

Article

Evolutionary History of RNA Modifications at N6-Adenosine Originating from the R-M System in Eukaryotes and Prokaryotes

Congshan Liu, Jianping Cao , Haobing Zhang and Jianhai Yin * 

National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention (Chinese Center for Tropical Diseases Research), NHC Key Laboratory of Parasite and Vector Biology, WHO Collaborating Center for Tropical Diseases, National Center for International Research on Tropical Diseases, Shanghai 200025, China; liucs@nipd.chinacdc.cn (C.L.); caojp@chinacdc.cn (J.C.); zhanghaobing2@163.com (H.Z.)

* Correspondence: chart2543@163.com or yinhj@nipd.chinacdc.cn

Simple Summary: The m⁶A is the most abundant and well-studied modification of mRNA, and plays an important role in transcription and translation. It is known to be evolutionarily conserved machinery present in the last eukaryotic common ancestor (LECA). The writers and erasers responsible for adding or removing m⁶A belong to specific protein families, respectively, suggesting that these members are evolutionarily related. However, only some of these mRNA m⁶A modification-associated proteins have been studied from an evolutionary perspective, while there has been no comprehensive and systematic analysis of the distributions and evolutionary history of N6mA-associated proteins in the three kingdoms of life. In this study, we identified orthologues of all the reported N6mA-associated proteins in 88 organisms from three kingdoms of life and comprehensively reconstructed the evolutionary history of the RNA N6mA modification machinery. The results demonstrate that RNA N6mA-MTases are derived from at least two different types of prokaryotic DNA MTases (class α and β MTases). As the m⁶A reader, YTH proteins may be acquired by LECA from an individual prokaryotic YTH-domain protein that evolved from the N-terminals of an R-M system endonuclease. In addition, the origin of eukaryotic ALKBH family proteins is inferred to be driven by at least two occasions of independent HTG from the bacterial ALKB family.



Citation: Liu, C.; Cao, J.; Zhang, H.; Yin, J. Evolutionary History of RNA Modifications at N6-Adenosine Originating from the R-M System in Eukaryotes and Prokaryotes. *Biology* **2022**, *11*, 214. <https://doi.org/10.3390/biology11020214>

Academic Editors: Brian McDonagh and Je-Hyun Yoon

Received: 8 November 2021

Accepted: 24 January 2022

Published: 28 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Methylation at the N6-position of adenosine (N6mA) on mRNA (m⁶A) is one of the most widespread, highly selective and dynamically regulated RNA modifications and plays an important role in transcription and translation. In the present study, a comprehensive analysis of phylogenetic relationships, conserved domain sequence characteristics and protein structure comparisons were employed to explore the distribution of RNA N6mA modification (m⁶A, m^{6,6}A, m⁶Am, m^{6,6}Am and m^{6,6}t⁶A)-associated proteins (writers, readers and erasers) in three kingdoms of life and reveal the evolutionary history of these modifications. These findings further confirmed that the restriction-modification (R-M) system is the origin of DNA and RNA N6mA modifications. Among them, the existing mRNA m⁶A modification system derived from the last eukaryotic common ancestor (LECA) is the evolutionary product of elements from the last universal common ancestor (LUCA) or driven by horizontal gene transfer (HGT) from bacterial elements. The subsequent massive gene gains and losses contribute to the development of unique and diverse functions in distinct species. Particularly, RNA methyltransferases (MTases) as the writer responsible for adding N6mA marks on mRNA and ncRNAs may have evolved from class α and β prokaryotic “orphan” MTases originating from the R-M system. The reader, YTH proteins that specifically recognize the m⁶A deposit, may be acquired by LECA from an individual prokaryotic YTH-domain protein that evolved from N-terminals of an R-M system endonuclease. The eraser, which emerged from the ALKB family (ALKBH5 and FTO) in eukaryotes, may be driven by independent HTG from bacterial ALKB proteins. The evolutionary history of RNA N6mA modifications was inferred in the present study, which will deepen our understanding of these modifications in different species.

Keywords: RNA modifications; epigenetics; m⁶A; RNA methyltransferases; demethylases; evolution

1. Introduction

As described by Francis Crick's central dogma, RNA is the only direct product of DNA and is subsequently exported from the nucleus to the cytoplasm, where mRNA is translated into proteins [1,2]. RNA is also known to be a key player in transcription, translation and cellular decision-making in the form of noncoding RNAs (ncRNAs), such as rRNAs, tRNAs, microRNAs, piRNAs and long noncoding RNAs [3,4]. During these processes, gene activity, expression and function can be regulated not only by alterations in RNA sequences but also by the mechanism that is not associated with any change in the nucleic acid sequence itself, which is referred to as RNA modifications or 'RNA epigenetics' [5]. Chemical modifications are found in almost all types of RNA, adding additional complexity to the information carried by RNA. These modifications influence the RNA structure and its interaction with other molecules, ultimately giving rise to the diverse functions of RNA molecules, especially within complex regulatory networks, where small, subtle structural changes can bring about significant changes in cellular metabolism by affecting mRNA stability, splicing, translational efficiency and pri-microRNA processing, among other processes [6,7].

RNA modifications occur across Eukaryotes, Bacteria and Archaea, indicating the conservation and importance of modifications at the RNA level [8]. Among a multitude of RNA modifications, m⁶A, one form of methylation at the N6 position of adenosine (N6mA) (Figure 1) on mRNA, was first discovered in the 1970s [9,10]. However, its functional importance and related enzymes have remained mysterious for decades until recently [8]. With a growing understanding of the functional significance of m⁶A modification in mRNA, the identification of m⁶A methyltransferase (METTL3/METTL14 complex) [11] and demethylase (FTO and ALKBH5) [12,13], and the development of next-generation sequencing (NGS), which allowed the m⁶A landscape to be delineated in the transcriptome [14,15], m⁶A has re-emerged at the forefront. It has been demonstrated that m⁶A is one of the most widespread internal modifications of messenger RNAs (mRNAs) and ncRNAs [6,16] and is highly selective and dynamically regulated in all aspects of physiological functions [17]. For instance, m⁶A play roles in protein homeostasis, proper cellular signaling and so on by affecting mRNA stability, mRNA structure, mRNA export, translation efficiency and translation [6,7,17], similar to other RNA modifications. Moreover, mutations in m⁶A modification enzymes have been linked to human diseases, including cancer, cardiovascular diseases, metabolic diseases and neurological disorders, highlighting their importance under both physiological and pathological conditions [18]. Recently, studies on other types of N6mA, such as m^{6,6}A, m⁶Am, m^{6,6}Am and m⁶t⁶A (Figure 1), have revealed their key roles in regulating mRNA stability, pre-mRNA splicing and the stress response [19–21], adding the complexity and diverse function of N6mA RNA modifications.

The pattern of m⁶A in mRNA relies on the complex interactions among three distinct classes of protein factors: "writers", which are m⁶A methyltransferases (MTases); "erasers", which are m⁶A-demethylating enzymes that mediate dynamic modifications; and "readers", which are m⁶A-binding proteins that recognize and regulate modifications, contributing to their cellular effects. The best-studied "writer" executes the deposition of m⁶A modifications via a multicomponent complex (MTC) composed of a METTL3-METTL14 heterodimer and regulatory factors, including WTAP, VIRMA, ZC3H13, RBM15 and HAKAI, by catalyzing the addition of a methyl group to adenosine in the context of a conserved consensus sequence (RRACH) [17,22]. In this complex, WTAP stabilizes the interaction between the two METTL proteins, while the other regulatory factors have been proposed to function in guiding methylation to targeted sites [22]. To date, only two demethylases, FTO and ALKBH5, have been known to remove mRNA m⁶A modifications [12,13]. Moreover, YTHs, HNRNPs, IGF2BPs, eIF3, FMRP1, ELAV1 and ribosomes affect the fate of RNAs by act-

ing as m^6A “readers” [17,23]. Together, these elements constitute the core mRNA m^6A modification components, regulating the functions of this modification. Recently, other MTases, such as METTL16, ZCCHC4 and METTL5, have been identified as m^6A MTases that can catalyze the deposition of m^6A marks on ncRNAs [2,24]. In addition, writers and erasers of other N6mA have been identified [20,21,25–29] (Figure 1), but no readers have yet been found.

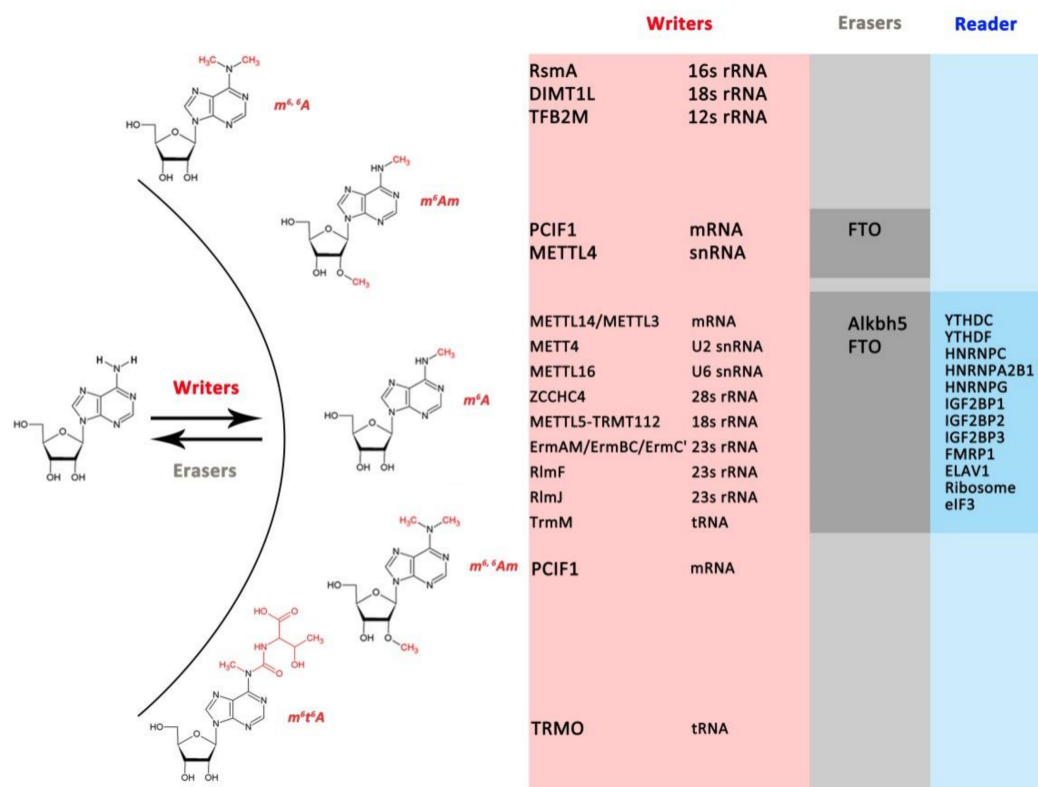


Figure 1. RNA methylation at N6 position. m^6Am : N6,2'-O-dimethyladenosine; m^6t^6m : N6-methyl-N6-threonylcarbamoyladenine; m^6A : N6-methyladenosine; $m^6,6A$: N6,N6-dimethyladenosine; $m^6,6Am$: N6,N6,2'-O-trimethyladenosine.

The mRNA m^6A modification is known to be an evolutionarily conserved machinery present in the last eukaryotic common ancestor (LECA), according to comparative analyzes of m^6A -associated proteins and their modification sites across closely related species [8,18,30] and very distant organisms [2]. Furthermore, the writers and erasers that regulate N6mA modifications belong to specific protein families, suggesting that these members are evolutionarily related. However, only some of them have been studied evolutionarily [31–36], and there has been no comprehensive and systematic analysis of the distributions and evolutionary history of N6mA-associated proteins in the three kingdoms of life. In this study, we identified homologues of N6mA-associated proteins (also referring to the components of N6mA modifications) in 88 organisms from three kingdoms of life and comprehensively reconstructed the evolutionary history of the RNA N6mA modification machinery, including writers (MTases), erasers (ALKHB family) and readers (YTH-domain-containing protein).

2. Materials and Methods

2.1. Sequence Similarity Searches

Candidate homologues of known m^6A , $m^6,6A$, m^6Am , $m^6,6Am$ and m^6t^6A writers, readers and erasers (Figure 1, Table S1) were identified by scanning a panel of 88 taxa (Table S2, including 17 Bacteria, 6 Archaea, 7 Discoba, 2 Metamonada, 14 SAR, 6 Archaeplastida and 36 Amorphea). The list of 47 reported and well-studied prokary-

ote and human protein sequences involved in these RNA modifications were used as an initial set (accession numbers for initial queries are provided in Table S1). Homologues were identified by using web-based PSI-BLASTP searches among NCBI non-redundant protein sequences (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>, accessed on 10 January 2021), with an E-value cut-off < 0.05. For *Euglena gracilis*, which lacks a high-quality genome in NCBI, the sequences were extracted from the proteome database (<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD009998>, accessed on 25 January 2021) with the basic local alignment search tool (BLAST 2.2.25) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>, accessed on 10 January 2021). All hits (E-value cut-off < 0.05) that were obtained for the same species were pooled, and redundant sequences were deleted using the CD-HIT suite [37] with a cut-off of 0.98. To determine the true sequences in the BLAST hits, only the sequences containing the domains as the initial queries (Table S3) were retained. The domain research was carried out by Batch CD-Search at NCBI (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>, accessed on 10 January 2021). There was one exception, ZC3H13, which lacked the reported domain in some species. Only the top hits were used for the phylogenetic analysis. The amino acid residues were visualized with Jalview Version 2 [38].

2.2. Alignment Generation and Phylogenetic Analyzes

Sequences were aligned using the web-based COBALT tool based on the conserved domain and local sequence similarity [39] with default parameters. To infer the evolutionary history of some protein families, such as the AdoMet_MTases superfamily, ALKBH family, YTHs, HNRNPs and IGF2BPs, the homologous sequences of all the species were pooled together before aligning the sequences. PCIF1 is not included in the phylogenetic tree of AdoMet_MTases because of its long branch attraction. These alignments were manually trimmed to retain only the columns with at least 30% letters in a column that were not gap characters and at least 10% of the pairs of letters in an alignment column that were conserved (Table S4). The alignments used for phylogenetic analyzes can be found in Supplementary File S1.

The online IQ-TREE platform [40], an algorithm to infer phylogenetic relationships by maximum likelihood, was used to find the best fitting model (Table S4) and build phylogenetic trees. The tree branches were tested with ultrafast bootstrapping (1000) and SH-like approximate likelihood ratio tests (SH-aLRT, 1000 replicates). The final trees were visualized with TreeGraph2 [41]. Gene gains and losses were inferred based on phylogenetic trees and conserved domains (sequence architecture). Using the core domains, the specific human tree of N6mA modifications related to AdoMet_MTases and the ALKBH family was also generated to facilitate visualization of the different families.

Orthologues, paralogues, xenologues and gene losses were further inferred according to the phylogenetic topologies. Homologous sequences that have diverged due to speciation events were identified as orthologies. The duplicated genes are searched through the identification of paralogous relationships. After comparing with the putative history of the species, xenologues were identified as genes in one species derived from a distantly related species through HGT. To minimize the genome quality effects, only the lineage-specific losses were defined as gene losses.

In order to ensure these gene losses are contributed by evolution rather than limitation of methods and genome quality, this study make the following endeavors: (1) In order to include as much sequences of studied species as possible, NCBI's non-redundant protein sequences, including all non-redundant GenBank CDS translations + PDB + SwissProt + PIR + PRF excluding environmental samples from WGS projects search sequences were used in the sequences searching; (2) All sequences with E-value > 0.05 rather than the top hits were included when we searched for homologues, and then a series of screenings and phylogenetic relationship analysis were used to ensure the integrity of the included sequences and the accuracy of classification in considered species; (3) The results have been validated through model organisms, including human, *D. melanogaster*, *C. elegans*

and *S. cerevisiae*, which has been reported as the m⁶A modification-associated elements and sequences; (4) Only the lineage-specific losses were identified as gene losses, so as to exclude sequence losses due to the genome quality; (5) NCBI and Ensemble are the two completed and qualified genome database, and also the main database used for orthology identification, such as OrthoMAM (<https://orthomam.mbb.cnrs.fr/>, accessed on 25 January 2021) and OrthoDB (https://www.orthodb.org/orthodb_userguide.html, accessed on 25 January 2021), so we searched the homologues of m⁶A associated protein in prokaryotes with Ensemble database (Table S5).

2.3. Structural Inferion and Similarity Comparison

Data for experimentally determined structures (Table S1) were retrieved from the PDB database (<https://www.rcsb.org>, accessed on 25 January 2021). Data for other structures without released crystal structures were extracted from the SIWSS-MODEL Repository database [42] (Table S1). Detailed information is summarized in Supplementary Table S1. These structures were visualized, and the similarity of protein structures, represented as the root mean square deviation (RMSD), was calculated using Open-Source PyMOL [43].

2.4. Structure-Based Alignments

For sequences with diverse amino acid sequences but similar topologies, PROMALS3D was used to accurately build multiple alignments based on the released 3D structure [44].

3. Results

3.1. The Distribution of N⁶mA Modification Components in Prokaryotes and Eukaryotes

3.1.1. mRNA m⁶A Modification Components in Prokaryotes and Eukaryotes

The homologues of mRNA m⁶A modification-associated proteins (Figure 1) are found in only eukaryotic genomes rather than prokaryotes (Figure 2), suggesting that the mRNA m⁶A modification system originated from LECA. As the core components, the writer complexes METTL3/METTL14, WTAP and RMB15, readers YTHDCs and YTHDFs, and erasers FTO and ALKBH5 are conserved in the major phylum of eukaryotes. However, some mRNA m⁶A modification-associated proteins are lineage-specific. For instance, HNRNPC and FMBP1 were identified in organisms that emerged evolutionarily after *Salpingoeca rosetta* and *Nematostella vectensis*, which are the closest single-celled relatives or progenitors of Metazoa, and were retained in the majority of Metazoa during subsequent evolution. HNRNPA2B1, HNRNPG, IGF2BPs and ELAV1 could be identified only in vertebrates, suggesting that they might be gene gains in the common ancestor of vertebrates. VIRMA and HAKAI were highly divergent between plants and animals and were not present in lower eukaryotes (Figure 2). Although most mRNA m⁶A modification-associated proteins were absent in prokaryotes, only with some sporadic m⁶A modification components identified in Bacteria or Archaea, such as METTL3/METTL14 in *Anabaena PCC 7122*, FTO in *Streptomyces subbrutis*, YTHDC in *Candidatus Bathyarchaeota archaeon* and YTHDF in *Escherichia coli*. Hence, the spread of these genes might be mediated by horizontal gene transfer (HGT) between eukaryotes and prokaryotes (Figures 2, S1, S9 and S16).

The METTL3-METTL14 heterodimer was lost in some species' lineages, such as kinetoplastida, heterokonts, fungus and nemathelminthes (Figure 2). Interestingly, the gene loss patterns of METTL3/METTL14, WTAP and the YTH family (especially YTHDC) were highly consistent (Figure 2), indicating that they coevolved as the core elements of mRNA m⁶A modification. In addition, both FTO and ALKBH5 were lost in most Amorphea (except for vertebrates) as well as in Metamonada, early Discoba and some SAR taxa (Figure 2).

3.1.2. The Distribution of Other N⁶mA-MTases in Prokaryotes and Eukaryotes

The majority of other N⁶mA-MTases, such as METTL4, METTL16, METTL5, TFB1M, TrmM and TRMO, might have originated in the early stages of evolution prior to the last universal common ancestor (LUCA), while DIMT1L/RsmA and PCIF1 were only found in eukaryotes, which were traceable to LECA (Figure 2). In addition, the distribution of some

N6mA-MTases was lineage-specific; for instance, ZCCHC4 was present in only Amorphea, TrmM was lost in Opsthoconts, and RlmJ was a bacteria-specific m⁶A methylase (Figure 2). Furthermore, unlike METTL3-METTL14, these N6mA-MTases are mainly modified ncRNAs rather than mRNAs, suggesting that ncRNA modifications are conserved machinery with prokaryotic origins and are retained in eukaryotes.

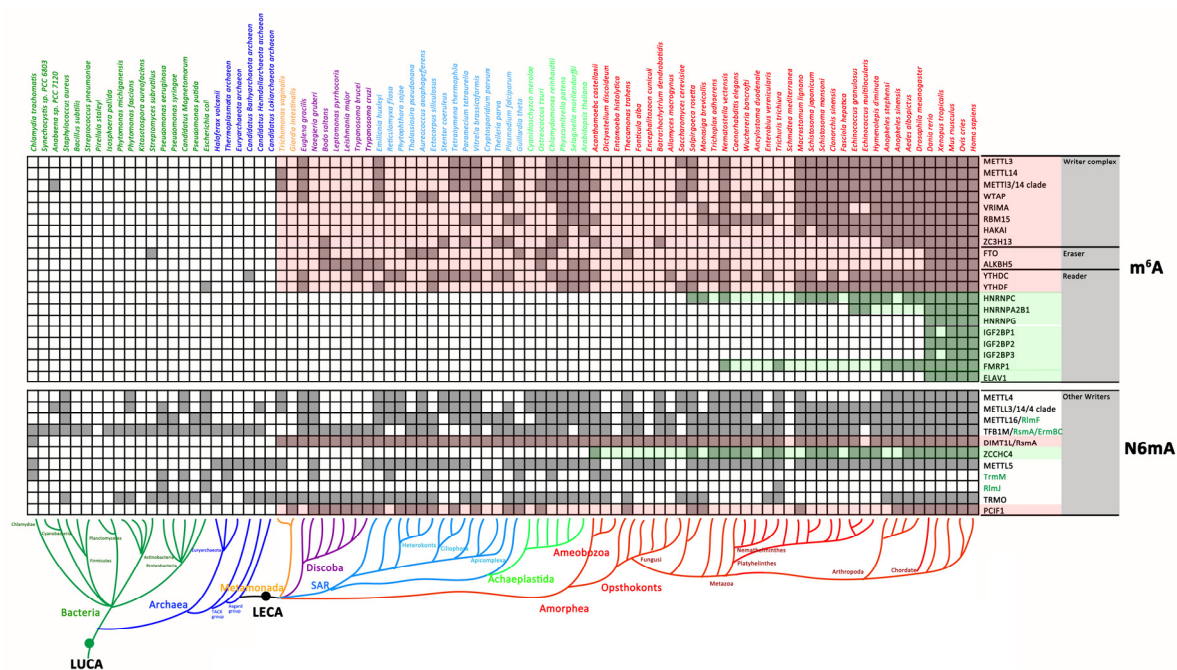


Figure 2. Distribution of RNA N6mA modification components in prokaryotes and eukaryotes. The landscape map was delineated based on Figures S1–S14. The red background highlights the gene gains before LECA; the green background highlights the gene gains in Amorphea.

3.2. The Evolutionary History of the RNA N6mA Modification Machinery

3.2.1. The Writer: RNA N6mA-MTases

AdoMet-dependent MTases are divided into five structurally distinct classes: I (Rossmann-Fold MTases, RFM), II, III, IV (SPOUT) and V [45]. Among them, only RFM and SPOUT MTases can methylate RNA [32,33]. TrmO, with a unique single-sheeted β -barrel structure close to that of *E. coli* yaeB, was identified to add m⁶t⁶A to tRNA (Thr) [46], while other reported N6mA-MTases exclusively belong to the RFM class [31,47,48] (Figure 3). In this study, the conserved N6mA-MTases from 88 taxa were clustered into different clades in the phylogenetic tree (Figure S1). We retained the well-studied eukaryotic and bacterial RNA N6mA-MTases in a simplified tree to easily clarify their evolutionary relationships (Figure 3A). Evolutionary analysis of N6mA MTases suggested that there was a common ancestor prior to LUCA for these RNA N6mA-MTases (except TrmO). Subsequently, they were independently acquired by LECA from separate prokaryotic ancestors (except ZCCHC4 and PCIF1), which might be products of HGT (Figure 3A).

Inspired by the consensus that RNA MTase and DNA MTases came from a common ancestor in the restriction-modification (R-M) system [31,34,35,49], bacterial “orphan” MTases (Class α DNA MTase Dam, Class β DNA MTase CrrM and EcoGII) were recruited in our study for the analysis of the evolutionary relationship of RNA N6mA-MTases based on their protein structures (Figure 3). Unlike MTases in the bacterial and archaeal R-M system that pair with a restriction endonuclease for self-non-self discrimination, “orphan” MTases are involved in chromosome replication, DNA repair and epigenetic gene regulation [31,50,51]. Furthermore, although both Dam and CrrM are RFMs and act on DNA N6mA, their methylation mechanisms are quite different, indicating that they evolved independently [31],

which is consistent with our finding that RNA N6mA-MTases originated from at least two types of DNA MTases, that is, Class β and Class α DNA MTases (Figure 3).

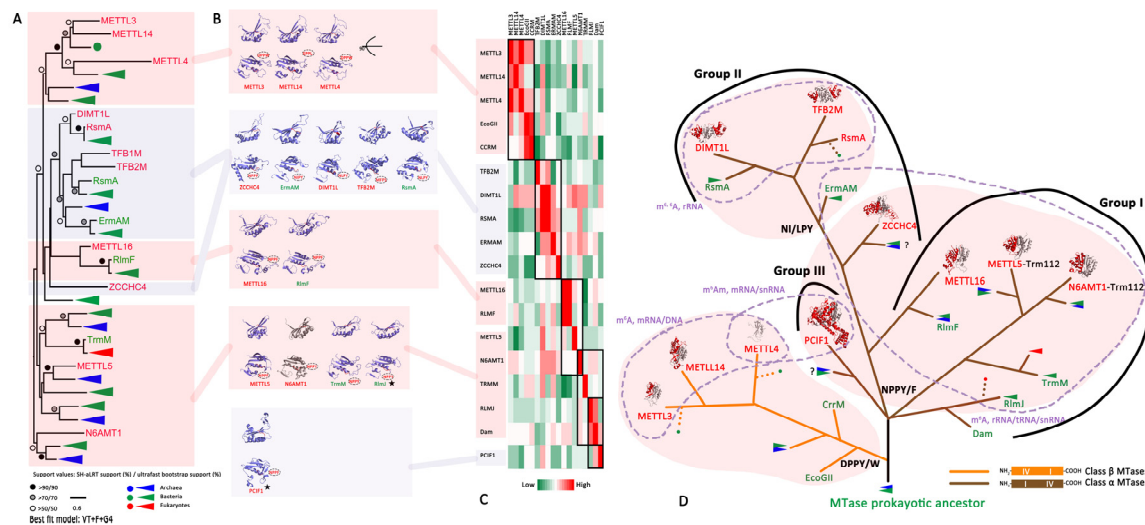


Figure 3. Evolution and characteristics of RNA N6mA methyltransferases. (A) The phylogenetic tree of RNA N6mA-MTases (details in Figure 1). (B) The structure of the catalytic domain of MTases. The conserved four-residue consensus is labeled in red. RFMs use AdoMet as the methyl group donor and show a conserved catalytic mechanism: a four-residue consensus (D/N/S/H-PP-W/F/Y) is generated in motif IV (methylation catalysis motif), where the bulky hydrophobic side chain (W/F/Y) stacks against the target base with π - π interactions, while the polar group of the first residue (D/N/S/H) abstracts a proton from the target NH_2 group. (C) Structural comparison of MTD in MTases. A heatmap was drawn based on the root mean square deviation (RMSD). (D) The hypothesis of the evolutionary history of N6mA MTases. RFMs can be categorized into Class α (DNA MTase Dam), β (DNA MTase CrrM) and γ according to their circular permutation in the order of motifs. Class α Dam, brown branch; Class β EcoGII, orange branch. There are three groups of Class α N6A MTases according to their structural similarity and four conserved residues. MTases in the same background retain similar catalytic motifs.

METTL3, METTL14 and METTL4, as the members of MT-A70 (Figure 3A), were inferred to originate from precursors of the DNA MTases CrrM and EcoGII for three reasons. First, all of these RNA and DNA MTases belong to class β MTases that can generate N6mA on double-stranded DNA (dsDNA), single-stranded DNA (ssDNA) or mRNA [31,52–54]. Second, there were more structural similarities of the methylation transferase domains (MTDs) between human METTL3/METTL14/METTL4 and bacterial CrrM/EcoGII than to other RNA N6mA-MTases (Figure 3C). Finally, there was a consistency of the motif order (IV-I) and the DPPY/W motif in their catalytic domain (Figure 3B,D). In addition, unlike METTL3-METTL14, which prefers single-stranded nucleic acids, another MT-A70 member, METTL4, is predicted to be the primary N6mA DNA MTase [55,56]. Hence, active METTL3 might represent by the gene duplication of METTL4 and enable a functional shift towards RNA methylation, followed by the emergence of inactive METTL14 as its partner.

Another group of Class α MTase (motif order is I-IV) [31,33,57], RNA N6mA-MTases, mostly containing N6mA modifications in the ncRNA, might have evolved from a precursor of DNA MTase Dam (Figure 3D). Although the EPPV motif of *E. coli* Dam differed from the NPPY motif of *E. coli* RlmJ, the root mean square deviation (RMSD) of their three-dimensional structures was 2.96 Å (Figure 3C, Table S6), indicating that these two proteins had close evolutionary relationships compared to other RNA MTases. We divided known RNA N6mA-MTases into three evolutionary groups according to their structure and functions (Figure 3D). Group I included RlmJ, TrmM, METTL5, N6AMT1, RlmF and METTL16, which contributed to m^6A modification on ncRNAs and showed the consistent

catalytic motifs NPPF/Y. Among them, some MTases have established special catalytic mechanisms and functions. For instance, METTL5 relies on Trm112 for stabilization [58], and N6AMT1/HemK-Trm112 is a protein MTase rather than a nucleic acid enzyme [59,60]. In Group II, ZCCHC4, ErmAM, RsmA, DIMT1L and TFB2M had high structural similarities (Figure 3B,C). ErmAM was thought to be the ancestor of RsmA, DIMT1L and TFB2M, all of which had consistent NI/LPY catalytic motifs. PCIF1 was the only member of Group III (Figure 3D) but was not included in the phylogenetic tree because of its long branch attraction (Figure 3A). Despite an absence of homologues in prokaryotic organisms, the NPPF motif and unique function of PCIF1 (catalyzing m⁶Am and m^{6,6}Am modifications of mRNA) [25,61] indicate that it may have been acquired by eukaryotes from prokaryotes prior to their radiation from LECA.

3.2.2. The Eraser: ALKBH Family Demethylases

ALKB homologues (ALKBHs) are a family of specific demethylases that depend on Fe²⁺ and α -ketoglutarate (α -KG) to catalyze demethylation on different substrates, including dsDNA, ssDNA, mRNA, tRNA and proteins [36]. Their catalytic domain contains a conserved double-stranded β -helix domain (DSBH) known as “jelly roll” fold, which forms the metal-(HXD-H) and cofactor-binding (N-Y-R-R) pockets (Figure 4C,D). The sequence characteristics, structures and functions of the ALKB family have been well described [36,62–64]. Here, their evolutionary relationships are carried out in a comprehensive way by integrating their sequences and structural characteristics of the active DBSH domain in the present study.

The phylogenetic analysis confirmed that the eukaryotic ALKBH family was derived from the bacterial DNA repair enzyme ALKB, and no ALKB homologues were found in Archaea [36,65] (Figures 4, S9 and S15). Importantly, we provide new insight that the eukaryotic ALKBH family evolved from different bacterial ALKB enzymes. That is, ALKBH1 and ALKBH10 originated from one bacterial ALKB group (Group I, Figure 4), while other ALKBHs originated from another bacterial ALKB group (Group II, Figure 4). In addition, we identified ALKBH10 as a novel eukaryotic ALKBH protein that was lost in metazoans during evolution (Figure S15). Considering that ALKBH10 is evolutionarily more closely related to *E. coli* ALKB than ALKBH1 (Figure 4C), we inferred that ALKBH1 arose from the gene duplication of ALKBH10 during eukaryogenesis and appeared before LECA. Therefore, the newly discovered ALKBH10 retained in lower eukaryotes might play a similar role in mediating DNA repair as human ALKBH1 and *E. coli* ALKB.

The ALKBHs predicted to evolve from Group II bacterial ALKB (brown branch in Figure 4D) were clustered into one clade in phylogenetic trees (Figure 4A). Among them, ALKBH2/ALKBH3 might emerge at earlier evolutionary stages, while others were derived via the gene duplication of ALKBH2 or ALKBH3 (Figures 4A, S2 and S15). These ALKBHs underwent functional differentiation to obtain certain functions, such as the direct reversal of alkylation damage in DNA (ALKBH2 and ALKBH3), regulation of epigenetic or epitranscriptomic nucleic acid methylation (ALKBH5 and FTO) [12,13], posttranslational modifications by acetylating proteins (ALKBH4 and ALKBH7) [62,66,67] or acting as a hydroxylating enzyme on tRNA rather than as a demethylase (ALKBH8) [68].

3.2.3. The Reader: YTH Family

The YTH domain, consisting of four α -helices and six β -strands, serves as the module for recognizing m⁶A in a methylation-dependent manner [69,70]. Three YTHDF and two YTHDC proteins have been identified in the human genome to recognize m⁶A markers rather than other RNA modifications [70]. These proteins were previously found only in eukaryotes [71,72], but their prokaryotic origin was recently indicated by two striking similar structures: the McrB-N domain (McrB, a methylation-dependent endonuclease in the R-M system) from the archaeon *Thermococcus gammatolerans* and MJECL36 from *Methanocaldococcus jannaschii* [73–75].

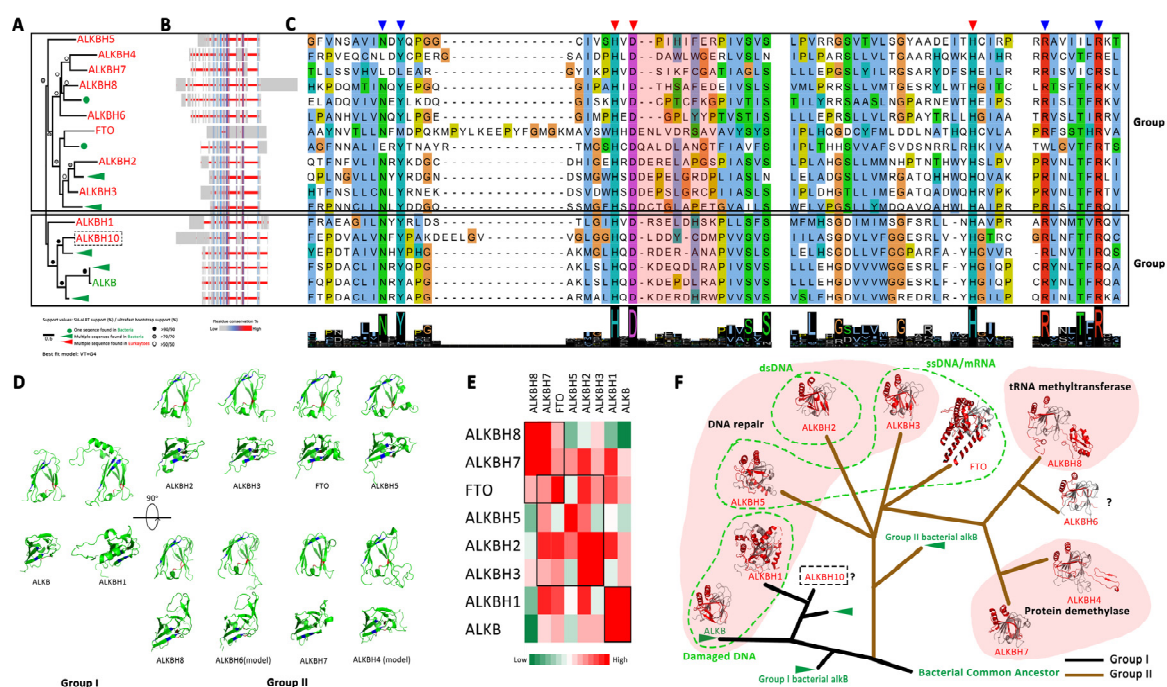


Figure 4. Evolution and characteristics of the ALKBH family. (A) The phylogenetic tree of the ALKBH family (details in Figure S9). (B) Alignments of the ALKBH family, ALKBH5 (NP_060228.3), ALKBH4 (NP_060091.1), ALKBH7 (NP_115682.1), ALKBH8 (NP_620130.2), *Pseudomonas syringae* KPW66023.1, ALKBH6 (NP_116267.3), FTO (NP_001073901.1), *Streptomyces subtrutilus* WP_069918251.1, ALKBH2 (NP_001138846.1), *P. syringae* WP_110682285.1, ALKBH3 (NP_631917.1), *P. aeruginosa* WP_021219048.1, ALKBH1 (NP_006011.2) *Physcomitrella patens* XP_024396199.1, *Kittatospora aureofaciens* WP_051826438.1, *P. aeruginosa* MXH34269.1, ALKB (NP_416716.1), *P. syringae* WP_191998247.1. (C) The amino acid in the catalytic domain. The conserved Fe²⁺-binding site residues that are marked by a red triangle are HXD-H, and the α-KG-binding site residues (N-Y-R-R) are marked by blue triangles. The red background indicates the active-site region that is critical for substrate recognition and demethylation specificity. (D) The structure of DBSH. HXD-H is labeled in red, and N-Y-R-R is labeled in blue. (E) Structural comparisons of DBSH in the ALKBH family. A heatmap was drawn based on RMSD. (F) The hypothesis concerning the evolutionary history of the ALKBH family, which is derived from the bacterial DNA repair enzyme ALKB but independently originates from different bacterial ALKB ancestors (enzymes with black branches originating from Group I bacterial ALKB, brown branches originating from Group II bacterial ALKB). ALKBHs in the same background show a similar catalytic function.

In the present study, YTHDF and YTHDC were found in the major phylum of eukaryotes and sporadic prokaryotes (Figure 2). The sequences of YTHDF and YTHDC in humans, basal eukaryotes and prokaryotes were well conserved (Figure 5A), while there were low amino acid sequence identities between their YTH domain and McrB-N domain or MJECL36 (Figure S16). However, the three hydrophobic amino acids forming an aromatic cage to accommodate nucleotides [73–75] were highly conserved among eukaryotic YTH, McrB-N domain and MJECL36 (Figure 5C, Figure S16), indicating their ancestral function of recognizing nucleotides. Furthermore, the topological structural similarities between TgMcrB and human YTH proteins (RMSD = 4.45 Å and 3.97 Å, Figure 5B) and between MJECL36 and human YTH proteins (RMSD = 3.90 Å and 3.29 Å, Figure 5B) indicated that they might have evolved from a common ancestor. In addition, the McrB-N domain can facilitate TgMcrB binding to dsDNA/ssDNA rather than RNA substrates [73], while human YTH proteins target ssDNA and mRNA. The functional transition from recognizing dsDNA to single-stranded nucleotides adds further evidence of the evolutionary origins of the m⁶A mRNA system.

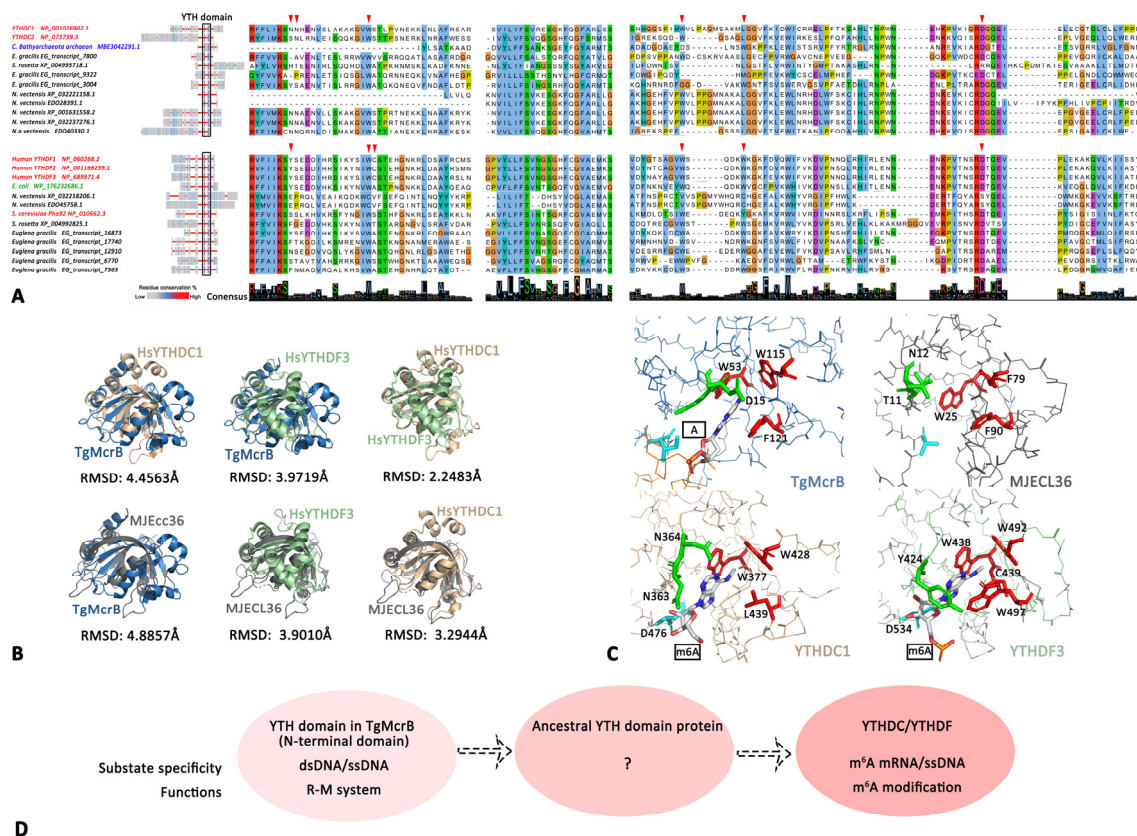


Figure 5. Evolution and characteristics of YTH domains. **(A)** Protein sequences of YTHDC and YTHDF in humans, prokaryotes and basal eukaryotes. The residues for recognition are marked by a red triangle. **(B)** The structural similarity of human YTHDC1 (6zcn_A), YTHDF3 (6zot_B), TgMcrB (6p0g_A) and MJECL36 (2p5d_A). **(C)** The m6A recognition aromatic cages of YTHDC1 and YTHDF3 and the corresponding amino acids in TgMcrB and MJECL36. **(D)** Hypothesis of evolution of the YTH domain containing proteins.

As the N-terminus of the R-M system endonuclease, TgMcrB might be one of the prokaryotic origins of the eukaryotic YTH domain, and it may have subsequently evolved to individual YTH domain-containing proteins such as MJECL36. After a long period of evolution, these YTH domain-containing proteins were acquired by LECA with adaptive functions for RNA recognition (Figure 5D).

4. Discussion

Phylogenetic trees are widely used to infer the evolutionary relationships of species and genes. For instance, many cases of bona fide functional inter-kingdom HGT events have been documented, validating the movement of genetic information between distant species. Among the approaches to identifying HGT, phylogenetic methods are the most general and potentially most sensitive methods [76]. These methods, as we used in this study, reconstruct the history of a family of homologous genes (a gene tree, Figures S1–S16), and compare it to a putative history of the species in which the genes are found (the species tree, Figure 2) [76,77]. Homologues are genes that share a common ancestry and are divided between orthologues (derived by speciation) and paralogues (derived by duplication). Based on this definition, the search for duplicated genes can be done through the identification of paralogue relationships [78–81]. However, the diversity of genetic information caused by evolution has introduced difficulties into analyzes, especially for genes with a relatively distant evolutionary relationship. Fortunately, the released protein structures make it possible to build evolutionary relationships of functions and sequence distinct proteins according to the topological similarity of structures. In this

study, phylogenetic trees and protein structures were integrated to reveal the distribution of N6mA-associated components in three kingdoms of life, predict possible modifications and functions in different species and explore the evolutionary history of mRNA m⁶A modification, providing a basis and motivation for in-depth studies of these modifications in distinctive species.

4.1. Eukaryotic mRNA m⁶A Modification Components Originating from LECA

The homologues of mRNA m⁶A modification-associated proteins are found in only eukaryotic genomes rather than prokaryotes (Figure 2), adding evidence to the hypothesis that the mRNA m⁶A modification system originated from LECA. The distribution of m⁶A modification core elements in distinct species suggests that the ancestral m⁶A machinery in LECA is a form composed of “ancient writer complex”, with one pair of METTL3/METTL14, one auxiliary WTAP and RBM15, ZC3H13 protein each; “ancient reader” with one YTHDC and one YTHDF; and “ancient eraser” with one FTO and one ALKBH5 (Figure 6B), which is a more complicated complex than indicated in a previously reported study [2]. This complex evolved to the extant complex in different species by recruiting additional proteins to perform distinct biological functions in heterogeneous and changing cellular environments. In parallel, gene expansion has led to the formation of lineage-specific characteristics, such as HNRNPC and FMBP1 in Metazoa, and HNRNPA2B1, HNRNPG, IGF2BPs, ELAV1 in vertebrates. VIRMA and HAKAI are highly divergent between plants and animals and are not present in lower eukaryotes. The emergence of these two proteins could represent a case of convergent evolution, similar to the occurrence of the origin of parental genomic imprinting [2].

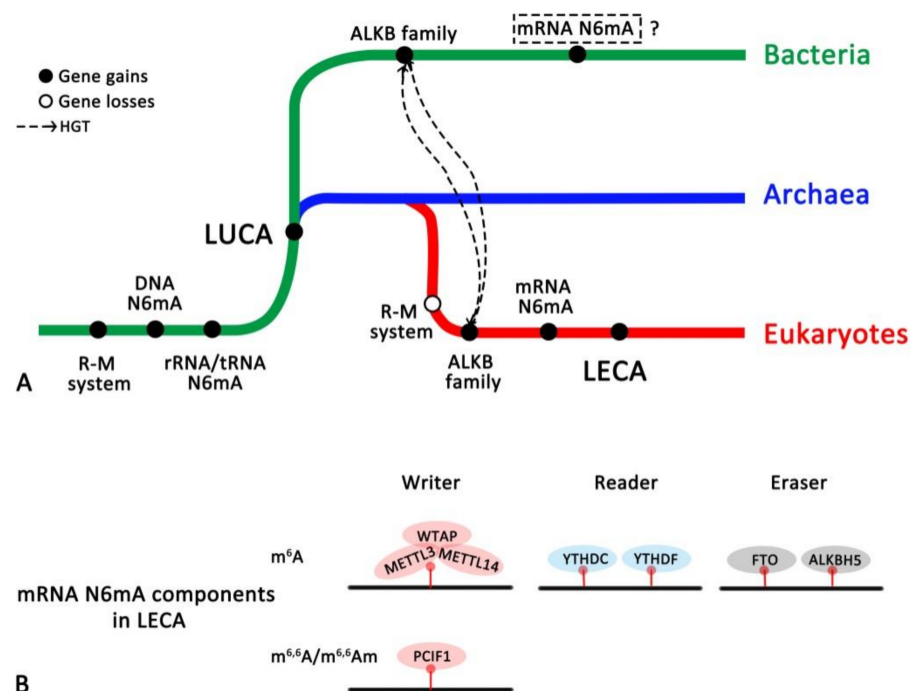


Figure 6. The hypothesis of the evolution of N6mA modifications. (A) Hypothesis of the evolution of the N6mA modification system. (B) Ancestral mRNA N6mA components in LECA.

4.2. The Concerto of Gene Losses and Gains Involved in the mRNA m⁶A Modification Machinery

Although a great deal of attention has been given to the mechanisms of evolution via gene duplication, gene loss is a crucial process [82]. Gene losses were also found in the evolution of mRNA m⁶A modification machinery, such as the absence of METTL3-METTL14 in some species lineages. One possible explanation for this phenomenon is the potential presence of substitute enzymes for adding m⁶A to mRNA because auxiliary components in the

“writer complex” were found in some species that lost the METTL3-METTL14 heterodimer. Another possibility is that these taxa have completely lost the mRNA m⁶A modification machinery, while the remaining auxiliary components play roles in processes other than m⁶A RNA modification. For the first and substitute hypothesis, METTL4, as a member of the MT-A70-like protein family, including METTL3 and METTL14 (Figure 2), is a potential substitute enzyme. Unlike METTL3 and METTL14, METTL4 is conserved in three kingdoms of life, especially in eukaryotes lacking METTL14/METTL3 proteins (Figure 2, lower panel). In addition, METTL4 can modify N⁶mA in DNA or snRNA [20,55,83,84], adding to its ability to functionally compensate for the absence of METTL3 and METTL14. Kinetoplastida lost all three enzymes, but m⁶A mRNA modification has been identified in *T. brucei* [85], also raising the possibility that other N⁶mA MTases might functionally compensate for the absence of the METTL3/METTL14 dyad in these species. Hence, in view of evolution, as the paralogues of METTL4, METTL3 and METTL14 might originate from the duplication of METTL4 before LECA, and thereby achieve functional expansion and diversity. In some species, the loss of METTL3/14 and the retention of METTL4 may be the result of their common ancestor’s adaptation to the environment. However, for some species, the varieties of enzymes for RNA modification are needed to maintain their functional diversity, which was supported by the reported distinct function of these MTases [20,55,83,84]. Therefore, it is inferred that the genes’ gains and loss of METTL3/14 may be the result of adaptive evolution.

Both FTO and ALKBH5 may be derived from LECAs, but their absence in early-branching eukaryotes raises the following questions that need to be answered: are there other m⁶A demethylases, and is m⁶A a dynamic modification in these species? These two enzymes belong to the ALKB family [86]. The conserved catalytic domain in this protein family makes other members potential substitutes for FTO/ALKBH5 (Figure S15), but no clear evidence has been obtained thus far.

Above all, the evolution of bacterial, archaeal and eukaryotic genomes is a dynamic process that involves extensive gains and losses of gene events [82,87]. In particular, gene duplication is a fundamental process in the evolution of species and plays a major role in the origination of novel gene functions [88]. Gene losses have often been associated with the loss of redundant gene duplicates driven by evolutionary forces but have recently been suggested to be beneficial for organisms, as they can be a pervasive source of genetic variation that gives rise to adaptive phenotypic diversity [82,89]. In the present study, “writer complex”, “erasers” and some “readers” have undergone specific gene losses during evolution, while some elements in “writer complex” and other “readers” were generated by gene gains. Consequently, gene gains and losses form specific combinations of m⁶A components to provide functional diversity in distinct species by finely tuning m⁶A modifications on mRNA. Hence, during the evolution, the mRNA m⁶A system becomes more complex or simpler in some species, for instance, Chordata contains the most complex mRNA m⁶A system by gene and function expansion, while Discoba, SAR and Fungi lost many m⁶A modification-related genes.

4.3. The Evolutionary History of the RNA N⁶mA Modification Machinery

The R-M system is the most widespread prokaryotic biological conflict system facilitating both discrimination of cellular self DNA from invasive non-self DNA and destruction of the latter [35]. The similar function and topological structure of MTase and YTH domain-containing proteins in prokaryotes and eukaryotes indicate that the R-M system in both Bacteria and Archaea is the ancestor of N⁶mA modification on DNA and RNA for three kingdoms of life (Figure 6A). N⁶mA modification of DNA is carried out by “orphan” MTases to mark and regulate replication origins of genomic and plasmid replicons, DNA repair and epigenetic gene regulation [31,35,50,51]. Subsequently, ncRNAs such as rRNA and tRNA make use of this mechanism to regulate translation. It is becoming apparent that before the split of bacteria and archaea, N⁶mA modifications on DNA, rRNA and tRNA, together with the R-M system, were retained in LUCA.

In the evolutionary history of eukaryotes, the emergence of mRNA N6mA components in LECA (Figure 6B) is the evolutionary product from LUCA (MTase and YTH domains) or driven by HGT from bacterial elements (ALKB family). Concomitantly, the physiological roles of mRNA N6mA were massively extended, which resulted in increases in mRNA N6mA modification-related genes in eukaryotic genomes and finally contributed to the complexity of the species. Furthermore, eukaryotes lost the R-M system in prokaryotes and retained N6mA modifications on rRNA and tRNA (Figure 6A). Although DNA methylation modification is likely coordinated with the recognition of specific histone methylation marks in lower eukaryotes [35], its existence and function are debatable issues regarding mammalian DNA [31,74].

Until now, RNA modifications in Archaea have remained mysterious, and mRNA N6mA modifications have been found in only eukaryotes and bacteria, but with quite different mechanisms because: (1) bacterial m⁶A motifs (GCCAUs) are distinct from those of mammals (RRACHs); (2) the enzymes responsible for adding m⁶A to bacterial mRNA have not been found [90]; and (3) no bacterial homologues of components associated with the mRNA m⁶A machinery were confirmed in our study.

5. Conclusions

The comprehensive evolutionary history of RNA N6mA modifications is delineated in the present study. The distribution of known RNA (mRNA, ncRNAs) N6mA (m⁶A, m^{6,6}A, m⁶Am, m^{6,6}Am and m⁶t⁶A) components in species from the three kingdoms of life indicates the possible RNA N6mA modifications involved in the specific biological function of distinct species (Figures S1–S14). Moreover, the hypothesis that RNA N6mA-MTases are derived from at least two different types of prokaryotic DNA MTases (class α and β MTases) was further confirmed. As the m⁶A reader, YTH proteins can specifically recognize the m⁶A tag and may be acquired by LECA from an individual prokaryotic YTH-domain protein that evolved from N-terminals of an R-M system endonuclease. In addition, the hypothesis that the origin of eukaryotic ALKBH family proteins is driven by at least two occasions of independent HTG from the bacterial ALKBHs was also supported. Therefore, all the findings in the present study will facilitate understanding of the evolution of RNA N6mA modifications and their important contributions to biological functions and diversity, with contributions from more experimental evidence in the future, especially on the HGT and gene duplication validation. However, one of the most commonly used and most sensitive analysis methods was used to analyze the public genome data of different species deposited in two databases in this study, and the genome with different quality may have an impact on the analysis. In the future, quality sequencing of the genomes of different species for systematic analysis and more analytical methods are recommended to validate our hypothesis.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biology11020214/s1>, Figure S1: Phylogenetic tree of N6mA MTases. Figure S2: Phylogenetic tree of PCIF1. Figure S3: Phylogenetic tree of TRMO. Figure S4: Phylogenetic tree of WTAP. Figure S5: Phylogenetic tree of VRIMA. Figure S6: Phylogenetic tree of RBM15. Figure S7: Phylogenetic tree of HAKAI. Figure S8: Phylogenetic tree of ZC3H13. Figure S9: Phylogenetic tree of ALKBH family. Figure S10: Phylogenetic tree of YTH family. Figure S11: Phylogenetic tree of HNRNPs. Figure S12: Phylogenetic tree of IGF2BPs. Figure S13: Phylogenetic tree of FMRP1. Figure S14: Phylogenetic tree of ELAV1. Figure S15: Distribution of ALKB family in prokaryotes and eukaryotes. Figure S16: Structure based alignments of YTH domain containing proteins (h = alpha-helix, e = beta-sheet,) Consensus amino acid symbols are: conserved amino acids are in bold and uppercase letters; aliphatic (I, V, L): l; aromatic (Y, H, W, F): @; hydrophobic (W, F, Y, M, L, I, V, A, C, T, H): h; alcohol (S, T): o; polar residues (D, E, H, K, N, Q, R, S, T): p; tiny (A, G, C, S): t; small (A, G, C, S, V, N, D, T, P): s; bulky residues (E, F, I, K, L, M, Q, R, W, Y): b; positively charged (K, R, H): +; negatively charged (D, E): −; charged (D, E, K, R, H): c.). Table S1: N6mA modification writers, readers and erasers protein families as initial queries. Table S2: Classification of taxa. Table S3: Domains used to filter the homologues of N6mA modification writers, readers

and erasers' protein families. Table S4: Alignments criteria and Best fit model. Table S5: The m⁶A associated protein of prokaryotes in Ensemble database. Table S6: RMSD of N⁶mA MTases and ALKBH family. Supplementary File S1: Alignments for building phylogenetic trees.

Author Contributions: All authors contributed extensively to the work presented in this paper. C.L., J.Y., H.Z. and J.C. conceived the study; C.L. analyzed the data, wrote and revised the manuscript; J.Y. supervised the work, wrote and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was sponsored by the Natural Science Foundation of Shanghai (20ZR1463600), Shanghai Municipal Health Commission (GWV-10.2-YQ31), China scholarship council (201908310036) and the National Natural Science Foundation of China (81702030).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All raw data related to this study have been deposited in the Supplementary Materials.

Acknowledgments: We thank Mark C. Field, Ning Zhang and Lael Barlow in the School of Life Science at the University of Dundee.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Crick, F. Central dogma of molecular biology. *Nature* **1970**, *227*, 561–563. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Balacco, D.L.; Solter, M. The m(6)A Writer: Rise of a Machine for Growing Tasks. *Biochemistry* **2019**, *58*, 363–378. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Jinek, M.; Doudna, J.A. A three-dimensional view of the molecular machinery of RNA interference. *Nature* **2009**, *457*, 405–412. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Joyce, G.F. RNA evolution and the origins of life. *Nature* **1989**, *338*, 217–224. [\[CrossRef\]](#)
5. Zhao, L.Y.; Song, J.; Liu, Y.; Song, C.X.; Yi, C. Mapping the epigenetic modifications of DNA and RNA. *Protein Cell* **2020**, *11*, 792–808. [\[CrossRef\]](#)
6. Boccaletto, P.; Machnicka, M.A.; Purta, E.; Piatkowski, P.; Baginski, B.; Wirecki, T.K.; de Crecy-Lagard, V.; Ross, R.; Limbach, P.A.; Kotter, A.; et al. MODOMICS: A database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* **2018**, *46*, D303–D307. [\[CrossRef\]](#)
7. Gilbert, W.V.; Bell, T.A.; Schaening, C. Messenger RNA modifications: Form, distribution, and function. *Science* **2016**, *352*, 1408–1412. [\[CrossRef\]](#)
8. Nachtergaele, S.; He, C. The emerging biology of RNA post-transcriptional modifications. *RNA Biol.* **2017**, *14*, 156–163. [\[CrossRef\]](#)
9. Dubin, D.T.; Taylor, R.H. The methylation state of poly A-containing messenger RNA from cultured hamster cells. *Nucleic Acids Res.* **1975**, *2*, 1653–1668. [\[CrossRef\]](#)
10. Desrosiers, R.; Friderici, K.; Rottman, F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 3971–3975. [\[CrossRef\]](#)
11. Liu, J.; Yue, Y.; Han, D.; Wang, X.; Fu, Y.; Zhang, L.; Jia, G.; Yu, M.; Lu, Z.; Deng, X.; et al. A METTL3-METTL14 complex mediates mammalian nuclear RNA N⁶-adenosine methylation. *Nat. Chem. Biol.* **2014**, *10*, 93–95. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Zheng, G.; Dahl, J.A.; Niu, Y.; Fedorcsak, P.; Huang, C.M.; Li, C.J.; Vagbo, C.B.; Shi, Y.; Wang, W.L.; Song, S.H.; et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **2013**, *49*, 18–29. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Jia, G.; Fu, Y.; Zhao, X.; Dai, Q.; Zheng, G.; Yang, Y.; Yi, C.; Lindahl, T.; Pan, T.; Yang, Y.G.; et al. N⁶-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **2011**, *7*, 885–887. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Meyer, K.D.; Saletore, Y.; Zumbo, P.; Elemento, O.; Mason, C.E.; Jaffrey, S.R. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **2012**, *149*, 1635–1646. [\[CrossRef\]](#)
15. Dominissini, D.; Moshitch-Moshkovitz, S.; Schwartz, S.; Salmon-Divon, M.; Ungar, L.; Osenberg, S.; Cesarkas, K.; Jacob-Hirsch, J.; Amariglio, N.; Kupiec, M.; et al. Topology of the human and mouse m⁶A RNA methylomes revealed by m⁶A-seq. *Nature* **2012**, *485*, 201–206. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Cao, G.; Li, H.B.; Yin, Z.; Flavell, R.A. Recent advances in dynamic m⁶A RNA modification. *Open Biol.* **2016**, *6*, 160003. [\[CrossRef\]](#)
17. Zaccara, S.; Ries, R.J.; Jaffrey, S.R. Reading, writing and erasing mRNA methylation. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 608–624. [\[CrossRef\]](#)
18. Jonkhout, N.; Tran, J.; Smith, M.A.; Schonrock, N.; Mattick, J.S.; Novoa, E.M. The RNA modification landscape in human disease. *RNA* **2017**, *23*, 1754–1769. [\[CrossRef\]](#)
19. Boo, S.H.; Kim, Y.K. The emerging role of RNA modifications in the regulation of mRNA stability. *Exp. Mol. Med.* **2020**, *52*, 400–408. [\[CrossRef\]](#)

20. Goh, Y.T.; Koh, C.W.Q.; Sim, D.Y.; Roca, X.; Goh, W.S.S. METTL4 catalyzes m6Am methylation in U2 snRNA to regulate pre-mRNA splicing. *Nucleic Acids Res.* **2020**, *48*, 9250–9261. [\[CrossRef\]](#)
21. Sendinc, E.; Valle-Garcia, D.; Dhall, A.; Chen, H.; Henriques, T.; Navarrete-Perea, J.; Sheng, W.; Gygi, S.P.; Adelman, K.; Shi, Y. PCIF1 Catalyzes m6Am mRNA Methylation to Regulate Gene Expression. *Mol. Cell* **2019**, *75*, 620–630.e9. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Garcias Morales, D.; Reyes, J.L. A birds'-eye view of the activity and specificity of the mRNA m(6) A methyltransferase complex. *Wiley Interdiscip. Rev. RNA* **2021**, *12*, e1618. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Shi, H.; Wei, J.; He, C. Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. *Mol. Cell* **2019**, *74*, 640–650. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Huang, H.; Weng, H.; Chen, J. m(6)A Modification in Coding and Non-coding RNAs: Roles and Therapeutic Implications in Cancer. *Cancer Cell* **2020**, *37*, 270–288. [\[CrossRef\]](#)
25. Pandey, R.R.; Delfino, E.; Homolka, D.; Roithova, A.; Chen, K.M.; Li, L.; Franco, G.; Vagbo, C.B.; Taillebourg, E.; Fauvarque, M.O.; et al. The Mammalian Cap-Specific m(6)Am RNA Methyltransferase PCIF1 Regulates Transcript Levels in Mouse Tissues. *Cell Rep.* **2020**, *32*, 108038. [\[CrossRef\]](#)
26. Wei, J.; Liu, F.; Lu, Z.; Fei, Q.; Ai, Y.; He, P.C.; Shi, H.; Cui, X.; Su, R.; Klungland, A.; et al. Differential m(6)A, m(6)Am, and m(1)A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm. *Mol. Cell* **2018**, *71*, 973–985.e5. [\[CrossRef\]](#)
27. Pinto, R.; Vagbo, C.B.; Jakobsson, M.E.; Kim, Y.; Baltissen, M.P.; O'Donohue, M.F.; Guzman, U.H.; Malecki, J.M.; Wu, J.; Kirpekar, F.; et al. The human methyltransferase ZCCHC4 catalyses N6-methyladenosine modification of 28S ribosomal RNA. *Nucleic Acids Res.* **2020**, *48*, 830–846. [\[CrossRef\]](#)
28. Rong, B.; Zhang, Q.; Wan, J.; Xing, S.; Dai, R.; Li, Y.; Cai, J.; Xie, J.; Song, Y.; Chen, J.; et al. Ribosome 18S m(6)A Methyltransferase METTL5 Promotes Translation Initiation and Breast Cancer Cell Growth. *Cell Rep.* **2020**, *33*, 108544. [\[CrossRef\]](#)
29. Bangs, J.D.; Crain, P.F.; Hashizume, T.; McCloskey, J.A.; Boothroyd, J.C. Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. *J. Biol. Chem.* **1992**, *267*, 9805–9815. [\[CrossRef\]](#)
30. Seo, K.W.; Kleiner, R.E. Mechanisms of epitranscriptomic gene regulation. *Biopolymers* **2021**, *112*, e23403. [\[CrossRef\]](#)
31. Woodcock, C.B.; Horton, J.R.; Zhang, X.; Blumenthal, R.M.; Cheng, X. Beta class amino methyltransferases from bacteria to humans: Evolution and structural consequences. *Nucleic Acids Res.* **2020**, *48*, 10034–10044. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Anantharaman, V.; Koonin, E.V.; Aravind, L. Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res.* **2002**, *30*, 1427–1464. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Mosquera-Rendon, J.; Cardenas-Brito, S.; Pineda, J.D.; Corredor, M.; Benitez-Paez, A. Evolutionary and sequence-based relationships in bacterial AdoMet-dependent non-coding RNA methyltransferases. *BMC Res. Notes* **2014**, *7*, 440. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Bujnicki, J.M.; Feder, M.; Radlinska, M.; Blumenthal, R.M. Structure prediction and phylogenetic analysis of a functionally diverse family of proteins homologous to the MT-A70 subunit of the human mRNA:m(6)A methyltransferase. *J. Mol. Evol.* **2002**, *55*, 431–444. [\[CrossRef\]](#)
35. Iyer, L.M.; Zhang, D.; Aravind, L. Adenine methylation in eukaryotes: Apprehending the complex evolutionary history and functional potential of an epigenetic modification. *BioEssays* **2016**, *38*, 27–40. [\[CrossRef\]](#)
36. Xu, B.; Liu, D.; Wang, Z.; Tian, R.; Zuo, Y. Multi-substrate selectivity based on key loops and non-homologous domains: New insight into ALKBH family. *Cell. Mol. Life Sci. CMLS* **2021**, *78*, 129–141. [\[CrossRef\]](#)
37. Degner, E.C.; Ahmed-Braimah, Y.H.; Borziak, K.; Wolfner, M.F.; Harrington, L.C.; Dorus, S. Proteins, Transcripts, and Genetic Architecture of Seminal Fluid and Sperm in the Mosquito *Aedes aegypti*. *Mol. Cell. Proteom. MCP* **2019**, *18* (Suppl. S1), S6–S22. [\[CrossRef\]](#)
38. Leismann, J.; Spagnuolo, M.; Pradhan, M.; Wacheul, L.; Vu, M.A.; Musheev, M.; Mier, P.; Andrade-Navarro, M.A.; Graille, M.; Niehrs, C.; et al. The 18S ribosomal RNA m(6) A methyltransferase Mettl5 is required for normal walking behavior in *Drosophila*. *EMBO Rep.* **2020**, *21*, e49443. [\[CrossRef\]](#)
39. Papadopoulos, J.S.; Agarwala, R. COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **2007**, *23*, 1073–1079. [\[CrossRef\]](#)
40. Trifinopoulos, J.; Nguyen, L.T.; von Haeseler, A.; Minh, B.Q. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **2016**, *44*, W232–W235. [\[CrossRef\]](#)
41. Stover, B.C.; Muller, K.F. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinform.* **2010**, *11*, 7. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [\[CrossRef\]](#)
43. Schrödinger, L. The PyMOL Molecular Graphics System, Version Open-Source. Available online: <https://github.com/schrodinger/pymol-open-source> (accessed on 13 March 2021).
44. Pei, J.; Kim, B.H.; Grishin, N.V. PROMALS3D: A tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **2008**, *36*, 2295–2300. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Schubert, H.L.; Blumenthal, R.M.; Cheng, X. Many paths to methyltransfer: A chronicle of convergence. *Trends Biochem. Sci.* **2003**, *28*, 329–335. [\[CrossRef\]](#)

46. Kimura, S.; Miyauchi, K.; Ikeuchi, Y.; Thiaville, P.C.; Crecy-Lagard, V.; Suzuki, T. Discovery of the beta-barrel-type RNA methyltransferase responsible for N6-methylation of N6-threonylcarbamoyladenine in tRNAs. *Nucleic Acids Res.* **2014**, *42*, 9350–9365. [\[CrossRef\]](#)
47. Bheemanaik, S.; Reddy, Y.V.; Rao, D.N. Structure, function and mechanism of exocyclic DNA methyltransferases. *Biochem. J.* **2006**, *399*, 177–190. [\[CrossRef\]](#)
48. Goedecke, K.; Pignot, M.; Goody, R.S.; Scheidig, A.J.; Weinhold, E. Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. *Nat. Struct. Biol.* **2001**, *8*, 121–125. [\[CrossRef\]](#)
49. Casadesus, J. Bacterial DNA Methylation and Methylomes. *Adv. Exp. Med. Biol.* **2016**, *945*, 35–61. [\[CrossRef\]](#)
50. Vasu, K.; Nagaraja, V. Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol. Mol. Biol. Rev. MMBR* **2013**, *77*, 53–72. [\[CrossRef\]](#)
51. Oliveira, P.H.; Touchon, M.; Rocha, E.P. Regulation of genetic flux between bacteria by restriction-modification systems. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 5658–5663. [\[CrossRef\]](#)
52. Murray, I.A.; Morgan, R.D.; Luyten, Y.; Fomenkov, A.; Correa, I.R., Jr.; Dai, N.; Allaw, M.B.; Zhang, X.; Cheng, X.; Roberts, R.J. The non-specific adenine DNA methyltransferase M.EcoGII. *Nucleic Acids Res.* **2018**, *46*, 840–848. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Bokar, J.A.; Shambaugh, M.E.; Polayes, D.; Matera, A.G.; Rottman, F.M. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA* **1997**, *3*, 1233–1247. [\[PubMed\]](#)
54. Woodcock, C.B.; Yu, D.; Hajian, T.; Li, J.; Huang, Y.; Dai, N.; Correa, I.R., Jr.; Wu, T.; Vedadi, M.; Zhang, X.; et al. Human MettL3-MettL14 complex is a sequence-specific DNA adenine methyltransferase active on single-strand and unpaired DNA in vitro. *Cell Discov.* **2019**, *5*, 63. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Greer, E.L.; Blanco, M.A.; Gu, L.; Sendinc, E.; Liu, J.; Aristizabal-Corrales, D.; Hsu, C.H.; Aravind, L.; He, C.; Shi, Y. DNA Methylation on N6-Adenine in *C. elegans*. *Cell* **2015**, *161*, 868–878. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Wang, X.; Li, Z.; Zhang, Q.; Li, B.; Lu, C.; Li, W.; Cheng, T.; Xia, Q.; Zhao, P. DNA methylation on N6-adenine in lepidopteran *Bombyx mori*. *Biochim. Biophys. Acta Gene Regul. Mech.* **2018**, *1861*, 815–825. [\[CrossRef\]](#)
57. Malone, T.; Blumenthal, R.M.; Cheng, X. Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.* **1995**, *253*, 618–632. [\[CrossRef\]](#)
58. van Tran, N.; Ernst, F.G.M.; Hawley, B.R.; Zorbas, C.; Ulryck, N.; Hackert, P.; Bohnsack, K.E.; Bohnsack, M.T.; Jaffrey, S.R.; Graille, M.; et al. The human 18S rRNA m6A methyltransferase METTL5 is stabilized by TRMT112. *Nucleic Acids Res.* **2019**, *47*, 7719–7733. [\[CrossRef\]](#)
59. Li, W.; Shi, Y.; Zhang, T.; Ye, J.; Ding, J. Structural insight into human N6amt1-Trm112 complex functioning as a protein methyltransferase. *Cell Discov.* **2019**, *5*, 51. [\[CrossRef\]](#)
60. Woodcock, C.B.; Yu, D.; Zhang, X.; Cheng, X. Human HemK2/KMT9/N6AMT1 is an active protein methyltransferase, but does not act on DNA in vitro, in the presence of Trm112. *Cell Discov.* **2019**, *5*, 50. [\[CrossRef\]](#)
61. Akichika, S.; Hirano, S.; Shichino, Y.; Suzuki, T.; Nishimasu, H.; Ishitani, R.; Sugita, A.; Hirose, Y.; Iwasaki, S.; Nureki, O.; et al. Cap-specific terminal N (6)-methylation of RNA by an RNA polymerase II-associated methyltransferase. *Science* **2019**, *363*, 6423. [\[CrossRef\]](#)
62. Van Deuren, V.; Plessers, S.; Robben, J. Structural determinants of nucleobase modification recognition in the AlkB family of dioxygenases. *DNA Repair* **2020**, *96*, 102995. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Alemu, A.E.; He, C.; Klungland, A. ALKBHs-facilitated RNA modifications and de-modifications. *DNA Repair* **2016**, *44*, 87–91. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Fedeles, B.I.; Singh, V.; Delaney, J.C.; Li, D.; Essigmann, J.M. The AlkB Family of Fe(II)/alpha-Ketoglutarate-dependent Dioxygenases: Repairing Nucleic Acid Alkylation Damage and Beyond. *J. Biol. Chem.* **2015**, *290*, 20734–20742. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Moore, C.; Meng, B. Prediction of the molecular boundary and functionality of novel viral AlkB domains using homology modelling and principal component analysis. *J. Gen. Virol.* **2019**, *100*, 691–703. [\[CrossRef\]](#)
66. Li, M.M.; Nilsen, A.; Shi, Y.; Fusser, M.; Ding, Y.H.; Fu, Y.; Liu, B.; Niu, Y.; Wu, Y.S.; Huang, C.M.; et al. ALKBH4-dependent demethylation of actin regulates actomyosin dynamics. *Nat. Commun.* **2013**, *4*, 1832. [\[CrossRef\]](#)
67. Wang, G.; He, Q.; Feng, C.; Liu, Y.; Deng, Z.; Qi, X.; Wu, W.; Mei, P.; Chen, Z. The atomic resolution structure of human AlkB homolog 7 (ALKBH7), a key protein for programmed necrosis and fat metabolism. *J. Biol. Chem.* **2014**, *289*, 27924–27936. [\[CrossRef\]](#)
68. van den Born, E.; Vagbo, C.B.; Songe-Moller, L.; Leihne, V.; Lien, G.F.; Leszczynska, G.; Malkiewicz, A.; Krokan, H.E.; Kirpekar, F.; Klungland, A.; et al. ALKBH8-mediated formation of a novel diastereomeric pair of wobble nucleosides in mammalian tRNA. *Nat. Commun.* **2011**, *2*, 172. [\[CrossRef\]](#)
69. Wang, X.; Lu, Z.; Gomez, A.; Hon, G.C.; Yue, Y.; Han, D.; Fu, Y.; Parisien, M.; Dai, Q.; Jia, G.; et al. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **2014**, *505*, 117–120. [\[CrossRef\]](#)
70. Liao, S.; Sun, H.; Xu, C. YTH Domain: A Family of N(6)-methyladenosine (m(6)A) Readers. *Genom. Proteom. Bioinform.* **2018**, *16*, 99–107. [\[CrossRef\]](#)
71. Zhang, Z.; Theler, D.; Kaminska, K.H.; Hiller, M.; de la Grange, P.; Pudimat, R.; Rafalska, I.; Heinrich, B.; Bujnicki, J.M.; Allain, F.H.; et al. The YTH domain is a novel RNA binding domain. *J. Biol. Chem.* **2010**, *285*, 14701–14710. [\[CrossRef\]](#)
72. Stoilov, P.; Rafalska, I.; Stamm, S. YTH: A new domain in nuclear proteins. *Trends Biochem. Sci.* **2002**, *27*, 495–497. [\[CrossRef\]](#)

73. Hosford, C.J.; Bui, A.Q.; Chappie, J.S. The structure of the *Thermococcus gammatolerans* McrB N-terminal domain reveals a new mode of substrate recognition and specificity among McrB homologs. *J. Biol. Chem.* **2020**, *295*, 743–756. [[CrossRef](#)]
74. Woodcock, C.B.; Horton, J.R.; Zhou, J.; Bedford, M.T.; Blumenthal, R.M.; Zhang, X.; Cheng, X. Biochemical and structural basis for YTH domain of human YTHDC1 binding to methylated adenine in DNA. *Nucleic Acids Res.* **2020**, *48*, 10329–10341. [[CrossRef](#)] [[PubMed](#)]
75. Xu, C.; Liu, K.; Ahmed, H.; Loppnau, P.; Schapira, M.; Min, J. Structural Basis for the Discriminative Recognition of N6-Methyladenosine RNA by the Human YT521-B Homology Domain Family of Proteins. *J. Biol. Chem.* **2015**, *290*, 24902–24913. [[CrossRef](#)]
76. Daubin, V.; Szollosi, G.J. Horizontal Gene Transfer and the History of Life. *Cold Spring Harb. Perspect. Biol.* **2016**, *8*, a018036. [[CrossRef](#)]
77. Husnik, F.; McCutcheon, J.P. Functional horizontal gene transfer from bacteria to eukaryotes. *Nat. Rev. Microbiol.* **2018**, *16*, 67–79. [[CrossRef](#)]
78. Lallemand, T.; Leduc, M.; Landes, C.; Rizzon, C.; Lerat, E. An Overview of Duplicated Gene Detection Methods: Why the Duplication Mechanism Has to Be Accounted for in Their Choice. *Genes* **2020**, *11*, 1046. [[CrossRef](#)]
79. A von der Dunk, S.H.; Snel, B. Recurrent sequence evolution after independent gene duplication. *BMC Evol. Biol.* **2020**, *20*, 98. [[CrossRef](#)]
80. De Grassi, A.; Lanave, C.; Saccone, C. Genome duplication and gene-family evolution: The case of three OXPHOS gene families. *Gene* **2008**, *421*, 1–6. [[CrossRef](#)]
81. Sheridan, P.O.; Raguideau, S.; Quince, C.; Holden, J.; Zhang, L.; Thames, C.; Williams, T.A.; Gubry-Rangin, C. Gene duplication drives genome expansion in a major lineage of Thaumarchaeota. *Nat. Commun.* **2020**, *11*, 5494. [[CrossRef](#)]
82. Albalat, R.; Canestro, C. Evolution by gene loss. *Nat. Rev. Genet.* **2016**, *17*, 379–391. [[CrossRef](#)] [[PubMed](#)]
83. Chen, H.; Gu, L.; Orellana, E.A.; Wang, Y.; Guo, J.; Liu, Q.; Wang, L.; Shen, Z.; Wu, H.; Gregory, R.I.; et al. METTL4 is an snRNA m(6)Am methyltransferase that regulates RNA splicing. *Cell Res.* **2020**, *30*, 544–547. [[CrossRef](#)] [[PubMed](#)]
84. Gu, L.; Wang, L.; Chen, H.; Hong, J.; Shen, Z.; Dhall, A.; Lao, T.; Liu, C.; Wang, Z.; Xu, Y.; et al. CG14906 (mettl4) mediates m(6)A methylation of U2 snRNA in *Drosophila*. *Cell Discov.* **2020**, *6*, 44. [[CrossRef](#)] [[PubMed](#)]
85. Liu, L.; Zeng, S.; Jiang, H.; Zhang, Y.; Guo, X.; Wang, Y. Differential m6A methylomes between two major life stages allows potential regulations in *Trypanosoma brucei*. *Biochem. Biophys. Res. Commun.* **2019**, *508*, 1286–1290. [[CrossRef](#)]
86. Lence, T.; Soller, M.; Roignant, J.Y. A fly view on the roles and mechanisms of the m(6)A mRNA modification and its players. *RNA Biol.* **2017**, *14*, 1232–1240. [[CrossRef](#)]
87. Iranzo, J.; Wolf, Y.I.; Koonin, E.V.; Sela, I. Gene gain and loss push prokaryotes beyond the homologous recombination barrier and accelerate genome sequence divergence. *Nat. Commun.* **2019**, *10*, 5376. [[CrossRef](#)]
88. Chauve, C.; Doyon, J.P.; El-Mabrouk, N. Gene family evolution by duplication, speciation, and loss. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **2008**, *15*, 1043–1062. [[CrossRef](#)]
89. Huelsmann, M.; Hecker, N.; Springer, M.S.; Gatesy, J.; Sharma, V.; Hiller, M. Genes lost during the transition from land to water in cetaceans highlight genomic changes associated with aquatic adaptations. *Sci. Adv.* **2019**, *5*, eaaw6671. [[CrossRef](#)]
90. Deng, X.; Chen, K.; Luo, G.Z.; Weng, X.; Ji, Q.; Zhou, T.; He, C. Widespread occurrence of N6-methyladenosine in bacterial mRNA. *Nucleic Acids Res.* **2015**, *43*, 6557–6567. [[CrossRef](#)]