Article - Discoveries

Evolution of methyltransferase like (METTL) proteins in Metazoa: A complex gene family involved in epitranscriptomic regulation and other epigenetic processes

Juliet M. Wong and Jose M. Eirin-Lopez *

Environmental Epigenetics Laboratory, Institute of Environment, Florida International University, Miami, FL, United States

* Corresponding author: Environmental Epigenetics Lab, Florida International University, Biscayne Bay Campus, 3000 NE 151 Street, office MSB-360, North Miami, FL 33181, USA, jeirinlo@fiu.edu, +1-305-919-4000

Abstract

The methyltransferase like (METTL) proteins constitute a family of seven-beta-strand methyltransferases with S-adenosyl methionine binding domains that modify DNA, RNA, and proteins. Methylation by METTL proteins contributes to the epigenetic, and in the case of RNA modifications, epitranscriptomic regulation of a variety of biological processes. Despite their functional importance, most investigations of the substrates and functions of METTLs within metazoans have been restricted to model vertebrate taxa. In the present work, we explore the evolutionary mechanisms driving the diversification and functional differentiation of 33 individual METTL proteins across Metazoa. Our results show that METTLs are nearly ubiquitous across the animal kingdom, with most having arisen early in metazoan evolution (i.e., occur in basal metazoan phyla). Individual METTL lineages each originated from single independent ancestors, constituting monophyletic clades, which suggests that each METTL was subject to strong selective constraints driving its structural and/or functional specialization. Interestingly, a similar process did not extend to the differentiation of nucleoside-modifying and protein-modifying METTLs (i.e., each METTL type did not form a unique monophyletic clade). The members of these two types of METTLs also exhibited differences in their rates of evolution. Overall, we provide evidence that the long-term evolution of METTL family members was driven by strong purifying selection, which in combination with adaptive selection episodes, led to the functional specialization of individual METTL lineages. This work contributes useful information regarding the evolution of a gene family that fulfills a variety of epigenetic functions, and can have profound influences on molecular processes and phenotypic traits.

Key words: methyltransferase, METTL, epigenetics, phylogenetics, selection, metazoan

Introduction

Methyltransferase enzymes catalyze the transfer of methyl groups to DNA, RNA, proteins, and other biomolecules (Cheng and Blumenthal 1999). An increasing interest in these proteins has been driven by epigenetics, formally defined as, "the study of phenomena and mechanisms that cause chromosome-bound, heritable changes to gene expression that are not dependent on changes to the DNA sequence" (Deans and Maggert 2015). In particular, work examining DNA methylation, as well as the enzymes responsible for this modification, has been conducted for decades across a broad range of contexts and model systems (Holliday 2006). In addition, other studies have recognized the structural and functional importance of methyltransferases that modify specific residues in proteins such as histones (Couture and Trievel 2006; Ng et al. 2009). More recently, there has been a growing interest in the posttranscriptional modification of RNA molecules, a concept known as RNA epigenetics or epitranscriptomics (He 2010; Saletore et al. 2012). Although over 170 different RNA modifications have been recorded (Machnicka et al. 2013), relatively little is known about the enzymes responsible for these modifications.

Depending on their protein structure, most methyltransferases are categorized into three large superfamilies: seven-beta-strand methyltransferases, SET methyltransferases, and SPOUT methyltransferases (Petrossian and Clarke 2009). Among the seven-beta-strand methyltransferases, the methyltransferase like (METTL) gene family (Table 1) encodes proteins characterized by a conserved S-adenosyl methionine (SAM or AdoMet) binding domain that is formed by part of the seven-beta-strand structure (Martin and McMillan 2002; Petrossian and Clarke 2009). In this study, we examine the METTL family because these enzymes have been demonstrated to modify DNA/RNA nucleosides as well as protein residues (Ignatova et al. 2019), leading to changes in gene expression and phenotype that can have profound effects on an organism's condition. For example, METTL3 and METTL14 are responsible for the formation of N^6-methyladenosine (m^6A) in RNA (Liu et al. 2014). In eukaryotes, m^6A is a very common RNA modification, and has been shown to play a role in mammalian temperature stress response by promoting translation initiation of heat shock response genes (Zhou et al. 2015). As another example, it has been demonstrated that METTL21D methylates a lysine residue of valosin-containing protein (VCP), and may be linked to diseases, including cancer, in humans (Thiele et al. 2011; Kernstock et al. 2012). However, the precise targets and functions of several METTL proteins remain unresolved. In addition to structure, methyltransferases can be characterized by the type of biomolecule that they modify. Here, METTLs are characterized into three types: those that modify DNA/RNA, those that modify proteins, and those for which the biomolecule type they modify is currently unknown.

We also chose to examine the METTL family because it has been studied considerably less than other families of methyltransferases, such as the DNA methyltransferase (DNMT) family. The structures,

activities, and functions of DNMTs have been studied across all domains of life (Lyko 2018; Bhattacharyya et al. 2020). In contrast, the examination of METTL proteins has remained extremely limited in non-model organisms despite the potential epigenetic functions of METTLs and their resulting influence on biological processes and phenotypic characteristics. Within Metazoa, studies of METTL proteins are often restricted to humans, mice, and rats. Thus, there remains a knowledge gap regarding the function of METTLs within other metazoan taxa, particularly within non-vertebrate phyla. Exploring how METTL proteins have evolved throughout Metazoa is important for understanding: 1) the biological functions of different METTLs, 2) the contribution of METTL functional specialization to the process of diversification among metazoan groups, and 3) the relative importance of different types of METTLs and their gene regions that vary in evolutionary conservation. The present work uses extensive data mining to collect available METTL sequences across metazoan taxa. From these, a select number of sequences from nine representative taxa across eight different metazoan phyla are used for multiple phylogenetic and evolutionary analyses. Lastly, evidence of diversifying selection episodes is assessed using vertebrate METTL sequences. By investigating METTL proteins across Metazoa, this work provides insight into the evolution of this epigenetically-relevant gene family.

Results and Discussion

METTLs are widespread across Metazoa

In this study we examined 33 genes that encode proteins within the METTL family (Table 1). These were broadly characterized into three types: 1) those that target and modify the nucleosides of DNA or RNA molecules, 2) those that target and modify the residues of proteins, and 3) those with unknown modification targets. Overall, we assessed 12 METTLs that have been demonstrated to modify DNA or RNA nucleosides, 15 METTLs that have been demonstrated to modify protein residues, and six METTLs of currently unknown function (Table 1). A detailed search of these methyltransferase-like genes was conducted within the NCBI GenBank database (https://www.ncbi.nlm.nih.gov/genbank). Importantly, this search was restricted to NCBI and may be limited by sequence data availability and annotation. Thus, any failure to detect a specific METTL gene within a particular phylum does not necessarily indicate that the gene is lost in that phylum. Nonetheless, the search found that METTL genes are present in most of the major metazoan phyla examined here including Porifera, Cnidaria, Brachiopoda, Mollusca, Platyhelminthes, Nematoda, Priapulida, Arthropoda, Echinodermata, Hemichordata, and Chordata (Fig. 1). Chordates possess the largest number of METTL types (i.e., 33 out of 33 examined in this current study), however, the diversification of METTL genes does not appear to be restricted to chordates and other deuterostomes. Indeed, among protostomes, our search detected that Mollusca possesses 28 unique METTL genes and Arthropoda has 26 unique METTLs. Basally to

triploblastic animals, diploblastic cnidarians have 25 METTLs and parazoan sponges, which lack specialized tissues, have 19 METTLs. The widespread presence of METTLs in ancient metazoan lineages such as Cnidaria and Porifera suggests that they arose quite early in the evolutionary history of Metazoa. Furthermore, it appears that the early functional diversification of these various METTL types was necessary to enable critical cellular functionality, and therefore, the evolution of these phyla (Carroll et al. 2013).

There were several sequences that were determined to be probable METTL sequences, but could not be definitively identified as specific METTL genes based on the data available (Fig. 1, represented as striped gray boxes). These sequences were tentatively recognized as METTLs via orthologous gene searches and Basic Local Alignment Search Tool (BLAST) searches, but did not have any predicted gene identity or function recorded in NCBI. Gene searches in NCBI only identified putative METTL orthologs within one species of annelid, the leech *Helobdella robusta* (NCBI: txid6412). For example, searches for orthologs to human METTL1 identified one gene in *H. robusta* (HELRODRAFT_91834), however, it is described as a, "hypothetical protein" and the reference sequence is defined as, "*Helobdella robusta* hypothetical protein partial mRNA" (XM_009013229.1). NCBI BLAST results of this sequence include hits to METTL1 sequences in other metazoan phyla, although with relatively low query coverage and percent identity. All search results of putative METTL orthologs within Annelida were similar "hypothetical" proteins in *H. robusta*. Upon including putative METTL protein sequences from *H. robusta* in a separate phylogenetic analysis, each sequence clustered with its prospective METTL orthologs from other metazoan taxa (supplementary Fig. S5, Supplementary Material). Thus, it is probable that METTLs occur in Annelida, however, based on the current data available in NCBI, we hesitate to state with certainty the precise identity of these *H. robusta* sequences.

Our search in NCBI was not able to locate any methyltransferase-like genes within the phylum Ctenophora. This is not to say that the methylation of DNA, RNA, or protein molecules does not occur in ctenophores. DNA methylation has been detected in the promoter and gene body regions of ctenophore genomes, and ctenophores have been shown to possess DNA methyltransferase 1 (DNMT1) (Dabe et al. 2015). DNMT1, a highly conserved member of the DNMT family of methyltransferases, acts to methylate cytosines in DNA and has been demonstrated to maintain methylation patterns following DNA replication (Kangaspeska et al. 2008; Lyko 2018). Ctenophore genomes also contain a DNA methyltransferase 2 (DNMT2) gene (Dabe et al. 2015). Although a member of the DNMT family, DNMT2 is a RNA methyltransferase also known as tRNA aspartic acid methyltransferase 1 (TRDMT1) (Goll et al. 2006; Lyko 2018). Thus, methylation in ctenophores may be driven by methyltransferases outside of the METTL family, such as by DNMTs. Alternatively, METTLs may be present within ctenophores, but they were undetected in our search due to a limited availability of genomic resources.

For instance, Ensembl Metazoa (https://metazoa.ensembl.org/index.html) detected at least one putative METTL in the ctenophore *Mnemiopsis leidyi*, which was identified as a potential ortholog of METTL1 (ML09559a). Upon including this putative METTL1 sequence from *M. leidyi* in a separate phylogenetic analysis, it clustered with METTL1 sequences from other metazoan taxa, although with low support (supplementary Fig. S5, Supplementary Material). This indicates that METTLs likely occur in Ctenophora. Nevertheless, evidence of this remains scant due to limited sequence data for multiple ctenophore species across available databases. Indeed, neither this nor similar METTL sequences were detected in Ctenophora by our NCBI search used for this study (Fig. 1). As sequence data continues to increase in quality and quantity, METTLs may later be identified within Ctenophora and we expect additional METTLs will be found within other phyla as well (e.g., Annelida).

METTL2, a methyltransferase that forms N3-methylcytidine (m^3C) in tRNAs (Xu et al. 2017), is found in the most phyla examined here (Fig. 1). Additionally, humans possess two paralogs of METTL2 that share 99% amino acid sequence identity, METTL2A and METTL2B (Arimbasseri et al. 2016; Lentini et al. 2020). Interestingly, the majority of METTLs found present in at least nine metazoan phyla are known to modify RNA (e.g., METTL1, METTL2, METTL3, METTL5, METTL6, METTL14, and METTL15). The ubiquity of RNA methyltransferases underscores the importance of RNA modifications in metazoan evolution. It is probable that RNA molecules required more than only four canonical nucleosides, and that the evolution of chemically distinct modified nucleosides enabled their multiple, necessary functions (Grosjean 2009).

Phylogeny of METTLs part I: Overview

Following our data search, we opted to limit our analyses to nine-preselected metazoan taxa (the sponge *Amphimedon queenslandica*, the stony coral *Acropora millepora*, the sea hare *Aplysia californica*, the priapulid *Priapulus caudatus*, the shrimp *Litopenaeus vannamei*, the sea star *Acanthaster planci*, the acorn worm *Saccoglossus kowalevskii*, the lancelet *Branchiostoma belcheri*, and humans *Homo sapiens*). This allowed us to analyze a reasonable quantity of data while including taxa that represented a diverse set of metazoan phyla. Furthermore, each of these taxa is the subject of numerous ecology and evolution studies and was selected due to its genome availability on NCBI. Thus, the phylogenetic and subsequent evolutionary analyses, with the exception of the episodic selection analysis that is limited to vertebrate METTLs, only include taxa from the phyla Porifera, Cnidaria, Mollusca, Priapulida, Arthropoda, Echinodermata, Hemichordata, and Chordata. Representatives from the phyla Ctenophora, Brachiopoda, Annelida, Platyhelminthes, and Nematoda are not included.

Phylogenetic reconstructions were performed based on an alignment of METTL protein sequences (alignment available in Supplementary Material) from the nine pre-selected metazoan taxa.

Escherichia coli prokaryotic rRNA dimethyltransferase (KsgA) was selected as the outgroup. KsgA, which was first identified in *E. coli* (Helser et al. 1972), dimethylates adenosines in the terminal helix (helix 45) near the 3' end of the small-subunit rRNA (Van Knippenberg et al. 1984). *Escherichia coli* KsgA is a member of the KsgA/Dim1 methyltransferase family, and like METTL proteins, it contains a seven-beta-strand sheet structure (Tu et al. 2009). KsgA, as well as the modifications it catalyzes, are nearly universally conserved throughout evolution (Xu et al. 2008), and have been described in archaea (O'Farrell et al. 2006), eubacteria (Helser et al. 1972; Van Buul et al. 1983), eukaryotes (Lafontaine et al. 1994; Housen et al. 1997), and in eukaryotic organelles (Tokuhisa et al. 1998; Seidel-Rogol et al. 2003). Using both in vivo and in vitro analyses, one study showed that archaeal and eukaryotic orthologs of KsgA were capable of complimenting for KsgA function in bacteria, demonstrating that the recognition elements and methyltransferase activity of the KsgA/Dim1 family have evolved little since the three domains diverged (O'Farrell et al. 2006). Thus, *E. coli* KsgA was selected as a methyltransferase outside of, but that shares structural similarities with, the METTL family, and has orthologs present throughout all domains of life but has demonstrated little evolutionary change. The selection of *E. coli* KsgA as the outgroup for our metazoan METTL phylogeny is further supported by a separate phylogenetic analysis that included several METTL orthologs from eukaryotic taxa outside of Metazoa (supplementary Fig. S5, Supplementary Material). This included METTL protein sequences from the choanoflagellate *Salpingoeca rosetta* and the yeast *Saccharomyces cerevisiae*, all of which grouped with their respective METTL lineages rather than by taxa or with *E. coli* KsgA.

The obtained tree showed that overall, the amino acid sequences cluster together by METTL lineage rather than by the taxa to which they belong (Fig. 2 and supplementary Fig. S1, Supplementary Material). Nearly all METTL lineages were each represented by a distinct monophyletic clade supported with high confidence values (e.g., all METTL1 orthologs from different metazoan taxa formed a monophyletic clade with 100% bootstrap). This observation suggests that the evolution of METTLs has been largely driven by selective constraints associated with the particular functional identity of each METTL protein, likely involved in specific outcomes of structural or functional specialization. Prior studies examining the structures, interactions, and functions of various METTLs have provided insight into their probable targets for methylation (e.g., (Cloutier et al. 2013; Xu et al. 2017; Ignatova et al. 2019). Accordingly, METTL proteins tend to cluster together by their modification targets (i.e., DNA/RNA nucleosides versus protein residues), although with several exceptions.

Despite generally grouping by type (i.e., METTLs that modify nucleic acids versus METTLs that modify proteins), it seems clear that not all of the DNA/RNA-modifying METTLs (Fig. 2, labeled in blue) exclusively possess a single, unique monophyletic origin and therefore do not appear to have evolved from a shared common ancestor. All of the protein-modifying METTLs (Fig. 2, labeled in red)

do not appear to share a single, unique monophyletic origin with one another either. Indeed, METTLs that target nucleosides for methylation are distributed across multiple clades within the phylogeny, and are occasionally grouped with METTLs that modify protein residues. This pattern could be explained by convergent molecular evolution that resulted in the emergence of METTLs with similar targets due to comparable selective pressures and modification requirements (Losos 2011). Alternatively, the diversification of the METTL family may have occurred very early in, or possibly prior to, the evolution of metazoans; this may have been driven by different selective pressures and constraints or even stochastic processes (Stayton 2015). The phylogenetic relationships among the various METTL types that target nucleic acids, proteins, or whose targets remain undetermined is further elaborated below.

Phylogeny of METTLs part II: DNA and RNA modifiers

DNA- and RNA-modifying METTLs target nucleosides for modification, resulting in the formation of methylcytidines, methylguanosines, methyladenosines, or methyluridines. Almost all of these METTLs appear to primarily target RNA molecules, although there is evidence that METTL4 may act to methylate DNA (Kweon et al. 2019; Zhang et al. 2020). METTL15, METTL6, METTL2, and METTL8 are all associated with the formation of methylcytidine. METTL15 falls most basally to the *E. coli* KsgA outgroup (Fig. 2) and forms N^4-methylcytidine (m^4C) in RNA (Van Haute et al. 2019). The remaining proteins that methylate cytidine, METTL2, METTL6, and METTL8, form a highly supported cluster (100% bootstrap). Both METTL2 and METTL6 form m^3C in specific tRNA molecules, whereas METTL8 forms m^3C in mRNA (Xu et al. 2017). Furthermore, of the METTL genes, our NCBI search detected that METTL2 and METTL6 are present across the greatest number of major metazoan phyla, while METTL8 was only found to be present in the phylum Chordata (Fig. 1). METTL1, the only METTL protein known to contribute methylguanosine, forms N^7-methylguanosine (m^7G) in tRNA, mRNA, and miRNA (Okamoto et al. 2014; Pandolfini et al. 2019; Zhang et al. 2019). In the phylogeny, METTL1 groups with methylcytidine-forming proteins as well as with METTL7A and METTL7B, whose precise function is unknown, although with very low confidence (bootstrap < 10%) (Fig. 2).

Several other nucleic acid-modifying METTLs target adenosine to form methyladenosine. These include METTL3, METTL4, METTL5, METTL14, METTL16, and METTL25B, which all function to form m^6A. However, despite this shared function, these METTLs do not form a single cluster in the METTL phylogeny (Fig. 2). Perhaps the best characterized of these are METTL3 and METTL14, which together form a heterocomplex with WTAP (Wilms' Tumor1-Associating Protein) to form m^6A in mRNA (Fu et al. 2014; Liu et al. 2014; Meyer and Jaffrey 2017). METTL3, METTL14, and WTAP are considered to be m^6A writers (i.e., factors that encode the chemical modification) (Lewis et al. 2017). While METTL3 is the primary m^6A-forming enzyme, METTL14 does not demonstrate enzymatic activity

and instead appears to bind substrate RNA and augment METTL3 activity (Sledz and Jinek 2016; P. Wang et al. 2016; X. Wang et al. 2016; Meyer and Jaffrey 2017). Unsurprisingly, METTL3 and METTL14 cluster together in the METTL phylogeny (86% bootstrap) (Fig. 2).

The precise function(s) of METTL4, which clusters closest with METTL14 (95% bootstrap), is less clear than that of the METTL3-METTL14 heterocomplex. There is evidence that METTL4 modifies RNA, DNA, or both nucleic acids. Studies have demonstrated that METTL4 forms m^6A in DNA (Kweon et al. 2019; Zhang et al. 2020), while others have shown that it forms N^6,2'O-dimethyladenosine (m^6Am) in snRNA (Chen et al. 2020; Goh et al. 2020). Interestingly, a recent study demonstrated that the human METTL3-METTL14 complex was active in vitro as a DNA N^6-adenine methyltransferase, although it is unclear if it functions in vivo (Woodcock et al. 2019).

Evidence suggests that METTL5, METTL16, and METTL25B also function in the formation of m^6A. METTL5 and METTL16 cluster near to each other in the METTL phylogeny (92% bootstrap). METTL5 forms a heterodimeric complex with TRMT112 (tRNA methyltransferase subunit 11-2) to provide metabolic stability in the formation of m^6A in 18S rRNA (van Tran et al. 2019; Leismann et al. 2020). The METTL5-TRMT112 complex has structural similarities to that of a m^6A DNA methyltransferase, and has been hypothesized to possess a RNA-binding mode unique to other m^6A methyltransferases (van Tran et al. 2019). METTL16, which forms m^6A in ncRNA, pre-mRNAs (Warda et al. 2017), and mRNA (Nance et al. 2020) appears to be structurally distinct from METTL5 (van Tran et al. 2019) as well as the METTL3-METTL14 heterocomplex (Ruszkowska et al. 2018). Furthermore, because METTL16 has been documented as both a nuclear protein and as a cytoplasmic methyltransferase, it has been suggested that its RNA binding targets may differ by the location of METTL16 within the cell (Nance et al. 2020). The function of METTL25B is less clear, although Gene Ontology (GO) annotations related to METTL25B infers it possesses rRNA (adenine-N6,N6-)-dimethyltransferase activity. Lastly, the function of METTL19, which does not cluster with other RNA-modifying METTLs in the phylogeny (Fig. 2), remains uncertain, but it has been reported to be a likely tRNA uracil-O(2)-methyltransferase that forms 2'-O-methyluridine (Leschziner et al. 2011). Overall, the majority of METTLs that modify nucleic acids appear to target primarily RNA, rather than DNA, nucleosides. While some of these DNA/RNA-modifying METTLs are closely related phylogenetically, they do not share a single monophyletic origin and are instead interspersed with METTLs that modify protein residues.

Phylogeny of METTLs part III: Protein modifiers

Multiple METTL proteins are responsible for the methylation of eukaryotic elongation factor 1 alpha (EEF1A). In addition to its role in protein synthesis as a translation factor that delivers aminoacyl-

tRNA to the ribosome, EEF1A has been ascribed to a wide variety of other functions in eukaryotes beyond protein synthesis (Mateyak and Kinzy 2010). METTL10, METTL21B, and METTL13 are responsible for the methylation of lysine residues within EEF1A (Li et al. 2014; Shimazu et al. 2014; Hamey et al. 2017; Malecki et al. 2017; Jakobsson et al. 2018). Interestingly, these METTLs do not cluster together within the phylogeny (Fig. 2). For example, METTL10 clusters closest with METTL19 (86% bootstrap), which is predicted to form 2'-O-methyluridine in RNA. On the other hand, METTL21B appears to share a close common ancestor with METTL21A (100% bootstrap), which also methylates lysine, although its target is within the molecular chaperone heat shock protein 70 (Hsp70) (Shimazu et al. 2014). METTL13 is unique in that it possesses two distinct methyltransferase domains and methylates the N-terminal as well as a lysine residue (Lys55) within EEF1A (Jakobsson et al. 2018; Liu et al. 2019).

Other METTL proteins methylate lysine residues of valosin-containing protein (VCP), also known as p97, an ATP-driven chaperone involved in numerous, independent cellular processes that is highly conserved across eukaryotes (Kernstock et al. 2012; Meyer et al. 2012). METTL21C and METTL21E, which form a highly supported cluster (100% bootstrap), are both responsible for lysine methylation in VCP (Wiederstein et al. 2018; Wang et al. 2019). Although METTL21D, which is also known to methylate lysine residues of VCP (Kernstock et al. 2012), forms a cluster with other METTL21 lineages (98% bootstrap), it is not as closely related to the other VCP methylases as they are to each other. Our search detected METTL21C and METTL21E only in chordates, whereas METTL21D is present in at least eight different metazoan phyla (Fig. 1). Thus, METTL21D may have evolved prior to METTL21C and METTL21E, although their functions appear to be conserved.

METTL22, which methylates a lysine residue of Kin17 (Cloutier et al. 2013), appears to share a common monophyletic origin with a handful of other lysine methyltransferases, METTL21A-E (Fig. 2). Kin17 functions in DNA replication and repair as well as RNA metabolism (Kannouche et al. 2000; Despras et al. 2003; Masson et al. 2003; Angulo et al. 2005). METTL22 may impact these processes because methylation of Kin17 affects its distribution between chromatin and the cytoplasm (Cloutier et al. 2014). METTL20 targets and trimethylates lysines of the electron transfer flavoprotein β-subunit (ETFβ) (Rhein et al. 2014; Małecki et al. 2015). Unlike most lysine methyltransferases, METTL20 is localized to the mitochondria where it appears to modulate the function of ETF. This difference in target and function may partially explain why METTL20 fails to cluster with other lysine protein methyltransferases (Fig. 2). METTL12 is yet another lysine methyltransferase, although it targets citrine synthase (Małecki et al. 2017; Rhein et al. 2017). Citrine synthase (CS) is localized within the mitochondrial matrix of eukaryotic cells and catalyzes the first step in the Krebs cycle (Wiegand and Remington 1986). Interestingly, METTL12 was not phylogenetically close to any other lysine methyltransferases (Fig. 2).

306 Not all protein-modifying METTL targets are restricted to lysine residues. For instance,
307 METTL18 appears to monomethylate a histidine residue of the ribosomal protein RPL3, and potentially
308 other proteins, to form a 3-methylhistidine (Webb, Zurita-Lopez, et al. 2010; Cloutier et al. 2013).
309 METTL18 is currently the only METTL for which there is evidence that its primary target is histidine,
310 although it shares a common phylogenetic origin with several lysine methyltransferases (Fig. 2). Other
311 protein-modifiers such as METTL11A and METTL11B appear to target N-terminal residues.
312 METTL11A and METTL11B are homologs and cluster closely together (100% bootstrap) (Fig. 2), and
313 while our search found that METTL11A is present in several metazoan phyla, METTL11B is only found
314 in chordates (Fig. 1). METTL11A recognizes the N-terminal X-Pro-Lys sequence motif (Webb, Lipson,
315 et al. 2010) and likely acts upon multiple protein targets (Petkowski et al. 2012; Faughn et al. 2018).
316 Furthermore, METTL11A is capable of mono-, di-, or trimethylating N-terminal residues (Tooley et al.
317 2010). Although it recognizes the same sequence motifs, METTL11B appears to be an N-terminal
318 monomethylase that increases the activity of METTL11A in chordates (Petkowski et al. 2013; Faughn et
319 al. 2018). METTL11A and METTL11B appear to be closely related to METTL9 (94% bootstrap) (Fig.
320 2). Although the precise identities of METTL9 target molecules is currently unknown, METTL9 has been
321 demonstrated to interact with multiple protein partners, including the membrane protein Calnexin
322 precursor (CANX) (Ignatova et al. 2019).
323 Lastly, METTL23 is localized in both the nucleus and cytoplasm, and is predicted to contain a
324 domain for either an adenosine-methionine or lysine methyltransferase (Reiff et al. 2014). Although its
325 precise target has not been definitely identified, METTL23 has been shown to associate with a
326 transcription factor subunit, GABPA (GA-binding protein transcription factor, alpha subunit), and a
327 molecular chaperone, GroEL (Bernkopf et al. 2014; Reiff et al. 2014). Here, METTL23 appears to be
328 phylogenetically related to several lysine methyltransferases that are associated with various molecular
329 chaperones and other stress responsive proteins (e.g., Hsp70, VCP, and possibly Kin17), including
330 METTL21A-E and METTL22 (Fig. 2). Future studies investigating molecular responses to stress may
331 benefit from examining these METTLs and their potential contributions to post-translational
332 modifications of stress-activated proteins. Overall, although the target substrates of many of the protein-
333 modifying METTLs have been identified, much less is known regarding the specific functional purpose
334 behind each modification. It is also worth noting that while there has been a sizable effort to examine
335 enzymes that methylate histones (Clarke 2013), the majority of protein-modifying METTLs discussed
336 here appear to act upon the lysine residues of non-histone proteins. Regardless, all of the protein-
337 modifying METTLs do not share a single, unique monophyletic origin.
338
339 *Phylogeny of METTLs part IV: Unknown modifiers*

The functions and biomolecule target types of several METTLs are currently unknown. This includes METTL7A, METTL7B, METTL17, METTL25, METTL26, and METTL27. The phylogenetic reconstruction may provide some insight into the potential function and target of these METTLs (Fig. 2, labeled in purple), although this should be interpreted with caution. METTL7A has been reported to be an integral membrane protein, and METTL7B has been shown to interact closely with TMEM126A, a mitochondrial membrane protein (Zehmer et al. 2009; Ignatova et al. 2019). Furthermore, both appear to function in lipid droplet formation in the endoplasmic reticulum (Turró et al. 2006; Zehmer et al. 2009). Here we found that METTL7A and METTL7B are closely related to one another (100% bootstrap), and appear to be most phylogenetically related to RNA-modifying METTLs that create m^3C, although with relatively low confidence (70% bootstrap, Fig. 2).

METTL17 has been shown to physically and functionally interact with estrogen receptors (ERs), ERα and ERβ, and acts as a coactivator, modulating their transcriptional activities (Du et al. 2015). This suggests that METTL17 is a protein modifier that targets ER proteins. However, another study proposed that METTL17 may function to form methylcytidine in RNA (Shi et al. 2019). This study found that METTL17 interacted with 12S mitochondrial ribosomal RNA (mt-rRNA) and small subunits of mitochondrial ribosome (MSSU). In particular, METTL17 appeared to regulate m^4C and m^5C modifications in 12S mt-rRNA, and its presence or absence altered mitochondrial ribosome function (Shi et al. 2019). Thus, it is unclear if METTL17 modifies protein residues or RNA nucleosides. In the phylogeny, METTL17 was most closely related to METTL18, a protein histidine methyltransferase, though with low support (Fig. 2).

Lastly, very little is known about METTL25, METTL26, and METTL27, although all three are predicted to have the characteristic seven-beta-strand structure METTL motif and possess S-adenosylmethionine-dependent methyltransferase (SAM) activity. METTL25 possibly functions in RNA methylation as it is phylogenetically very similar to METTL25B (100% bootstrap, Fig. 2), which forms m^6A in rRNA. METTL26 forms a lowly supported cluster with METTL11A, METTL11B, and METTL9, which may suggest its function as a protein-modifying METTL. Based on the phylogeny, there is no clear indication of the potential function or target of METTL27, which forms a lowly supported cluster with a protein-modifying METTL (METTL12), several RNA-modifying METTLs (METTL2, METTL6, and METTL8), and other METTLs of unknown function (METTL7A and METTL7B) (Fig. 2). The phylogenetic tree may provide some indication of the potential substrates for the METTLs of currently unknown functions. However, given that the DNA/RNA-modifying METTLs and protein-modifying METTLs are not separated into distinct clades, we must once again stress that these findings should be interpreted with caution. Additional studies will be required to identify the specific purpose of

373 METTL7A, METTL7B, METTL17, METTL25, METTL26, and METTL27, but the phylogeny does
374 provide some insight into their evolution.
375
376 *Phylogeny of METTLs part V: Conserved domains*
377 All conserved domains (CDs) that were identified within each METTL fell within one of two CD
378 superfamilies: the AdoMet MTases superfamily or the MT-A70 superfamily (Table 2 and supplementary
379 Tables S1 and S2, Supplementary Material). Only METTL3, METTL14, and METTL4 contain CDs
380 within the MT-A70 superfamily (CDD Accession cl01947) while all other METTLs contain CDs within
381 the AdoMet MTases superfamily (CDD Accession cl17173). *S*-adenosyl-L-methionine (SAM or AdoMet)
382 methyltransferases (MTases) use SAM/AdoMet as a substrate for methyl transfer (Schubert et al. 2003).
383 MT-A70 refers specifically to the 70-kDa SAM-binding subunit first identified in human mRNA (N6-
384 adenosine)-methyltransferase (i.e., human METTL3) (Bokar et al. 1997; Bujnicki et al. 2002). Thus, the
385 CD Search successfully identified the conserved SAM binding domains formed by the central seven-beta-
386 strand motif that is structurally conserved across all METTLs.
387 For several METTLs, no specific CD was recognized (i.e., only the CD superfamily was
388 identified). A second phylogenetic reconstruction was performed using only the CD-containing regions of
389 the protein sequences (alignment available in Supplementary Material). Interestingly, although some
390 METTLs that share CDs cluster together (e.g., METTL3, METTL14, and METTL4), this is not true for
391 all METTLs with shared CDs (Fig. 3 and supplementary Fig. S2, Supplementary Material). For instance,
392 eight different METTLs share the CD Methyltransf_25 (Table 2), but these METTLs form separate
393 clusters in the phylogenetic reconstruction (Fig. 3). Thus, even within relatively well-conserved areas of
394 the amino acid sequence, there exists enough sequence-level diversity to support the specialization of
395 multiple METTLs with different and highly specific targets and functions.
396 Overall the full sequence phylogeny and the conserved domain phylogeny were similar, which
397 indicates that METTL evolution was largely determined by the selective constraints operating on these
398 CDs. However, there were some notable differences between the full sequence and the CD phylogenies.
399 For instance, in the phylogeny represented by full protein sequences (Fig. 2), METTL17 was most closely
400 related to METTL18, a protein histidine methyltransferase, whereas in the CD phylogeny (Fig. 3),
401 METTL17 forms a lowly conserved cluster with several RNA-modifying METTLs: METTL6, METTL2,
402 METTL8, and METTL1 (62% bootstrap). As previously mentioned, there is conflicting evidence
403 regarding whether METTL17 acts on protein residues or RNA nucleosides. Another difference between
404 the phylogenetic reconstructions worth noting is the relationship of METTL27 to other METTLs, which
405 remained unclear in the full sequence phylogeny (Fig. 2). In the conserved domain phylogeny, however,
406 METTL27 is most closely related to METTL11A and METTL11B (78% bootstrap, Fig. 3), which modify

protein N-terminal residues. Thus, restricting the protein sequences to only the CD regions provided some additional insight into the potential targets of METTLs of unknown function.

Molecular evolution and selection in METTL proteins

Despite shared similarities in protein structure (i.e., a SAM binding domain and seven-beta-strand motif), METTL evolution throughout Metazoa appears to have been largely driven by functional constraints that led to the specialization of each METTL lineage. To further examine the evolutionary mechanisms underlying their differentiation, variation within each METTL lineage was evaluated using two separate alignments (nucleotide and protein) of all METTLs from each of the nine pre-selected metazoan taxa (Table 3) (alignments available in Supplementary Material). While variation in METTLs with only one representative taxa/sequence in the selected dataset was not assessed (i.e., METTL8, METTL11B, METTL21B, METTL21C, and METTL21E), the obtained results showed that METTL27, a METTL of unknown function, displayed the greatest variation (mean amino acid distance, $d_{AA} = 0.69 \pm 0.03$) (Table 3). METTL27 also displayed slightly higher levels of codon bias (ENC = 48.47). However, codon bias was relatively low overall, with average ENC values per METTL ranging from 45.13 (METTL21B) to 54.92 (METTL7B). The second-most variation in protein sequences was observed in METTL12 ($d_{AA} = 0.63 \pm 0.03$), a lysine methyltransferase found in several different metazoan phyla (Fig. 1) that was not closely related to other lysine methyltransferases in the phylogenetic reconstruction (Fig. 2). The METTL that exhibited the least diversity was METTL14 ($d_{AA} = 0.32 \pm 0.02$), a m^6A writer that forms a heterocomplex with METTL3 and WTAP. The next lowest levels of variation were in METTL1, METTL2, METTL3, METTL5, and METTL6 (average $d_{AA} = 0.38 \pm 0.02$). Thus, the six METTLs that showed the least amount of variation are each associated with the methylation of RNA nucleosides, indicating that high sequence conservation may be necessary to preserve their function.

Many conserved multigene families are thought to be subject to concerted evolution, in which gene family members evolve as a unit (i.e., in concert) via gene conversion or unequal crossing-over (Eirín-López et al. 2012). In the case of concerted evolution, we might expect to observe roughly the same level of synonymous variation (d_S) and nonsynonymous variation (d_N) because gene conversion and crossover processes should affect synonymous and nonsynonymous sites equally (Nei and Rooney 2005). However, the nucleotide variation underlying the diversity observed among METTLs is primarily synonymous. For all METTLs, d_S was significantly greater than d_N and all Z-tests for selection (i.e., HA: $d_N < d_S$) were highly significant (p-value < 0.001) (Table 3). Therefore, strong purifying selection appears to have operated on the different METTLs. This finding was additionally supported by the results from Fisher's exact tests of selection (supplementary Tables S3 and S4, Supplementary Material). Indications of strong purifying selection have been reported in many other protein coding gene families, including

highly conserved histone and ubiquitin gene families (Nei et al. 2000; Piontkivska et al. 2002; Rooney et al. 2002; Eirín-López et al. 2004; González-Romero et al. 2008; González-Romero et al. 2010; González-Romero et al. 2012). These studies argued that, rather than concerted evolution, these gene families are subject to the birth-and-death model of evolution in which gene family members evolve independently via gene duplication followed by their maintenance or loss (Nei and Hughes 1992). Indeed, the birth-and-death concept may be the primary mechanism directing the long-term evolution of most multigene families (Eirín-López et al. 2012). Although detecting birth-and-death was not a goal of the present work, future studies of this gene family might investigate intra- and interspecific gene duplication and the presence of pseudogenes, which would further support that METTLs are subject to this model of evolution.

Rates of METTL evolution

In order to further explore functional constraints across METTLs that vary in target and function, the present study conducted estimations of the rates of evolution in METTL proteins. Results show that several RNA-modifying METTLs exhibited a relatively slow rate of evolution (Fig. 4). In particular, METTL1, the only METTL known to form methylguanosine, had the slowest rate of protein evolution at 3.91×10^{-4} amino acid substitutions/site/million years (MY) (supplementary Table S5, Supplementary Material). METTL14 and METTL3, which together form a heterocomplex to form m^6A in RNA, also had relatively slow rates of evolution. In fact, the six METTLs with the slowest evolution rates (i.e., METTL1, METTL14, METTL2, METTL6, METTL3, and METTL5) all function to methylate RNA molecules (Fig. 4 and supplementary Table S5, Supplementary Material). These same METTLs are also found in multiple metazoan phyla, including basal metazoans (Fig. 1). Thus, the formation of m^7G, m^3C, and m^6A modifications in RNA appears to be a fundamental epitranscriptomic mechanism that has been preserved throughout metazoan evolution. In contrast, METTL12, otherwise known as citrate synthase-lysine N-methyltransferase (CSKMT), exhibited the fastest rate of evolution with 7.95×10^{-4} amino acid substitutions/site/MY (Fig. 4 and supplementary Table S5, Supplementary Material). Methylation modifications have been identified in numerous diverse, structurally complex, and specialized proteins (Paik and Kim 1971; Clarke 2013). Therefore, a comparatively rapid evolution and diversification of protein-modifying methyltransferases may have been necessary to target these proteins and fulfill a wide variety of highly specific roles. Protein-modifying METTLs being subjected to different selective pressures than those that modify DNA/RNA nucleosides may explain their different rates of evolution. Also exhibiting relatively rapid rates of evolution were METTL7A and METTL17, although the precise functions of these proteins are unknown. All remaining METTLs that exhibited a relatively intermediate rate of evolution were a combination of RNA-modifying METTLs, protein-modifying METTLs, and

475 METTLs of unknown function (Fig. 4). These METTLs exhibited highly similar rates of evolution at an
476 average of $6.64 \times 10^{-4} \pm 3.48 \times 10^{-5}$ amino acid substitutions/site/MY (supplementary Table S5,
477 Supplementary Material).
478
479 *Episodic selection within METTL lineages*
480 We wanted to explore what drives the variation in evolutionary rates that we calculated across the
481 METTL family, and we searched for indications of adaptive selective episodes that may have occurred
482 during METTL evolution. Since METTL lineages each have a monophyletic origin (Fig. 2), additional
483 evolutionary analyses using only vertebrate METTLs were implemented to further explore how selection
484 drove the differentiation of METTL types (alignment available in Supplementary Material). Vertebrates
485 were selected because all members of the METTL family are present within this subphylum of
486 metazoans. Only two representative orthologous sequences per METTL were included in the analysis due
487 to computational requirements and constraints. Accordingly, we examined ortholog sequences of all
488 METTLs from each of two classes of vertebrates: Mammalia and Amphibia. METTL7B was not included
489 in this analysis as orthologous gene searches in NCBI failed to clearly identify METTL7B orthologs in
490 amphibians. The obtained results revealed that METTL proteins diverged from a homogeneous evolution
491 pattern as expected by the global molecular clock hypothesis (ln*L* with clock = -10239.1, ln*L* without
492 clock = -9715.854, *p*-value < 0.001). This result was also supported when testing the molecular clock
493 hypothesis using nucleotide sequences of the vertebrate METTLs (ln*L* with clock = -109637.787, ln*L*
494 without clock = -108792.321, *p*-value < 0.001). These findings, along with the estimated rates of
495 evolution reported above, reaffirm that various METTL proteins appear to have evolved at different rates
496 from one another.
497 Given the observed heterogeneous rates of evolution across the METTL family, the presence of
498 diversifying (i.e., adaptive) selection episodes was examined across vertebrate METTLs, uncovering
499 traces of diversifying selection ($\omega > 1$) on specific branches of the phylogeny (*p*-value < 0.05) (Fig. 5A).
500 The terminal branch leading to METTL6, which forms m^3C in tRNA (Xu et al. 2017), was significant (*p*-
501 value < 0.01) (Fig. 5A). The terminal branch of METTL9, a probable protein methyltransferase (Ignatova
502 et al. 2019), also showed significant traces of diversifying selection (*p*-value < 0.001) (Fig. 5A). Several
503 subtrees that included closely related METTLs (i.e., METTL11A and METTL11B; METTL21A,
504 METTL21B, METTL21C, METTL21D, and METTL21E; METTL25 and METTL25B) were collapsed
505 into a single branch for figure readability.
506 Several individual sites subject to diversifying selection in METTLs were also identified using
507 the mixed effects model of evolution (MEME) model (Murrell et al. 2012) (Fig. 5B-C), a method we have
508 previously used to explore the evolution of high mobility group (HMG) proteins (González-Romero et al.

2015) and sex-determining proteins (Eirín-López and Sánchez 2015). As expected, the number and identities of these sites varied depending on whether analyses included all METTL types, or if analyses discriminated among METTLs based on their target molecule type (i.e., those that target DNA or RNA nucleosides versus those that target protein residues). Upon analyzing all METTL sequences available for vertebrates regardless of type (alignment available in Supplementary Material), 19 codon sites were identified, most of which were subjected to episodic positive selection (Fig. 5B). Four sites of predominantly positive selection (codons 89, 114, 747, and 785) were common to the majority of METTLs, spanning a variety of different targets and functions (Fig. 5C). Analysis of these codon positions within the context of the METTL phylogeny suggests that mutations at these sites led to the differentiation of multiple METTL lineages (Fig. 5C).

Evidence of positive selection at codon 89 was indicated at several internal nodes of the phylogeny, as well as the terminal branches leading to METTL1, METTL3, and METTL14 (Fig. 5C). Codon 89 was also identified as a site of positive selection when the MEME analysis was restricted to only METTLs that modify DNA or RNA nucleosides (Fig. 6A). Analysis of codon 114 indicated instances of both positive and negative (purifying) selection (Fig. 5C). Specifically, although the majority of METTL lineages were subject to positive selection at codon 114, a few terminal branches (e.g., METTL3 and METTL14) indicate the presence of more synonymous than nonsynonymous mutations. Codon 114 was additionally identified by both MEME analyses that separately analyzed DNA/RNA-modifying METTLs and protein-modifying METTLs (Fig. 6). Thus, episodic selection at codon 114 may have been particularly important to the diversification of METTLs, both within and across different METTL types. Mutations at codon 747 were also common to many different METTL lineages within the phylogeny. Interestingly, this position is located within the conserved domain regions of all METTLs that possess CDs belonging to the AdoMet MTases superfamily (supplemental Table 2, Supplementary Material). Codon 747 is not, however, located within the CDs belonging to the MT-A70 superfamily, which are found in METTL3, METTL14, and METTL4. Codon 785 had the most substitutions of the sites identified from the MEME analysis of all vertebrate METTLs (Fig. 5B), and appears to have contributed to the differentiation of many of the METTL lineages. This site was also identified as a site of positive selection for the MEME analysis of only DNA/RNA-modifying METTLs, and therefore, appears to have played a significant role in the diversification of METTLs that bind to nucleosides. Although only a few prominent sites identified from the MEME analysis of all METTLs are discussed here, additional substitutions at over a dozen more positions were identified (Fig. 5B and 5C).

Several additional sites subject to diversifying selection that were not recognized in the indiscriminate MEME analysis of all vertebrate METTLs were identified when performing MEME analyses that were separated by METTL type (Fig. 6A). The analysis of only METTLs that modify DNA

or RNA nucleosides identified additional sites of predominantly positive selection at codons 33, 53, 365, 1295, 1520, 1793, and 1825 (Fig. 6A). Sites at codons 89, 114, 414, and 785 were previously identified in the METTL analysis that included all vertebrate METTLs (Fig. 5B). Upon examining these sites within the context of the phylogeny, a mutation at codon 89 shows evidence of positive selection on the internal node leading to METTL2, METTL6, METTL8, and METTL15 (Fig. 6B). However, additional substitutions at codons 33, 53, 114, 785, 1520, and 1825 were necessary for the differentiation of these METTLs (Fig. 6B), all of which are responsible for the formation of methylcytidine in RNA (Xu et al. 2017; Van Haute et al. 2019). Although the evolution of METTLs that target DNA/RNA nucleosides included several sites that were primarily subjected to positive selection, there were also evidence of neutral evolution and purifying selection for certain METTL lineages at a few sites (i.e., codon 114, 1793, and 1825) (Fig. 6B).

Although sites at codons 114 and 541 had been identified in the METTL analysis that included all vertebrate METTLs (Fig. 5B), by separately analyzing only METTLs that modify protein residues, MEME identified additional sites of predominantly positive selection at codons 69, 72, 91, 243, 719, 768, and 997 (Fig. 6A and 6C). Nearly all of the site mutations identified using MEME in protein-modifying METTLs were subjected to positive selection. However, there was evidence of purifying selection at codon 69 (METTL20), codon 719 (METTL23), and codon 768 (METTL10 and METTL22) (Fig. 6C). There were also evidence of equal synonymous and nonsynonymous substitutions (i.e., neutral evolution) for METTL22 at codons 72 and 997 when analyzed within the context of the phylogeny (Fig. 6C).

Conclusions

METTLs are widespread throughout the animal kingdom, with some having emerged quite early in the evolution of Metazoa. Individual METTL lineages formed independent monophyletic clades, and while various METTLs grouped together roughly by methylation target type, those with similar targets did not appear to have evolved from the same, recent common ancestor (e.g., all METTLs that modify nucleosides did not share an exclusive, single monophyletic origin). Evidence indicates that the long-term evolution of the METTL family is primarily driven by strong purifying selection and exhibits heterogeneous rates of evolution across the different METTL lineages. Functional specialization of the various METTLs seems to have occurred via episodes of adaptive selection at specific evolutionary times and codon sites. Lastly, given their presence in basal metazoans, comparatively low sequence-level variation, and slow estimated rates of evolution, several METTLs that target and methylate RNA nucleosides seem to be more conserved than those that modify protein residues, possibly due to their biological necessity and particular functional constraints. In view of the ubiquity of METTLs throughout metazoans, it is clear that these proteins fulfill a diverse set of essential biological functions across an

expansive range of animal taxa, and that the epigenetic mechanisms associated with this gene family contributed to metazoan evolution and complexity. We therefore stress the importance of investigating these methyltransferases in species beyond the traditional metazoan model systems. This present work provides important insight into the evolution of this gene family and constitutes a valuable resource for studying the epigenetic functions of METTL proteins, and their impacts on biological processes and phenotype.

Materials and Methods

Molecular Data Mining

Extensive data mining was performed using the GenBank database (https://www.ncbi.nlm.nih.gov/genbank) to collect METTL sequences (available as of June 2020). METTLs across metazoan taxa were located using orthologous gene searches as well as the Basic Local Alignment Search Tool (BLAST) in NCBI. Each METTL gene (NCBI IDs provided in Table 1) was used to search for NCBI orthologs and similar genes. These are calculated using protein sequence similarity, local synteny information, and similarity of protein architectures (CDD domains defined by NCBI SPARCLE) by NCBI's Eukaryotic Genome Annotation pipeline and Gene database. Each METTL gene listed in Table 1 was searched in BLAST using the somewhat similar sequences (blastn) algorithm and limiting the organism search set to Porifera (NCBI: txid6040), Ctenophora (NCBI: txid10197), Cnidaria (NCBI: txid6073), Brachiopoda (NCBI: txid7568), Mollusca (NCBI: txid6447), Annelida (NCBI: txid6340), Platyhelminthes (NCBI: txid6157), Priapulida (NCBI: txid33467), Nematoda (NCBI: txid6231), Arthropoda (NCBI: txid6656), Hemichordata (NCBI: txid10219), and Echinodermata (NCBI: txid7586). For any phyla for which search results only located sequences that were similar to a METTL, but did not have any predicted gene identity or function recorded in NCBI, they were noted as probable METTL sequences for that phyla.

Ultimately, sequences were selected from nine representative taxa across eight different metazoan phyla for phylogenetic and evolution analyses. These taxa included the sponge (phylum Porifera) *Amphimedon queenslandica* (NCBI: txid400682), the stony coral (phylum Cnidaria) *Acropora millepora* (NCBI: txid45264), the sea hare (phylum Mollusca) *Aplysia californica* (NCBI: txid6500), the priapulid (phylum Priapulida) *Priapulus caudatus* (NCBI: txid3762), the shrimp (phylum Arthropoda) *Litopenaeus vannamei* (NCBI: txid6689), the sea star (phylum Echinodermata) *Acanthaster planci* (NCBI: txid133434), the acorn worm (phylum Hemichordata) *Saccoglossus kowalevskii* (NCBI: txid10224), the lancelet (phylum Chordata) *Branchiostoma belcheri* (NCBI: txid7741), and humans (phylum Chordata) *Homo sapiens* (NCBI: txid9606). A total of 207 nucleotide coding sequences (CDS) and their corresponding amino acid sequences were collected across 33 METTLs (supplementary Tables S6 and

S7, Supplementary Material). Sequence alignments for both CDS nucleotide sequences and amino acid sequences were performed using MAFFT version 7.309 (Katoh 2002; Katoh and Standley 2013) in Geneious version 9.1.8 (https://www.geneious.com). All METTL CDS nucleotide sequences, using a translation alignment, and all METTL amino acid sequences were aligned. Both the CDS nucleotide and amino acid alignments were aligned using the E-INS-I strategy algorithm and the BLOSUM62 scoring matrix with a 1.53 gap open penalty. Conserved domains (CDs) were identified within each of 207 CDS nucleotide sequences and 207 protein sequences using the CD-Search and Batch CD-Search tools (Marchler-Bauer and Bryant 2004; Marchler-Bauer et al. 2011) in NCBI. All sequences were searched across the Conserved Domain Database (CDD) (Lu et al. 2020) using an expect value (E-value) threshold of 0.01.

Phylogenetic analyses

The ribosomal RNA small subunit methyltransferase A (KsgA) from *Escherichia coli* O157:H7 strain Sakai (NCBI: txid386585) was included as an outgroup for the phylogenetic analyses. Best-fit substitution models for nucleotide and protein sequence alignments were calculated using a maximum likelihood approach in MEGA version 10.1.8 (Tamura et al. 2013), and included all used sites (i.e., no data was excluded). Maximum likelihood phylogenetic trees were reconstructed using IQ-TREE version 1.6.12 (Nguyen et al. 2015) using the best-fit substitution models determined from MEGA. Accordingly, the LG model (Le and Gascuel 2008) corrected for discrete Gamma model (Yang 1994) with four rate categories (LG+G) was used for the reconstruction of METTL protein phylogenies across the nine pre-selected metazoan taxa. Similarly, a protein tree was also reconstructed for all METTL proteins across the nine pre-selected metazoan taxa but using only conserved domain amino acid sequences identified using the Batch CD-Search tools (Marchler-Bauer and Bryant 2004; Marchler-Bauer et al. 2011) in NCBI. The reliability of the reconstructed topologies was contrasted by an ultra-fast bootstrap approximation (UFBoot) (Hoang et al. 2018) with 1,000 replicates. Additional phylogenetic reconstructions were performed using the CDS nucleotide alignment as well as for individual METTL lineage protein sequences (supplementary Figs. S3 and S4, Supplementary Material).

Molecular evolution and selection analyses

Several molecular evolutionary analyses were performed in MEGA version 10.1.8 (Tamura et al. 2013). Sequences from each METTL gene were analyzed based on CDS nucleotide and amino acid alignments of all METTL sequences from each of the nine pre-selected metazoan taxa (alignments available in Supplementary Materials). The nucleotide alignment includes 9,684 nucleotide sites (3,228 codons) and the protein alignment includes 3,124 amino acid sites. The transition/transversion ratio (R)

was calculated for all METTLs except for those in which only one sequence across the nine pre-selected taxa was available (i.e., METTL8, METTL11B, METTL21B, METTL21C, and METTL21E). Due to their smaller variance (Nei and Kumar 2000), nucleotide (d_{NT}) and protein (d_{AA}) sequence distances were estimated using uncorrected differences (*p*-distances) for each METTL in which sequences were obtained from more than one representative taxa (i.e., all METTLs with the exception of METTL8, METTL11B, METTL21B, METTL21C, and METTL21E). These estimations, along with their standard errors, were calculated using bootstrap variance estimation (1,000 replicates) and the proportion of different nucleotide sites (i.e. transitions + transversions). Nucleotide and amino acid distances were computed using uniform rates among sites. Gaps and missing data were removed prior to the analysis if a site had higher than 95% of ambiguous sites (i.e., partial deletion). The rates of protein evolution were estimated for each METTL based on the distances obtained, in those cases in which there were representatives across at least three of the pre-selected metazoan taxa. METTLs that did not fulfill this requirement and were therefore excluded from the rate of evolution analysis included METTL7B, METTL8, METTL11B, METTL21B, METTL21C, METTL21E, and METTL27. Evolution rates were estimated by correlating pairwise protein divergences between pairs of metazoan taxa with their corresponding divergence as defined in the TimeTree database (Kumar et al. 2017) (supplementary Table S8, Supplementary Material). Linear regression analyses were implemented using Microsoft Excel version 16.42.

The footprint of selection on METTL genes was studied using two major approaches. First, descriptive analyses of nucleotide variation and the mode of evolution displayed by METTLs were carried out using the CDS nucleotide alignment of METTLs from the nine pre-selected taxa (alignment available in Supplementary Material). Accordingly, the numbers of synonymous (d_S) and non-synonymous (d_N) substitutions per site, as well as their standard errors, were computed using bootstrap variance estimations (1,000 replications) and the modified Nei-Gojobori method (Nei and Gojobori 1986) with the corresponding transition/transversion ratio (R), uniform rates among sites, and partial deletion (95% coverage cutoff) of gaps and missing data. Codon-based Z-tests were performed to gauge the presence and nature of selection by comparing the estimated number of synonymous (d_S) versus nonsynonymous (d_N) nucleotide differences per site. In this case, estimations of nucleotide substitutions were performed using uncorrected differences (*p*-distances) considering heterogeneity in transition/transversion ratio, as these encompass smaller variance than other methods (Nei and Kumar 2000). In addition, the large number of taxa analyzed further contributes to reducing the potential effects of multiple substitutions by breaking long branches (Lartillot and Philippe 2008). For these tests, neutral evolution was set as the null hypothesis (H_0: $d_N = d_S$) with purifying selection set as the alternative hypothesis (H_A: $d_N < d_S$). The variance of the difference between d_N and d_S was estimated using the bootstrap method with 1,000 replications and a modified Nei-Gojobori method (Nei and Gojobori 1986)

with the corresponding transition/transversion ratio (R). Codon-based Fisher's exact tests of selection were also performed (supplementary Tables S3 and S4, Supplementary Material), but with positive selection set as the alternative hypothesis (H_A: $d_S < d_N$). The amount of codon usage bias and the presence of molecular clocks were investigated using the programs DnaSP version 5 (Librado and Rozas 2009) and HyPhy (Pond et al. 2005), respectively.

Second, the presence of lineages displaying evidence of diversifying (adaptive) selection episodes ($\omega > 1$) was examined across vertebrate METTL evolution by using the branch-site Random Effects Likelihood (REL) model (Pond and Frost 2005). To this end, a total of 1,856 codon positions were examined using a maximum likelihood phylogeny that was reconstructed using vertebrate METTL nucleotide coding regions as a reference (alignment available in Supplementary Material). The alignment included two ortholog sequences (i.e., one from Mammalia and one from Amphibia) of each METTL that were downloaded from GenBank (supplementary Table S9, Supplementary Material). In this case, nucleotide substitution models incorporating multiple substitutions were used as indicated in Table 3. For this instance, the best-fit model of evolution was defined as a general time reversible model (Tavaré 1986) with a discrete Gamma model that allows for a proportion of invariable sites (Gu et al. 1995) (GTR+G+I). No prior assumptions about which lineages have been subject to diversifying selection were made. The proportion of sites inferred to be evolving under diversifying selection at each branch was estimated using likelihood ratio tests (LRTs), resulting in a *p*-value for episodic selection. The strength of selection was partitioned for descriptive purposes into three categories ($\omega > 5$, $\omega = 1$, and $\omega = 0$), using three different significance levels (*p*-value < 0.001, *p*-value < 0.01, and *p*-value < 0.05) to assess the obtained results. Additionally, the presence of selection at individual sites was assessed by using a mixed effects model of evolution (MEME), modeling variable ω (d_N/d_S) across lineages at individual sites (Murrell et al. 2012). Codons subject to significant episodes of diversifying selection (*p*-value < 0.05) were detected using MEME, and analyzed in the context of the METTL phylogeny, providing information on internal branches accumulating higher numbers of non-synonymous mutations. In addition to conducting this analysis across all vertebrate METTLs, additional analyses using MEME were performed separately for vertebrate DNA/RNA-modifying METTLs only and for vertebrate protein-modifying METTLs only. All analyses in this section were carried out using the HyPhy program (Pond et al. 2005) and the Datamonkey web server (Poon et al. 2009; Delport et al. 2010)

Data Availability

The data underlying this article are available within the National Center for Biotechnology Information (NCBI) GenBank database (https://www.ncbi.nlm.nih.gov/genbank), and all GenBank accession numbers

712 are listed in supplementary Tables S6, S7, and S9 (Supplementary Material). Alignments analyzed in the
713 present work are available in the article's online Supplementary Material.
714
715 **Supplementary Material**
716 Supplementary alignments, Newick tree files, episodic selection results, tables, and figures are included
717 as Supplementary Material available at *Molecular Biology and Evolution* online.
718
719 **Acknowledgments**
720 This work was supported by grants from the National Science Foundation (Grant No. EF-1921402)
721 awarded to JEL. JMW was supported by the College of Arts, Science and Education Distinguished
722 Postdoctoral Program at Florida International University, and by the NSF Postdoctoral Research
723 Fellowships in Biology Program under Grant No. DBI-2010791. This material is also based upon work
724 supported by the National Science Foundation under Grant No. HRD-1547798. This NSF Grant was
725 awarded to Florida International University as part of the Centers of Research Excellence in Science and
726 Technology (CREST) Program. This is contribution number 266 from the Coastlines and Oceans
727 Division of the Institute of Environment at Florida International University.

References

Angulo JF, Mauffirey P, Pinon-Lataillade G, Miccoli L, Biard DSF. 2005. Putative Roles of kin17, a Mammalian Protein Binding Curved DNA, in Transcription. In: DNA conformation and transcription. p. 75–89.

Arimbasseri AG, Iben J, Wei F-Y, Rijal K, Tomizawa K, Hafner M, Maraia RJ. 2016. Evolving specificity of tRNA 3-methyl-cytidine-32 (m3C32) modification: a subset of tRNAsSer requires N6-isopentenylation of A37. *RNA* 22:1400–1410.

Bernkopf M, Webersinke G, Tongsook C, Koyani CN, Rafiq MA, Ayaz M, Müller D, Enzinger C, Aslam M, Naeem F, et al. 2014. Disruption of the methyltransferase-like 23 gene METTL23 causes mild autosomal recessive intellectual disability. *Hum. Mol. Genet.* 23:4015–4023.

Bhattacharyya M, De S, Chakrabarti S. 2020. Origin and Evolution of DNA methyltransferases (DNMT) along the tree of life: A multi-genome survey. *bioRxiv* [Internet]. Available from: http://dx.doi.org/10.1101/2020.04.09.033167

Bokar JA, Shambaugh ME, Polayes D, Matera AG, Rottman FM. 1997. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA* 3:1233–1247.

Bujnicki JM, Feder M, Radlinska M, Blumenthal RM. 2002. 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.* 55:431–444.

Carroll SB, Grenier JK, Weatherbee SD. 2013. From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design. John Wiley & Sons

Cheng X, Blumenthal R. 1999. S-adenosylmethionine-dependent Methyltransferases: Structures and Functions. World Scientific

Chen H, Gu L, Orellana EA, Wang Y, Guo J, Liu Q, Wang L, Shen Z, Wu H, Gregory RI, et al. 2020. METTL4 is an snRNA m6Am methyltransferase that regulates RNA splicing. *Cell Research* [Internet] 30:544–547. Available from: http://dx.doi.org/10.1038/s41422-019-0270-4

Clarke SG. 2013. Protein methylation at the surface and buried deep: thinking outside the histone box. *Trends Biochem. Sci.* 38:243–252.

Cloutier P, Lavallée-Adam M, Faubert D, Blanchette M, Coulombe B. 2013. A Newly Uncovered Group of Distantly Related Lysine Methyltransferases Preferentially Interact with Molecular Chaperones to Regulate Their Activity. *PLoS Genetics* [Internet] 9:e1003210. Available from: http://dx.doi.org/10.1371/journal.pgen.1003210

Cloutier P, Lavallée-Adam M, Faubert D, Blanchette M, Coulombe B. 2014. Methylation of the DNA/RNA-binding protein Kin17 by METTL22 affects its association with chromatin. *J. Proteomics* 100:115–124.

Couture J-F, Trievel RC. 2006. Histone-modifying enzymes: encrypting an enigmatic epigenetic code. *Curr. Opin. Struct. Biol.* 16:753–760.

Dabe EC, Sanford RS, Kohn AB, Bobkova Y, Moroz LL. 2015. DNA Methylation in Basal Metazoans:

766 Insights from Ctenophores. *Integr. Comp. Biol.* 55:1096–1110.

767 Deans C, Maggert KA. 2015. What do you mean, "epigenetic"? *Genetics* 199:887–896.

768 Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL. 2010. Datamonkey 2010: a suite of
769 phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26:2455–2457.

770 Despras E, Miccoli L, Créminon C, Rouillard D, Angulo JF, Biard DSF. 2003. Depletion of KIN17, a
771 human DNA replication protein, increases the radiosensitivity of RKO cells. *Radiat. Res.* 159:748–
772 758.

773 Du P, Yuan B, Cao J, Zhao J, Ding L, Chen L, Ying S, Jiang L, Lin J, Xu X, et al. 2015.
774 Methyltransferase-like 17 physically and functionally interacts with estrogen receptors. *IUBMB Life*
775 67:861–868.

776 Eirín-López JM, González-Tizón AM, Martínez A, Méndez J. 2004. Birth-and-death evolution with
777 strong purifying selection in the histone H1 multigene family and the origin of orphon H1 genes.
778 *Mol. Biol. Evol.* 21:1992–2003.

779 Eirín-López JM, Rebordinos L, Rooney AP, Rozas J. 2012. The birth-and-death evolution of multigene
780 families revisited. *Genome Dyn.* 7:170–196.

781 Eirín-López JM, Sánchez L. 2015. The comparative study of five sex-determining proteins across insects
782 unveils high rates of evolution at basal components of the sex determination cascade. *Dev. Genes*
783 *Evol.* 225:23–30.

784 Faughn JD, Dean WL, Schaner Tooley CE. 2018. The N-terminal methyltransferase homologs NRMT1
785 and NRMT2 exhibit novel regulation of activity through heterotrimer formation. *Protein Sci.*
786 27:1585–1599.

787 Fu Y, Dominissini D, Rechavi G, He C. 2014. Gene expression regulation mediated through reversible
788 m6A RNA methylation. *Nature Reviews Genetics* [Internet] 15:293–306. Available from:
789 http://dx.doi.org/10.1038/nrg3724

790 Goh YT, Koh CWQ, Sim DY, Roca X, Goh WSS. 2020. METTL4 catalyzes m6Am methylation in U2
791 snRNA to regulate pre-mRNA splicing. *Nucleic Acids Res.* 48:9250–9261.

792 Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh C-L, Zhang X, Golic KG, Jacobsen SE, Bestor TH.
793 2006. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science* 311:395–
794 398.

795 González-Romero R, Ausió J, Méndez J, Eirín-López JM. 2008. Early Evolution of Histone Genes:
796 Prevalence of an "Orphon" H1 Lineage in Protostomes and Birth-and-Death Process in the H2A
797 Family. *Journal of Molecular Evolution* [Internet] 66:505–518. Available from:
798 http://dx.doi.org/10.1007/s00239-008-9109-1

799 González-Romero R, Eirín-López JM, Ausió J. 2015. Evolution of high mobility group nucleosome-
800 binding proteins and its implications for vertebrate chromatin specialization. *Mol. Biol. Evol.*
801 32:121–131.

802 González-Romero R, Rivera-Casas C, Ausió J, Méndez J, Eirín-López JM. 2010. Birth-and-death long-
803 term evolution promotes histone H2B variant diversification in the male germinal cell line. *Mol.*

Biol. Evol. 27:1802–1812.

González-Romero R, Rivera-Casas C, Frehlick LJ, Méndez J, Ausió J, Eirín-López JM. 2012. Histone H2A (H2A.X and H2A.Z) Variants in Molluscs: Molecular Characterization and Potential Implications For Chromatin Dynamics. *PLoS ONE* [Internet] 7:e30006. Available from: http://dx.doi.org/10.1371/journal.pone.0030006

Grosjean H. 2009. DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution. CRC Press

Gu X, Fu YX, Li WH. 1995. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Mol. Biol. Evol.* 12:546–557.

Hamey JJ, Wienert B, Quinlan KGR, Wilkins MR. 2017. METTL21B Is a Novel Human Lysine Methyltransferase of Translation Elongation Factor 1A: Discovery by CRISPR/Cas9 Knockout. *Mol. Cell. Proteomics* 16:2229–2242.

He C. 2010. Grand challenge commentary: RNA epigenetics? *Nat. Chem. Biol.* 6:863–865.

Helser TL, Davies JE, Dahlberg JE. 1972. Mechanism of kasugamycin resistance in Escherichia coli. *Nat. New Biol.* 235:6–9.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35:518–522.

Holliday R. 2006. Epigenetics: a historical overview. *Epigenetics* 1:76–80.

Housen I, Demonté D, Lafontaine D, Vandenhaute J. 1997. Cloning and characterization of the KlDIM1 gene from Kluyveromyces lactis encoding the m2(6)A dimethylase of the 18S rRNA. *Yeast* 13:777–781.

Ignatova VV, Jansen PWTC, Baltissen MP, Vermeulen M, Schneider R. 2019. The interactome of a family of potential methyltransferases in HeLa cells. *Sci. Rep.* 9:6584.

Jakobsson ME, Małecki JM, Halabelian L, Nilges BS, Pinto R, Kudithipudi S, Munk S, Davydova E, Zuhairi FR, Arrowsmith CH, et al. 2018. The dual methyltransferase METTL13 targets N terminus and Lys55 of eEF1A and modulates codon-specific translation rates. *Nat. Commun.* 9:3411.

Kangaspeska S, Stride B, Métivier R, Polycarpou-Schwarz M, Ibberson D, Carmouche RP, Benes V, Gannon F, Reid G. 2008. Transient cyclical methylation of promoter DNA. *Nature* 452:112–115.

Kannouche P, Mauffrey P, Pinon-Lataillade G, Mattei MG, Sarasin A, Daya-Grosjean L, Angulo JF. 2000. Molecular cloning and characterization of the human KIN17 cDNA encoding a component of the UVC response that is conserved among metazoans. *Carcinogenesis* 21:1701–1710.

Katoh K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* [Internet] 30:3059–3066. Available from: http://dx.doi.org/10.1093/nar/gkf436

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30:772–780.

Kernstock S, Davydova E, Jakobsson M, Moen A, Pettersen S, Mælandsmo GM, Egge-Jacobsen W,

Falnes PØ. 2012. Lysine methylation of VCP by a member of a novel human protein methyltransferase family. *Nat. Commun.* 3:1038.

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* 34:1812–1819.

Kweon S-M, Chen Y, Moon E, Kvederaviciutė K, Klimasauskas S, Feldman DE. 2019. An Adversarial DNA N6-Methyladenine-Sensor Network Preserves Polycomb Silencing. *Molecular Cell* [Internet] 74:1138–1147.e6. Available from: http://dx.doi.org/10.1016/j.molcel.2019.03.018

Lafontaine D, Delcour J, Glasser AL, Desgrès J, Vandenhaute J. 1994. The DIM1 gene responsible for the conserved m6(2)Am6(2)A dimethylation in the 3'-terminal loop of 18 S rRNA is essential in yeast. *J. Mol. Biol.* 241:492–497.

Lartillot N, Philippe H. 2008. Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363:1463–1472.

Leismann J, Spagnuolo M, Pradhan M, Wacheul L, Vu MA, Musheev M, Mier P, Andrade-Navarro MA, Graille M, Niehrs C, et al. 2020. The 18S ribosomal RNA m A methyltransferase Mettl5 is required for normal walking behavior in Drosophila. *EMBO Rep.* 21:e49443.

Lentini JM, Alsaif HS, Faqeih E, Alkuraya FS, Fu D. 2020. DALRD3 encodes a protein mutated in epileptic encephalopathy that targets arginine tRNAs for 3-methylcytosine modification. *Nat. Commun.* 11:2510.

Leschziner GD, Coffey AJ, Andrew T, Gregorio SP, Dias-Neto E, Calafato M, Bentley DR, Kinton L, Sander JW, Johnson MR. 2011. Q8IYL2 is a candidate gene for the familial epilepsy syndrome of Partial Epilepsy with Pericentral Spikes (PEPS). *Epilepsy Res.* 96:109–115.

Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25:1307–1320.

Lewis CJT, Pan T, Kalsotra A. 2017. RNA modifications and structures cooperate to guide RNA–protein interactions. *Nature Reviews Molecular Cell Biology* [Internet] 18:202–210. Available from: http://dx.doi.org/10.1038/nrm.2016.163

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.

Liu J, Yue Y, Han D, Wang X, Fu Y, Zhang L, Jia G, Yu M, Lu Z, Deng X, et al. 2014. A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* 10:93–95.

Liu S, Hausmann S, Carlson SM, Fuentes ME, Francis JW, Pillai R, Lofgren SM, Hulea L, Tandoc K, Lu J, et al. 2019. METTL13 Methylation of eEF1A Increases Translational Output to Promote Tumorigenesis. *Cell* 176:491–504.e21.

Li Z, Gonzalez PA, Sasvari Z, Kinzy TG, Nagy PD. 2014. Methylation of translation elongation factor 1A by the METTL10-like See1 methyltransferase facilitates tombusvirus replication in yeast and plants. *Virology* 448:43–54.

Losos JB. 2011. Convergence, adaptation, and constraint. *Evolution* 65:1827–1840.

879 Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH,
880 Song JS, et al. 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res*.
881 48:D265–D268.

882 Lyko F. 2018. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nature*
883 *Reviews Genetics* [Internet] 19:81–92. Available from: http://dx.doi.org/10.1038/nrg.2017.80

884 Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, Januszewski W,
885 Kalinowski S, Dunin-Horkawicz S, Rother KM, et al. 2013. MODOMICS: a database of RNA
886 modification pathways--2013 update. *Nucleic Acids Res*. 41:D262–D267.

887 Malecki J, Aileni VK, Ho AYY, Schwarz J, Moen A, Sørensen V, Nilges BS, Jakobsson ME, Leidel SA,
888 Falnes PØ. 2017. The novel lysine specific methyltransferase METTL21B affects mRNA translation
889 through inducible and dynamic methylation of Lys-165 in human eukaryotic elongation factor 1
890 alpha (eEF1A). *Nucleic Acids Res*. 45:4370–4389.

891 Małecki J, Ho AYY, Moen A, Dahl H-A, Falnes PØ. 2015. Human METTL20 Is a Mitochondrial Lysine
892 Methyltransferase That Targets the β Subunit of Electron Transfer Flavoprotein (ETFβ) and
893 Modulates Its Activity. *Journal of Biological Chemistry* [Internet] 290:423–434. Available from:
894 http://dx.doi.org/10.1074/jbc.m114.614115

895 Małecki J, Jakobsson ME, Ho AYY, Moen A, Rustan AC, Falnes PØ. 2017. Uncovering human
896 METTL12 as a mitochondrial methyltransferase that modulates citrate synthase activity through
897 metabolite-sensitive lysine methylation. *Journal of Biological Chemistry* [Internet] 292:17950–
898 17962. Available from: http://dx.doi.org/10.1074/jbc.m117.808451

899 Marchler-Bauer A, Bryant SH. 2004. CD-Search: protein domain annotations on the fly. *Nucleic Acids*
900 *Res*. 32:W327–W331.

901 Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY,
902 Geer RC, Gonzales NR, et al. 2011. CDD: a Conserved Domain Database for the functional
903 annotation of proteins. *Nucleic Acids Res*. 39:D225–D229.

904 Martin JL, McMillan FM. 2002. SAM (dependent) I AM: the S-adenosylmethionine-dependent
905 methyltransferase fold. *Curr. Opin. Struct. Biol*. 12:783–793.

906 Masson C, Menaa F, Pinon-Lataillade G, Frobert Y, Chevillard S, Radicella JP, Sarasin A, Angulo JF.
907 2003. Global genome repair is required to activate KIN17, a UVC-responsive gene involved in DNA
908 replication. *Proc. Natl. Acad. Sci. U. S. A*. 100:616–621.

909 Mateyak MK, Kinzy TG. 2010. eEF1A: thinking outside the ribosome. *J. Biol. Chem*. 285:21209–21213.

910 Meyer H, Bug M, Bremer S. 2012. Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin
911 system. *Nat. Cell Biol*. 14:117–123.

912 Meyer KD, Jaffrey SR. 2017. Rethinking m6A Readers, Writers, and Erasers. *Annual Review of Cell and*
913 *Developmental Biology* [Internet] 33:319–342. Available from: http://dx.doi.org/10.1146/annurev-
914 cellbio-100616-060758

915 Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. 2012. Detecting
916 individual sites subject to episodic diversifying selection. *PLoS Genet*. 8:e1002764.

Nance DJ, Satterwhite ER, Bhaskar B, Misra S, Carraway KR, Mansfield KD. 2020. Characterization of METTL16 as a cytoplasmic RNA binding protein. *PLoS One* 15:e0227647.

Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3:418–426.

Nei M, Hughes AL. 1992. Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. In: Tsuji K, Aizawa M, Sasazuki T, editors. 11th Histocompatibility workshop and conference. Oxford University Press. Available from: https://ci.nii.ac.jp/naid/10015427478/

Nei M, Kumar S. 2000. Molecular Evolution and Phylogenetics. Oxford University Press

Nei M, Rogozin IB, Piontkivska H. 2000. Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc. Natl. Acad. Sci. U. S. A.* 97:10866–10871.

Nei M, Rooney AP. 2005. Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* 39:121–152.

Ng SS, Yue WW, Oppermann U, Klose RJ. 2009. Dynamic protein methylation in chromatin biology. *Cellular and Molecular Life Sciences* [Internet] 66:407–422. Available from: http://dx.doi.org/10.1007/s00018-008-8303-z

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32:268–274.

O'Farrell HC, Pulicherla N, Desai PM, Rife JP. 2006. Recognition of a complex substrate by the KsgA/Dim1 family of enzymes has been conserved throughout evolution. *RNA* 12:725–733.

Okamoto M, Fujiwara M, Hori M, Okada K, Yazama F, Konishi H, Xiao Y, Qi G, Shimamoto F, Ota T, et al. 2014. tRNA modifying enzymes, NSUN2 and METTL1, determine sensitivity to 5-fluorouracil in HeLa cells. *PLoS Genet.* 10:e1004639.

Paik WK, Kim S. 1971. Protein Methylation. *Science* 174:114–119.

Pandolfini L, Barbieri I, Bannister AJ, Hendrick A, Andrews B, Webster N, Murat P, Mach P, Brandi R, Robson SC, et al. 2019. METTL1 Promotes let-7 MicroRNA Processing via m7G Methylation. *Mol. Cell* 74:1278–1290.e9.

Petkowski JJ, Bonsignore LA, Tooley JG, Wilkey DW, Merchant ML, Macara IG, Schaner Tooley CE. 2013. NRMT2 is an N-terminal monomethylase that primes for its homologue NRMT1. *Biochem. J* 456:453–462.

Petkowski JJ, Schaner Tooley CE, Anderson LC, Shumilin IA, Balsbaugh JL, Shabanowitz J, Hunt DF, Minor W, Macara IG. 2012. Substrate specificity of mammalian N-terminal α-amino methyltransferase NRMT. *Biochemistry* 51:5942–5950.

Petrossian T, Clarke S. 2009. Bioinformatic identification of novel methyltransferases. *Epigenomics* [Internet] 1:163–175. Available from: http://dx.doi.org/10.2217/epi.09.3

Piontkivska H, Rooney AP, Nei M. 2002. Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol. Biol. Evol.* 19:689–697.

Pond SLK, Frost SDW. 2005. A genetic algorithm approach to detecting lineage-specific variation in

selection pressure. *Mol. Biol. Evol.* 22:478–485.

Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21:676–679.

Poon AFY, Frost SDW, Kosakovsky Pond SL. 2009. Detecting Signatures of Selection from DNA Sequences Using Datamonkey. *Methods in Molecular Biology* [Internet]:163–183. Available from: http://dx.doi.org/10.1007/978-1-59745-251-9_8

Reiff RE, Ali BR, Baron B, Yu TW, Ben-Salem S, Coulter ME, Schubert CR, Hill RS, Akawi NA, Al-Younes B, et al. 2014. METTL23, a transcriptional partner of GABPA, is essential for human cognition. *Hum. Mol. Genet.* 23:3456–3466.

Rhein VF, Carroll J, Ding S, Fearnley IM, Walker JE. 2017. Human METTL12 is a mitochondrial methyltransferase that modifies citrate synthase. *FEBS Lett.* 591:1641–1652.

Rhein VF, Carroll J, He J, Ding S, Fearnley IM, Walker JE. 2014. Human METTL20 methylates lysine residues adjacent to the recognition loop of the electron transfer flavoprotein in mitochondria. *J. Biol. Chem.* 289:24640–24651.

Rooney AP, Piontkivska H, Nei M. 2002. Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. *Mol. Biol. Evol.* 19:68–75.

Ruszkowska A, Ruszkowski M, Dauter Z, Brown JA. 2018. Structural insights into the RNA methyltransferase domain of METTL16. *Sci. Rep.* 8:5311.

Saletore Y, Meyer K, Korlach J, Vilfan ID, Jaffrey S, Mason CE. 2012. The birth of the Epitranscriptome: deciphering the function of RNA modifications. *Genome Biology* [Internet] 13:175. Available from: http://dx.doi.org/10.1186/gb-2012-13-10-175

Schubert HL, Blumenthal RM, Cheng X. 2003. Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci.* 28:329–335.

Seidel-Rogol BL, McCulloch V, Shadel GS. 2003. Human mitochondrial transcription factor B1 methylates ribosomal RNA at a conserved stem-loop. *Nat. Genet.* 33:23–24.

Shimazu T, Barjau J, Sohtome Y, Sodeoka M, Shinkai Y. 2014. Selenium-based S-adenosylmethionine analog reveals the mammalian seven-beta-strand methyltransferase METTL10 to be an EF1A1 lysine methyltransferase. *PLoS One* 9:e105394.

Shi Z, Xu S, Xing S, Yao K, Zhang L, Xue L, Zhou P, Wang M, Yan G, Yang P, et al. 2019. Mettl17, a regulator of mitochondrial ribosomal RNA modifications, is required for the translation of mitochondrial coding genes. *FASEB J.* 33:13040–13050.

Sledz P, Jinek M. 2016. Crystal structure of the human METTL3-METTL14 complex bound to SAM. Available from: http://dx.doi.org/10.2210/pdb5l6e/pdb

Stayton CT. 2015. What does convergent evolution mean? The interpretation of convergence and its implications in the search for limits to evolution. *Interface Focus* 5:20150039.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30:2725–2729.

991 Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures*
992 *on mathematics in the life sciences* 17:57–86.

993 Thiele W, Novac N, Mink S, Schreiber C, Plaumann D, Fritzmann J, Cremers N, Rothley M, Schwager
994 C, Regiert T, et al. 2011. Discovery of a novel tumour metastasis-promoting gene, NVM-1. *J.*
995 *Pathol*. 225:96–105.

996 Tokuhisa JG, Vijayan P, Feldmann KA, Browse JA. 1998. Chloroplast development at low temperatures
997 requires a homolog of DIM1, a yeast gene encoding the 18S rRNA dimethylase. *Plant Cell* 10:699–
998 711.

999 Tooley CES, Petkowski JJ, Muratore-Schroeder TL, Balsbaugh JL, Shabanowitz J, Sabat M, Minor W,
1000 Hunt DF, Macara IG. 2010. NRMT is an alpha-N-methyltransferase that methylates RCC1 and
1001 retinoblastoma protein. *Nature* 466:1125–1128.

1002 van Tran N, Ernst FGM, Hawley BR, Zorbas C, Ulryck N, Hackert P, Bohnsack KE, Bohnsack MT,
1003 Jaffrey SR, Graille M, et al. 2019. The human 18S rRNA m6A methyltransferase METTL5 is
1004 stabilized by TRMT112. *Nucleic Acids Res*. 47:7719–7733.

1005 Tu C, Tropea JE, Austin BP, Court DL, Waugh DS, Ji X. 2009. Structural basis for binding of RNA and
1006 cofactor by a KsgA methyltransferase. *Structure* 17:374–385.

1007 Turró S, Ingelmo-Torres M, Estanyol JM, Tebar F, Fernández MA, Albor CV, Gaus K, Grewal T, Enrich
1008 C, Pol A. 2006. Identification and characterization of associated with lipid droplet protein 1: A novel
1009 membrane-associated protein that resides on hepatic lipid droplets. *Traffic* 7:1254–1269.

1010 Van Buul CP, Damm JB, Van Knippenberg PH. 1983. Kasugamycin resistant mutants of Bacillus
1011 stearothermophilus lacking the enzyme for the methylation of two adjacent adenosines in 16S
1012 ribosomal RNA. *Mol. Gen. Genet*. 189:475–478.

1013 Van Haute L, Hendrick AG, D'Souza AR, Powell CA, Rebelo-Guiomar P, Harbour ME, Ding S,
1014 Fearnley IM, Andrews B, Minczuk M. 2019. METTL15 introduces N4-methylcytidine into human
1015 mitochondrial 12S rRNA and is required for mitoribosome biogenesis. *Nucleic Acids Res*. 47:10267–
1016 10281.

1017 Van Knippenberg PH, Van Kimmenade JM, Heus HA. 1984. Phylogeny of the conserved 3' terminal
1018 structure of the RNA of small ribosomal subunits. *Nucleic Acids Res*. 12:2595–2604.

1019 Wang C, Zhang B, Ratliff AC, Arrington J, Chen J, Xiong Y, Yue F, Nie Y, Hu K, Jin W, et al. 2019.
1020 Methyltransferase-like 21e inhibits 26S proteasome activity to facilitate hypertrophy of type IIb
1021 myofibers. *FASEB J*. 33:9672–9684.

1022 Wang P, Doxtader KA, Nam Y. 2016. Structural Basis for Cooperative Function of Mettl3 and Mettl14
1023 Methyltransferases. *Mol. Cell* 63:306–317.

1024 Wang X, Feng J, Xue Y, Guan Z, Zhang D, Liu Z, Gong Z, Wang Q, Huang J, Tang C, et al. 2016.
1025 Structural basis of N(6)-adenosine methylation by the METTL3-METTL14 complex. *Nature*
1026 534:575–578.

1027 Warda AS, Kretschmer J, Hackert P, Lenz C, Urlaub H, Höbartner C, Sloan KE, Bohnsack MT. 2017.
1028 Human METTL16 is a -methyladenosine (mA) methyltransferase that targets pre-mRNAs and
1029 various non-coding RNAs. *EMBO Rep*. 18:2004–2014.

1030 Webb KJ, Lipson RS, Al-Hadid Q, Whitelegge JP, Clarke SG. 2010. Identification of Protein N-Terminal
1031 Methyltransferases in Yeast and Humans. *Biochemistry* [Internet] 49:5225–5235. Available from:
1032 http://dx.doi.org/10.1021/bi100428x

1033 Webb KJ, Zurita-Lopez CI, Al-Hadid Q, Laganowsky A, Young BD, Lipson RS, Souda P, Faull KF,
1034 Whitelegge JP, Clarke SG. 2010. A novel 3-methylhistidine modification of yeast ribosomal protein
1035 Rpl3 is dependent upon the YIL110W methyltransferase. *J. Biol. Chem.* 285:37598–37606.

1036 Wiederstein JL, Nolte H, Günther S, Piller T, Baraldo M, Kostin S, Bloch W, Schindler N, Sandri M,
1037 Blaauw B, et al. 2018. Skeletal Muscle-Specific Methyltransferase METTL21C Trimethylates p97
1038 and Regulates Autophagy-Associated Protein Breakdown. *Cell Rep.* 23:1342–1356.

1039 Wiegand G, Remington SJ. 1986. Citrate synthase: structure, control, and mechanism. *Annu. Rev.*
1040 *Biophys. Biophys. Chem.* 15:97–117.

1041 Woodcock CB, Yu D, Hajian T, Li J, Huang Y, Dai N, Corrêa IR, Wu T, Vedadi M, Zhang X, et al. 2019.
1042 Human MettL3–MettL14 complex is a sequence-specific DNA adenine methyltransferase active on
1043 single-strand and unpaired DNA in vitro. *Cell Discovery* [Internet] 5. Available from:
1044 http://dx.doi.org/10.1038/s41421-019-0136-4

1045 Xu L, Liu X, Sheng N, Oo KS, Liang J, Chionh YH, Xu J, Ye F, Gao Y-G, Dedon PC, et al. 2017. Three
1046 distinct 3-methylcytidine (mC) methyltransferases modify tRNA and mRNA in mice and humans. *J.*
1047 *Biol. Chem.* 292:14695–14703.

1048 Xu Z, O'Farrell HC, Rife JP, Culver GM. 2008. A conserved rRNA methyltransferase regulates ribosome
1049 biogenesis. *Nat. Struct. Mol. Biol.* 15:534–536.

1050 Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates
1051 over sites: approximate methods. *J. Mol. Evol.* 39:306–314.

1052 Zehmer JK, Bartz R, Bisel B, Liu P, Seemann J, Anderson RGW. 2009. Targeting sequences of UBXD8
1053 and AAM-B reveal that the ER has a direct role in the emergence and regression of lipid droplets.
1054 *Journal of Cell Science* [Internet] 122:3694–3702. Available from:
1055 http://dx.doi.org/10.1242/jcs.054700

1056 Zhang L-S, Liu C, Ma H, Dai Q, Sun H-L, Luo G, Zhang Z, Zhang L, Hu L, Dong X, et al. 2019.
1057 Transcriptome-wide Mapping of Internal N7-Methylguanosine Methylome in Mammalian mRNA.
1058 *Molecular Cell* [Internet] 74:1304–1316.e8. Available from:
1059 http://dx.doi.org/10.1016/j.molcel.2019.03.036

1060 Zhang Z, Hou Y, Wang Y, Gao T, Ma Z, Yang Y, Zhang P, Yi F, Zhan J, Zhang H, et al. 2020.
1061 Regulation of Adipocyte Differentiation by METTL4, a 6 mA Methylase. *Sci. Rep.* 10:8285.

1062 Zhou J, Wan J, Gao X, Zhang X, Jaffrey SR, Qian S-B. 2015. Dynamic m6A mRNA methylation directs
1063 translational control of heat shock response. *Nature* [Internet] 526:591–594. Available from:
1064 http://dx.doi.org/10.1038/nature15377

1065 **Figure Legends**

1066 **Fig. 1.** METTLs in major metazoan phyla according to NCBI search results. The metazoan cladogram is
1067 modified from the Tree of Life web project (http://tolweb.org/tree/phylogeny.html). The matrix displays
1068 the detection (gray), probable detection but with uncertain identification (striped gray), or failure of

detection (white) by the search for each METTL gene within each phylum. The number of METTL genes out of 33 total is noted in parenthesis with asterisks indicating phyla in which probable METTL sequences were either located but could not be definitively identified as specific METTL genes, or were located outside of our search in the NCBI GenBank database (e.g., Ensembl Metazoa). METTLs are listed in order of decreasing number of representatives in each phylum (left to right).

Fig. 2. Maximum likelihood tree describing the evolutionary relationships across metazoan METTL proteins. Bootstrap values (1,000 replicates) are displayed at each internal node. Branch labels displayed in blue indicate METTLs that have been shown to target DNA or RNA nucleosides, labels in red are METTLs shown to target protein residues, and labels in purple represent METTLs whose function and targets are unknown. The METTL proteins comprise nine different taxa representing eight different phyla: a sponge (*Amphimedon queenslandica*, Porifera), a stony coral (*Acropora millepora*, Cnidaria), a sea hare (*Aplysia californica*, Mollusca), a priapulid (*Priapulus caudatus*, Priapulida), a shrimp (*Litopenaeus vannamei*, Arthropoda), a sea star (*Acanthaster planci*, Echinodermata), an acorn worm (*Saccoglossus kowalevskii*, Hemichordata), a lancelet (*Branchiostoma belcheri*, Chordata), and human (*Homo sapiens*, Chordata). Subtrees with the same METTL gene from multiple taxa have been collapsed for figure readability.

Fig. 3. Maximum likelihood phylogeny reconstructed based on the conserved domains of each METTL protein across Metazoa. Bootstrap values (1,000 replicates) are displayed at each node. Branch label colors are identical to Fig. 2.

Fig. 4. Estimated rates of evolution of METTL proteins. Rates were only estimated for METTLs in which amino acid differences per site could be calculated by comparing the protein between three or more of the pre-selected metazoan phyla.

Fig. 5. Episodes of diversifying selection shaping the evolution of METTL proteins in vertebrates. For figure readability, "METTL" has been removed from the names of each tip label (e.g., "METTL1" is simply displayed as "1"). (*A*) The strength of selection at significant branches is represented in red ($\omega > 5$), gray ($\omega = 1$), and blue ($\omega = 0$), with the proportion of sites within each class represented by the color width. Thicker branches have been classified as undergoing episodic diversifying selection at corrected *p*-value < 0.001 (thicker branches) and *p*-value < 0.01 (thinner branches). (*B*) The physical positions of adaptive selection episodes involved in the diversification of METTL genes. Numbers of synonymous (blue bars) and non-synonymous (red bars) substitutions at codon positions that are subject to significant

episodes of diversifying selection in vertebrates (p-value < 0.05). (*C*) Phylogenetic location of the mutations involved in such episodes. Branches in red account for higher numbers of nonsynonymous mutations, branches in blue indicate higher numbers of synonymous mutations, and branches in green represent cases with the same numbers of nonsynonymous and synonymous mutations. Light blue, light red, and light purple backgrounds behind each branch indicate METTLs that modify DNA/RNA, proteins, or unknown targets, respectively. Subtrees that include several closely related METTLs (i.e., METTL11A and METTL11B; METTL21A, METTL21B, METTL21C, METTL21D, and METTL21E; METTL25 and METTL25B) were collapsed into a single branch for figure readability.

Fig. 6. Episodes of adaptive selection identified upon separately analyzing vertebrate DNA/RNA-modifying METTLs and vertebrate protein-modifying METTLs. (*A*) The physical positions of adaptive selection episodes involved in the diversification of METTLs that modify DNA/RNA nucleosides (top) and METTLs that modify protein residues (bottom). Numbers of synonymous (blue bars) and non-synonymous (red bars) substitutions at codon positions that are subject to significant episodes of diversifying selection in vertebrates (p-value < 0.05). (*B*) Phylogenetic location of the mutations involved in such episodes for DNA/RNA-modifying METTLs, indicated by a light blue background, and (*C*) protein-modifying METTLs, indicated by a light red background. For figure readability, "METTL" has been removed from the names of each tip label (e.g., "METTL1" is simply displayed as "1"). Branches in red account for higher numbers of nonsynonymous mutations, branches in blue indicate higher numbers of synonymous mutations, and branches in green represent cases with the same numbers of nonsynonymous and synonymous mutations.

Tables

Table 1. Methyltransferase like (METTL) genes separated by type (i.e., DNA/RNA-modifying, protein-modifying, and unknown). The gene names, NCBI IDs, aliases, and preferred names are vertebrate METTL genes.

Gene Name	NCBI ID	Aliases	Preferred Name	Target
DNA/RNA-modifying METTLs				
METTL1	4234	C12orf1, TRM8, TRMT8, YDL201w	tRNA (guanine-N(7)-)-methyltransferase	m7G (tRNA, mRNA, miRNA)
METTL2 (METTL2A/B)	339175/ 55798	METL, PSENIP1	tRNA N(3)-methylcytidine methyltransferase METTL2; methyltransferase-like protein 2	m3C (tRNA)
METTL3	56339	IME4, M6A, MT-A70, Spo8, hMETTL3	N6-adenosine-methyltransferase catalytic subunit	m6A (mRNA)
METTL4	64863	HsT661	N(6)-adenine-specific methyltransferase METTL4	m6A (DNA) / m6Am (snRNA)
METTL5	29081	HSPC133, MRT72	rRNA N6-adenosine-methyltransferase METTL5	m6A (rRNA)
METTL6	131965	none	tRNA N(3)-methylcytidine methyltransferase METTL6	m3C (tRNA)
METTL8	79828	TIP	mRNA N(3)-methylcytidine methyltransferase METTL8	m3C (mRNA)
METTL14	57721	none	N6-adenosine-methyltransferase non-catalytic subunit	m6A (mRNA)
METTL15	196074	METT5D1	12S rRNA N4-methylcytidine (m4C) methyltransferase	m4C (rRNA)
METTL16	79066	METT10D	RNA N6-adenosine-methyltransferase METTL16	m6A (ncRNA, pre-mRNA, mRNA)
METTL19	152992	TRMT44, C4orf23, TRM44	probable tRNA (uracil-O(2)-)-methyltransferase	2′-O-methyluridine (tRNA)
METTL25B	51093	RRNAD1, C1orf66, CGI-41	ribosomal RNA adenine dimethylase domain containing 1	m6A (rRNA)
Protein-modifying METTLs				
METTL9	51108	CGI-81, DREV, DREV1, PAP1	methyltransferase-like protein 9	unknown
METTL10	399818	EEF1AKMT2, C10orf138, Efm4	EEF1A lysine methyltransferase 2	lysine (EEF1A)
METTL11A	28989	NTMT1, AD-003, C9orf32, HOMT1A, NRMT, NRMT1, NTM1A	N-terminal Xaa-Pro-Lys N-methyltransferase 1	N-terminal
METTL11B	149281	C1orf184, HOMT1B, NTM1B	alpha N-terminal protein methyltransferase 1B	N-terminal
METTL12	751071	CSKMT, CS-KMT, U99HG	citrate synthase-lysine N-methyltransferase CSKMT, mitochondrial	lysine (citrate synthase)
METTL13	51603	EEF1AKNMT, 5630401D24Rik, CGI-01, DFNB26, DFNB26M, DFNM1, KIAA0859, feat	eEF1A lysine and N-terminal methyltransferase	lysine (EEF1A)
METTL18	92342	AsTP2, C1orf156, HPM1	histidine protein methyltransferase 1 homolog	histidine
METTL20	254013	ETFBKMT, C12orf72, ETFB-KMT	electron transfer flavoprotein beta subunit lysine methyltransferase	lysine (ETFβ)
METTL21A	151194	FAM119A, HCA557b, HSPA-KMT	protein N-lysine methyltransferase METTL21A	lysine (HSPA)
METTL21B	25895	EEF1AKMT3, FAM119B	EEF1A lysine methyltransferase 3	lysine (EEF1A)
METTL21C	196541	C13orf39	protein-lysine methyltransferase METTL21C	lysine (VCP)
METTL21D	79609	VCPKMT, C14orf138, VCP-KMT	protein-lysine methyltransferase METTL21D	lysine (VCP)
METTL21E	403183	4832428D23Rik, Gm991	protein-lysine methyltransferase METTL21E	lysine (VCP)
METTL22	79091	C16orf68	methyltransferase-like protein 22	lysine (Kin17)
METTL23	124512	C17orf95, MRT44	methyltransferase-like protein 23	unknown
METTLs of unknown function				
METTL7A	25840	AAM-B, AAMB	methyltransferase-like protein 7A	unknown
METTL7B	196410	ALDI	methyltransferase-like protein 7B	unknown
METTL17	64745	METT11D1	methyltransferase-like protein 17, mitochondrial	unknown
METTL25	84190	C12orf26	methyltransferase-like protein 25	unknown
METTL26	84326	C16orf13, JFP2	methyltransferase-like 26	unknown
METTL27	155368	WBSCR27	methyltransferase-like protein 27	unknown

Table 2. Conserved domains within each METTL gene based on results from CD-Search and Batch CD-Search (NCBI).

Superfamily	CD	Gene
AdoMet MTases superfamily	COG2263	METTL5
	DREV	METTL9
	DUF938	METTL26
	Methyltransf_11	METTL7A, METTL7B
	Methyltransf_12	METTL2, METTL6
	Methyltransf_16	METTL21A, METTL21D
	Methyltransf_25	METTL2, METTL6, METTL7A, METTL8, METTL10, METTL12, METTL13, METTL27
	Methyltransf_31	METTL10
	Methyltransf_32	METTL25B
	Methyltransf_4	METTL1
	Methyltransf_PK	METTL11A
	Nnt1	METTL20
	PRK00050	METTL15
	No specific CD (superfamily only)	METTL11B, METTL16, METTL17, METTL18, METTL19, METTL21B, METTL21C, METTL21E, METTL22, METTL23, METTL25
MT-A70 superfamily	MT-A70	METTL3, METTL14
	No specific CD (superfamily only)	METTL4

Table 3. Selection results, variance estimations[a], and best-fit ML models for all METTLs and each individual METTL gene.

Gene	R	Z-test	d_{NT} ± SE	d_{AA} ± SE	d_S ± SE	d_N ± SE	ENC	$model_{NT}$	$model_{AA}$
All	0.5	5.8***	0.65 ± 0.01	0.82 ± 0.02	0.74 ± 0.005	0.62 ± 0.02	52.45	GTR+G+I	LG+G
METTL1	0.8	23.38***	0.36 ± 0.01	0.32 ± 0.02	0.83 ± 0.02	0.23 ± 0.02	53.60	GTR+G+I	WAG+G
METTL2	0.7	22.02***	0.4 ± 0.01	0.39 ± 0.02	0.79 ± 0.01	0.28 ± 0.02	51.09	GTR+G+I	LG+G
METTL3	0.7	27.28***	0.4 ± 0.01	0.4 ± 0.02	0.82 ± 0.01	0.28 ± 0.01	53.89	GTR+G	LG+G
METTL4	0.6	12.93***	0.49 ± 0.01	0.55 ± 0.02	0.78 ± 0.02	0.4 ± 0.02	50.78	T92+G+I	LG+G+I
METTL5	0.8	18.17***	0.4 ± 0.01	0.42 ± 0.02	0.81 ± 0.02	0.28 ± 0.02	53.16	T92+G	LG+G
METTL6	0.8	20.0***	0.39 ± 0.01	0.4 ± 0.02	0.81 ± 0.02	0.27 ± 0.02	51.25	GTR+G	LG+G
METTL7A	0.6	12.02***	0.52 ± 0.01	0.61 ± 0.02	0.78 ± 0.02	0.44 ± 0.02	54.10	K2+G+I	LG+G+I
METTL7B	0.6	10***	0.53 ± 0.01	0.61 ± 0.03	0.8 ± 0.02	0.45 ± 0.02	54.92	JC+I	LG+I
METTL8	vertebrates only						50.53		
METTL9	0.7	14.15***	0.49 ± 0.01	0.54 ± 0.02	0.81 ± 0.01	0.39 ± 0.02	52.31	K2+G+I	LG+G+I
METTL10	0.7	12.85***	0.47 ± 0.01	0.55 ± 0.02	0.75 ± 0.02	0.38 ± 0.02	54.78	T92+G+I	LG+G+I
METTL11A	0.7	18.84***	0.45 ± 0.01	0.49 ± 0.02	0.8 ± 0.01	0.35 ± 0.02	51.79	K2+G+I	LG+G+I
METTL11B	vertebrates only						49.58		
METTL12	0.6	9.02***	0.54 ± 0.01	0.63 ± 0.03	0.76 ± 0.02	0.46 ± 0.02	53.23	K2+I	LG+G+I
METTL13	0.7	24.82***	0.48 ± 0.01	0.56 ± 0.01	0.78 ± 0.01	0.39 ± 0.01	50.84	GTR+G+I	LG+G
METTL14	0.7	29.64***	0.37 ± 0.01	0.32 ± 0.02	0.83 ± 0.01	0.23 ± 0.01	53.61	GTR+G	LG+G
METTL15	0.7	18.11***	0.45 ± 0.01	0.48 ± 0.02	0.79 ± 0.01	0.34 ± 0.02	51.84	HKY+G+I	LG+G+I
METTL16	0.6	18.45***	0.49 ± 0.01	0.57 ± 0.02	0.79 ± 0.01	0.41 ± 0.01	53.99	HKY+G+I	LG+G+I
METTL17	0.7	17.89***	0.53 ± 0.01	0.61 ± 0.02	0.79 ± 0.01	0.44 ± 0.01	54.27	GTR+G	LG+G+I
METTL18	0.7	13.1***	0.5 ± 0.01	0.57 ± 0.02	0.78 ± 0.02	0.41 ± 0.02	52.54	T92+G+I	LG+G+I
METTL19	0.7	13.24***	0.49 ± 0.01	0.56 ± 0.02	0.77 ± 0.01	0.41 ± 0.02	51.68	GTR+G+I	LG+G+I
METTL20	0.7	14.86***	0.45 ± 0.01	0.48 ± 0.03	0.77 ± 0.01	0.35 ± 0.02	52.27	K2+G+I	LG+G+I
METTL21A	0.6	17.06***	0.48 ± 0.01	0.54 ± 0.02	0.78 ± 0.01	0.39 ± 0.02	54.38	K2+G+I	LG+G
METTL21B	vertebrates only						45.13		
METTL21C	vertebrates only						54.33		
METTL21D	0.7	13.23***	0.47 ± 0.01	0.54 ± 0.02	0.76 ± 0.02	0.38 ± 0.02	54.89	T92+G+I	LG+G+I
METTL21E	chordates only						53.19		
METTL22	0.6	15.33***	0.5 ± 0.01	0.58 ± 0.02	0.79 ± 0.01	0.42 ± 0.02	53.77	T92+G+I	LG+G+I
METTL23	0.6	18.63***	0.45 ± 0.01	0.5 ± 0.02	0.78 ± 0.01	0.35 ± 0.02	52.42	K2+G+I	LG+G
METTL25	0.6	16.26***	0.5 ± 0.01	0.58 ± 0.02	0.76 ± 0.01	0.42 ± 0.01	50.58	T92+G+I	LG+G+I
METTL25B	0.7	17.41***	0.5 ± 0.01	0.56 ± 0.02	0.79 ± 0.01	0.4 ± 0.01	51.20	GTR+G+I	LG+G+I
METTL26	0.7	13.38***	0.44 ± 0.01	0.49 ± 0.03	0.8 ± 0.02	0.34 ± 0.02	50.61	K2+I	LG+G+I
METTL27	0.5	4.12***	0.55 ± 0.02	0.69 ± 0.03	0.68 ± 0.04	0.5 ± 0.03	48.47	JC	LG

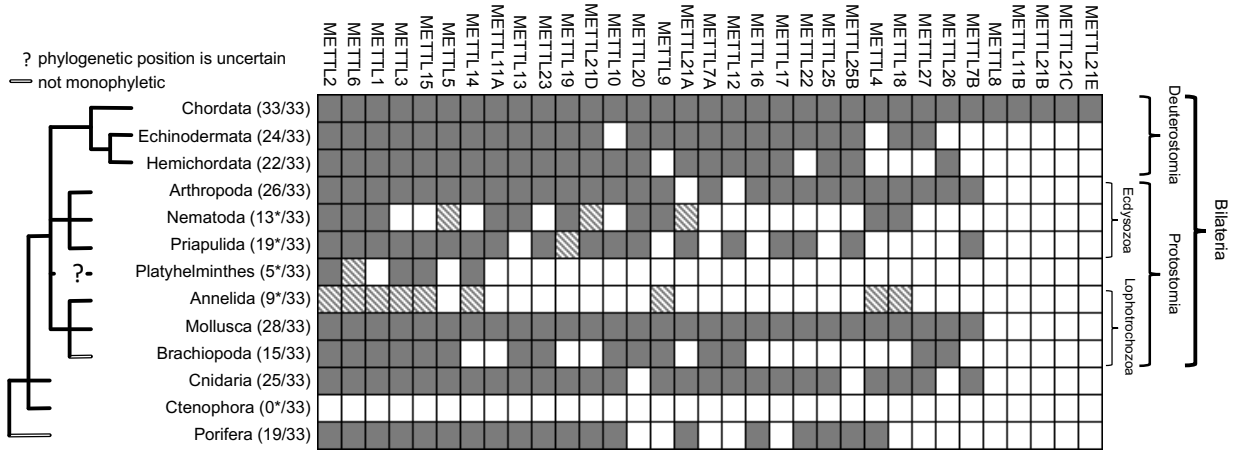
NOTE.—SE, standard error; R, average transition/transversion ratio; d_{NT}, mean nucleotide distance, d_{AA}, mean amino acid distance; d_S, synonymous substitution distance; d_N, non-synonymous substitution distance; ENC, effective number of codons (codon bias) ranging from 61 (i.e., no bias) to 20 (i.e., maximum bias); $model_{NT}$, best-fit maximum likelihood (ML) DNA model; $model_{AA}$, best-fit ML protein model.

[a]All SE values were calculated using the Bootstrap method with 1,000 replications.

***p-value < 0.001 in Z-test of purifying selection (i.e., HA: $d_N < d_S$).

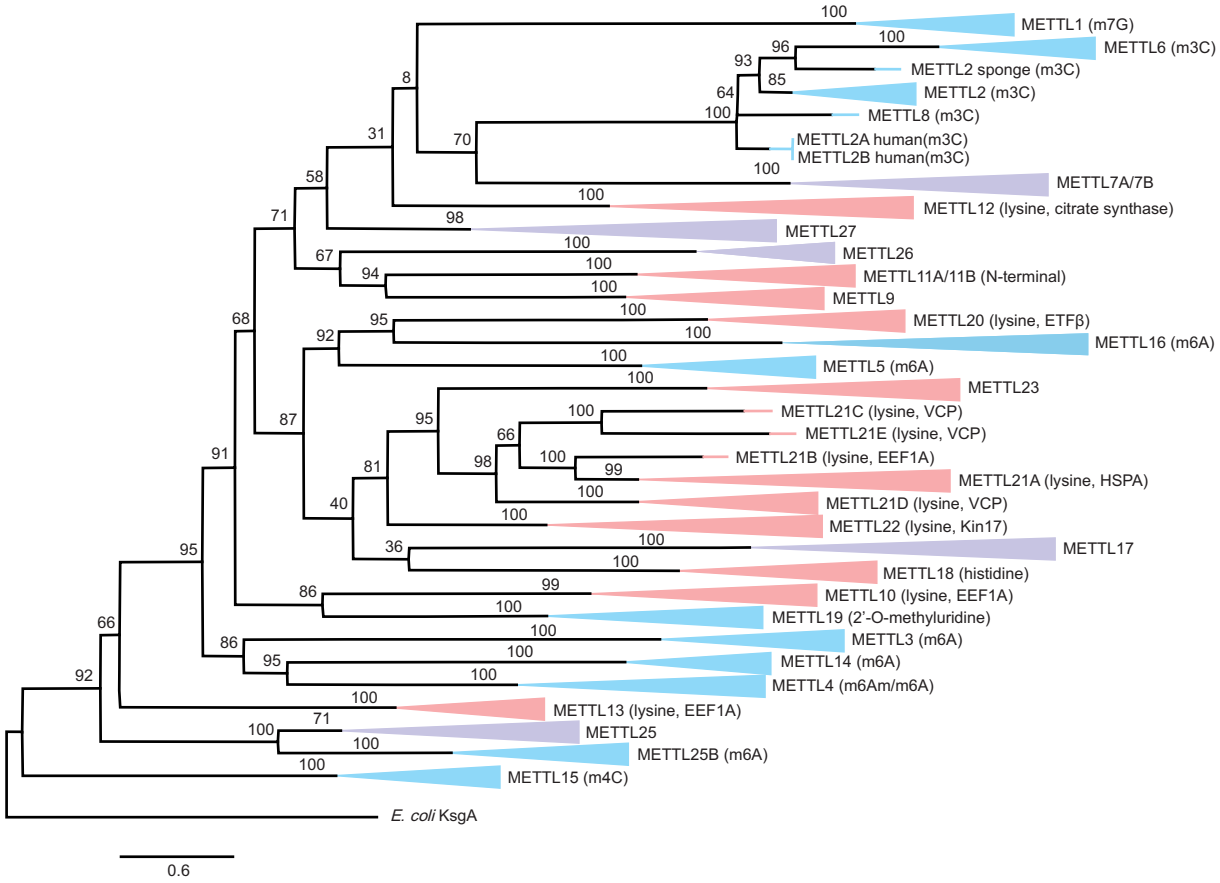
1144 **Figures**

1145 **Fig. 1.**



1146

1147 **Fig. 2.**

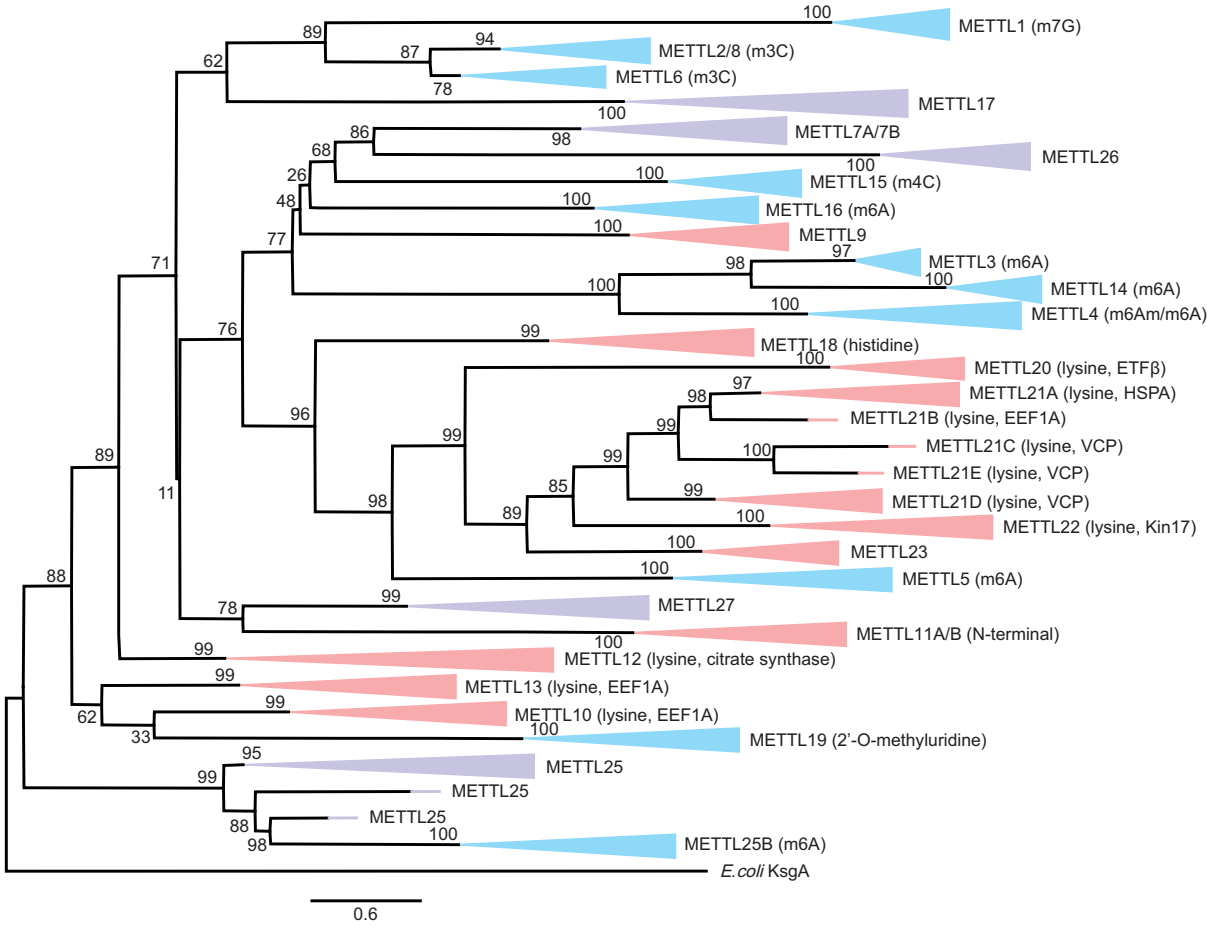


1148

1149

1150

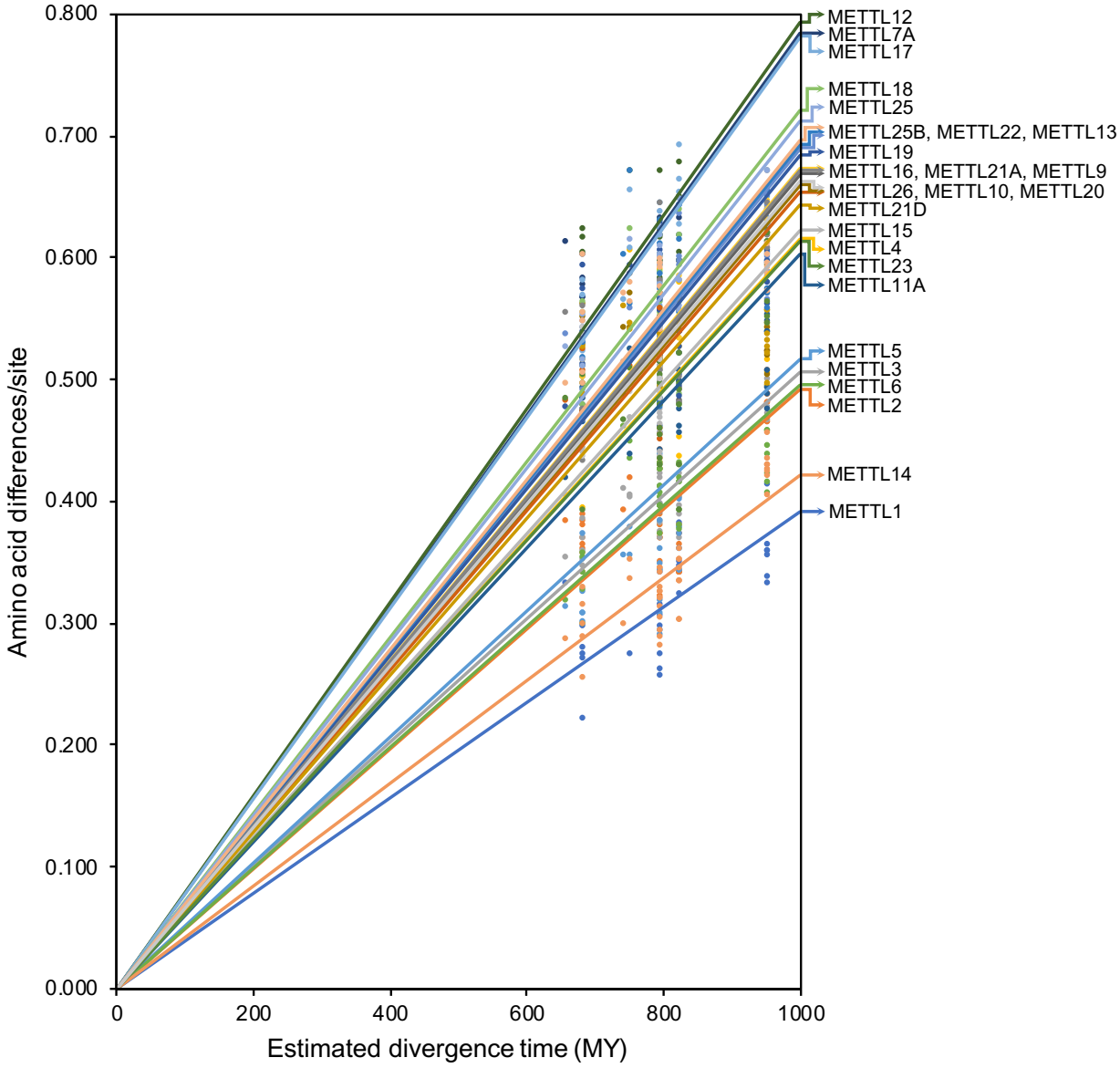
1151 **Fig. 3.**



1152

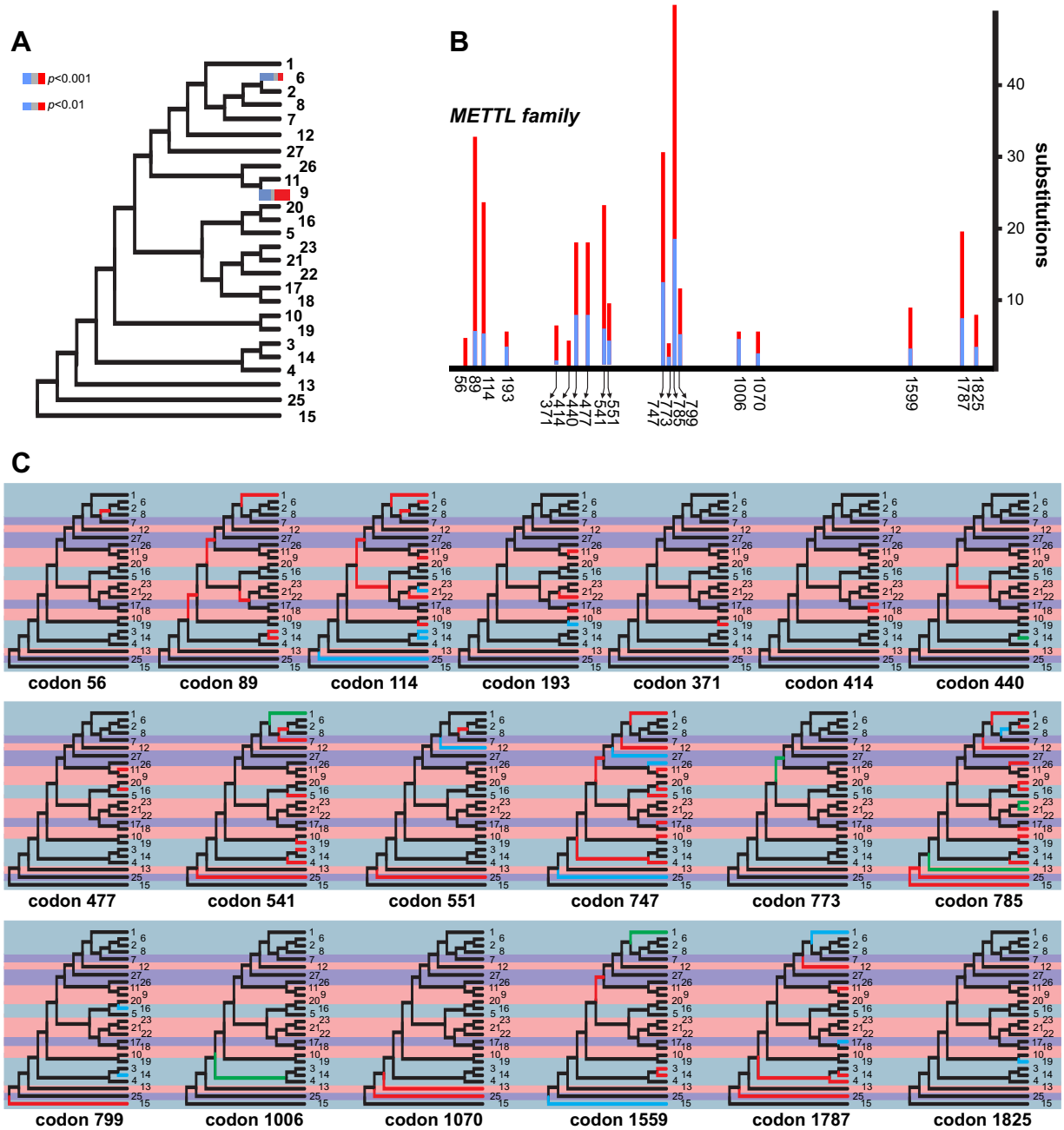
1153

1154 **Fig. 4.**



1155

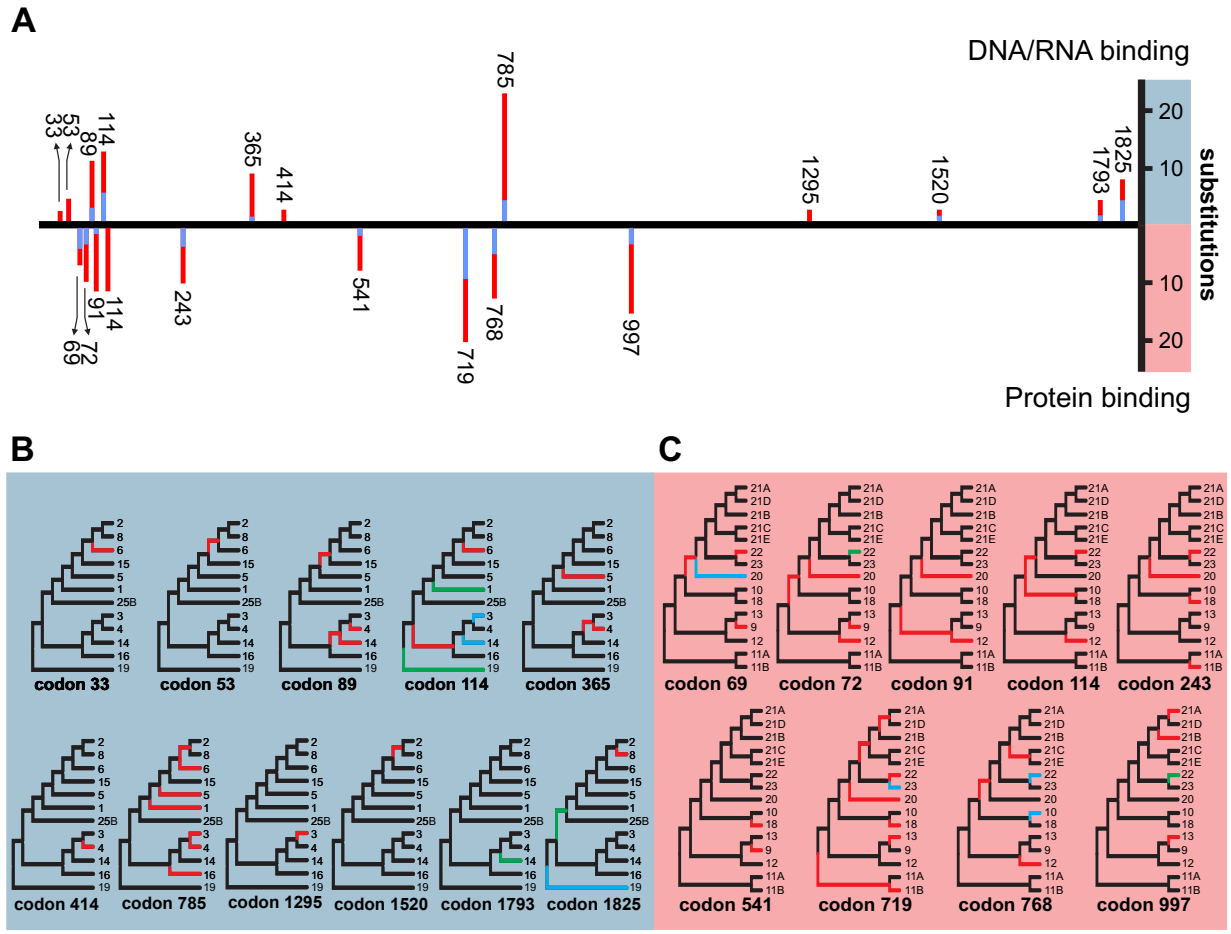
1156 **Fig. 5.**



1157

1158

1159 **Fig. 6.**



1160

1161