Occurrence and quantities of DNA modifications across the tree of life

Sreejith Jayasree Varma[†1], Enrica Calvani[†2,3], Nana-Maria Grüning[1], Christoph Messner[2,3], Nicholas Grayson[5], Floriana Capuano[2], Michael Mülleder[4], Markus Ralser* [1,2,3]

[1] Department of Biochemistry, Charité - Universitätsmedizin Berlin, Germany.

[2] Department of Biochemistry and Cambridge Systems Biology Center, University of Cambridge, UK.

[3] The Molecular Biology of Metabolism Laboratory, The Francis Crick Institute, London, UK.

[4] Core Facility – High Throughput Mass Spectrometry, Charité - Universitätsmedizin Berlin, Germany.

[5] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom.

[†]These authors contributed equally to the work

*To whom correspondence should be addressed. M. Ralser, Charité Universitätsmedizin Berlin. markus.ralser@charite.de

Abstract

Enzymatic DNA modifications like methylcytosine (5mdC), methyladenine (N6mdA), or hydroxymethylcytosine (5hmdC) are key for chromatin function, gene expression regulation, and antiviral defense, but they remain understudied in non-model organisms. We established a mass spectrometric method for the sensitive and accurate quantification of enzymatic DNA modifications, and analyzed 85 bacterial genomes, 19 plant samples, 41 tissues from 12 animal species, 6 yeast species, and two archaeal species. We report no or only very low concentrations of DNA modifications in yeast and insects, but find DNA modifications universal to both bacteria and higher eukaryotes. Specifically for prokaryotes, our dataset indicates that evolutionary relationships and host–pathogen interactions, but not the ecological niche in general, select for a similar degree of DNA modification. In higher eukaryotes, largest concentration differences between tissues are detected for 5hmdC. Our dataset further reveals unique biological cases that warrant attention in the study of DNA modifications. For instance, while our data shows that most species contain just one dominating DNA modification, we detect all dominianting DNA modifications (5mdC, N6mdA, and 5hmdC) to exist in parallel in *Raphanus sativus.* Other plant species, like onion, sunflower, or the grass big bluestem, can have more than 35% of cytosines methylated. Finally, 5hmdC, so far mostly studied in the vertebrate central nervous system, was identified to reach a concentration of up to 8% of all cytosines in the Oman garra brain, and was also detected in several plants, like *Lepidium sativum*. The present study underscores the exploitation of biological diversity for studying DNA modifications.

Introduction

Enzyme-catalyzed DNA modifications are studied for their roles in chromatin structure, gene-expression regulation, prevention of viral DNA integration, epigenetic inheritance, cell–environment interactions, developmental biology, immunity, memory, aging, and cancer (1–10). The methylation of the 5th carbon (C5) of the cytosine ring to yield 5-methyl-2′-deoxycytidine (5mdC) was the first nucleotide modification to be discovered (11), and has remained the most intensively studied (12,13). 5mdC can be enzymatically oxidized into 5-hydroxymethyl-2′-deoxycytidine (5hmdC) and further into 5-formyl-2′-deoxycytidine (fdC) and 5-carboxyl-2′-deoxycytidine (cadC) (14–16). Although these modifications have been described as transient intermediates of 5mdC demethylation, at least one (5hmdC) has been found to accumulate in the mammalian brain, specifically in the large Purkinje neurons, indicating a regulatory function (17). N4-methyl-2′-deoxycytidine (4mdC), found in bacteria, is yet another form of cytosine modification (18,19). Cytosine thus exists in multiple chemical states (dC, 5mdC, 5hmdC, fdC, cadC, 4mdC, as well as the rare 4,5-dimethyl-2′-deoxycytidine (4,5dmdC)) (12,20). Another important modification is the N6 methylation of adenine. N6-methyl-2′-deoxyadenosine (N6mdA) was initially discovered in bacterial genomes (21) and later also in archaea, plants, and nematodes (22,23). Although N6mdA is not essential in microbial model organisms, this modification has been increasingly associated with functions that promote virulence or viral DNA integration (24,25). Indeed, it seems likely that DNA modifications play different roles in different species, as indicated by the varying amounts of DNA modifications across model organisms. For instance, *Arabidopsis thaliana* has orders of magnitude higher levels 5mdC compared to the dominant insect model *Drosophila melanogaster,* while the dominant yeast model organism *Saccharomyces cerevisiae* lacks this modification altogether (26,27).

Since until recently studying DNA modifications was technically challenging, information concerning their content and function is still scarce for species other than model organisms, several crops, and humans. Therefore, it is rather difficult to translate the knowledge derived from those intensively studied species into a broader biological context. For instance, it is hard to judge from the current literature if the low amount of DNA modifications in laboratory yeast and *D. melanogaster*, or the high amount in *A. thaliana (27),* represent the rule or the exception in their respective phylogenetic group without a broader multi-species dataset for comparison.

Sequencing technologies are the dominating method for studying the role of DNA modifications in humans and model organisms. They provide only relative quantitative values about the consent of DNA modifications, but position-specific information, which is required to understand the specific molecular functions of modification, for instance whether they influence gene-expression regulation in a specific locus (28,29). On the other hand, it is equally important to absolutely quantify DNA modifications at the genome-wide level. Quantitative values help to address more-general questions, like evolution and activity of the biochemical pathways that modify nucleic acids, their role in viral immunity, relationship between different modifications, and for comparing (non-model) species still lacking high-quality reference genomes. We and others (27,30–34) have shown previously that targeted mass spectrometry is an ideal technology to determine absolute quantities of DNA modifications, specifically, if they are low abundant and in the noise range of sequencing technologies. Mass spectrometry further is suitable for studying poorly characterized species, as no prior knowledge about the genome is required for data analysis. Aside from that, targeted mass spectrometry is economical, with running costs per sample amounting to single-digit dollars. For these reasons, mass-spectrometric quantification is well suited for identifying interesting patterns in the amount and relative abundances of DNA modifications, specifically within understudied species.

Results and discussion

Quantification of a panel of enzymatic DNA modification using liquid chromatography/multiple reaction monitoring

In order to quantify multiple enzymatic DNA modifications in a single analysis, we expanded a previous method based on liquid chromatography–multiple reaction monitoring (LC–MRM) and designed for the quantification of 5mdC (35). This method is characterized by a sensitivity down to attomoles and a broad dynamic range, and discriminates between RNA and DNA modifications, clarifying the previously debated content of 5mdC in several yeast species (27). In this method, isolated DNA is first enzymatically digested to obtain the corresponding nucleosides using a nuclease enzyme mixture (DNA Degradase Plus, Zymo Research). The resulting digests are then filtered and directly analyzed by a targeted assay using liquid chromatography - multiple reaction monitoring (LC–MRM) using a triple quadrupole (QQQ) mass spectrometer. Distinguishing the nucleosides arising from a DNA monomer from a potentially co-purified RNA monomer occurs on the basis of the precursor mass difference of the sugar moiety. Such a strategy ensures the measured nucleosides are free from RNA contamination as many base modifications are more frequently observed in RNA (27,35). For quantifying other DNA modifications, namely 5hmdC, N6mdA, cadC and fdC, we obtained synthetic standards for these molecules (Methods, Table S1) and optimized the instrumental and chromatography parameters accordingly. Moreover, we supplemented the method by a neutral loss scan to confirm the MRM results, as well as to detect additional modifications like 4mdC, or to screen for yet unknown modifications (Methods). Combined with the high sensitivity offered by a triple quadrupole mass spectrometer (Agilent 6470), we were able to achieve detection limits in picomolar ranges (Fig. 1A).

Upon setting up the method, we sampled cells or tissues for a large number of species across the three domains of life. Because our method does not include any amplification steps, and detects modifications on the DNA directly, it requires clean DNA at microgram levels, at least for the detection of the lowly concentrated DNA modifications. Unfortunately for some rare specimens, we only had limited sample amounts, and in many cases, standard DNA preparation protocols did not yield DNA of sufficient quality or concentration for our assay. However, by combining

different protocols, we were able to extract clean DNA at microgram levels for 286 distinct cases. These are derived from 124 different species, including 85 bacterial species, 6 yeast species, 2 archeal species, 19 plant species, and 18 tissue and cell-culture samples from 12 animal species, including human and mouse. The collection included both the typical model organisms, and specifically for bacteria, vertebrates, and plants we included a significant number of species that have been barely characterized at the molecular level so far (Fig 1B). Furthermore, for a number of vertebrates, including human, the model organisms mouse (*Mus musculus)*, African clawed frog (*Xenopus laevis*), but also for some less studied species, the opossum *(Monodelphis domestica*), the Alpine marmot (*Marmota marmota*), and the Oman garra (*Garra barreimiae*), we obtained DNA from multiple tissues and/or cell lines in order to quantify tissue differences in the absolute DNA modification content. For plants, we focused on seedlings that were germinated in the lab (Methods). The seedlings not only allowed for efficient DNA extraction, which can be hampered by high concentrations of plant polymers in fully differentiated plant tissues, but also for direct comparison between the plants at a similar developmental stage.

While multiple lower eukaryotes lack DNA modifications, N6mdA dominates in bacteria, and 5mdC is the dominating DNA modification across higher eukaryotes

Our results reveal major global differences in the nature and total concentration of DNA modifications when comparing the domains of life (Fig. 1C–D). First, despite the broad coverage, high sensitivity, and precision of our method, we did not detect significant levels of fdC and cadC in any of the genomes measured (limits of detection were 238 pM and 251 pM, respectively). These oxidized forms of 5-methyl-2′-deoxycytidine have been associated with the degradation of 5mdC (15), and according to our results they seem to remain transient across species as they do not accumulate to significant, genome-wide-scale levels. Next, we detected hardly any DNA modification in any of the unicellular fungi studied (Table S3). Hence it is not merely 5mdC (27,36,37), but also its oxidized form 5hmdC along with N6mdA that are very low if not absent in typical yeast species. Is interesting in this context that the insects *Trichoplusia ni, Spodoptera frugiperda and D. melanogaster* (Table S3) all had very low amounts of these DNA modifications as well. Indeed, the fruit fly *D. melanogaster* has so far been considered an unusual case among the laboratory model organisms, as it contains only trace amounts, if any, of cytosine methylation

(27,38,39). The presence of other DNA modifications in *D. melanogaster* like N6mdA has also been contested due to the presence of an appreciable gut microbiome, which could confound the results (40). We assessed this situation, comparing the genomic DNA obtained from fruit flies that possessed a functioning gut microbiome versus ones grown under germ-free conditions. N6mdA was also detected in germ-free *D. melanogaster* (~0.04%, Fig. S1). In a recent study comparing DNA adenine methylation levels in multiple eukaryotic species, the bacterial contamination affected the N6mdA measurements. However, it was possible to distinguish the N6mdA in *Drosophila* tissue from microbial contamination using quantitative deconvolution (41). While the adult *D. melanogaster* contained methylated adenine as a DNA building block, ovarian cells collected from two moth species (*T. ni* and *S. frugiperda*) principally contained methylated cytosine as the preferred base modification (0.2% and 0.1%, respectively).

Which conclusions can be drawn from the low concentrations of DNA modifications in yeasts and insects? First, these results support the notion that enzymatic DNA modifications are not universal, which could have peculiar evolutionary consequences. Studies in yeast have concluded that DNA modifications could have been specifically lost during yeast evolution (42). However, our result that insects can have similarly low DNA modification levels raises another possibility: that DNA modifications could have evolved in higher eukaryotes and bacteria, after yeasts and insects branching off.

As a rule, most genomes contained a single modification type. Some exceptions to this were however encountered. A subset of the eukaryotes and a subset of prokaryotic species contained low concentrations also of a second modification, which could be either 5mdC, N6mdA, or 5hmdC (Fig. 2, Table S3, Fig S2). For instance, *Diplotaxis tenuifolia* had low amounts of N6mdA (0.1%, Table S3) next to high amounts of 5mdC. Of particular interest was *Raphanus sativus*, which was the only species among those analyzed that possessed all the three modifications at detectable levels and in parallel. Among prokaryotes, we observed a maximum of two modifications (5mdC and N6mdA), with 5hmdC entirely missing. Our study further featured two archeal genomes (*Sulfolobus acidocaldarius* and *Halobacterium salinarum*), which shared a similar level of the cytosine modification but differed in their levels of adenosine modification. While we detected N6mdA in *Halobacterium*, no adenosine modification was observed for *Sulfolobus* (Table S3).

Tissue divergence of 5mdC concentrations in vertebrate and plant genomes

Among the DNA modifications, 5mdC had the highest abundance and was specifically abundant in plants. Most vertebrate genomes studied had a 5mdC content of around 5% (mean 4.66, SD 2.17) of the cytosine residues. Some species, including the model organisms *Danio rerio* and *Xenopus laevis,* had higher levels consistent with early observations (43). In plants however, 5mdC concentrations of 10% (mean 20.34, SD 9.81) and higher were typical (Fig. 1D). Extremely high values for cytosine methylation were observed in *Andropogon gerardii* and *Allium cepa*, where more than 35% of cytosines were methylated (Fig. 1D, Table S3). Given that very low levels or no 5mdC were detected in yeast and insects, cytosine 5 methylation levels hence differ by several orders of magnitude within the eukaryotic kingdom.

In multicellular organisms, DNA modifications are important for development, and tissue differences between DNA modification patterns are observed (44–46). Our data suggests that a change in the modification pattern or sequence context does not necessarily have a strong impact on the total concentrations of the DNA modifications however. We analyzed spleen, muscle, lung, liver, kidney, heart, and central nervous system (CNS) samples from five animal species, of which two are model organisms *(Xenopus laevis*, *Mus musculus),* and three non-model organisms *(Garra barreimiae*, *Monodelphis domestica*, *Marmota marmota*). From *M. musculus* we further examined tissues from multiple inbred laboratory lines: BALB/c, FVB/N, Hsd/Ola/MF1, B6SJL/CD451/CD452, BALB/cAnN, 129S8, and F1/CBAxB6. In parallel, we analyzed multiple human cell lines (Table S3). The obtained data was consistent, in the sense that the values for 5mdC levels were highly similar, as long as the tissues were derived from the same species (Fig. 3A, left). For instance, most tissues in *G. barreimiae*, *M. marmota,* and *M. musculus* tissues had 5mdC levels of around 5–6% (Fig 3A). Between the different mouse lines, there were no significant differences in 5mdC levels (Table S3). We noted, however, some small but notable differences between specific tissues. Heart tissue presented a broad cytosine methylation level and brain tissue had a higher median value for percentage methylation compared to other tissues (5.3% vs 4.9%) (Fig 3B). We then tested whether different nutritional conditions would change the picture. Therefore, we grew a commonly used mammalian cell line (HeLa) under different growth conditions. The different growth conditions affected 5mdC levels, and the detected differences were in a similar magnitude as the small differences detected between tissues (Fig S3).

Overall 5mdC concentrations in opossum and *Xenopus*, respectively, were different to the aforementioned species. In opossum, we detected much lower levels (2%) of 5mdC in all tissues examined. Conversely, in *X. laevis*, all tissues had much higher concentrations (about 9.4%). However, also here, in both cases the tissue differences in the 5mdC concentrations were minimal, at least when compared to the differences that exist between species. Although we tested fewer cases in plants, our data suggest the situation could be similar there too. We tested different tissues (roots, leaf, stem, and seed cotyledon) from *Phaseolus vulgaris*, and obtained consistently high (16.7%) 5mdC concentrations in all measured tissues (Table S3). Hence, the several tissues examined from animal species, cell lines, and *Phaseolus vulgaris* provided a largely consistent picture: in a given organism, several tissues exhibit similar levels of 5mdC, and, that within-tissue differences are typically smaller compared to the differences that can be detected between species.

Tissue specificity of 5-hydroxymethyl deoxycytidine in the vertebrate CNS

Tissue specificity was, however, detected for another modification, 5-hydroxymethyl deoxycytidine (5hmdC). Indeed, 5hmdC was previously discovered in mammalian brain tissue, where it is formed via oxidation of 5mdC by TET enzymes (16,47). Our dataset shows that 5hmdC is detected in a broad range of vertebrate tissues except for spleen, but reaches significantly higher concentrations specifically in samples from the CNS. Although the spleen tissues had similar 5mdC levels as other mouse tissues, 5hmdC was not detected in these tissues (Fig. 3B). Interestingly, our data reveals that the highest 5hmdC levels were not detected in the mammalian brain. The presence of this modification could reach up to 8% of cytosine residues in *G. barreimiae*. *M. musculus* (3.3%) and *X. laevis* (2%) too had high levels of 5hmdC specifically in brain tissue relative to other tissues in those organisms (Fig. 3A, right). An interesting exception was in opossum, the only vertebrate species analyzed, in which 5hmdC levels were not higher in the brain compared to peripheral tissue.

Apart from vertebrates, 5hmdC was also observed in *A. thaliana* and *Oryza sativa* (48). Our data shows that the presence of 5hmdC is by no means universal in plants. Indeed we did not

detect it in the majority of plant samples. However, our data adds several species (*Allium cepa, Laurus nobilis, Lepidium sativum*, and *R. sativus*) in which we confirmed low concentrations of 5hmdC. Further, we did not detect 5hmdC in any of the bacterial or fungal genomes analyzed. Our results support the fact that the modification of 5hmdC is more widespread in biological systems as previously assumed, but are not universal or specific, to any part of the phylogenetic tree.

Variations in DNA modification across different bacterial species

In prokaryotes, high amounts of DNA modifications all concerned N6mdA, with the highest levels detected in *Mobiluncus curtisii* (~1.4%) and *M. thermoacetica* (~1.1%). In total, the prokaryotic genomes hence contained higher amounts of DNA modifications compared to lower eukaryotes such as yeasts and insects, but lower amounts of DNA modifications compared to higher eukaryotes—plants and vertebrates in particular.

Typical bacterial species contain only one modification type—mostly N6mdA (Fig. 4A). Our data reveals some exceptions. Certain genera such as *Campylobacter* contain trace quantities of 5mdC (<0.1%) next to the dominating N6mdA modification (Table S4). In general, the observed trend was that the occurrence of one type of modification limits the occurrence of the other. For instance, *M. curtisii* with ~1.4% of its adenine residues methylated shows only 0.3% 5mdC, while *Sebaldella termitidis*, with unusually high cytidine methylation (~2.4%), has only 0.1% of its adenines methylated. Interestingly, we observed that median values for 5mdC dominate over N6mdA in those bacteria that colonize or enter mutualistic relationships with higher eukaryote species that carry 5mdC as their main modification (Fig. S4, Table S4). This included the genus *Neisseria*, mucosal-surface-colonizing bacteria, which showed 1.4% and 2% (*N. gonorrhoeae, N. lactamica* respectively) of cytosine residues were methylated while containing only <0.3% N6mdA, and *Faecalicoccus pleomorphus* and *Bifidobacterium adolescentis*, with >1.5% of 5mdC without any detectable levels of N6mdA modification. Indeed, others made a similar observation in single-cell fungi. While the environmental yeasts studied lacked any modifications (27), the most frequent commensal yeast pathogen *Candida albicans* contains 5mdC (49). This result is interesting, because it is believed that the main role of DNA modifications in single cellular organisms is a defense against viruses, while higher organisms adapted methylation for gene regulation. In that light, the increase of DNA methylation in the pathogen in host–pathogen

interactions would be hard to explain. Our data suggests hence that the picture about the functions of DNA modifications in prokaryotes is not complete.

Finally, we also observed a third modification, 4mdC, to be frequent in prokaryotes. 4mdC occurred as an exclusive cytosine modification in *Legionella fairfieldensis, Bacteroides caccae, Acinetobacter baumannii, Listeria innocua, Bifidobacterium breve, Moorella thermoacetica, Thermocrinis ruber, Bacteroides vulgatus*, and *Caldilinea aerophila*. Moreover, it existed in tandem with 5mdC as a second modification in *Shewanella putrefaciens, Stenotrophomonas maltophilia, Bifidobacterium dentium, Mobiluncus curtisii*, and *Gallibacterium anatis* (not quantified).

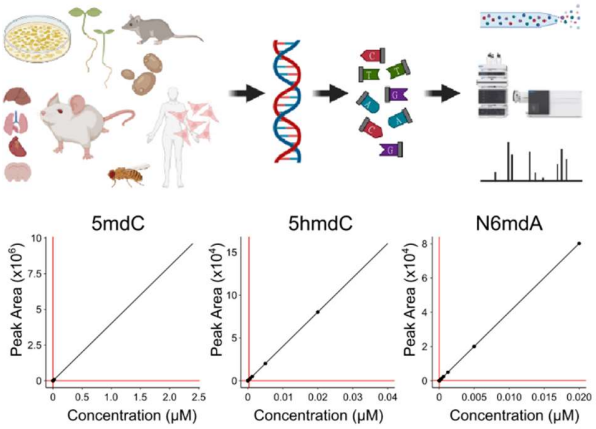
Having analyzed 85 species, we were able to ask if bacterial species with a close evolutionary relationship or similar habitat or genome properties also have a more similar modification makeup. We did not detect any relationship between nature and level of modification and genome size or GC content (Fig. S5). Similarly, we detected no significant correlation between factors such as pathogenicity, temperature of growth, or tolerance to oxygen and the amount of modifications per unit genome size. We did however observe obvious patterns at the different taxonomic levels once we grouped the different bacterial strains according to phylum, class, and genus. Similarities are detected at the genus level (Fig. 4B–C). Members of the same genus often displayed similar values for a given modification. For example, species of the *Vibrio* genus presented similar quantities of N6mdA. At the class level we observed trends between the different classes and the amount of modification. α- and γ-Proteobacteria had the highest N6mdA content among different classes present while bacteroidetes presented with more cytidine methylation than adenosine methylation. At the phylum level the patterns were more prominent in Proteobacteria, containing more N6mdA than 5mdC, while a reverse trend of more 5mdC than N6mdA was observed for Bacteroidetes and Firmicutes. Combined, these results suggest that differences in the modifications do not reflect basic structural genome features such as size or GC content, but the observation that more-closely evolutionarily related species have higher similarities in DNA modification suggests that gene drift and gene function are key drivers in the evolution of DNA modifications.

Acknowledgements

We thank Biological Research Facility at Francis Crick Institute for *Mus musculus*, *Danio rerio*, *Xenopus laevis*, samples, Bryony Lee (Turner lab, The Francis Crick Institute) for opossum samples, Annick Sawala (Gould Lab) for *Drosophila* samples, Cell Services (The Francis Crick Institute) for animal cell lines, National Yeast Collection for yeast samples, Felix Forest (Kew Gardens) and (Nell Jones) Chelsea Physic Garden for plant samples, Barbara Tautsher, Elisabeth Haring, Luise Kruckenhauser, for *Cepaea hortensis* and *Garra barremiae* samples (Natural History Museum of Vienna), Florian Winkler, Heinrich Aukenthaler, Erhard Seehauser, and Gottfried Hopfgartner (Forestry and Hunting Authorities South Tyrol, or Jagdrevier Mauls, Bolzano Province, Italy) for their support in obtaining tissue samples from alpine marmot in their wild habitats of Mauls and Gsies (Italy). This work was supported by the Francis Crick Institute which receives its core funding from Cancer Research UK (FC001134), the UK Medical Research Council (FC001134), and the Wellcome Trust (FC001134), and received specific support from the Wellcome Trust (200829/Z/16/Z, 101503/Z/13/Z) and the German Ministry of Education and Research (BMBF) as part of the National Research Node "Mass spectrometry in Systems Medicine (MSCoresys)", under grant agreement 031L0220A.

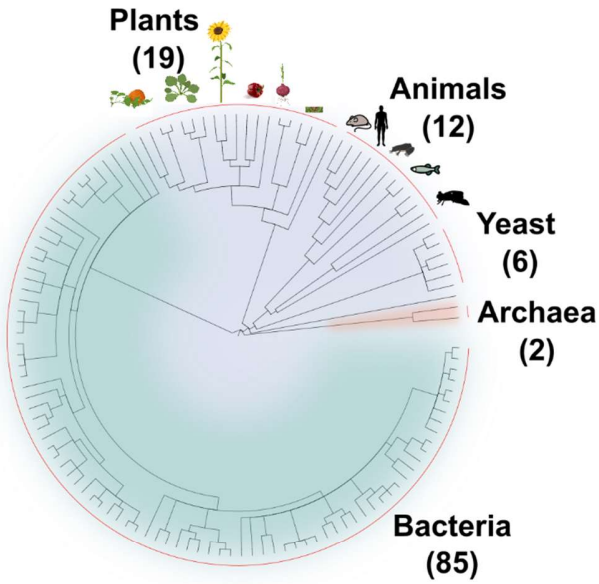
Figures

A

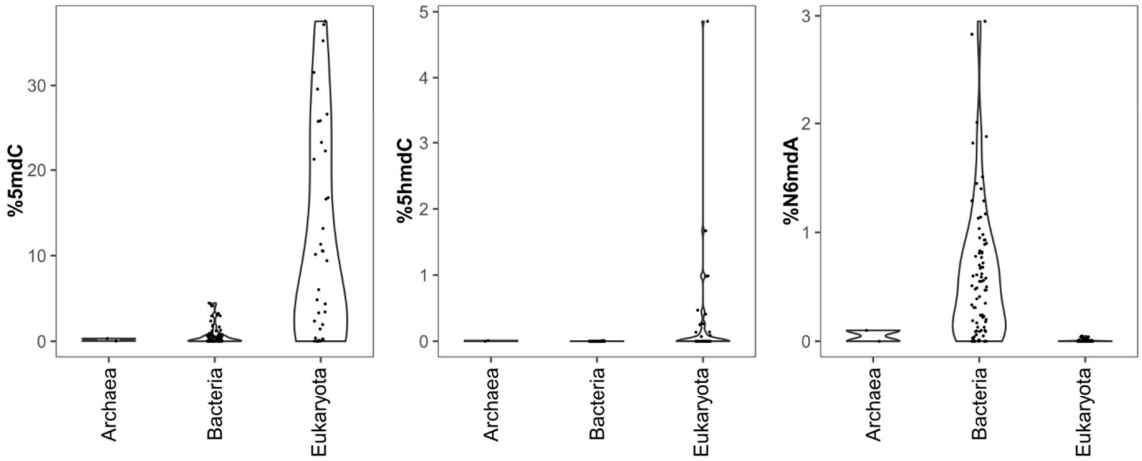


Modification	Equation	R^2	LOD
5mdC	y = 464800x-1052	0.99	104 pM
5hmdC	y = 3092x^0.64	0.99	316 pM
N6mdA	y = 4539238x+46	0.99	19 pM

B



C



D

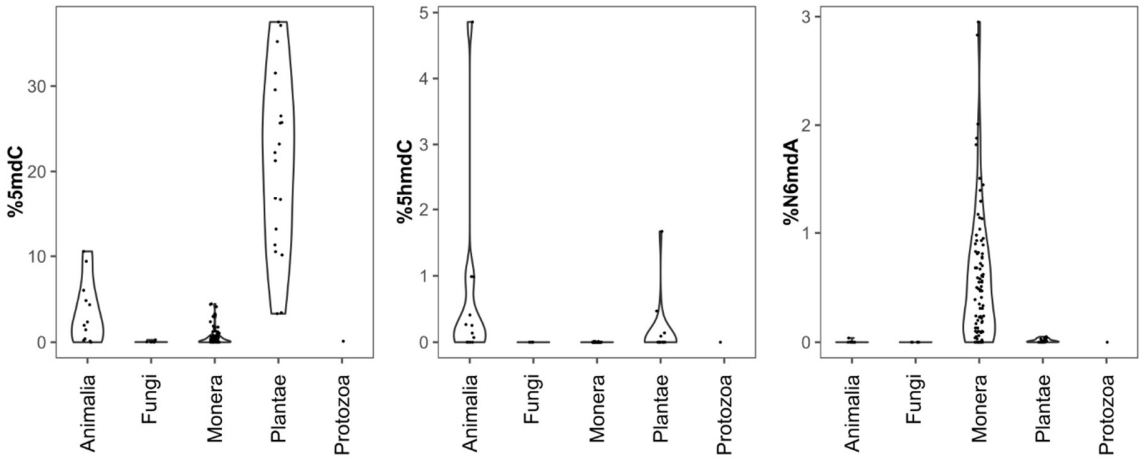


Figure 1. (A) Multiplex analysis of various genomic DNA modifications using liquid chromatography multiple reaction monitoring (LC-MRM) following enzymatic digestion of DNA. The regression curves and limit of detection (LOD) for modifications 5mdC, 5hmdC, and N6mdA are represented. Although our method also quantifies cadC and fdC, we did not detect significant concentrations of these in any of the measured samples; these modifications were hence omitted from the graphical illustrations. (B) 286 samples from 124 species were analyzed in the present study: 19 species from plants, 12 from animals, 6 from yeast, 2 from archaea, and 85 from bacteria. (C–D) Distribution of 5mdC, 5hmdC, and N6mdA across (C) archaeal, bacterial, and eukaryotic domains, and (D) animal, fungi, monera, plant, and protozoan kingdoms. The values depict percentage of cytosine residues bearing either methyl (%5mdC) or hydroxymethyl (%5hmdC) modification and percentage of adenine residues bearing methyl modification (N6mdA).

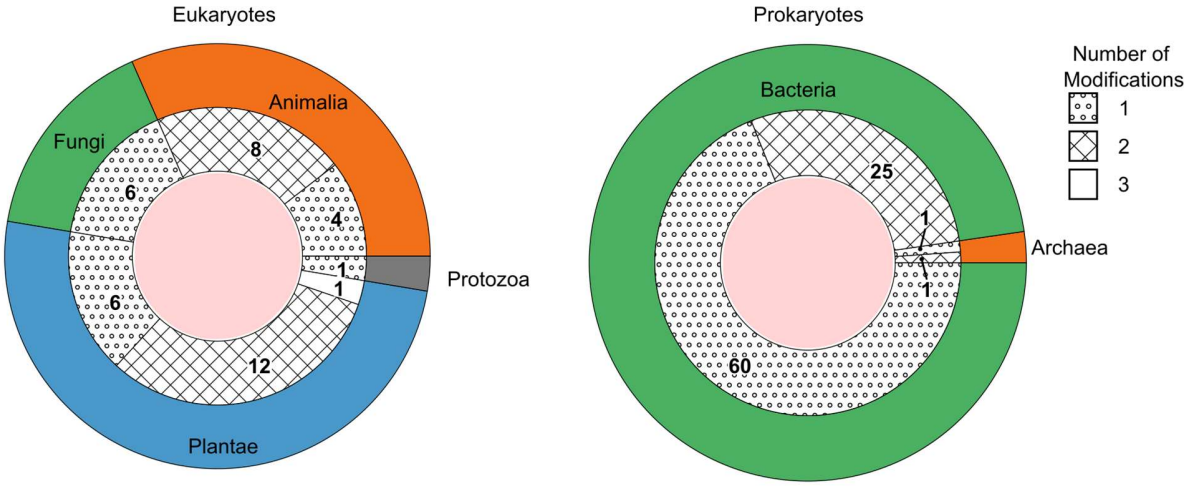


Figure 2. The number of species possessing one, two, or three DNA modification types grouped as eukaryotes (left) and prokaryotes (right). The outer ring represents the kingdoms present within these domains. The groupings per number of modifications are shown as fill patterns on the inner ring, where dots represent species in which only one among 5mdC, 5hmdC, and N6mdA were found; crosses represent species bearing two modifications simultaneously; and no fill represents species carrying all three modifications.

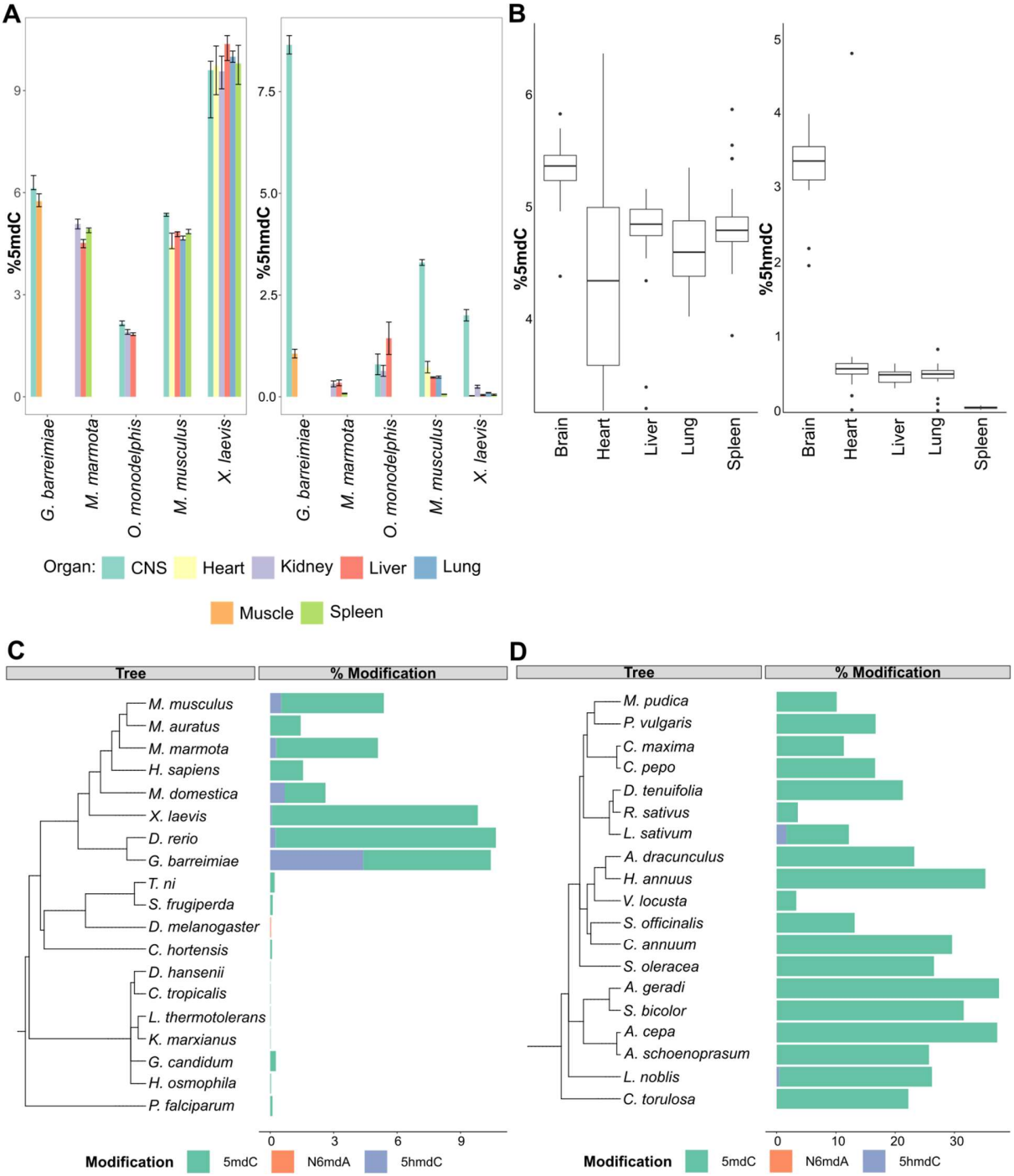


Figure 3. (A) The concentration of 5-methyl deoxycytidine (left) and 5-hydroxymethyl deoxycytidine (right) in different vertebrate genomes. (B) Distribution of 5-methyl deoxycytidine (left) and 5-hydroxymethyl deoxycytidine (right) in different mouse tissues. Variations in percentage modification across different (C) non-plant eukaryotes including representatives from

vertebrates like mammals, amphibians, and fish, invertebrates like insects and molluscs, and unicellular fungi and protozoa (D) plants species comprising both gymnosperms and angiosperms.

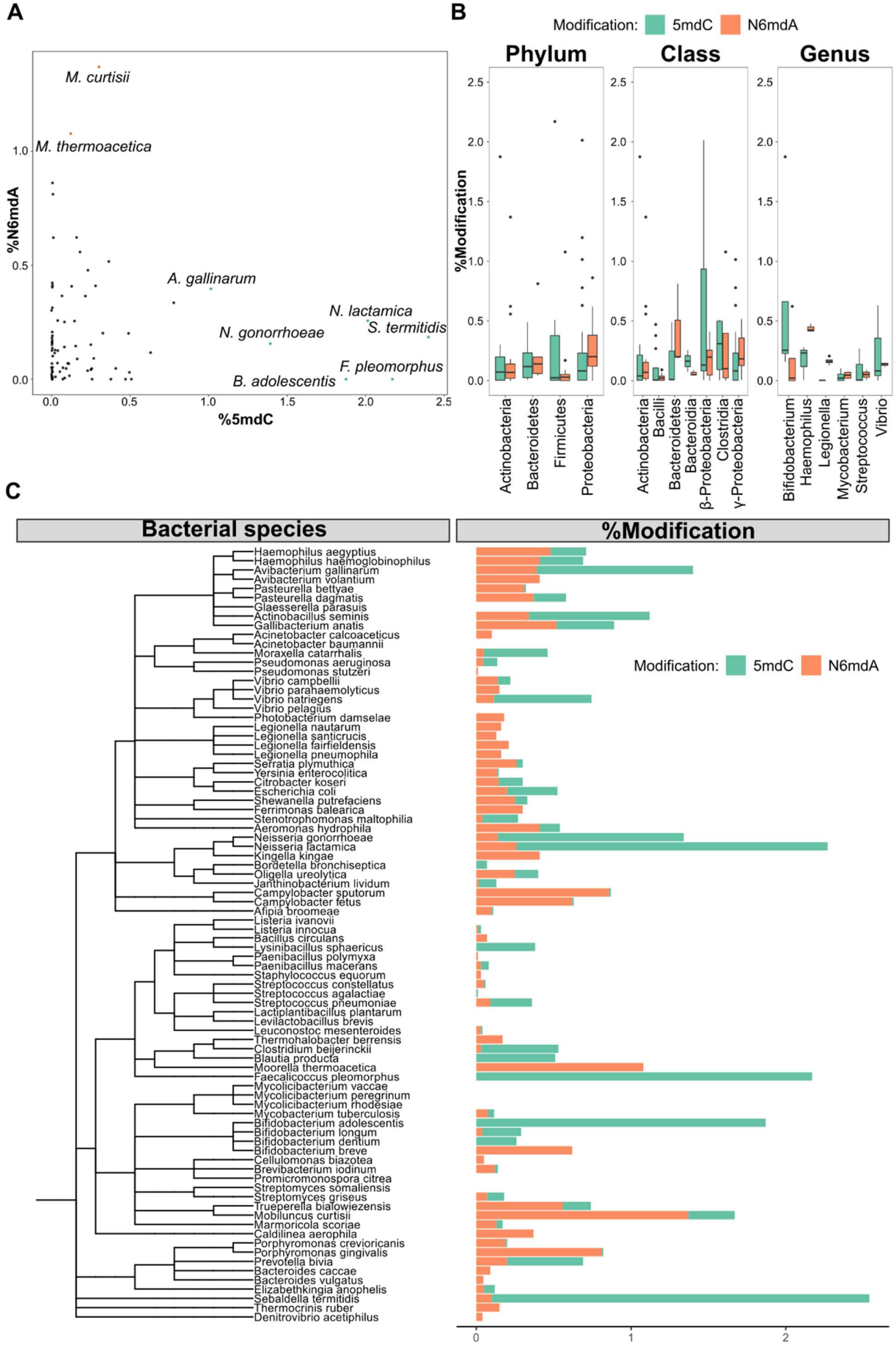

A
M. curtisii
M. thermoacetica
A. gallinarum
N. lactamica
S. termitidis
N. gonorrhoeae
F. pleomorphus
B. adolescentis
%N6mdA
%5mdC
B
Modification: 5mdC N6mdA
Phylum
Class
Genus
%Modification
Actinobacteria
Bacteroidetes
Firmicutes
Proteobacteria
Actinobacteria
Bacilli
Bacteroidetes
Bacteroidia
β-Proteobacteria
Clostridia
γ-Proteobacteria
Bifidobacterium
Haemophilus
Legionella
Mycobacterium
Streptococcus
Vibrio
C
Bacterial species
%Modification
Modification: 5mdC N6mdA
Haemophilus aegyptius
Haemophilus haemoglobinophilus
Avibacterium gallinarum
Avibacterium volantium
Pasteurella bettyae
Pasteurella dagmatis
Glaesserella parasuis
Actinobacillus seminis
Gallibacterium anatis
Acinetobacter calcoaceticus
Acinetobacter baumannii
Moraxella catarrhalis
Pseudomonas aeruginosa
Pseudomonas stutzeri
Vibrio campbellii
Vibrio parahaemolyticus
Vibrio natriegens
Vibrio pelagius
Photobacterium damselae
Legionella nautarum
Legionella santicrucis
Legionella fairfieldensis
Legionella pneumophila
Serratia plymuthica
Yersinia enterocolitica
Citrobacter koseri
Escherichia coli
Shewanella putrefaciens
Ferrimonas balearica
Stenotrophomonas maltophilia
Aeromonas hydrophila
Neisseria gonorrhoeae
Neisseria lactamica
Kingella kingae
Bordetella bronchiseptica
Oligella ureolytica
Janthinobacterium lividum
Campylobacter sputorum
Campylobacter fetus
Afipia broomeae
Listeria ivanovii
Listeria innocua
Bacillus circulans
Lysinibacillus sphaericus
Paenibacillus polymyxa
Paenibacillus macerans
Staphylococcus equorum
Streptococcus constellatus
Streptococcus agalactiae
Streptococcus pneumoniae
Lactiplantibacillus plantarum
Levilactobacillus brevis
Leuconostoc mesenteroides
Thermohalobacter berrensis
Clostridium beijerinckii
Blautia producta
Moorella thermoacetica
Faecalicoccus pleomorphus
Mycolicibacterium vaccae
Mycolicibacterium peregrinum
Mycolicibacterium rhodesiae
Mycobacterium tuberculosis
Bifidobacterium adolescentis
Bifidobacterium longum
Bifidobacterium dentium
Bifidobacterium breve
Cellulomonas biazotea
Brevibacterium iodinum
Promicromonospora citrea
Streptomyces somaliensis
Streptomyces griseus
Trueperella bialowiezensis
Mobiluncus curtisii
Marmoricola scoriae
Caldilinea aerophila
Porphyromonas crevioricanis
Porphyromonas gingivalis
Prevotella bivia
Bacteroides caccae
Bacteroides vulgatus
Elizabethkingia anophelis
Sebaldella termitidis
Thermocrinis ruber
Denitrovibrio acetiphilus

Figure 4. DNA modifications in bacteria. (A) Percentage of cytidine methylated against the percentage of adenine methylated in bacterial species. (B) Variation of % 5-methyl deoxycytidine and % N6-methyl deoxyadenosine among taxonomic divisions: phylum, class, and genus. (C) Distribution of 5mdC and N6mdA among 87 bacterial species measured against their phylogenetic diversity.

References

1. Miller JL, Grant PA. The role of DNA methylation and histone modifications in transcriptional regulation in humans. Subcell Biochem. 2013;61:289–317.

2. Breiling A, Lyko F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. Epigenetics Chromatin. 2015 Jul 21;8:24.

3. Guo JU, Ma DK, Mo H, Ball MP, Jang M-H, Bonaguidi MA, et al. Neuronal activity modifies the DNA methylation landscape in the adult brain. Nat Neurosci. 2011 Aug 28;14(10):1345–51.

4. Jessop P, Ruzov A, Gering M. Developmental Functions of the Dynamic DNA Methylome and Hydroxymethylome in the Mouse and Zebrafish: Similarities and Differences. Front Cell Dev Biol. 2018 Mar 20;6:27.

5. Calle-Fabregat C de la, Morante-Palacios O, Ballestar E. Understanding the Relevance of DNA Methylation Changes in Immune Differentiation and Disease. Genes [Internet]. 2020 Jan 18;11(1). Available from: http://dx.doi.org/10.3390/genes11010110

6. Day JJ, Sweatt JD. DNA methylation and memory formation. Nat Neurosci. 2010 Nov;13(11):1319–23.

7. Masser DR, Hadad N, Porter H, Stout MB, Unnikrishnan A, Stanford DR, et al. Analysis of DNA modifications in aging research. Geroscience. 2018 Feb;40(1):11–29.

8. Han M, Jia L, Lv W, Wang L, Cui W. Epigenetic Enzyme Mutations: Role in Tumorigenesis and Molecular Inhibitors. Front Oncol. 2019 Mar 29;9:194.

9. Cusack M, King HW, Spingardi P, Kessler BM, Klose RJ, Kriaucionis S. Distinct contributions of DNA methylation and histone acetylation to the genomic occupancy of transcription factors. Genome Res. 2020 Oct;30(10):1393–406.

10. Day JJ. DNA Modifications and Memory [Internet]. DNA Modifications in the Brain. 2017.

p. 95–111. Available from: http://dx.doi.org/10.1016/b978-0-12-801596-4.00006-x

11. Hotchkiss RD. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J Biol Chem. 1948 Aug;175(1):315–32.

12. Umer M, Herceg Z. Deciphering the epigenetic code: an overview of DNA methylation analysis methods. Antioxid Redox Signal. 2013 May 20;18(15):1972–86.

13. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet. 2013 Mar;14(3):204–20.

14. Hu C-W, Chen J-L, Hsu Y-W, Yen C-C, Chao M-R. Trace analysis of methylated and hydroxymethylated cytosines in DNA by isotope-dilution LC--MS/MS: first evidence of DNA methylation in Caenorhabditis elegans. Biochem J. 2015;465(1):39–47.

15. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011 Sep 2;333(6047):1300–3.

16. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009 May 15;324(5929):930–5.

17. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009 May 15;324(5929):929–30.

18. Janulaitis A, Klimasauskas S, Petrusyte M, Butkus V. Cytosine modification in DNA by BcnI methylase yields N4-methylcytosine. FEBS Lett. 1983 Sep 5;161(1):131–4.

19. Ehrlich M, Wilson GG, Kuo KC, Gehrke CW. N4-methylcytosine as a minor base in bacterial DNA. J Bacteriol. 1987 Mar;169(3):939–43.

20. Klimasauskas S, Gerasimaite R, Vilkaitis G, Kulakauskas S. N4,5-dimethylcytosine, a novel hypermodified base in DNA. Nucleic Acids Res Suppl. 2002;(2):73–4.

21. Dunn DB, Smith JD. Occurrence of a new base in the deoxyribonucleic acid of a strain of Bacterium coli. Nature. 1955 Feb 19;175(4451):336–7.

22. Couturier M, Lindås A-C. The DNA Methylome of the Hyperthermoacidophilic Crenarchaeon. Front Microbiol. 2018 Feb 8;9:137.

23. Liang Z, Shen L, Cui X, Bao S, Geng Y, Yu G, et al. DNA N-Adenine Methylation in Arabidopsis thaliana. Dev Cell. 2018 May 7;45(3):406–16.e3.

24. Heusipp G, Fälker S, Schmidt MA. DNA adenine methylation and bacterial pathogenesis. Int J Med Microbiol. 2007 Feb;297(1):1–7.

25. O'Brown ZK, Greer EL. N6-Methyladenine: A Conserved and Dynamic DNA Mark. Adv

Exp Med Biol. 2016;945:213–46.

26. Münzel M, Globisch D, Carell T. 5-Hydroxymethylcytosine, the sixth base of the genome. Angew Chem Int Ed. 2011;50(29):6460–8.

27. Capuano F, Mülleder M, Kok R, Blom HJ, Ralser M. Cytosine DNA methylation is found in Drosophila melanogaster but absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe, and other yeast species. Anal Chem. 2014 Apr 15;86(8):3697–702.

28. Chen L-Q, Zhao W-S, Luo G-Z. Mapping and editing of nucleic acid modifications. Comput Struct Biotechnol J. 2020 Mar 19;18:661–7.

29. Liu L, Zhang Y, Jiang D, Du S, Deng Z, Wang L, et al. Recent Advances in the Genomic Profiling of Bacterial Epigenetic Modifications. Biotechnol J. 2019 Jan;14(1):e1800001.

30. Le T, Kim K-P, Fan G, Faull KF. A sensitive mass spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples. Anal Biochem. 2011 May 15;412(2):203–9.

31. Chowdhury B, Cho I-H, Irudayaraj J. Technical advances in global DNA methylation analysis in human cancers. J Biol Eng. 2017 Mar 1;11:10.

32. Tang Y, Zheng S-J, Qi C-B, Feng Y-Q, Yuan B-F. Sensitive and simultaneous determination of 5-methylcytosine and its oxidation products in genomic DNA by chemical derivatization coupled with liquid chromatography-tandem mass spectrometry analysis. Anal Chem. 2015 Mar 17;87(6):3445–52.

33. Chilakala S, Feng Y, Li L, Mahfouz R, Quteba E, Saunthararajah Y, et al. Tracking Decitabine Incorporation into Malignant Myeloid Cell DNA in vitro and in vivo by LC-MS/MS with Enzymatic Digestion. Sci Rep. 2019 Mar 14;9(1):4558.

34. Gosselt HR, van Zelst BD, de Rotte MCFJ, Hazes JMW, de Jonge R, Heil SG. Higher baseline global leukocyte DNA methylation is associated with MTX non-response in early RA patients. Arthritis Res Ther. 2019 Jun 26;21(1):157.

35. Tsuji M, Matsunaga H, Jinno D, Tsukamoto H, Suzuki N, Tomioka Y. A validated quantitative liquid chromatography-tandem quadrupole mass spectrometry method for monitoring isotopologues to evaluate global modified cytosine ratios in genomic DNA. J Chromatogr B Analyt Technol Biomed Life Sci. 2014 Mar 15;953-954:38–47.

36. Binz T, D'Mello N, Horgen PA. A comparison of DNA methylation levels in selected isolates of higher fungi. Mycologia. 1998 Sep;90(5):785.

37. Nai Y-S, Huang Y-C, Yen M-R, Chen P-Y. Diversity of Fungal DNA Methyltransferases and Their Association With DNA Methylation Patterns. Front Microbiol. 2020;11:616922.

38. Lyko F, Ramsahoye BH, Jaenisch R. DNA methylation in Drosophila melanogaster. Nature. 2000 Nov 30;408(6812):538–40.

39. Zhang G, Huang H, Liu D, Cheng Y, Liu X, Zhang W, et al. N6-methyladenine DNA modification in Drosophila. Cell. 2015 May 7;161(4):893–906.

40. O'Brown ZK, Boulias K, Wang J, Wang SY, O'Brown NM, Hao Z, et al. Sources of artifact in measurements of 6mA and 4mC abundance in eukaryotic genomic DNA. BMC Genomics. 2019 Jun 3;20(1):445.

41. Kong Y, Cao L, Deikus G, Fan Y, Mead EA, Lai W, et al. Critical assessment of DNA adenine methylation in eukaryotes using quantitative deconvolution. Science. 2022 Feb 4;375(6580):515–22.

42. Bhattacharyya M, De S, Chakrabarti S. Origin and Evolution of DNA methyltransferases (DNMT) along the tree of life: A multi-genome survey [Internet]. bioRxiv. 2020 [cited 2021 Sep 21]. p. 2020.04.09.033167. Available from: https://www.biorxiv.org/content/10.1101/2020.04.09.033167v1

43. Colwell M, Drown M, Showel K, Drown C, Palowski A, Faulk C. Evolutionary conservation of DNA methylation in CpG sites within ultraconserved noncoding elements. Epigenetics. 2018 Feb 6;13(1):49–60.

44. De Bustos C, Ramos E, Young JM, Tran RK, Menzel U, Langford CF, et al. Tissue-specific variation in DNA methylation levels along human chromosome 1. Epigenetics Chromatin. 2009 Jun 8;2(1):7.

45. He Y, Hariharan M, Gorkin DU, Dickel DE, Luo C, Castanon RG, et al. Spatiotemporal DNA methylome dynamics of the developing mouse fetus. Nature. 2020 Jul;583(7818):752–9.

46. Zhu Q, Stöger R, Alberio R. A Lexicon of DNA Modifications: Their Roles in Embryo Development and the Germline. Front Cell Dev Biol. 2018 Mar 27;6:24.

47. Globisch D, Münzel M, Müller M, Michalakis S, Wagner M, Koch S, et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS One. 2010 Dec 23;5(12):e15367.

48. Mahmood AM, Dunwell JM. Evidence for novel epigenetic marks within plants. AIMS Genet. 2019 Dec 24;6(4):70–87.

49. Mishra PK, Baum M, Carbon J. DNA methylation regulates phenotype-dependent transcriptional activity in Candida albicans. Proc Natl Acad Sci U S A. 2011 Jul 19;108(29):11965–70.