Dynamic Evolution of De Novo DNA Methyltransferases in Rodent and Primate Genomes

Antoine Molaro [1,*], Harmit S. Malik [1,2] and Deborah Bourc'his [3]

[1]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA

[2]Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA

[3]Institut Curie, PSL University, Inserm, CNRS, Paris, France

***Corresponding author:** E-mail: amolaro@fredhutch.org.

Associate editor: Yoko Satta

Abstract

Transcriptional silencing of retrotransposons via DNA methylation is paramount for mammalian fertility and reproductive fitness. During germ cell development, most mammalian species utilize the de novo DNA methyltransferases DNMT3A and DNMT3B to establish DNA methylation patterns. However, many rodent species deploy a third enzyme, DNMT3C, to selectively methylate the promoters of young retrotransposon insertions in their germline. The evolutionary forces that shaped DNMT3C's unique function are unknown. Using a phylogenomic approach, we confirm here that *Dnmt3C* arose through a single duplication of *Dnmt3B* that occurred ~60 Ma in the last common ancestor of muroid rodents. Importantly, we reveal that DNMT3C is composed of two independently evolving segments: the latter two-thirds have undergone recurrent gene conversion with *Dnmt3B*, whereas the N-terminus has instead evolved under strong diversifying selection. We hypothesize that positive selection of *Dnmt3C* is the result of an ongoing evolutionary arms race with young retrotransposon lineages in muroid genomes. Interestingly, although primates lack DNMT3C, we find that the N-terminus of DNMT3A has also evolved under diversifying selection. Thus, the N-termini of two independent de novo methylation enzymes have evolved under diversifying selection in rodents and primates. We hypothesize that repression of young retrotransposons might be driving the recurrent innovation of a functional domain in the N-termini on germline DNMT3s in mammals.

Key words: **DNA methylation, retrotransposons, gene conversion, diversifying selection, chromatin modifications.**

Article

Introduction

The deposition of methylation on DNA is a deeply conserved process. In mammals, it is crucial for genome stability, development, genomic imprinting, and chromosome-wide epigenetic silencing such as X-inactivation (Smith and Meissner 2013). Mammalian DNA methyltransferases (DNMTs) are enzymes that catalyze the addition of a methyl group onto cytosines (Lyko 2018). Most mammals encode three catalytically active enzymes (DNMT1, DNMT3A, and DNMT3B) and one nonenzymatic germ cell-specific cofactor (DNMT3L) (Bestor 2000; Lees-Murdock et al. 2004; Ponger and Li 2005; Lyko 2018). Although DNMT1 targets hemimethylated cytosines (maintenance DNA methyltransferase) (Gruenbaum et al. 1982; Bestor et al. 1988; Song et al. 2011), DNMT3A and DNMT3B are classified as de novo methyltransferases that target unmodified sites (Okano et al. 1998, 1999; Jia et al. 2007; Zhang et al. 2018). In mice, constitutive genetic knock-outs (KO) of *Dnmt1*, *Dnmt3A*, or *Dnmt3B* are lethal, whereas *Dnmt3L* mutations lead to sterility (Li et al. 1992; Okano et al. 1999; Bourc'his et al. 2001).

Phylogenetic analyses have suggested that the DNMT enzymes belong to the clade of 5-cytosine methyltransferases, which likely predated the origin of eukaryotes (Ponger and Li 2005; Law and Jacobsen 2010). Although both *Dnmt1* and *Dnmt3A* were present in the common ancestor of all metazoans, *Dnmt3B* is believed to have arisen by a gene duplication event close to the origin of tetrapods (Ponger and Li 2005; Nguyen et al. 2018). Closer phylogenetic analyses in several taxa have revealed mammalian lineage-specific duplications, including the duplication and diversification of several *Dnmt1* paralogs in marsupials (Alvarez-Ponce et al. 2018) and the evolution of *Dnmt3L* from *Dnmt3A* in eutherian mammals (Yokomine et al. 2006). Similarly, a gene duplication of *Dnmt3B* gave rise to *Dnmt3C* in muroid rodents where it has acquired a distinct, non-redundant role in retrotransposon repression during spermatogenesis (Barau et al. 2016; Jain et al. 2017). Thus, a series of ancient and recent gene duplications have led to the current repertoires of mammalian DNMTs.

Retrotransposons are selfish genetic elements that propagate within host genomes at the cost of optimal reproductive fitness. The silencing of retrotransposons by DNA methylation is critical for mammalian germline development (Yoder et al. 1997). This is because germ cell development is particularly vulnerable to retrotransposon activity in mammals, as many chromatin marks that otherwise repress retrotransposons—like DNA methylation—are transiently erased (Reik and Surani 2015). It can be a challenge, however, to silence retrotransposons as they exhibit rapid sequence divergence

and belong to many evolutionarily distinct families (Molaro and Malik 2016). This sequence heterogeneity means that conserved DNA motifs may not systematically mark problematic retrotransposons. To cope with this, mice use two distinct waves of de novo methylation during male fetal germ cell development to silence retrotransposons according to their age (Molaro et al. 2014). During the first wave, evolutionarily old retrotransposons gain methylation together with the rest of the genome. However, evolutionarily young and transcriptionally active retrotransposons are refractory to this wave and require the piRNA pathway—a small RNA-based defense system—to target DNA methylation to their promoters (Aravin et al. 2008; Molaro et al. 2014). Accessible heterochromatin states characterize young retrotransposons prior to piRNA-directed DNA methylation (Yamanaka et al. 2019).

Two recent studies showed that DNMT3C is crucial to the silencing of young retrotransposons (Barau et al. 2016; Jain et al. 2017). *Dnmt3C* KO males are sterile and their germ cell methylation profiles are similar to those of piRNA mutants, with a 1 to 4% drop in genome-wide DNA methylation content that selectively affects the promoters of young copies of LINE and ERVK retrotransposons (Molaro et al. 2014; Manakov et al. 2015; Barau et al. 2016). This contrasts with germ cell-specific *Dnmt3B* KO, which has no impact on male fertility (Kaneda et al. 2004), whereas constitutive *Dnmt3B* KO shows embryonic lethality (Okano et al. 1999). On the other hand, germ cell-specific *Dnmt3*A KO males are infertile but only display mild alteration in the methylation levels of SINE retrotransposons (Kaneda et al. 2004; Kato et al. 2007). This suggests that *Dnmt3A* might act nonredundantly with *Dnmt3C* for methylating the male germ cell genome.

Catalytically active DNMT3s have three well-defined domains. The most C-terminal region encodes the methyltransferase domain (MTase), which includes highly conserved protein motifs that catalyze the addition of methyl groups (Posfai et al. 1989; Timinskas et al. 1995). The central portion encodes two chromatin-reading domains, ADD (ATRX–DNMT3–DNMT3L) and PWWP (Pro–Trp–Trp–Pro), that play important roles in their targeting and regulation (Jeltsch et al. 2018). ADD domain binding to nucleosomes is inhibited by trimethylation of lysine 4 of histone H3 (H3K4me3) (Ooi et al. 2007; Otani et al. 2009; Zhang et al. 2010; Guo et al. 2015), whereas the PWWP domain anchor DNMT3 proteins to methylated H3K36 residues (Qiu et al. 2002; Chen et al. 2004; Ge et al. 2004; Dhayalan et al. 2010; Rondelet et al. 2016). Interestingly, mouse *Dnmt3C* lost the two exons coding for the PWWP domain, making it unique among catalytically active DNMT3s (Barau et al. 2016; Jain et al. 2017). In contrast to the central and C-terminal segments, the N-terminal portion of DNMT3s remains largely uncharacterized.

Based on both its recent origin and its function in silencing young, potentially rapidly adapting retrotransposon families, we speculated that *Dnmt3C* might be participating in an ongoing evolutionary arms race, or genetic conflict, with these genetic parasites (Molaro and Malik 2016). We therefore performed a detailed phylogenetic survey of rodent genomes to investigate *Dnmt3C*'s age and the evolutionary forces that shape its unique function. Extending previous findings, we dated *Dnmt3C*'s evolutionary origin in the common ancestor of muroids ~60 Ma. We provide evidence for a pattern of gene conversion between *Dnmt3B* and *Dnmt3C* paralogs throughout muroid evolution. Gene conversion recurrently homogenizes the latter two-thirds of DNMT3B and DNMT3C but does not extend to their N-terminal domains. Interestingly, we found strong diversifying selection in the N-terminal tail of DNMT3C, but not DNMT3B, consistent with an ongoing genetic conflict. Although *Dnmt3C* is not present outside rodents, we found that the N-terminal tail of *DNMT3A* has similarly evolved under diversifying selection in primates. Thus, two distinct DNMT3 enzymes display hallmarks of ongoing genetic conflicts—potentially with endogenous retrotransposons—in two separate mammalian lineages.

Results

Evolutionary Origins and Dynamics of *Dnmt3C* in Rodents

To investigate the evolutionary age and dynamics of *Dnmt3C*, we retrieved and annotated DNMT3 sequences in partially or fully assembled genomes of 19 species of Glires—which include rodents and lagomorphs (fig. 1A, see Materials and Methods). Like other mammals, most species of Glires encode unique *Dnmt3A*, *Dnmt3L*, and *Dnmt3B* genes within syntenic loci present in all placental mammals (fig. 1A). However, in the subgroup of muroid species, the syntenic locus containing *Dnmt3B* also encodes *Dnmt3C* (fig. 1A) (Barau et al. 2016).

We investigated genomes from 11 muroid and 8 "outgroup" species and used available transcriptome or de novo gene assemblies to annotate coding sequences (CDS, see Materials and Methods). In some cases, genome assemblies allowed us to tentatively assign gene orthology using shared synteny. However, in most cases, genome assemblies were too fragmented to reconstruct genomic contexts. Instead, we focused on retrieving partial or full-length sequences of putative *Dnmt3* genes. We then constructed a multiple alignment and used maximum likelihood methods to build a gene phylogeny (see Materials and Methods for details). Using this approach, we were able to resolve all retrieved sequences into distinct clades of DNMT3s (fig. 1B).

If *Dnmt3C* arose from *Dnmt3B* in the last common ancestor of all muroids, we would expect 1) *Dnmt3C* sequences to branch inside the *Dnmt3B* clade and 2) form two independent lineages following the split of muroids from other rodents and lagomorphs. Our first expectation was met; all putative *Dnmt3C* sequences branched within the *Dnmt3B* clade, supporting the close relatedness of these two genes relative to other *Dnmt3s* (fig. 1B). Moreover, a detailed phylogeny including all *Dnmt3B* and *Dnmt3C* orthologs was consistent with a single duplication event (fig. 1C). Based on the presence of *Dnmt3C* in mountain blind mole rats (*Nannospalax galili*), but not beavers or guinea pigs (*Castor canadensis* and *Cavia porcellus*), we estimate that the

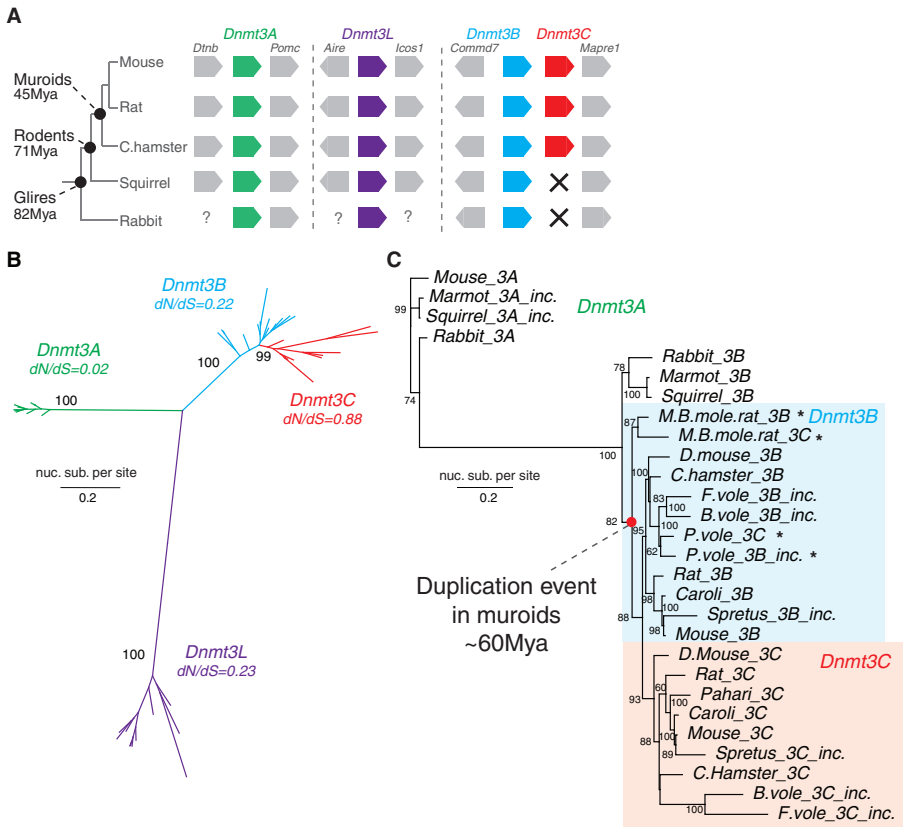

FIG. 1. *Dnmt3C* duplicated in the last common ancestor of all muroids. (*A*) Schematic of the genomic locus encoding *Dnmt3* genes in representative species of Glires. Estimated divergence times based on Timetree analyses (Hedges et al. 2015) are indicated at each major node. *Dnmt3* genes (colors) and corresponding neighboring genes (gray) are indicated with arrows. "?" denotes incomplete assembly, whereas "X" symbols denote absence of coding sequence. (*B*) Maximum likelihood nucleotide phylogeny of *Dnmt3* genes in Glires. Bootstrap values and average pairwise dN/dS are indicated for each clade. (*C*) Maximum likelihood phylogeny of all identified *Dnmt3B* and *Dnmt3C* genes. Incomplete sequences are indicated with "inc," whereas cases where *Dnmt3B* and *Dnmt3C* orthologs from the same species unexpectedly group together are highlighted with "*." Bootstrap support values >50% are reported. In addition included are *Dnmt3Bs* from rabbit (*Oryctolagus cuniculus*), marmot (*Marmota marmota*), 13-lined ground squirrel (*Ictidomys tridecemlineatus*), and 3 *Mus* species (mouse, *Mus musculus*; caroli, *Mus caroli*; spretus, *Mus spretus*) as well as selected *Dnmt3As* as an outgroup. Abbreviations and species names: Rat, *Rattus norvegicus*; M.B.mole.rat, mountain blind mole rat (*Spalax judaei*); D.mouse, deer mouse (*Peromyscus maniculatus*); C.hamster, Chinese hamster (*Cricetulus griseus*); F.vole, field vole (*Microtus agrestis*); B.vole, bank vole (*Myodes glareolus*); P.vole, prairie vole (*Microtus ochrogaster*).

duplication occurred before the radiation of muroids between 45 and 71 Ma (Hedges et al. 2015).

However, our second expectation—that *Dnmt3B* and *Dnmt3C* genes evolve independently—was not met. Although most *Dnmt3B* and *Dnmt3C* genes grouped into two distinct clades according to the accepted muroid species phylogeny (Steppan and Schenk 2017), both the prairie vole (*Microtus ochrogaster*) and the mountain blind mole rat (*N. galili*) had *Dnmt3B* and *Dnmt3C* paralogs that were more closely related to each other than to their respective orthologs (fig. 1C, asterisks). This pattern could indicate separate origins of *Dnmt3C* in these species or, alternatively, recent gene conversion. It is also possible that partial gene conversion between *Dnmt3B* and *Dnmt3C* occurred in other muroid species but was not evident in this phylogenetic analysis, perhaps because full-length gene sequences obscured this signal.

We therefore used a likelihood-based method, GARD, to map putative recombination breakpoints between *Dnmt3B* and *Dnmt3C* (see Materials and Methods; Kosakovsky Pond et al. 2006). Such analyses aim to identify recombination breakpoints based on segments of multiple alignments that have clearly discordant phylogenetic histories from each other. We identified three high-confidence breakpoints in muroid *Dnmt3B* and *Dnmt3C* sequences, partitioning the aligned sequences into four segments with distinct evolutionary histories—A, B, C, and D (fig. 2A). Upon generating phylogenies of each segment independently, we observed that discordance between these segments was not limited to prairie vole and mountain blind mole rat (fig. 2B). Gene conversion between *Dnmt3B* and *Dnmt3C* therefore occurred in many muroid lineages.

Next, we investigated the individual evolutionary trajectories of the distinct recombination segments within *Dnmt3C*. Consistent with rampant gene conversion, nucleotide phylogenies showed that segments B and C—encoding the ADD and part of the MTase (fig. 2A)—grouped *Dnmt3C* and *Dnmt3B* paralogs by species (fig. 2B and supplementary fig. S1, Supplementary Material online) rather than by orthology groups. We also found evidence for gene conversion in

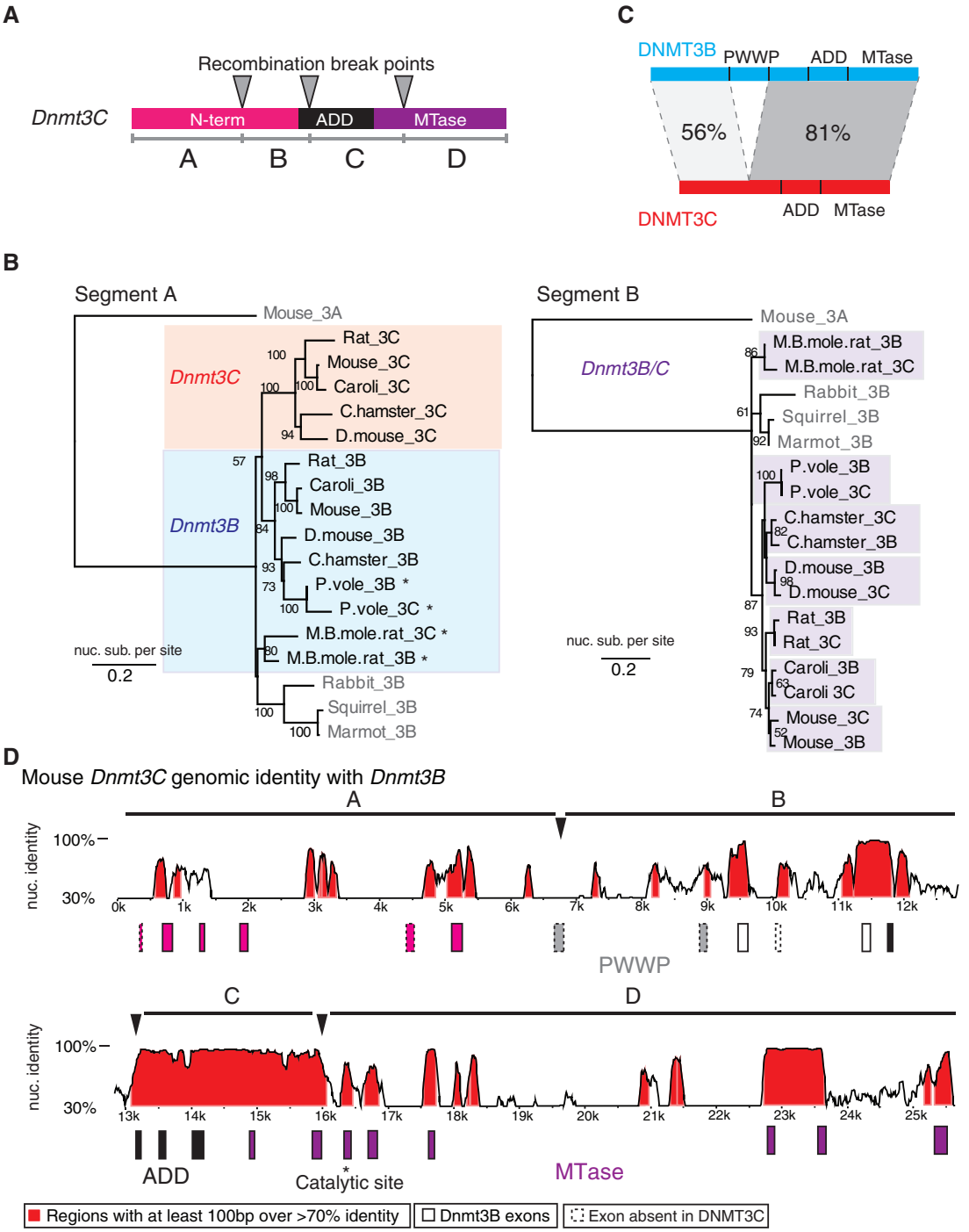

Fig. 2. Gene conversion between *Dnmt3C* and *Dnmt3B* in muroids. (*A*) Recombination map of *Dnmt3C* CDS. Breakpoints (arrowheads) identified by GARD (see Materials and Methods) and recombination segments ("A" through "D") are indicated on the *Dnmt3C* CDS with its predicted protein domains. (*B*) Maximum likelihood phylogenies of recombination segments A and B. Bootstrap supports >50% are shown. (*C*) Average pairwise percent identities between muroid DNMT3C and DNMT3B proteins before and after the PWWP domain. (*D*) Map of the mouse *Dnmt3B* genomic locus with annotated exons (boxes). Exons missing in *Dnmt3C* are shown with dashed lines. Exons are colored according to the domain they encode: pink, N-terminus; gray, PWWP; black, ADD; purple, methyltransferase. Percent nucleotide identity with *Dnmt3C* is plotted (*y* axis) using mVISTA (see Materials and Methods). Recombination segments (top) and predicted domains (bottom) are also shown.

segment D, which encodes the rest of the MTase (supplementary fig. S1*B*, Supplementary Material online). With the possible exception of prairie voles and mountain blind mole rats, we found no evidence for gene conversion in segment A, which encodes the N-terminal tail of DNMT3C (fig. 2*A*). Indeed, a phylogeny based on segment A alone almost perfectly separated the *Dnmt3B* and *Dnmt3C* paralogs based on orthology groups, consistent with divergent evolution of the two genes following duplication (fig. 2*B*). Thus, in most muroids and in contrast to the rest of the gene, the 5′ ends

Table 1. Summary of Selection Tests across Muroid *Dnmt3* Genes.

	Seg.	nb. Species	Length (bp)	Tree Length	PAML—M7 vs. M8 *P* value	PAML—M8a vs. M8 *P* value	M(0) dN/dS	% Sites dN/dS > 1 (avg. dN/dS)	Sites BEB ≥ 90%
Dnmt3C	A	11	471	3.19	0.002	0.004	0.886	49 (1.62)	54 (T), 57 (Q), 95 (P), 96 (L)
	B	8	135	1.35	0.509	0.965	0.176	—	N/A
	C	8	546	1.2	1.000	0.827	0.114	—	N/A
	D	10	666	1.46	1.000	0.907	0.184	—	N/A
Dnmt3B	All	8	2,052	1.32	0.170	0.475	0.116	1 (1.59)	N/A
Dnmt3A	All	9	2,718	0.83	0.823	0.463	0.022	—	N/A
Dnmt3L	All	9	1,218	2.17	0.546	0.732	0.271	—	N/A

NOTE.—Recombination segments of *Dnmt3C* were analyzed independently, whereas full-length sequences were used for other *Dnmt3s*. *P* values are for likelihood ratio tests between substitution models allowing or not allowing for positive selection using codeml (PAML). Colored boxes highlight *P* values <0.05. See text and Materials and Methods for details.

of *Dnmt3B* and *Dnmt3C* do not appear to have engaged in recent gene conversion. Consistent with these findings, DNMT3B and DNMT3C protein sequences shared much higher homology in their C-terminal compared with their N-terminal domains (fig. 2C).

To further confirm our findings, we investigated the coding and noncoding genomic sequences of *Dnmt3B* and *Dnmt3C* for signatures of high-sequence identity. High-nucleotide identities between mouse *Dnmt3C* and *Dnmt3B* were evident not only in coding exons but also across many introns (fig. 2D). More specifically, all introns displayed >70% identity in segment C but not in segment A (fig. 2D), consistent with the recombination breakpoint analysis (fig. 2A). Similarly, we identified high identity in several introns of segment D (fig. 2D and supplementary fig. S1C and D, Supplementary Material online). We found an even more evident pattern of sequence homogenization between *Dnmt3B* and *Dnmt3C* in genomes of rats and mountain blind mole rats (supplementary fig. S1C and D, Supplementary Material online, respectively). In particular, the high-sequence identity between the *Dnmt3B* and *Dnmt3C* loci in mountain blind mole rats (supplementary fig. S1D, Supplementary Material online) supports the hypothesis that this species, as well as prairie voles (fig. 2B), engaged in gene conversion more recently that other muroids. Taken together, these results suggest that following duplication, *Dnmt3B* and *Dnmt3C* have been subject to extensive gene conversion, except in their 5′ ends. Thus, DNMT3C N-termini evolve under distinct evolutionary trajectories from their DNMT3B counterparts, whereas the central domains and C-termini of *Dnmt3B*, and *Dnmt3C* exchange sequences to remain similar within each genome.

We took advantage of our recombination analyses to get a more precise estimate of when *Dnmt3C* first evolved in rodents. Using segment A, which we estimate has not been subject to gene conversion following the origin of *Dnmt3C* in rodents, we calculated the rate of synonymous substitutions (dS) between rabbit and mouse *Dnmt3B* to be 0.81, which is remarkably similar (as expected) to the dS of 0.79 between rabbit *Dnmt3B* and mouse *Dnmt3C*. Similarly, we calculated the dS between mouse *Dnmt3B* and *Dnmt3C* as 0.60. Based on an estimated divergence time of 80 Ma between rabbit and mouse (Hedges et al. 2015), we infer that *Dnmt3C* first arose in muroids ~60 Ma (fig. 1C).

DNMT3C N-Terminal Domain Evolve under Positive Selection

Gene conversion has homogenized several segments of DNMT3C and DNMTB, but not their N-terminal domains. We hypothesized that this could be to retain the functional divergence of DNMT3B and DNMT3C in their N-terminal domains. For example, loss of the ancestral PWWP domain in DNMT3C may have allowed it to specialize for functions distinct from DNMT3B. If this were the case, we might expect to find additional differences in the selective constraints that act on *Dnmt3B* versus *Dnmt3C*, especially in their N-terminal domains. We therefore investigated how the DNMT3B and DNMT3C N-terminal domains may have diverged in their selective constraints.

As the depth of species divergence is similar in all subtrees (fig. 1B), *Dnmt3C* appears to be the most divergent of all *Dnmt3* genes in muroid rodents based on the branch lengths of the DNMT3 phylogeny, followed by *Dnmt3L*, *Dnmt3B*, and finally *Dnmt3A*, which is the most highly conserved. To evaluate selective constraints, we calculated the rates of nonsynonymous (amino-acid altering, d*N*) and synonymous (silent, dS) substitutions across orthologous sequences of all *Dnmt3* genes. *Dnmt3C* displays the highest average pairwise d*N*/dS of all *Dnmt3* genes (0.88) compared with *Dnmt3L* (0.23), *Dnmt3B* (0.22), and *Dnmt3A* (0.02) (fig. 1B). Higher d*N*/dS values could reflect relaxation of selective constraint. Alternatively, these higher values could be the result of diversifying selection acting on *Dnmt3C*.

To distinguish between these possibilities, we used likelihood methods implemented in the PAML package to detect signatures of positive selection (Yang 1997). Muroideae are an ideal species set for these analyses because they span a short evolutionary time (~40 My) with low saturation of dS (Steppan and Schenk 2017). We separately analyzed each of the four recombination segments across all orthologs identified in muroids. Because some *Dnmt3C* genes are based on incomplete gene models, each segment alignment contained between 8 and 11 species (table 1). We then used PAML to identify site that were subject to positive selection (see Materials and Methods) (Yang 1997). We found no evidence of positive selection having acted on *Dnmt3B* or the other *Dnmt3s*. In contrast, we found strong support for positive selection having acted on segment A of *Dnmt3C*, but not on segments B, C, or D (table 1).

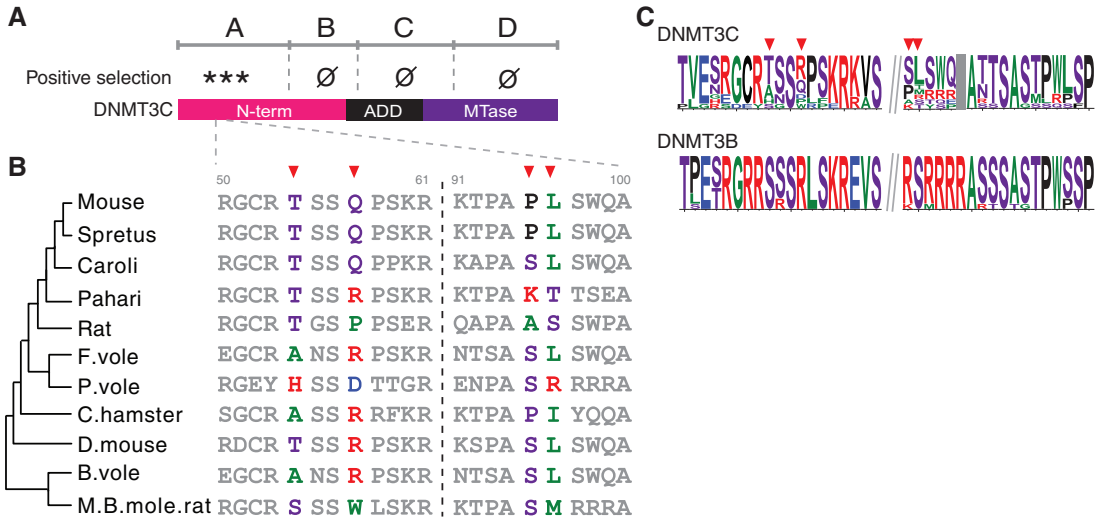

FIG. 3. DNMT3C N-terminus is subject to positive selection. (*A*) Schematic representation of positive selection test results for all recombination segments of DNMT3C CDS. "***" denotes the finding of positive selection in segment A (see text and Materials and Methods for details), whereas "∅" indicates no support for positive selection. (*B*) Amino-acid alignments (positions 50–61 and 91–100) of muroid DNMT3Cs showing four positively selected sites identified with PAML (red arrowheads). Sequences are arranged according to segment A phylogeny with species names on the right. (*C*) Amino-acid logos of DNMT3C and DNMT3B around the positively selected sites (arrowheads). Backslashes indicate sequences not shown. The gray box denotes an alignment gap between DNMT3C and DNMT3B.

In segment A of *Dnmt3C*, PAML analyses estimated 49% of sites that evolved with an average $dN/dS > 1$ indicative of potential diversifying selection; their average dN/dS was estimated to be 1.6. Of these, four sites were highlighted with a high posterior probability of having evolved under positive selection (Bayes Empirical Bayes [P] ≥90%, table 1, and fig. 3*B*). These sites (codons 54, 57, 95, and 96 in mouse *Dnmt3C*) all cluster within the most 5′ end of the gene (first 300 bp of the CDS) and display extensive diversification in both charge and hydrophobicity across muroids (fig. 3*B*). For sites 95 and 96, rapid evolution disrupts a highly conserved arginine patch of unknown function, which is highly conserved among muroid DNMT3B proteins (fig. 3*C*). Thus, in addition to the loss of the PWWP domain, DNMT3B and DNMT3C differ in the selective constraints to which they are subject. The signature of positive selection and loss of the PWWP domains make *Dnmt3C* unique among all *Dnmt3* genes.

As an alternate means to detect positive selection, we used the branch-site unrestricted statistical test for episodic diversification (BUSTED) method as implemented in the HyPhy server (Murrell et al. 2015; Weaver et al. 2018). Consistent with our PAML analyses, we found strong evidence for positive selection using this method on rodent *Dnmt3C* ($P < 0.0001$) but not on *Dnmt3A* ($P = 1.00$), *Dnmt3B* ($P = 0.95$), or *Dnmt3L* ($P = 0.62$) in rodents.

Selective Constraints Acting on DNMT3 Proteins in Primate Genomes

The evolutionary birth of *Dnmt3C* afforded muroid rodents a unique opportunity to silence young, active retrotransposon families by DNA methylation. However, most mammalian genomes face a similar pressure by young retrotransposon lineages and yet do not encode *Dnmt3C*. We therefore hypothesized that non-rodent mammalian species might deploy alternative mechanisms, possibly other DNMT3 enzymes, to achieve DNMT3C-like repression of active retrotransposons. If true, we might expect *Dnmt3* genes to be locked in these molecular arms races, and therefore subject to similar selective pressures (i.e., diversifying selection) as *Dnmt3C* in muroid rodents.

To investigate this possibility, we analyzed the evolutionary constraints that act on *DNMT3* genes in primates, a distinct lineage of mammals that have substantial genomic resources across multiple species in a comparable evolutionary timespan to muroid rodents (Hedges et al. 2015). Using maximum likelihood-based analyses, we found strong evidence of diversifying selection acting on *DNMT3A* and marginal evidence of positive selection in the catalytically inactive cofactor *DNMT3L* (table 2). In contrast, we found no evidence of diversifying selection in primate *DNMT3B* (table 2) or muroid *DNMT3A* (table 1). Similarly, BUSTED analyses also revealed a signature of episodic positive selection in *DNMT3A* ($P = 0.016$) but not in *DNMT3B* ($P = 0.47$) or *DNMT3L* ($P = 0.68$) in primates.

As in muroid *Dnmt3C*, the diversifying selection signature also primarily mapped to the N-terminal domain of primate *DNMT3A* (codons 61 and 81, table 2, and fig. 4). To rule out that this signature could be due to unaccounted recombination, we performed GARD analyses (Kosakovsky Pond et al. 2006). This identified a single break point (within the first 1 kb of the CDS), however, whereas a maximum likelihood phylogeny of the first segment (including the rapidly evolving sites) had strong bootstrap support, the second segment did not (supplementary fig. S2, Supplementary Material online). Further inspecting this second segment showed a high rate of CpG mutations which prevent appropriate reconstruction of its evolution and accurate selection analyses.

Table 2. Summary of Selection Tests across Primate *DNMT3s*.

	Segment	nb. Species	Length (bp)	Tree Length	PAML—M7 vs. M8 *P* Value	PAML—M8a vs. M8 *P* Value	M(0) dN/dS	% Sites dN/dS > 1 (avg. dN/dS)	Sites BEB ≥ 90%
DNMT3A	Whole	20	2,724	0.509	0.022	≤0.001	0.047	2.02 (1.8)	66 (P), 81 (A)
DNMT3B	Whole	21	2,481	0.828	0.149	0.430	0.080	1.81 (1.7)	—
DNMT3L	Whole	20	1,155	1.7	0.018	0.124	0.204	5.29 (1.6)	—

NOTE.—Codeml (PAML) analyses using the accepted species phylogeny. *P* values are for likelihood ratio tests between substitution models allowing or not allowing for positive selection. Colored boxes highlight *P* values <0.05.

FIG. 4. Positive selection of N-terminal domain of primate DNMT3A. Amino-acid alignments (positions 58–90) of primate DNMT3As showing the two positively selected sites identified with PAML (arrowheads). Sequences are arranged according to the accepted species phylogeny with species names on the right.

We therefore conclude that there is insufficient evidence for gene conversion affecting *DNMT3A* evolution in primates. In spite of this, PAML analysis of *DNMT3A* putative first (N-terminal) segment also identifies sites 61 and 81 as evolving under positive selection (not shown).

Unlike their DNA-methyltransferase and ADD domains, primate *DNMT3A* and rodent *DNMT3C* share only 15% of their N-terminal residues. This level of homology is so low that BLAST searches between the N-terminal domains only return an E-value of 0.78. Thus, although we cannot make any strong statements about functional homology between these domains, we note that the region under positive selection in primate DNMT3A does appear to overlap with one patch of positive selection found in DNMT3C (supplementary fig. S3, Supplementary Material online).

Overall, we find evidence of diversifying selection on distinct *DNMT3* genes in rodent and primate genomes (tables 1 and 2). Our findings could imply that the N-terminal portions of DNMT3 proteins wage evolutionary arms races for DNA methylation of young, active retrotransposons in different mammalian lineages. They further raise the possibility that DNMT3A, which is universal to all mammals, may be the original DNMT3 that targets young retrotransposons. The subsequent birth of *Dnmt3C* in muroid rodents may have absolved DNMT3A of this role, which could be why we cannot detect any signatures of diversifying selection in *Dnmt3A* in rodent species.

Discussion

Retrotransposons activity poses a significant fitness challenge to host genomes. To protect themselves, host genomes deploy multipronged strategies to curb retrotransposon activity. Here, we identified the selective forces shaping the function of a recently duplicated DNA methyltransferase, DNMT3C, that specifically targets evolutionarily young retrotransposons in muroid rodents. We found that *Dnmt3C* has undergone recurrent gene conversion with its parental gene *Dnmt3B*, except for the N-terminal domain. These findings are reminiscent of previous studies of gene families subject to genetic conflicts (Daugherty and Zanders 2019). For example, the true evolutionary histories of the mammalian antiviral *IFIT1/IFIT1B* paralogs, which diverged 100 Ma, were also confounded by recurrent gene conversion (Daugherty et al. 2016). Similarly, recurrent gene conversion affected the histone-fold domain but not the distinct N-terminal tails of centromeric histone paralogs in *Drosophila* species (Kursel and Malik 2017). In all these cases, as well as several additional examples (Daugherty and Zanders 2019), natural selection maintains gene conversion within the core functional domain of the paralogs while it selects against gene conversion in the domain that drives their functional diversification. Mechanistically, we speculate that the close proximity of the paralogs following gene duplication—as it is the case for *Dnmt3B* and *Dnmt3C*—facilitated multiple episodes of gene conversion during meiotic recombination.

We found that the N-terminal domain of *Dnmt3C*, but not its parental gene *Dnmt3B*, has evolved under strong diversifying selection. Diversifying selection—especially in a host "defense" gene—is a signature of an evolutionary arms race between host genomes and retrotransposons (Molaro and Malik 2016). As host genomes deploy repressive chromatin strategies, retrotransposons must adapt to ward off host repression, in turn spurring host adaptation. The evolutionary arms race model further makes the prediction that residues or domains that directly engage in the antagonism should be rapidly evolving. Thus, one possibility is that the positive selection in *Dnmt3* genes results from active antagonism by an RNA or protein expressed by young retrotransposons. Under this model, positive selection in DNMT3 proteins allows them to evade binding and antagonism by young retrotransposons.

An alternative model is that positive selection shapes the targeting of DNMT3 proteins to young retrotransposons to mediate their silencing. This predicted activity would be

similar to the KZNF (KRAB domain containing Zinc Finger) proteins, which use rapid evolution of their DNA-binding domains to keep pace with a changing nucleotide landscape of retrotransposon families (Thomas and Schneider 2011). We hypothesize that similar evolutionary dynamics could drive the diversifying selection of the N-terminal domains in rodent DNMT3C and primate DNMT3A proteins. Interestingly, DNMT3A exists both as a long A1 isoform, and a short A2 isoform that lacks the N-terminal domain (Chen et al. 2002). We posit that the long DNMT3A1 isoform may target young retrotransposons in male germ cells in DNMT3C-less mammalian species, such as primates. The recurrent signature of rapid evolution within the N-termini of two different DNMT3 proteins in different mammalian lineages may highlight a novel functional domain that may be key to DNMT3 targeting to retrotransposons. Unlike the canonical PWWP, ADD and MTase domains, however, this domain may be characterized by its rapid evolution rather than conservation. How this domain engages with retrotransposons remains to be determined. In contrast to KZNF proteins, there is no suggestion that DNMT3 proteins have DNA sequence-binding specificity. Instead, it is possible that this region mediates interaction with components of the piRNA pathway—some of which are rapidly evolving in other animals (Simkin et al. 2013; Yi et al. 2014).

In sum, the DNMT3C N-terminal domains can be distinguished from other DNMT3 proteins by its diversifying selection and loss of a coding PWWP domain. The PWWP domain is essential for coupling de novo DNA methylation to local chromatin environment, via recognition of H3K36 methylated histones (Ge et al. 2004; Dhayalan et al. 2010; Rondelet et al. 2016). In DNMT3B, the PWWP domain binds H3K36me3 marks, which are typical of transcribed gene bodies (Baubec et al. 2015). In DNMT3A, the PWWP domain is intact and was recently shown to mediate DNMT3A-dependent methylation of intergenic sequences (Weinberg et al. 2019). We hypothesize here that DNMT3C's N-terminal domain may be required to substitute for PWWP-dependent chromatin-targeting function. However, the mode of targeting of DNMT3C to young retrotransposon promoters remains to be determined.

In conclusion, our evolutionary studies identified a new functional domain in DNMT3C, a DNA methyltransferase enzyme whose exclusive function is to silence the most active, rapidly adapting retrotransposon families in rodent genomes (Barau et al. 2016). Furthermore, based on our findings of diversifying selection in primate DNMT3As, we suggest that diversifying selection of enzymes that methylate retrotransposons in developing germ cells might be pervasive across mammalian genomes, although this targeting may be mediated by distinct DNMT3 paralogs.

Materials and Methods

Identification of *DNMT3* Orthologs

To identify *Dnmt3* orthologs, we performed TBLASTN searches on the NCBI nonredundant nucleotide database (Altschul et al. 1990; NCBI Resource Coordinators 2016), using reference protein sequences of mouse DNMT3A (NP_031898.1), DNMT3B (XP_006498745.1), DNMT3L (NP_001075164.1) as well as the predicted protein sequence from the *Dnmt3C* cDNA cloned from male fetal gonads (Barau et al. 2016). Although most *Dnmt3s* have predicted sequences in reference databases, *Dnmt3C* genes are not annotated in most muroid genomes. In these cases, we queried genomes directly using TBLASTN, and predicted gene models from contigs using GeneWise (Birney et al. 2004). CDSs were annotated based on the longest mouse gene model.

Queried Genomes

We used the following genome assemblies to predict *Dnmt3B* and *Dnmt3C* gene models. Muroids: *Mus musculus* (UCSC mm10), *Mus spretus* (Sanger, SPRET_EiJ), *Mus caroli* (Sanger, CAROLI_EiJ), *Mus pahari* (Sanger, Pahari_EiJ), *Apodemus sylvaticus* (NCBI, GCA_001305905.1_ASM130590v1), *Rattus norvegicus* (UCSC, rn6), *Peromyscus maniculatus* (NCBI, GCF_000500345.1_Pman_1.0), *Myodes glareolus* (NCBI, GCA_001305785.1_ASM130578v1), *Microtus agrestis* (NCBI, GCA_001305995.1_ASM130599v1), *M. ochrogaster* (NCBI, MicOch1), *Mesocricetus auratus* (NCBI, MesAur1), *Cricetulus griseus* (UCSC, criGri1), and *N. galili* (NCBI, GCF_000622305.1_S.galili_v1.0).

Glires: *C. canadensis* (NCBI, C.can genome v1.0), *Oryctolagus cuniculus* (UCSC, oryCun2), *Marmota marmota* (NCBI, GCF_001458135.1_marMar2.1), *Ictidomys tridecemlineatus* (UCSC, speTri2), and *Cav. porcellus* (Broad Institute cavPor3).

Species Divergence Times

Divergence time estimates were obtained from using timetree.org, last accessed February 28, 2020 (Hedges et al. 2015), by specifying sister taxa that belong to either Glires, rodents, or muroids. Timetree outputs a range of estimated divergence times summarizing phylogenetic and fossil dating.

Synteny Analysis

Shared synteny blocks were identified using the online server Genomicus (V95.1), last accessed February 28, 2020 (Nguyen et al. 2018). Mouse was used as a reference locus and individual synteny blocks were inspected using the UCSC genome browser (Kent et al. 2002).

Alignments and Phylogenies

All sequence alignments are available as Supplementary Material online. Alignments were generated using ClustalW v2.1 (IUB cost matrix; Larkin et al. 2007) or MAFFT v7.388 (Katoh and Standley 2013). Maximum likelihood phylogenies were built using PHyML v3.0 with 100 bootstraps (Guindon et al. 2010). Trees were visualized using the software Geneious Prime (Biomatters Ltd). In all cases, we used nucleotide alignments of the CDS and the HKY85 substitution model.

Detection of Recombination

To test for recombination, we used an alignment of *Dnmt3C* and *Dnmt3B* CDS from six species with nearly complete gene models (mouse, *Mus caroli*, rat, prairie vole, Chinese hamster,

and mountain blind mole rat). Assembly gaps were removed. To detect recombination breakpoints, we used GARD with the general discrete model of site to site variation and three rate classes (Kosakovsky Pond et al. 2006). We kept breakpoints with right and left P values <0.01. We subsequently segmented the *Dnmt3C* alignment according to these breakpoints. Similarly, recombination in primate *DNMT3A* was tested using an alignment of all primate CDS.

Genomic Alignments

To identify region of homology between *Dnmt3C* and *Dnmt3B* genomic loci, we extracted the regions from assembled genomes of the mouse and rat and contigs of mountain blind mole rat and aligned them using mVista (Frazer et al. 2004). Exon annotations were based on reference alignments with the species CDS.

Selection Analyses

We measured overall d*N*/d*S* rates with codeml, PAMLX V1.3.1 (Yang 1997), under model 0 and average pairwise with SNAP V2.1.1 (Korber et al. 2000). We tested for positive selection using codon alignments generated with PAL2NAL (Suyama et al. 2006) free of any gaps and stop codons and with either accepted species or gene phylogenies. We compared "NSsites" evolutionary models that do not allow d*N*/d*S* to exceed 1 (M7 or M8a) to a model that does (M8). We tested for statistical significance using a χ^2 test of the twice difference in log-likelihoods between M8 and matched null model M7 or M8a, with the degrees of freedom reflecting the difference in number of parameters between the two models compared (Yang 1997). Positively selected sites were classified as those sites with M8 Bayes Empirical Bayes posterior probability $>90\%$. The results we present are from codeml runs using the F3x4 codon frequency model, and initial Omega 0.4. Analyses were robust to use of different starting parameters (codon frequency model F61; starting Omega 1.5). In parallel, we also carried out analyses to detect episodic positive selection on a gene by gene basis using the BUSTED method (Murrell et al. 2015) as implemented in the HyPhy online server, datamonkey.org, last accessed February 28, 2020 (Weaver et al. 2018).

DNMT3C and DNMT3B Logo Plots

Logo plots were generated using weblogo.threeplusone.com, last accessed February 28, 2020 (Crooks et al. 2004); using all muroid species with alignable sequences over these exons: mouse (*Mus musculus*), *Mus spretus*, *Mus caroli*, rat, deer mouse, field vole, prairie vole, bank vole, Chinese hamster, and mountain blind mole rat.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We would like to thank members of the Bourc'his and Malik laboratories, especially Joan Barau, Tera Levin, and Janet Young for technical help and critical reading of this article. This work was supported by a postdoctoral fellowship from the Damon Runyon Cancer Research Foundation (to A.M.), National Institutes of Health (NIH) Grant R01 GM074108 (to H.S.M.), and an Howard Hughes Medical Institute (HHMI) Investigator Award (to H.S.M.). The laboratory of D.B. is part of the Laboratoire d'Excellence (LABEX) entitled DEEP (11-LBX0044).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215(3):403–410.

Alvarez-Ponce D, Torres-Sanchez M, Feyertag F, Kulkarni A, Nappi T. 2018. Molecular evolution of DNMT1 in vertebrates: duplications in marsupials followed by positive selection. *PLoS One* 13(4):e0195162.

Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, Bestor T, Hannon GJ. 2008. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell*. 31(6):785–799.

Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Hérault Y, Guillou F, Bourc'his D. 2016. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* 354(6314):909–912.

Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, Akalin A, Schubeler D. 2015. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* 520(7546):243–247.

Bestor T, Laudano A, Mattaliano R, Ingram V. 1988. Cloning and sequencing of a CDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J Mol Biol*. 203(4):971–983.

Bestor TH. 2000. The DNA methyltransferases of mammals. *Hum Mol Genet*. 9(16):2395–2402.

Birney E, Clamp M, Durbin R. 2004. Genewise and genomewise. *Genome Res*. 14(5):988–995.

Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. 2001. DNMT3L and the establishment of maternal genomic imprints. *Science* 294(5551):2536–2539.

Chen T, Tsujimoto N, Li E. 2004. The PWWP domain of DNMT3A and DNMT3B is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. *Mol Cell Biol*. 24(20):9048–9058.

Chen T, Ueda Y, Xie S, Li E. 2002. A novel DNMT3A isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. *J Biol Chem*. 277(41):38746–38754.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res*. 14(6):1188–1190.

Daugherty MD, Schaller AM, Geballe AP, Malik HS. 2016. Evolution-guided functional analyses reveal diverse antiviral specificities encoded by IFIT1 genes in mammals. *Elife* 5.

Daugherty MD, Zanders SE. 2019. Gene conversion generates evolutionary novelty that fuels genetic conflicts. *Curr Opin Genet Dev*. 58–59:49–54.

Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, Jeltsch A. 2010. The DNMT3A PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J Biol Chem*. 285(34):26114–26120.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. Vista: computational tools for comparative genomics. *Nucleic Acids Res*. 32(Web Server):W273–W279.

Ge YZ, Pu MT, Gowher H, Wu HP, Ding JP, Jeltsch A, Xu GL. 2004. Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. *J Biol Chem*. 279(24):25447–25454.

Gruenbaum Y, Cedar H, Razin A. 1982. Substrate and sequence specificity of a eukaryotic DNA methylase. *Nature* 295(5850):620–622.
Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 59(3):307–321.
Guo X, Wang L, Li J, Ding Z, Xiao J, Yin X, He S, Shi P, Dong L, Li G, et al. 2015. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* 517(7536):640–644.
Hedges SB, Marin J, Suleski M, Paymer M, Kumar S. 2015. Tree of life reveals clock-like speciation and diversification. *Mol Biol Evol*. 32(4):835–845.
Jain D, Meydan C, Lange J, Claeys Bouuaert C, Lailler N, Mason CE, Anderson KV, Keeney S. 2017. Rahu is a mutant allele of DNMT3C, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline. *PLoS Genet*. 13(8):e1006964.
Jeltsch A, Broche J, Bashtrykov P. 2018. Molecular processes connecting DNA methylation patterns with DNA methyltransferases and histone modifications in mammalian genomes. *Genes* 9(11):566.
Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X. 2007. Structure of DNMT3A bound to DNMT3L suggests a model for de novo DNA methylation. *Nature* 449(7159):248–251.
Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H. 2004. Essential role for de novo DNA methyltransferase DNMT3A in paternal and maternal imprinting. *Nature* 429(6994):900–903.
Kato Y, Kaneda M, Hata K, Kumaki K, Hisano M, Kohara Y, Okano M, Li E, Nozaki M, Sasaki H. 2007. Role of the DNMT3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum Mol Genet*. 16(19):2272–2280.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.
Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res*. 12(6):996–1006.
Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, Hahn BH, Wolinsky S, Bhattacharya T. 2000. Timing the ancestor of the HIV-1 pandemic strains. *Science* 288(5472):1789–1796.
Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. 2006. GARD: a genetic algorithm for recombination detection. *Bioinformatics* 22(24):3096–3098.
Kursel LE, Malik HS. 2017. Recurrent gene duplication leads to diverse repertoires of centromeric histones in *Drosophila* species. *Mol Biol Evol*. 34(6):1445–1462.
Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21):2947–2948.
Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet*. 11(3):204–220.
Lees-Murdock DJ, McLoughlin GA, McDaid JR, Quinn LM, O'Doherty A, Hiripi L, Hack CJ, Walsh CP. 2004. Identification of 11 pseudogenes in the DNA methyltransferase gene family in rodents and humans and implications for the functional loci. *Genomics* 84(1):193–204.
Li E, Bestor TH, Jaenisch R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69(6):915–926.
Lyko F. 2018. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat Rev Genet*. 19(2):81–92.
Manakov SA, Pezic D, Marinov GK, Pastor WA, Sachidanandam R, Aravin AA. 2015. MIWI2 and MILI have differential effects on piRNA biogenesis and DNA methylation. *Cell Rep*. 12(8):1234–1243.
Molaro A, Falciatori I, Hodges E, Aravin AA, Marran K, Rafii S, McCombie WR, Smith AD, Hannon GJ. 2014. Two waves of de novo methylation during mouse germ cell development. *Genes Dev*. 28(14):1544–1549.
Molaro A, Malik HS. 2016. Hide and seek: how chromatin-based pathways silence retroelements in the mammalian germline. *Curr Opin Genet Dev*. 37:51–58.
Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM, et al. 2015. Gene-wide identification of episodic selection. *Mol Biol Evol*. 32(5):1365–1371.
NCBI Resource Coordinators. 2016. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 44(D1):D7–D19.
Nguyen NTT, Vincens P, Roest Crollius H, Louis A. 2018. Genomicus 2018: karyotype evolutionary trees and on-the-fly synteny computing. *Nucleic Acids Res*. 46(D1):D816–D822.
Okano M, Bell DW, Haber DA, Li E. 1999. DNA methyltransferases DNMT3A and DNMT3B are essential for de novo methylation and mammalian development. *Cell* 99(3):247–257.
Okano M, Xie S, Li E. 1998. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet*. 19(3):219–220.
Ooi SKT, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin S-P, Allis CD, et al. 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448(7154):714–717.
Otani J, Nankumo T, Arita K, Inamoto S, Ariyoshi M, Shirakawa M. 2009. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep*. 10(11):1235–1241.
Ponger L, Li WH. 2005. Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Mol Biol Evol*. 22(4):1119–1128.
Posfai J, Bhagwat AS, Posfai G, Roberts RJ. 1989. Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res*. 17(7):2421–2435.
Qiu C, Sawada K, Zhang X, Cheng X. 2002. The PWWP domain of mammalian DNA methyltransferase DNMT3B defines a new family of DNA-binding folds. *Nat Struct Biol*. 9(3):217–224.
Reik W, Surani MA. 2015. Germline and pluripotent stem cells. *Cold Spring Harb Perspect Biol*. 7(11):a019422.
Rondelet G, Dal Maso T, Willems L, Wouters J. 2016. Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. *J Struct Biol*. 194(3):357–367.
Simkin A, Wong A, Poh YP, Theurkauf WE, Jensen JD. 2013. Recurrent and recent selective sweeps in the piRNA pathway. *Evolution* 67(4):1081–1090.
Smith ZD, Meissner A. 2013. DNA methylation: roles in mammalian development. *Nat Rev Genet*. 14(3):204–220.
Song J, Rechkoblit O, Bestor TH, Patel DJ. 2011. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* 331(6020):1036–1040.
Steppan SJ, Schenk JJ. 2017. Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. *PLoS One* 12(8):e0183070.
Suyama M, Torrents D, Bork P. 2006. Pal2Nal: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 34(Web Server):W609–W612.
Thomas JH, Schneider S. 2011. Coevolution of retroelements and tandem zinc finger genes. *Genome Res*. 21(11):1800–1812.
Timinskas A, Butkus V, Janulaitis A. 1995. Sequence motifs characteristic for DNA [cytosine-N4] and DNA [adenine-N6] methyltransferases. Classification of all DNA methyltransferases. *Gene* 157(1–2):3–11.
Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. 2018. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol*. 35(3):773–777.
Weinberg DN, Papillon-Cavanagh S, Chen H, Yue Y, Chen X, Rajagopalan KN, Horth C, McGuire JT, Xu X, Nikbakht H, et al. 2019. The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* 573(7773):281–286.
Yamanaka S, Nishihara H, Toh H, Eijy Nagai LA, Hashimoto K, Park SJ, Shibuya A, Suzuki AM, Tanaka Y, Nakai K, et al. 2019. Broad heterochromatic domains open in gonocyte development prior to de novo DNA methylation. *Dev Cell*. 51(1):21–34. e25.
Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*. 13(5):555–556.

Yi M, Chen F, Luo M, Cheng Y, Zhao H, Cheng H, Zhou R. 2014. Rapid evolution of piRNA pathway in the teleost fish: implication for an adaptation to transposon diversity. *Genome Biol Evol*. 6(6):1393–1407.

Yoder JA, Walsh CP, Bestor TH. 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet*. 13(8):335–340.

Yokomine T, Hata K, Tsudzuki M, Sasaki H. 2006. Evolution of the vertebrate DNMT3 gene family: a possible link between existence of DNMT3L and genomic imprinting. *Cytogenet Genome Res*. 113(1–4):75–80.

Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I, Rathert P, Brandt O, Reinhardt R, Fischle W, et al. 2010. Chromatin methylation activity of DNMT3A and DNMT3A/3L is guided by interaction of the add domain with the histone H3 tail. *Nucleic Acids Res*. 38(13):4246–4253.

Zhang ZM, Lu R, Wang P, Yu Y, Chen D, Gao L, Liu S, Ji D, Rothbart SB, Wang Y, et al. 2018. Structural basis for DNMT3A-mediated de novo DNA methylation. *Nature* 554(7692):387–391.