6mA.** Genome browser snapshots of four species with 5mCG on gene bodies. MNase and RNA-seq data are normalized using reads per million (RPM).

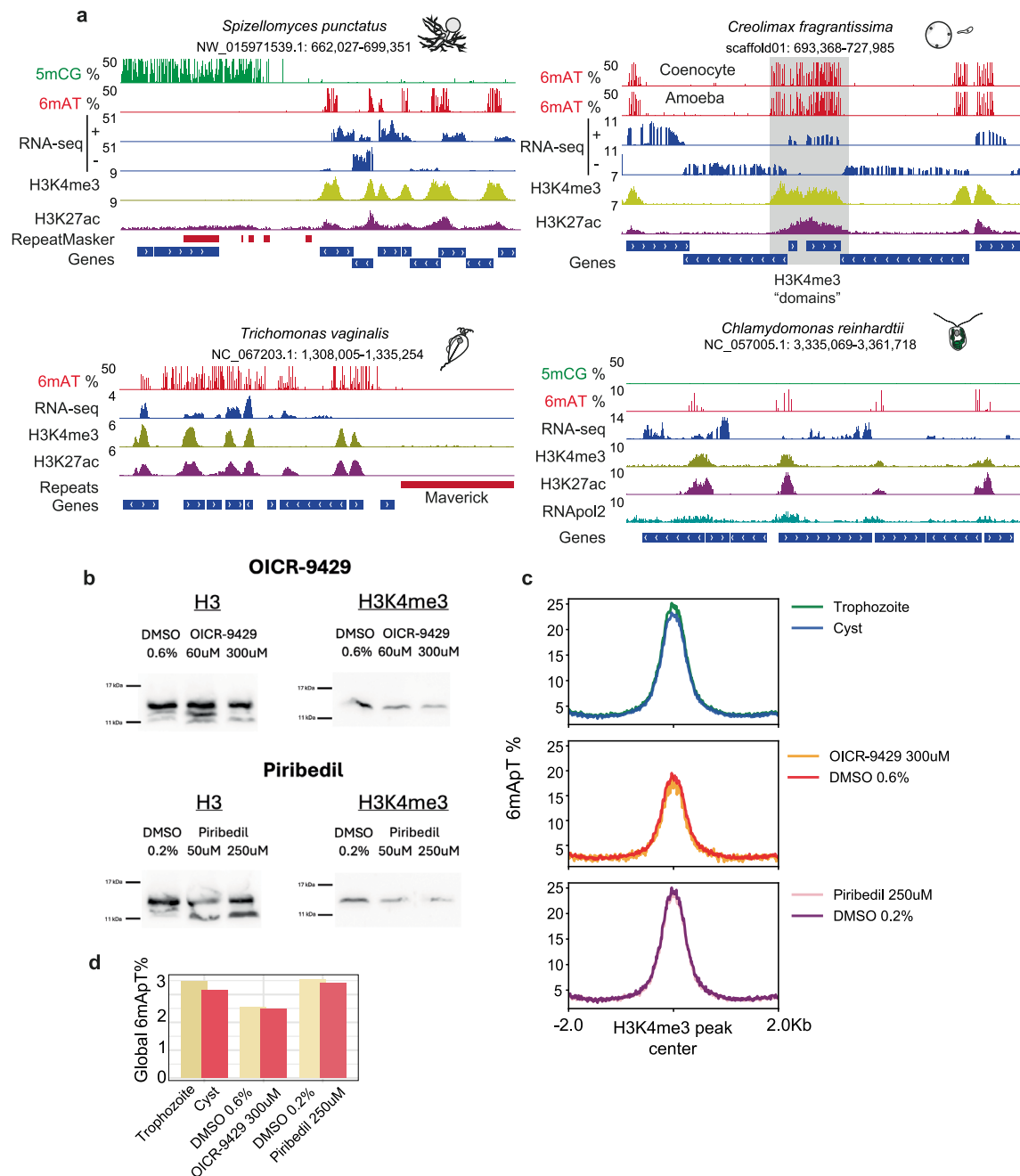


Extended Data Fig. 9 | See next page for caption.

**Extended Data Fig. 9 | 6mA is located within H3K4me3-demarcated regions.**

**a.** Fraction of reads falling within MACS3 narrow peaks for all H3K4me3 datasets used in this study. **b.** Distribution of peak sizes called as narrow peaks. **c.** Broad peaks, in each species. **d.** Heatmap showing the % of 6mA at ApT sites and H3K4me3 ChIP-seq coverage on both broad ( $\geq 1200$  bp) and narrow ( $< 1200$  bp) H3K4me3 peaks (as defined in **c**, with average metaplots summarizing the data). ChIP-seq data are presented as reads per million (RPM), and each plot is centered on the H3K4me3 peak midpoint. **e.** Boxplot of the percentage of

6mA at ApT sites on H3K4me3 peaks compared to genomic regions that do not overlap H3K4me3 peaks (spanning less than 3500 bp). Boxplots show the median (centerline), interquartile range (box), and the smallest and largest values within  $1.5 \times$  IQR (whiskers); points outside this range are plotted as outliers. An asterisk indicates a Wilcoxon ranked test (one-sided) p-value  $< 2.2 \times 10^{-16}$ . Cfra, *Creolimax fragrantissima*; Spun, *Spizellomyces punctatus*; Acas, *Acanthamoeba castellanii*; Crei, *Chlamydomonas reinhardtii*; Ngru, *Naegleria gruberi*; Tvag, *Trichomonas vaginalis*.



**Extended Data Fig. 10 | Pharmacological H3K4me3 depletion does not result in lower 6mA.** **a**, Genome browser snapshots from four species, with genomic coordinates indicated at the top. RNA-seq (displayed for both positive and negative strands) and ChIP-seq datasets are normalized using reads per million (RPM). Non-stranded RNA-seq data are shown as a single track. For *Chlamydomonas reinhardtii*, RNA polymerase II ChIP-seq data are also displayed. **b**, Western blot analysis showing total histone H3 and H3K4me3 modifications in

*A. castellanii* cells treated with the H3K4me3 inhibitors OICR-9429 and Piribedil. Equal amounts of total protein were loaded in each well. This experiment was repeated twice. **c**, Average 6mA at ApT (%) levels measured on H3K4me3 peaks in wild-type cells, comparing OICR-9429 treatment versus control and Piribedil versus control. DMSO concentrations for controls are matched to the final DMSO levels in the inhibitor samples. **d**, Global 6mA at ApT (%) levels in each sample as described in **c**.

## Reporting Summary

Nature Portfolio 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 Portfolio 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** Nanopore MinKNOW app was used to sequence the Nanopore libraries. This is the constantly updated version of Oxford Nanopore available to all users.

**Data analysis** For sequence search and phylogenetic analysis HMMER3 (3.3.1), BLASTP(2.9.0+), MAFFT (7.475), trimAL (v1.4.rev15), IQ-TREE(2.1.2). For Nanopore sequencing, basecalling and methylation analysis Guppy v6.5.7, modbam2bed (0.9.1), Dorado(0.7.2), Modkit (0.2.5), BEDTools (v2.29.2), UCSC bedGraphToBigWig, DeepTools2(3.5.0) and Bioconductor bsseq package for R were used. For RNA-seq analysis HISAT2 (64-bit), Stringtie (2.1.2) and Kallisto (0.46.0) were used. For ChIP-seq library preparation and analysis fastp (0.20.0), bowtie2 (1.2), Sambamba (0.6.6), MACS2 (3.0.0a7) and DeepTools2 were used. For genome assembly and re-annotation Guppy, Flye(v2.9-b1768), HyPo (v1.0.3), RagTag (v2.1.0), HISAT2 (2.2.1), Stringtie(v2.1.2), Trinity (v2.8.5), gmap (2020-10-14), Mikado (v2.3.0), Augustus (3.4.0), Liftoff (v1.6.3), PASA(v2.0.1), RepeatModeler2 (2.0.2a) and RepeatMasker (4.1.2-p1) were used. The code used for the analysis of this paper can be found in [https://github.com/AlexdeMendoza/6mA\\_evolution](https://github.com/AlexdeMendoza/6mA_evolution)

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 Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Nanopore raw sequencing data has been uploaded to ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>) under the accession number S-BSST1363, and the ChIP-seq data can be found in the GEO submission GSE261870. The ChIP-seq for *C. reinhardtii* was obtained from GSE59629, and *T. vaginalis* from GSE89662. Publicly available RNA-seq was obtained from PRJNA794325 (*A. castellanii*), PRJNA360056 (*A. whisleri*), SAMD00394225 (*A. limacinum*), PRJNA1091032 (*C. perkinsii*), PRJNA285347 (*C. reinhardtii*), PRJNA849385 (*T. vaginalis*), PRJNA262632 (*C. limacisporum*), PRJNA210187 (*C. variabilis*), GSE46692 (*M. pusilla*), GSE155535 (*O. tauri*), PRJNA642022 (*N. fowleri*), GSE249241 (*A. parasiticum*) and GSE68616 (*C. fragrantissima*). Annotation and other analysis files associated with this article can be found in: [https://github.com/AlexdeMendoza/6mA\\_evolution](https://github.com/AlexdeMendoza/6mA_evolution). IGV genome browser session can be accessed for *Creolimax fragrantissima* (<https://tinyurl.com/22zs884x>), *Spizellomyces punctatus* (<https://tinyurl.com/2d7jg2pz>), *Trichomonas vaginalis* (<https://tinyurl.com/2arl7o56>), *Acanthamoeba castellanii* (<https://tinyurl.com/22wjj72j>), *Naegleria gruberi* (<https://tinyurl.com/mr44r47h>), *Chlamydomonas reinhardtii* (<https://tinyurl.com/25aww7qk>).

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	<input type="text" value="n/a"/>
Reporting on race, ethnicity, or other socially relevant groupings	<input type="text" value="n/a"/>
Population characteristics	<input type="text" value="n/a"/>
Recruitment	<input type="text" value="n/a"/>
Ethics oversight	<input type="text" value="n/a"/>

Note that full information on the approval of the study protocol must also be provided in 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/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

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

Sample size	<input type="text" value="We sequenced all libraries to the highest coverage possible, aiming for &gt;20x in all species, yet for some the gDNA was obtained from culture collections and due to pore clogging more limited amounts were obtained. This is explicit in the methods, figure legends and supplementary material."/>
Data exclusions	<input type="text" value="No data was excluded."/>
Replication	<input type="text" value="Technical replicates were obtained splitting in half the amount of reads for modification base-calling, and the results were compared to verify that each set of reads provided the same pattern."/>
Randomization	<input type="text" value="No randomization was needed."/>
Blinding	<input type="text" value="Blinding is not relevant to the study."/>

## 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	Involvement in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

## Methods

n/a	Involvement in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used	Anti-Histone H3 (acetyl K27) antibody - CHIP Grade (abcam, Cat.No. ab4729) and Anti-trimethyl-Histone H3 (Lys4) Antibody (merckmillipore, Cat.No. 07-473) (Missing clone name and lot number), for western blot: Anti-Histone H3 (Abcam, ab1791), Histone H3K4me3 (Active Motif AB_2615077)
Validation	These antibodies have been commonly used in the Sebe-Pedros laboratory for profiling histone modifications in many species, and were selected for specificity and signal to noise ratio. Western blot antibodies are regularly used in Paul Hurd laboratory for working on honeybees.

## Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	The study did not involve laboratory animals
Wild animals	The study did not involve wild animals
Reporting on sex	The study did not require any sex-based analysis as asexual single cell eukaryotes were used
Field-collected samples	The study did not involve samples collected from the field
Ethics oversight	No ethical approval was required as single cell eukaryotes were used

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

## Plants

Seed stocks	n/a
Novel plant genotypes	n/a
Authentication	n/a

## ChIP-seq

### Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links <i>May remain private before publication.</i>	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE261870">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE261870</a> Token: kdkbqsewtngtrtaf
Files in database submission	raw fastq files and coverage bigwigs.

Genome browser session  
(e.g. [UCSC](#))

We provide 6 genome browser links: Creolimax fragrantissima: <https://tinyurl.com/22zs884x>  
Spizellomyces punctatus: <https://tinyurl.com/2d7jgzpz>  
Trichomonas vaginalis: <https://tinyurl.com/2arl7o56>  
Acanthamoeba castellanii: <https://tinyurl.com/22wj72j>  
Naegleria gruberi: <https://tinyurl.com/mr44r47h>  
Chlamydomonas reinhardtii: <https://tinyurl.com/25aww7qk>

## Methodology

Replicates	Each ChIP-seq condition was obtained in duplicates (per histone mark & per species).
Sequencing depth	This is available in Supplementary Tables.
Antibodies	Anti-Histone H3 (acetyl K27) antibody - ChIP Grade (abcam, Cat.No. ab4729) and Anti-trimethyl-Histone H3 (Lys4) Antibody (merckmillipore, Cat.No. 07-473) (Missing clone name and lot number)
Peak calling parameters	macs3 callpeaks -t sample.bam -q 0.01 -g species_assembly_size
Data quality	Fraction of Reads in Peaks is >60% for all samples.
Software	All ChIP-seq data was analysed using fastp to trim the reads, and these were mapped to the genomes using bowtie2, allowing a maximum insert size of 2000 base pairs (-I 2000). Duplicate reads were removed using Sambamba. DeepTools2 was used to generate bigwig files and to visualise epigenomic data, as well as Integrative Genome Viewer.