

RESEARCH ARTICLE

# Molecular evolution of *DNMT1* in vertebrates: Duplications in marsupials followed by positive selection

David Alvarez-Ponce<sup>1\*</sup>, María Torres-Sánchez<sup>1,2</sup>, Felix Feyertag<sup>1</sup>, Asmita Kulkarni<sup>1</sup>, Taylen Nappi<sup>1</sup>

**1** Department of Biology, University of Nevada, Reno, Nevada, United States of America, **2** Department of Biodiversity, Ecology and Evolution, Complutense University of Madrid, Madrid, Spain

\* [dap@unr.edu](mailto:dap@unr.edu)



## OPEN ACCESS

**Citation:** Alvarez-Ponce D, Torres-Sánchez 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. <https://doi.org/10.1371/journal.pone.0195162>

**Editor:** Albert Jeltsch, Universität Stuttgart, GERMANY

**Received:** January 15, 2018

**Accepted:** March 16, 2018

**Published:** April 5, 2018

**Copyright:** © 2018 Alvarez-Ponce et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This work was supported by a Pilot Grant from the Smooth Muscle Plasticity COBRE of the University of Nevada, Reno, funded by the National Institutes of Health (grant 5P30GM110767-04). MTS was supported by a FPI predoctoral fellowship (BES-2013- 062723) and a travel grant (EEBB-I-16-11395) from the Ministry of Economy and Competitiveness of Spain. The funders had no

## Abstract

DNA methylation is mediated by a conserved family of DNA methyltransferases (Dnmts). The human genome encodes three active Dnmts (Dnmt1, Dnmt3a and Dnmt3b), the tRNA methyltransferase Dnmt2, and the regulatory protein Dnmt3L. Despite their high degree of conservation among different species, genes encoding Dnmts have been duplicated and/or lost in multiple lineages throughout evolution, indicating that the DNA methylation machinery has some potential to undergo evolutionary change. However, little is known about the extent to which this machinery, or the methylome, varies among vertebrates. Here, we study the molecular evolution of Dnmt1, the enzyme responsible for maintenance of DNA methylation patterns after replication, in 79 vertebrate species. Our analyses show that all studied species exhibit a single copy of the *DNMT1* gene, with the exception of tilapia and marsupials (tammar wallaby, koala, Tasmanian devil and opossum), each of which displays two apparently functional *DNMT1* copies. Our phylogenetic analyses indicate that *DNMT1* duplicated before the radiation of major marsupial groups (i.e., at least ~75 million years ago), thus giving rise to two *DNMT1* copies in marsupials (copy 1 and copy 2). In the opossum lineage, copy 2 was lost, and copy 1 recently duplicated again, generating three *DNMT1* copies: two putatively functional genes (copy 1a and 1b) and one pseudogene (copy 1ψ). Both marsupial copies (*DNMT1* copies 1 and 2) are under purifying selection, and copy 2 exhibits elevated rates of evolution and signatures of positive selection, suggesting a scenario of neofunctionalization. This gene duplication might have resulted in modifications in marsupial methylomes and their dynamics.

## Introduction

In vertebrate genomes, cytosine methylation is widespread (e.g., 60–90% of CpGs are methylated in mammals [1,2]) and plays pivotal roles in the silencing of gene expression and transposable elements, gene imprinting, and X-chromosome inactivation [3]. DNA methylation is mediated by a conserved family of DNA methyltransferases (Dnmts). The human genome

role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

encodes five members of this family: Dnmt1, Dnmt2, Dnmt3a, Dnmt3b and Dnmt3L. Dnmt3a and Dnmt3b are responsible for *de novo* DNA methylation in germ cells and early embryos [4,5]. An additional member of the Dnmt3 group, Dnmt3L, does not exhibit catalytic activity, but acts as a regulator of Dnmt3a and Dnmt3b. Once established by Dnmt3a and Dnmt3b, methylation patterns are maintained by Dnmt1, which copies them to the daughter DNA strand after replication [6]. Despite its sequence and structural similarity to Dnmt1 and Dnmt3s, Dnmt2 methylates the anticodon loop of aspartic acid transfer RNA, rather than DNA [7,8].

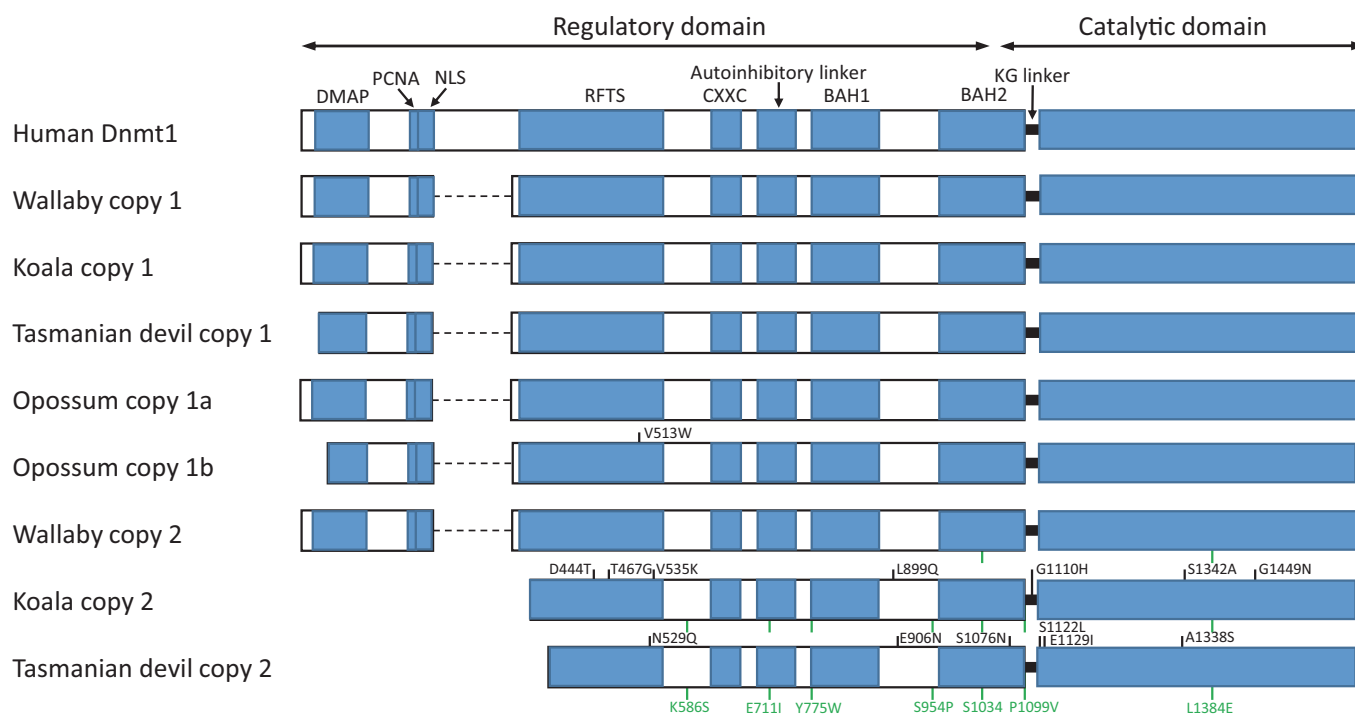
Prior comparative analyses of distantly related organisms have revealed a number of gene duplications and losses in the evolutionary history of the genes encoding Dnmts. A number of organisms lack such genes (and cytosine methylation), including the yeast *Saccharomyces cerevisiae* and the nematode *Caenorhabditis elegans*, and the number of Dnmts of each kind varies among lineages [2,9–15]. For instance, *DNMT3C*, a mouse retrogene that evolved by duplication of *DNMT3B*, has been recently shown to be responsible for silencing young retrotransposons in the male germ line [16]. Genes encoding all three Dnmt classes present in animals (classes 1, 2 and 3) are duplicated in some insect groups and completely absent from others [13]. Some insects, including Diptera, have lost cytosine methylation (however, evidence in *Drosophila* has been controversial [17]), and insects with a methylome include some lacking Dnmt1s or Dnmt3s, indicating that neither of the enzymes individually is essential for DNA methylation [13]. A phylogenetic analysis of prokaryotic and eukaryotic Dnmts revealed that the last universal eukaryotic ancestor contained members of classes 1, 2 and 3, and suggested that Dnmt2s evolved from DNA-methylating Dnmts, and that eukaryotic Dnmt1s and Dnmt3s originated independently from prokaryotic Dnmts [18].

Little is known about the extent to which the DNA methylation machinery, or the methylome, may vary among vertebrates. All jawed vertebrates characterized so far share a number of features, including global hypermethylation, a negative correlation between methylation levels at transcription start sites and gene expression levels, and widespread methylation of transposable elements and gene bodies, which results in repression of transposable elements and spurious gene transcription and exon splicing [19,20]. In contrast, in invertebrates methylation is sparse and shows a mosaic distribution, with unmethylated regions being interspersed with hypermethylated regions, there is no correlation between methylation at transcription start sites and expression levels, and methylation does not necessarily correlate with the position of genes or transposable elements [20]. Interestingly, the genome of the sea lamprey *Petromyzon marinus*, a basal jawless vertebrate, is intermediately methylated, with a methylome displaying intermediate characteristics between those of jawed vertebrates and invertebrates [21]. Vertebrates also differ in the link between X chromosome inactivation and methylation: in eutherian females, one of the X female chromosomes is inactivated thanks to methylation of promoter-associated CpG islands [22–24]; in marsupials, transcription start sites are equally methylated in active and inactive X chromosomes, but the flanking regions are hypomethylated in the inactive X chromosome [24,25]; in birds and monotremes, active and inactive X/Z chromosomes do not differ in their methylation levels [24,25].

Molecular evolution studies of the DNA methylation machinery in vertebrates include some comparative analyses of members of the Dnmt3 group [26,27], but less is known about the evolution of the *DNMT1* gene. The human *DNMT1* gene has 40 exons and encodes a full, 1616-amino acid somatic isoform (named Dnmt1s) and a truncated isoform expressed in oocytes (Dnmt1o), which lacks the first 118 amino acids. Dnmt1 proteins contain an N-terminal regulatory region and a C-terminal catalytic domain, separated by a KG repeat. In the Dnmt1s isoform, the regulatory region comprises a DNA methyltransferase associated protein (DMAP) binding domain, a nuclear localization signal (NLS), a replication foci targeting

sequence (RFTS), a cysteine-rich DNA binding domain (CXXC), an autoinhibitory linker that prevents *de novo* methylation, and two bromo-adjacent homology domains (BAH1 and BAH2), among other protein-interaction domains (Fig 1; for a comprehensive review, see ref. [28]). A direct interaction between the N-terminal and the C-terminal domains seems to be necessary for enzyme activation [29]. Activated Dnmt1 shows high preference for the methylation of hemimethylated CpG sites, which allows it to maintain methylation states after replication [30,31]. *DNMT1*-null mouse embryos die soon after implantation and display delayed development and structural abnormalities [32], and overexpression of *DNMT1* has been observed in multiple cancer tissues [33–36].

Here, with the aim of identifying potential differences among the methylation machineries of vertebrates, we study the molecular evolution of *DNMT1* in 79 vertebrate species. Our analyses reveal that all studied species have a single *DNMT1* copy, with the only exception of tilapia and marsupials (tammar wallaby, koala, Tasmanian devil and opossum), each of which exhibit two putatively functional *DNMT1* copies. Our phylogenetic analyses indicate that *DNMT1* duplicated before the radiation of major marsupial groups (at least ~75 million years ago), thus giving rise to two *DNMT1* copies (copies 1 and 2) in marsupials. Copy 2 was subsequently lost in the opossum lineage, whereas copy 1 recently duplicated again twice in the opossum lineage, to generate three genes in this species: two putatively functional ones (copies 1a and 1b) and one pseudogene (copy 1ψ). Both marsupial copies (*DNMT1* copies 1 and 2) are under purifying selection, and copy 2 displays signatures of positive selection, suggesting a scenario



**Fig 1. Structure of Dnmt1 proteins in human and in marsupials.** The human Dnmt1s isoform is represented. Sites under positive selection specific to one of the sequences are represented in black. Sites under positive selection shared across multiple sequences (due to positive selection in an internal branch) are represented in green, and their coordinates are only indicated for the last sequence. Amino acid coordinates refer to the human protein. Dashed lines represent missing parts. DMAP, DNA methyltransferase associated protein-binding domain; PCNA, proliferating cell nuclear antigen-binding domain; NLS, nuclear localization signal; RFTS, replication foci targeting sequence; CXXC, cysteine-rich DNA binding domain; BAH, bromo-adjacent homology domains.

<https://doi.org/10.1371/journal.pone.0195162.g001>

of neofunctionalization. We discuss how the presence of two *DNMT1* copies in marsupials might have affected their methylome.

## Results

### *DNMT1* duplicated in a marsupial ancestor and one of the resulting copies further duplicated in an opossum ancestor

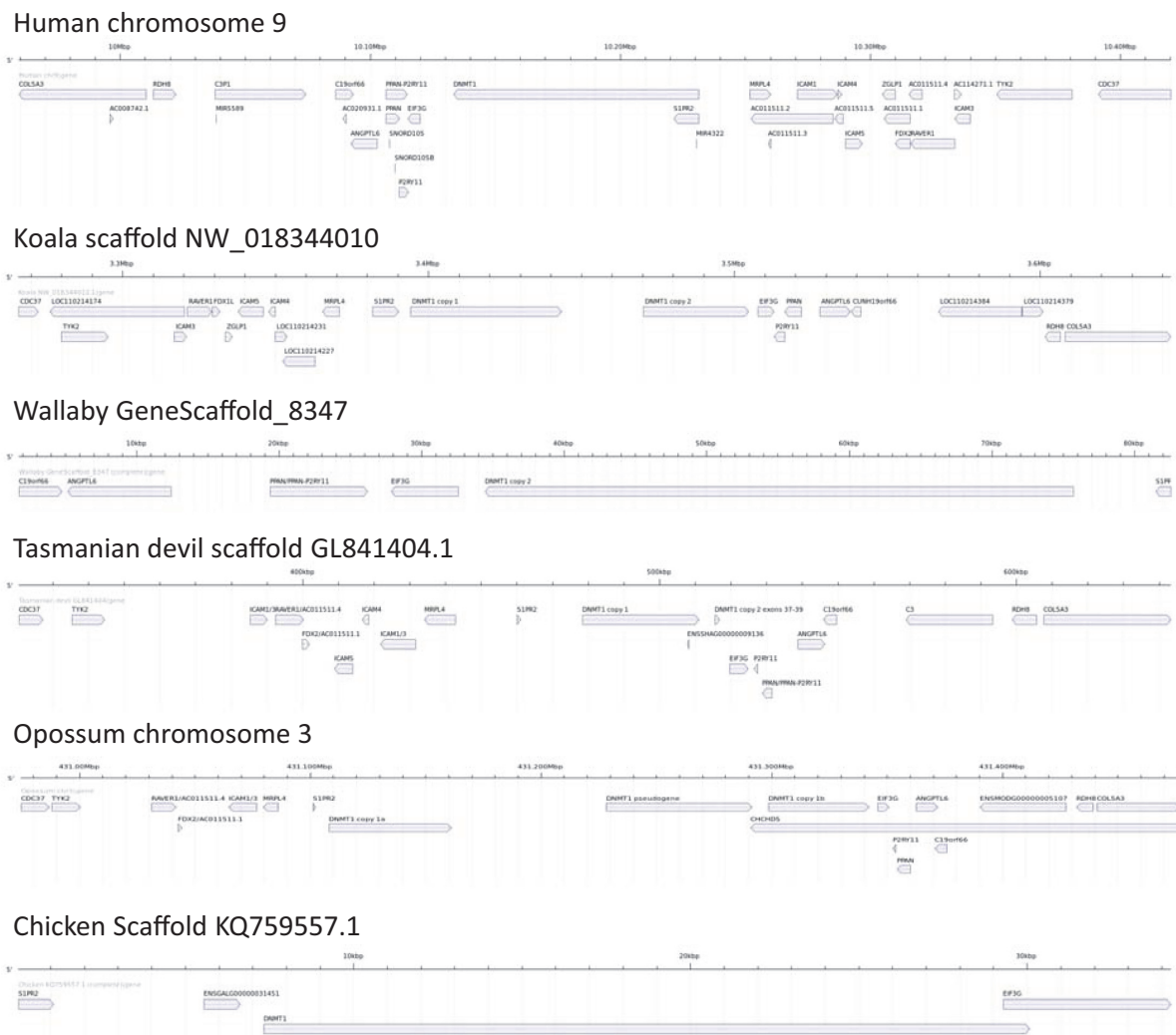
We searched the complete genomes of 58 mammals, 5 birds, 2 reptiles, one amphibian and 13 fish for orthologs of the human *DNMT1* gene. The studied mammalian species included 53 eutherians, four marsupials (tammar wallaby [37], koala [38], Tasmanian devil [39] and opossum [40]) and one monotreme (platypus [41]) (S1 Table). All studied genomes exhibit a single *DNMT1* copy, with the exception of tilapia and the four marsupials, each of which displays two putatively functional copies. In addition, 8 of the studied genomes (dog, lesser Egyptian jerboa, marmoset, opossum, alpaca, hyrax, mouse lemur, and Northern American deer mouse) contain pseudogenes maintaining homology to a substantial length of human *DNMT1*.

According to the annotations of the Ensembl database [42], the tilapia genome contains two *DNMT1* copies (Ensembl gene IDs: ENSONIG00000001574 and ENSONIG00000007221). The first copy encodes a full Dnmt1 protein (1505 amino acids). The second copy is located in a very small scaffold (AERX01074151.1, 3084 nucleotides), which only covers exons 36–40 (184 amino acids; throughout this manuscript, exons for non-human *DNMT1* genes are numbered based on the homologous exons in the human *DNMT1*, using the transcript encoding the Dnmt1s isoform). These exons are identical between both copies, but many differences (single-point mutations and indels) are observed in the introns. These observations indicate a very recent duplication of *DNMT1* in tilapia, but the fact that only a small portion of one of the copies is available prevents further analysis. Thus, the evidence cannot exclude the possibility that one of the copies is a pseudogene, or in the process of pseudogenization.

Some of the *DNMT1* copies identified were unannotated, or their exon/intron structure was incorrectly annotated in the Ensembl [42] and nr databases. Where necessary, marsupial and platypus sequences were re-annotated manually using the human *DNMT1* as reference (see Methods), and incomplete sequences (due to their location in partially sequenced genomic regions) were completed using available RNA-seq data [43–45]. The resulting protein sequences are shown in S1 and S2 Figs.

In the case of opossum, the three *DNMT1* copies (two putatively functional genes and one pseudogene) are located in tandem in chromosome 3 (Fig 2), suggesting two recent duplication events. The two koala sequences are also located in the same scaffold (NW\_018344010.1, 26.8 Kb apart; Fig 2). Tasmanian devil's scaffold GL841404.1 contains copy 1 and part (exons 37–39) of the copy 2, 4.6 Kb apart; the other exons of the second copy are located in another two scaffolds (GL841374.1 contains exons 16–24 and GL843446.1 contains exons 25–36), most likely due to assembly errors (see Methods; S3 Fig). The wallaby copies are located on different scaffolds (copy 1 is located in GeneScaffold\_10206 and copy 2 in GeneScaffold\_8347); however, these scaffolds are small (45.9 and 90.7 Kb, respectively; Fig 2), and therefore we cannot discard the possibility that both wallaby copies are also closely linked.

A high degree of synteny was observed when comparing the genomic regions surrounding *DNMT1*s in human, koala, opossum and chicken (Fig 2). Tasmanian devil's scaffold GL841404.1 and wallaby's GeneScaffold\_8347 also exhibit a similar gene order (Fig 2). In contrast, Tasmanian devil's scaffold GL841374.1 is devoid of such syntenic structure (S3 Fig). Due to their small size, Tasmanian devil's scaffold GL843446.1, wallaby's GeneScaffold\_10206, and platypus' Contig12710 and Contig19880 only contain one *DNMT1* copy (or part of a

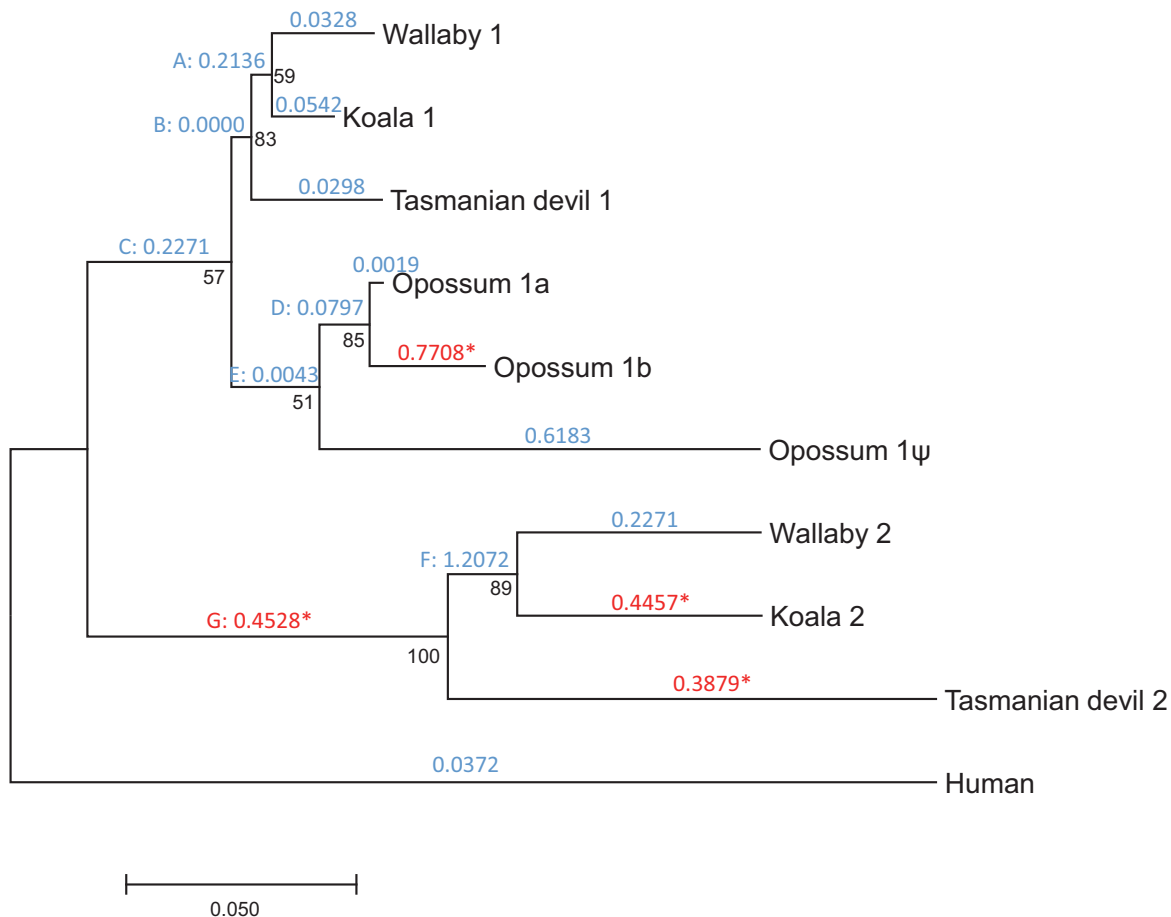


**Fig 2. Synteny analysis of the genomic regions including *DNMT1* copies in human, koala, wallaby, Tasmanian devil, opossum and chicken.** Wallaby's GeneScaffold\_10206, Tasmanian devil's scaffold GL843446.1 and platypus' Contig12710 and Contig19880 are not shown, as they only contain a single *DNMT1* gene, or part of the gene. For unnamed non-human genes, the name of the human ortholog (according to Ensembl's annotations) is shown. Gene coordinates were extracted from the Ensembl database, except for marsupial *DNMT1* genes, for which we used our manually refined annotations. Genome visualizations were generated using GenomeTools [46].

<https://doi.org/10.1371/journal.pone.0195162.g002>

*DNMT1* copy; see [Methods](#)), and thus synteny could not be assessed in the corresponding genomic regions.

The three opossum copies display high sequence similarity (copy 1a vs. copy 1b:  $d_N = 0.022$ ;  $d_S = 0.047$ ; copy 1a vs. copy 1ψ:  $d_N = 0.081$ ;  $d_S = 0.201$ ; copy 1b vs. copy 1ψ:  $d_N = 0.091$ ;  $d_S = 0.205$ ; measures of divergence calculated using the Nei-Gojobori method [47] and the Jukes-Cantor correction [48] as implemented in DnaSP version 5.10.01 [49]; analyses were restricted to the 826 codons present and available in all sequences), whereas the wallaby, koala and Tasmanian devil copies are much more divergent (wallaby's copy 1 vs. copy 2:  $d_N = 0.114$ ;  $d_S = 0.440$ ; koala's copy 1 vs. copy 2:  $d_N = 0.120$ ;  $d_S = 0.395$ ; Tasmanian devil's copy 1 vs. copy 2:  $d_N = 0.146$ ;  $d_S = 0.521$ ) (S1 and S2 Figs). These observations, combined with the results of our phylogenetic analysis (Fig 3), and the known marsupial phylogeny (among the studied species, wallaby and koala are the most closely related, followed by Tasmanian devil and opossum [50–



**Fig 3. Phylogenetic tree showing the duplication of *DNMT1* in marsupials.** Numbers in black represent bootstrap values. Numbers in blue or red above each branch represent  $d_N/d_S$  values according to the free-ratios model. For branches under positive selection according to the branch-site test,  $d_N/d_S$  ratios are represented in red and are followed by an asterisk. Internal branches are labelled with capitals letters.

<https://doi.org/10.1371/journal.pone.0195162.g003>

52]), suggest a scenario in which: (a) *DNMT1* duplicated in a common ancestor of marsupials, giving rise to copies 1 and 2; (b) copy 2 was lost from the opossum lineage; and (c) copy 1 was recently duplicated twice in the opossum lineage, giving rise to two putatively functional copies and one pseudogene. The relative order of the latter two events is unclear. Based on this inferred scenario, we named the three opossum copies as copy 1a (chromosome 3, positions 431,108,118–431,161,113), copy 1b (positions 431,298,625–431,342,040) and copy 1ψ (pseudogene, positions 431,228,446–431,291,545). Copy 1a was already reported by Ding et al. [53], and the presence of a second copy in opossum was noted by Mikkelsen et al. [40].

All marsupial and monotreme sequences lack exons 7–12, consistent with the opossum sequence reported by Ding et al. [53] (Fig 1). A BLASTP search ( $E$ -value  $< 10^{-3}$ ) against all proteomes available in the Ensembl database failed to find any significant hit in non-eutherians, indicating that these exons, which encode amino acids 201–320, were acquired in eutherians. These amino acids overlap with the following regions: region of interaction with the PRC2/EED-EZH2 complex (amino acids 1–606), region of interaction with Dnmt3b (positions 149–217), NLS (positions 177–205) and homodimerization region (positions 310–502). In addition, koala's copy 2 lacks exons 1–14 (first 347 amino acids), and Tasmanian devil's copy 2 lacks exons 1–15 (amino acids 1–374). Thus, the encoded proteins lack the regions of



interaction with DMAP (positions 18–103), Dnmt3a (positions 1–148), Dnmt3b (positions 149–217), and PCNA (positions 163–174), the NLS (positions 177–205), part of the homodimerization (positions 310–502) and RFTS (positions 331–550) regions, and the region of interaction with the PRC2/EED-EZH2 complex (positions 1–606). Nonetheless, all marsupial and monotreme Dnmt1s appear to include a complete CXXC domain, an autoinhibitory linker, the BAH1 and BAH2 domains, and the catalytic domain (Fig 1), thus being potentially functional. The opossum pseudogene (copy 1 $\psi$ ) lacks exons 1–16, 20, and 30–31, and contains five stop codons (two in exon 21, one in exon 23, one in exon 28, and one in the codon shared between exons 39 and 40) and two frameshift mutations (exons 18 and 26).

## Marsupial *DNMT1* copies are differentially expressed

We next attempted to determine in which tissues, and to what extent, each copy is expressed. First, we searched the transcriptomes of a number of koala tissues [54] for transcripts corresponding to copy 1 and copy 2, finding only transcripts for copy 1. Second, we searched two Tasmanian devil transcriptomic datasets (lymph and spleen) for sequences similar to *DNMT1*, finding only reads for copy 1. Third, we mined RNA-seq data for 5 wallaby tissues (testes, male liver, female liver, male blood and female blood; ref. [44]), and identified 11,267 reads specific to copy 1 and only 5 reads specific to copy 2 (another 735 reads matched both copies; Table 1). Finally, we mined RNA-seq data for 11 opossum tissues (testis and male and female brain, cerebral cortex, heart, kidney and liver; ref. [43]). A total of 3831, 194 and 290 reads matched opossum's copies 1a, 1b and 1 $\psi$ , respectively (Table 2).

## Both marsupial *DNMT1* copies are under purifying selection

We used PAML [55] to estimate the non-synonymous to synonymous divergence ratio ( $d_N/d_S$ ) in each of the branches of the gene tree. We restricted this analysis to human and the four marsupials, as incomplete genomic data and annotation errors in many of the other species would have hindered our analyses. This ratio was substantially below one in all branches of the phylogeny, except in the internal branch leading to the most recent common ancestor (MRCA) of wallaby's and koala's copy 2 (Fig 3). This indicates that nonsynonymous changes are under substantial purifying selection in all the sequences studied, suggesting that all copies are functional, or that they pseudogenized only recently—which is the case for opossum's copy 1 $\psi$  ( $d_N/d_S = 0.618$ ).

The  $d_N/d_S$  ratios varied substantially among the different branches (Fig 3). Indeed, the free-ratios model fit the data significantly better than the one-ratio model M0 ( $2\Delta\ell = 438.07$ ,  $P = 3.71 \times 10^{-83}$ ), indicating significant heterogeneity in the  $d_N/d_S$  ratios. Remarkably,  $d_N/d_S$  was substantially higher in copy 2 than in copy 1 (Fig 3). In addition,  $d_N/d_S$  was 0.0019 in the branch leading to opossum's copy 1a, and 0.7708 in the branch leading to opossum's copy 1b. This increase in the  $d_N/d_S$  ratios of copy 2 (wallaby, koala and Tasmanian devil) and copy 1b

**Table 1. Number of RNA-seq reads matching wallaby's copies 1 and 2.**

Tissue	Run accession number	Copy 1	Copy 2
Male liver	SRR1041778	502	0
Female liver	SRR1552212	1340	1
Male blood	SRR1552202	371	0
Female blood	SRR1552210	233	2
Testis	SRR1041779	8821	2
	Total:	11267	5

<https://doi.org/10.1371/journal.pone.0195162.t001>

**Table 2. Number of RNA-seq reads matching opossum's copies 1a, 1b and 1ψ.**

Tissue	Run accession number	Copy 1a	Copy 1b	Copy 1ψ
Male brain	SRR306744	158	4	11
Male cerebral cortex	SRR306746	279	9	37
Male heart	SRR306750	318	8	14
Male kidney	SRR306752	190	40	40
Male liver	SRR306754	87	1	8
Testis	SRR306756	1739	52	24
Female brain	SRR306743	106	5	11
Female cerebral cortex	SRR306745	265	20	51
Female heart	SRR306748	142	5	3
Female kidney	SRR306751	344	45	83
Female liver	SRR306753	203	5	8
Total:		3831	194	290

<https://doi.org/10.1371/journal.pone.0195162.t002>

(opossum) could be explained by a relaxation of purifying selection acting on protein sequences and/or by positive selection in these copies.

### Marsupials' copy 2 of *DNMT1* is under positive selection

We then used PAML to test for signatures of positive selection. The M8 vs. M7 test was significant ( $2\Delta\ell = 8.36$ ,  $P = 0.015$ ), indicating that a fraction of codons were under positive selection. We then used a branch-site test (model A vs. null model A1; refs. [56,57]) to infer the action of positive selection at each of the branches of the phylogeny, except the branch leading to the opossum pseudogene. The test was significant for the external branches leading to koala's copy 2, Tasmanian devil's copy 2, and opossum's copy 1b, and for the internal branch leading to the MRCA of the copy 2 of wallaby, koala and Tasmanian devil. The  $d_N/d_S$  values for these branches are represented in red and marked with an asterisk in Fig 3, and more detailed results are provided in Table 3.

A total of 9 codons were detected to be under positive selection: one in the opossum's copy 1b, three in koala's copy 2, three in Tasmanian devil's copy 2, and 2 in the internal branch leading to the MRCA of copy 2 of wallaby, koala and Tasmanian devil. Sites under positive selection were different in each branch, and affected the catalytic domain (4 codons), the site of interaction with the PRC2/EED-EZH2 complex (6 codons), the BAH2 domain (2 codons) and the homodimerization domain (1 codon; Fig 1; Table 3).

### Reanalysis removing incomplete sequences

The marsupial *DNMT1* coding sequences (CDSs) used in this study are complete or almost complete (S1 and S2 Figs). The only notable exceptions are wallaby's copy 2, for which 409 codons (in exons 5–6, 13–17 and 19–24) remain unsequenced due to limited genome coverage (2×; ref. [37]), and the opossum pseudogene, which lacks exons 1–14, 20 and 30–31. This means that our natural selection analyses were limited to only 826 codons. We repeated our analysis after removing these sequences, rendering 1172 codons analyzable (present in all sequences). We obtained similar results: First, the  $d_N/d_S$  ratio was substantially higher in copy 2 than in copy 1, and in opossum's copy 1b ( $d_N/d_S = 0.808$ ) than in opossum's copy 1a ( $d_N/d_S = 0.000$ ; Fig 4). Second, positive selection was detected in the external branches leading to opossum's copy 1b, koala's copy 2 and Tasmanian devil's copy 2, and in the internal branch leading to the MRCA of koala's copy 2 and Tasmanian devil's copy 2 (Fig 4; S2 Table). This



**Table 3. Branch-site tests of positive selection.**

Branch <sup>a</sup>	Log-likelihood model A	Log-likelihood null model A1	2Δℓ	P-value	ω <sub>s</sub> <sup>b</sup>	Selected codons <sup>c</sup>
Wallaby 1	−10,537.29	−10,537.29	5.2×10 <sup>−4</sup>	0.497	1.000	
Koala 1	−10,537.29	−10,537.29	0.00	0.500	1.000	
Tasmanian devil 1	−10,537.29	−10,537.29	0.00	0.500	1.000	
Opossum 1a	−10,537.29	−10,537.29	0.00	0.500	1.000	
Opossum 1b	−10,527.28	−10,531.13	7.70	0.003*	11.328	V513W
Wallaby 2	−10,535.71	−10,535.71	0.00	0.500	1.000	
Koala 2	−10,526.31	−10,527.99	3.35	0.034*	2.960	T467G, S1342A, G1449N
Tasmanian devil 2	−10,517.81	−10,523.87	12.12	2.5×10 <sup>−4***</sup>	5.101	E906N, S1076N, A1338S
Human	−10,537.12	−10,537.29	0.34	0.279	43.846	
A	−10,537.29	−10,537.29	0.00	0.500	1.000	
B	−10,537.29	−10,537.29	0.00	0.500	1.000	
C	−10,537.13	−10,537.13	0.00	0.500	1.000	
D	−10,537.29	−10,537.29	0.00	0.500	1.000	
E	−10,537.13	−10,537.13	0.00	0.500	1.000	
F	−10,532.75	−10,533.68	1.85	0.087	150.054	
G	−10,526.76	−10,529.87	6.22	0.006*	3.717	S1034, L1384E

<sup>a</sup>Internal branches are represented with letters as in Fig 3.

<sup>b</sup> $d_N/d_S$  for the class of codons under positive selection.

<sup>c</sup>For each mutation, the first letter and the number correspond to the amino acid in human Dnmt1s, and the last letter corresponds to the mutation observed in the sequence(s) of interest. For the mutation S1034, the final amino acid is not provided because it is not the same in all the descendants of branch G.

\*,  $P < 0.05$

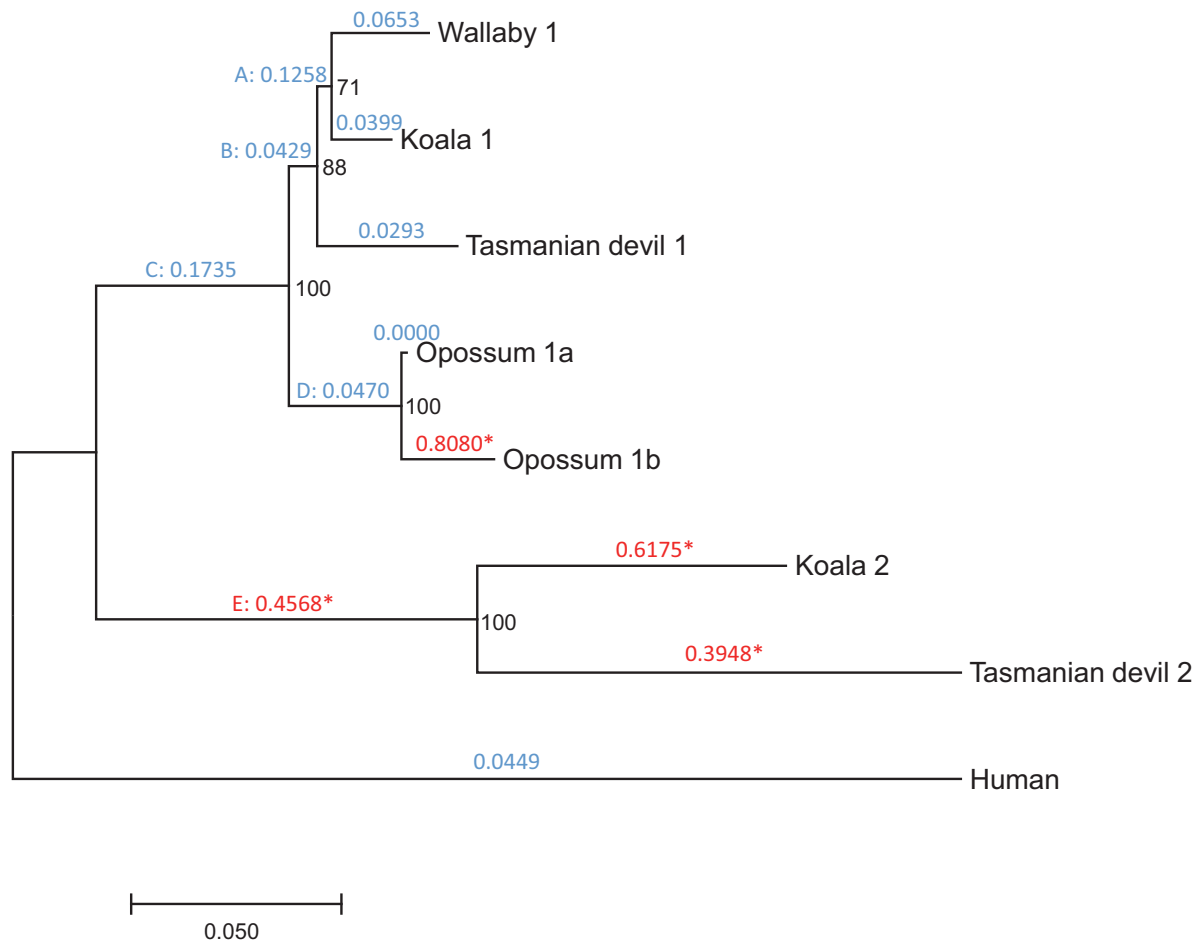
\*\*\*,  $P < 0.001$ .

<https://doi.org/10.1371/journal.pone.0195162.t003>

analysis detected a total of 21 codons under positive selection (including the 9 ones detected before), which affected the catalytic domain (4 codons), the site of interaction with the PRC2/EED-EZH2 complex (10 codons), the BAH2 domain (3 codons), the homodimerization domain (2 codons), the autoinhibitory linker (1 codon), and the KG linker (1 codon; Fig 1; S2 Table).

## Reanalysis using 44 outgroup sequences

We repeated our tests of positive selection using as outgroup not only human, but a total of 44 species, including 33 eutherians (bushbaby, cat, chimpanzee, Chinese hamster, cow, degu, elephant, ferret, gibbon, golden hamster, gorilla, guinea pig, horse, human, kangaroo rat, long-tailed chinchilla, macaque, marmoset, microbat, mouse, mouse lemur, naked mole-rat, Northern American deer mouse, panda, pig, prairie vole, rat, Ryukyu mouse, sheep, shrew mouse, tarsier, Upper Galilee Mountains blind mole rat and vervet), platypus, anole lizard, *Xenopus*, and 8 fish (Amazon molly, fugu, medaka, platyfish, spotted gar, stickleback, tetraodon, and zebrafish). These were the species for which exons 16–39 of *DNMT1* (the ones included in our positive selection analyses) were correctly annotated in the Ensembl database. In agreement with our results using only human as outgroup (Fig 3; Table 3), signatures of positive selection were detected in the external branches leading to koala's copy 2, Tasmanian devil's copy 2, and opossum's copy 1b, and in the internal branch leading to the MRCA of the copy 2 of wallaby, koala and Tasmanian devil. In addition, positive selection was detected in the branch leading to the MRCA of the copy 2 of wallaby and koala. A total of 59 codons were detected to be under positive selection (S3 Table).



**Fig 4. Phylogenetic tree showing the duplication of *DNMT1* in marsupials, removing wallaby's copy 2 and opossum's copy 1 $\psi$ .** Numbers in black represent bootstrap values. Numbers in blue or red above each branch represent  $d_N/d_S$  values according to the free-ratios model. For branches under positive selection according to the branch-site test,  $d_N/d_S$  ratios are represented in red and are followed by an asterisk. Internal branches are labelled with capitals letters.

<https://doi.org/10.1371/journal.pone.0195162.g004>

## Discussion

Our analyses indicate that the *DNMT1* gene duplicated in a common ancestor of marsupials, giving rise to two copies (copies 1 and 2). The opossum lineage and the wallaby/koala/Tasmanian devil lineage diverged ~75 million years ago [50,51], implying that the *DNMT1* duplication occurred prior to that time. Copy 2 was subsequently lost in the opossum lineage. Copy 2 is expressed at very low, or even undetectable levels, at least in the wide range of wallaby (Table 1), koala [54] and Tasmanian devil tissues examined. However, both copies exhibit  $d_N/d_S$  ratios lower than one (Figs 3 and 4), and none display signatures of pseudogenization (premature stop codons or frameshift mutations) indicating that they are likely expressed—perhaps in tissues not included in our analyses, in early developmental stages or under certain environmental conditions—and functional. Otherwise, signatures of pseudogenization and a  $d_N/d_S$  close to 1 would be expected. Part of the regulatory region of koala's and Tasmanian devil's copy 2 appear to have been lost; however, all *DNMT1* copies retain the catalytic domain and a significant fraction of the regulatory region, suggesting that they are functional—of note, the human *Dnmt1o* isoform is functional despite also lacking part of the regulatory region.

Remarkably, copy 2 exhibits a high  $d_N/d_S$  ratio compared to copy 1, in addition to signatures of positive selection. These results suggest a scenario of neofunctionalization, in which copy 1 may have retained the function of the ancestral *DNMT1*, and copy 2 may have acquired a new or modified function. Signatures of positive selection can be detected in the branch leading to the MRCA of wallaby's, koala's, and Tasmanian devil's copy 2, and in the external branches leading to koala's and Tasmanian devil's copy 2 (Figs 3 and 4; Table 3; S2 and S3 Tables). These observations indicate that neofunctionalization occurred both before and after the divergence of wallaby, koala and Tasmanian devil (i.e., both before and after ~60 million years ago; refs. [50,51]). Substitutions under positive selection affect different domains, making it difficult to predict how they may have affected the function of copy 2.

Copy 1 recently underwent another two duplication events in the opossum lineage, which resulted in three genes (copies 1a, 1b and the pseudogene 1ψ) located in tandem in chromosome 3. Their high degree of similarity, along with our phylogenetic analyses (Fig 3), indicate that these sequences are the result of the duplication of the copy 1 of *DNMT1*, and that they are not remnants of the ancestral duplication identified in the other marsupials. Opossum's copy 1b also displays an elevated  $d_N/d_S$  (compared to copy 1a) and signatures of positive selection, which would also suggest neofunctionalization in the copy 1b. However, in this case we are skeptical about our inference of positive selection, because the only codon inferred to be under positive selection with high probability (V513 in the human protein, a tryptophan in opossum's copy 1b) is located near an unsequenced region of the opossum genome (S1 Fig), and such regions are prone to sequencing errors. Opossum's copy 1b is expressed at lower levels than copy 1a in the tissues included in our analyses (Table 2).

It is currently not possible to infer the functions of marsupial *DNMT1* derived duplicates (copy 2 of wallaby, koala and Tasmanian devil and copy 1b of opossum). We propose three different possible scenarios. First, as both marsupial *DNMT1* copies seem to be expressed in different sets of tissues (Tables 1 and 2), positive selection in the derived *DNMT1* copies may simply reflect subtle adjustments to the biochemistry of the tissue or tissues in which they are expressed. Second, assuming that the function of both marsupial *DNMT1* copies is similar to that of the ancestral *DNMT1*—maintenance of methylation patterns throughout the life of the animal after each DNA replication event—it is possible that an increased *Dnmt1* abundance may cause marsupial methylomes to be particularly stable during aging—in other mammals methylation patterns change during the lifespan of an organism [58]. This, however would only apply to the unknown tissue or tissues (or developmental stages or environmental conditions) in which the derived copies are expressed at substantial levels. Third, the duplication of *DNMT1* may have caused marsupial genomes to be hypermethylated. Given that methylated cytosines have an increased mutation rate [59], this scenario might explain the low GC content of marsupial genomes [37,39,40,60]. However, this scenario would require that the derived *DNMT1* copies would act as *de novo* Dnmts rather than maintenance Dnmts, which is at odds with the presence of an autoinhibitory linker in the proteins encoded by both copies. Additional functional studies of marsupial *Dnmt1*s, and methylome data for Australian marsupials—which are currently unavailable—will be required to establish their functions.

## Conclusions

Our analyses of 79 vertebrate genomes reveal that all studied species exhibit a single *DNMT1* gene, with the exception of tilapia and marsupials (wallaby, koala, Tasmanian devil and opossum), each of which display two apparently functional *DNMT1* copies. Our phylogenetic analyses indicate that *DNMT1* duplicated before the radiation of major marsupial groups (at least ~75 million years ago), thus giving rise to *DNMT1* copies 1 and 2. Copy 2 was lost in the

opossum lineage, and copy 1 recently duplicated again to generate three opossum genes: two putatively functional ones and one pseudogene. Both *DNMT1* copies are under purifying selection, and copy 2 is under positive selection. These results suggest a scenario of neofunctionalization.

## Methods

### Gene identification and annotation

In order to identify *DNMT1* orthologs in the studied vertebrate genomes, we conducted TBLASTN searches against the Ensembl database (release 90; ref. [42]), using the human *Dnmt1*s protein sequence as query and an *E*-value cut-off of  $10^{-10}$  and all other parameters set as default. The koala genome was queried in the nr database, as it is not represented in Ensembl. Only scaffolds with at least 450 identities (added across the different TBLASTN hits) were considered. Pseudogenes were inferred from the presence of premature stop codons.

Where necessary, wallaby, koala, Tasmanian devil, opossum and platypus sequences were manually re-annotated using the intron/exon structure of human *DNMT1* as reference. For that purpose, incorrectly annotated exons (those not showing significant similarity to the human sequence) were removed, and missing exons were searched for using TBLASTN and BLASTN searches. Putative stop codons and frameshift mutations were confirmed by visualization of the corresponding original reads in the trace archive database.

In the case of Tasmanian devil's *DNMT1* copy 2 and platypus' *DNMT1*, exons present on different scaffolds were combined into a single gene annotation. The platypus *DNMT1* exons are distributed along two small contigs: Contig19880 (17.7 Kb; exons 1–13) and Contig12710 (18.1 Kb; exons 19–25 and 27–38) (S1 Table). In the current Tasmanian devil assembly, the exons of copy 2 are distributed across three different scaffolds: exons 16–24 are located in scaffold GL841374.1 (4.0 Mb), exons 25–36 are located in GL843446.1 (17.2 Kb), and exons 37–39 are located in GL841404.1 (1.6 Mb); this is probably the result of assembly errors.

Some of the exons of wallaby's copies 1 and 2, opossum's copy 1b, and the single copy of platypus, could not be recovered (or completely recovered) from available genome assemblies because they were located in unsequenced regions. We thus attempted to recover these exons from available RNA-seq datasets [43–45]. For each unsequenced exon, we retrieved the sequence of the end of the prior exon or the beginning of the next exon and searched for RNA-seq reads that exactly matched these sequences. In the case of wallaby's copy 2, this was not possible due to the very few reads available (Table 1), and in the case of opossum's copy 1b it was not possible either due to the high similarity between copies 1a and 1b.

### Gene expression levels in different tissues

We used koala's copy 2 as query in a TBLASTN search against the koala transcriptome [54]; all retrieved copies, however, corresponded to copy 1. Similarly, we used Tasmanian devil's copy 2 as query in a TBLASTN search against all the RNA-seq reads available for two Tasmanian devil tissues (lymph and spleen; SRA accession numbers: ERR695583 and ERR695584), finding again only reads corresponding to copy 1. Default parameters were used in TBLASTN searches.

We next mined RNA-seq datasets for a number of tissues of wallaby [44] and opossum [43], in order to measure expression levels of each of the *DNMT1* copies in the different tissues. For each read, it was determined whether it perfectly matched (it was contained in) one or more of the copies in the genome of interest, using an in-house PERL script. Reads that matched more than one copy were not used to compute expression levels.

## Phylogenetic analyses

The CDSs of human, wallaby, koala, Tasmanian devil, opossum and platypus were translated in silico into protein sequences. The protein sequences were aligned using ProbCons version 1.12 [61], and the resulting sequences were used to guide the alignment of the CDSs. Alignments were visualized and, where necessary, manually edited using BioEdit version 7.2.5 [62]. A phylogenetic tree was obtained using the maximum-likelihood method implemented in MEGA7 [63], using the Tamura-Nei model [64] and 1000 bootstraps.

## Natural selection analyses

The codeml program in the PAML package, version 4.4d [55] was used to conduct natural selection analyses. The free-ratios model was used to calculate a separate  $d_N/d_S$  for each of the branches of the gene tree. Heterogeneity of  $d_N/d_S$  among branches was tested by comparing the likelihoods of the free-ratios model and model 0, which assumes a homogeneous  $d_N/d_S$  across all sites and branches. This comparison was conducted using a likelihood ratio test [65], assuming that twice the difference between the log-likelihoods of both models,  $2\Delta\ell = 2 \times (\ell_{FR} - \ell_{M0})$ , where  $\ell_i$  is the log-likelihood of model  $i$ , followed a chi-squared distribution with a number of degrees of freedom equivalent to the difference between the number of parameters of both nested models.

To infer the presence of codons under positive selection, we first compared the likelihoods of models M8 and M7. Positive selection was inferred if model M8 (which allows for a class of codons with  $d_N/d_S > 1$ ) fitted the data significantly better than model M7 (which allows  $d_N/d_S$  to vary between 0 and 1). The statistic  $2\Delta\ell = 2 \times (\ell_{M8} - \ell_{M7})$  was assumed to follow a chi-squared distribution with two degrees of freedom. Next, for each of the branches in the gene tree, a branch-site test of positive selection (Test 2; refs. [56,57]) was conducted. Positive selection was inferred if model A fitted the data significantly better than null model A1. The statistic  $2\Delta\ell = 2 \times (\ell_{MA} - \ell_{MA1})$  was assumed to follow a 50%:50% mixture of a point of mass 0 and a chi-squared distribution with one degree of freedom. The Bayes Empirical Bayes approach [56] was used to identify codons under positive selection (posterior probability  $\geq 95\%$ ).

## Supporting information

### S1 Fig. Alignment for the N-terminal part of Dnmt1 in human, marsupials and platypus.

The human sequence corresponds to the Dnmt1s isoform. Dashes represent alignment gaps or missing regions. Stretches of “X” symbols represent unsequenced regions. Single “X” symbols represent incomplete codons (e.g., due to frameshift mutations).

(PDF)

### S2 Fig. Alignment for the C-terminal part of Dnmt1 in human, marsupials and platypus.

The human sequence corresponds to the Dnmt1s isoform. Dashes represent alignment gaps or missing regions. Stretches of “X” symbols represent unsequenced regions. Single “X” symbols represent incomplete codons (e.g., due to frameshift mutations).

(PDF)

### S3 Fig. Tasmanian devil’s scaffold GL841374.1. The scaffold contains part of *DNMT1* copy 2.

For unnamed non-human genes, the name of the human ortholog (according to Ensembl’s annotations) is shown. Gene coordinates were extracted from the Ensembl database, except for *DNMT1* copy 2, for which we used our manually refined annotations. Genome visualizations were generated using GenomeTools [46].

(PDF)

**S1 Table. DNMT1 copies in vertebrate genomes.**  
(XLSX)

**S2 Table. Branch-site tests of positive selection excluding wallaby's copy 2 and opossum's copy 1ψ.**  
(XLSX)

**S3 Table. Branch-site tests of positive selection using 44 outgroup species.**  
(XLSX)

## Acknowledgments

The authors are grateful to Soojin Yi, Paul Waters and Julio Rozas for helpful feedback. Computational Resources were provided by Information Technology Operations of the University of Nevada, Reno.

## Author Contributions

**Conceptualization:** David Alvarez-Ponce.

**Data curation:** David Alvarez-Ponce, María Torres-Sánchez, Felix Feyertag, Asmita Kulkarni, Taylen Nappi.

**Formal analysis:** David Alvarez-Ponce, María Torres-Sánchez.

**Funding acquisition:** David Alvarez-Ponce.

**Investigation:** David Alvarez-Ponce, María Torres-Sánchez.

**Methodology:** David Alvarez-Ponce, María Torres-Sánchez, Felix Feyertag.

**Project administration:** David Alvarez-Ponce.

**Resources:** David Alvarez-Ponce.

**Software:** David Alvarez-Ponce.

**Supervision:** David Alvarez-Ponce.

**Validation:** David Alvarez-Ponce.

**Visualization:** David Alvarez-Ponce.

**Writing – original draft:** David Alvarez-Ponce.

**Writing – review & editing:** David Alvarez-Ponce, María Torres-Sánchez, Felix Feyertag, Asmita Kulkarni, Taylen Nappi.

## References

1. Bird AP. CpG-rich islands and the function of DNA methylation. *Nature*. 1986; 321: 209–213. <https://doi.org/10.1038/321209a0> PMID: 2423876
2. Hodges E, Smith AD, Kendall J, Xuan Z, Ravi K, Rooks M, et al. High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. *Genome Res*. 2009; 19: 1593–1605. <https://doi.org/10.1101/gr.095190.109> PMID: 19581485
3. Lee JT. Molecular links between X-inactivation and autosomal imprinting: X-inactivation as a driving force for the evolution of imprinting? *Curr Biol*. 2003; 13: R242–254. PMID: 12646153
4. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*. 1999; 99: 247–257. PMID: 10555141
5. Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet*. 1998; 19: 219–220. <https://doi.org/10.1038/890> PMID: 9662389



6. Leonhardt H, Page AW, Weier HU, Bestor TH. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell*. 1992; 71: 865–873. PMID: [1423634](#)
7. Jeltsch A, Nellen W, Lyko F. Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. *Trends Biochem Sci*. 2006; 31: 306–308. <https://doi.org/10.1016/j.tibs.2006.04.005> PMID: [16679017](#)
8. Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X, et al. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science*. 2006; 311: 395–398. <https://doi.org/10.1126/science.1120976> PMID: [16424344](#)
9. Borkovich KA, Alex LA, Yarden O, Freitag M, Turner GE, Read ND, et al. Lessons from the genome sequence of *Neurospora crassa*: tracing the path from genomic blueprint to multicellular organism. *Microbiol Mol Biol Rev*. 2004; 68: 1–108. <https://doi.org/10.1128/MMBR.68.1.1-108.2004> PMID: [15007097](#)
10. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, et al. Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci U S A*. 2010; 107: 8689–8694. <https://doi.org/10.1073/pnas.1002720107> PMID: [20395551](#)
11. Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*. 2010; 328: 916–919. <https://doi.org/10.1126/science.1186366> PMID: [20395474](#)
12. Jeltsch A. Molecular biology. Phylogeny of methylomes. *Science*. 2010; 328: 837–838. <https://doi.org/10.1126/science.1190738> PMID: [20466912](#)
13. Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ. Evolution of DNA methylation across insects. *Mol Biol Evol*. 2017; 34: 654–665. <https://doi.org/10.1093/molbev/msw264> PMID: [28025279](#)
14. Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem*. 2005; 74: 481–514. <https://doi.org/10.1146/annurev.biochem.74.010904.153721> PMID: [15952895](#)
15. Ponger L, Li WH. Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Mol Biol Evol*. 2005; 22: 1119–1128. <https://doi.org/10.1093/molbev/msi098> PMID: [15689527](#)
16. Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Hérault Y, et al. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science*. 2016; 354: 909–912. <https://doi.org/10.1126/science.aah5143> PMID: [27856912](#)
17. Dunwell TL, Pfeifer GP. *Drosophila* genomic methylation: new evidence and new questions. *Epigenomics*. 2014; 6: 459–461. <https://doi.org/10.2217/epi.14.46> PMID: [25431937](#)
18. Jurkowski TP, Jeltsch A. On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. *PloS One*. 2011; 6: e28104. <https://doi.org/10.1371/journal.pone.0028104> PMID: [22140515](#)
19. Peat JR, Ortega-Recalde O, Kardailsky O, Hore TA. The elephant shark methylome reveals conservation of epigenetic regulation across jawed vertebrates. *F1000Research*. 2017; 6: 526.
20. Hendrich B, Tweedie S. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet*. 2003; 19: 269–277. [https://doi.org/10.1016/S0168-9525\(03\)00080-5](https://doi.org/10.1016/S0168-9525(03)00080-5) PMID: [12711219](#)
21. Zhang Z, Liu G, Zhou Y, Lloyd JP, McCauley DW, Li W, et al. Genome-wide and single-base resolution DNA methylomes of the Sea Lamprey (*Petromyzon marinus*) Reveal Gradual Transition of the Genomic Methylation Pattern in Early Vertebrates. *BioRxiv*. 2015: 033233.
22. Lock LF, Takagi N, Martin GR. Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. *Cell*. 1987; 48: 39–46. PMID: [3791414](#)
23. Norris DP, Brockdorff N, Rastan S. Methylation status of CpG-rich islands on active and inactive mouse X chromosomes. *Mamm Genome*. 1991; 1: 78–83. PMID: [1799791](#)
24. Waters SA, Livernois AM, Patel H, O'Meally D, Craig JM, Marshall Graves JA, et al. Landscape of DNA methylation on the marsupial X. *Mol Biol Evol*. 2017; 35: 431–439.
25. Rens W, Wallduck MS, Lovell FL, Ferguson-Smith MA, Ferguson-Smith AC. Epigenetic modifications on X chromosomes in marsupial and monotreme mammals and implications for evolution of dosage compensation. *Proc Natl Acad Sci U S A*. 2010; 107: 17657–17662. <https://doi.org/10.1073/pnas.0910322107> PMID: [20861449](#)
26. Yokomine T, Hata K, Tsudzuki M, Sasaki H. Evolution of the vertebrate DNMT3 gene family: a possible link between existence of DNMT3L and genomic imprinting. *Cytogenet Genome Res*. 2006; 113: 75–80. <https://doi.org/10.1159/000090817> PMID: [16575165](#)
27. Campos C, Valente LM, Fernandes JM. Molecular evolution of zebrafish dnmt3 genes and thermal plasticity of their expression during embryonic development. *Gene*. 2012; 500: 93–100. <https://doi.org/10.1016/j.gene.2012.03.041> PMID: [22450363](#)
28. Qin W, Leonhardt H, Pichler G. Regulation of DNA methyltransferase 1 by interactions and modifications. *Nucleus*. 2011; 2: 392–402. <https://doi.org/10.4161/nucl.2.5.17928> PMID: [21989236](#)

29. Fatemi M, Hermann A, Pradhan S, Jeltsch A. The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. *J Mol Biol.* 2001; 309: 1189–1199. <https://doi.org/10.1006/jmbi.2001.4709> PMID: 11399088
30. Gruenbaum Y, Cedar H, Razin A. Substrate and sequence specificity of a eukaryotic DNA methylase. *Nature.* 1982; 295: 620–622. PMID: 7057921
31. Bestor TH, Ingram VM. Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proc Natl Acad Sci U S A.* 1983; 80: 5559–5563. PMID: 6577443
32. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell.* 1992; 69: 915–926. PMID: 1606615
33. el-Deiry WS, Nelkin BD, Celano P, Yen RW, Falco JP, Hamilton SR, et al. High expression of the DNA methyltransferase gene characterizes human neoplastic cells and progression stages of colon cancer. *Proc Natl Acad Sci U S A.* 1991; 88: 3470–3474. PMID: 2014266
34. Oh BK, Kim H, Park HJ, Shim YH, Choi J, Park C, et al. DNA methyltransferase expression and DNA methylation in human hepatocellular carcinoma and their clinicopathological correlation. *Int J Mol Med.* 2007; 20: 65–73. PMID: 17549390
35. Robertson KD, Uzvolgyi E, Liang G, Talmadge C, Sumegi J, Gonzales FA, et al. The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. *Nucleic Acids Res.* 1999; 27: 2291–2298. PMID: 10325416
36. Hermann A, Gowher H, Jeltsch A. Biochemistry and biology of mammalian DNA methyltransferases. *Cell Mol Life Sci.* 2004; 61: 2571–2587. <https://doi.org/10.1007/s00018-004-4201-1> PMID: 15526163
37. Renfree MB, Papenfuss AT, Deakin JE, Lindsay J, Heider T, Belov K, et al. Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. *Genome Biol.* 2011; 12: R81. <https://doi.org/10.1186/gb-2011-12-8-r81> PMID: 21854559
38. Johnson RN, Hobbs M, Eldridge MD, King AG, Colgan DJ, Wilkins MR, et al. The koala genome consortium. *Tech Rep Aust Mus.* 2014; 24: 91–92.
39. Murchison EP, Schulz-Trieglaff OB, Ning Z, Alexandrov LB, Bauer MJ, Fu B, et al. Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. *Cell.* 2012; 148: 780–791. <https://doi.org/10.1016/j.cell.2011.11.065> PMID: 22341448
40. Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, et al. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature.* 2007; 447: 167–177. <https://doi.org/10.1038/nature05805> PMID: 17495919
41. Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, Grützner F, et al. Genome analysis of the platypus reveals unique signatures of evolution. *Nature.* 2008; 453: 175–183. <https://doi.org/10.1038/nature06936> PMID: 18464734
42. Aken BL, Achuthan P, Akanni W, Amode MR, Bernsdröff F, Bhai J, et al. Ensembl 2017. *Nucleic Acids Res.* 2017; 45: D635–D642. <https://doi.org/10.1093/nar/gkw1104> PMID: 27899575
43. Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, et al. The evolution of gene expression levels in mammalian organs. *Nature.* 2011; 478: 343–348. <https://doi.org/10.1038/nature10532> PMID: 22012392
44. Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, et al. Origins and functional evolution of Y chromosomes across mammals. *Nature.* 2014; 508: 488–493. <https://doi.org/10.1038/nature13151> PMID: 24759410
45. Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, et al. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature.* 2014; 505: 635–640. <https://doi.org/10.1038/nature12943> PMID: 24463510
46. Gremme G, Steinbiss S, Kurtz S. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans Comput Biol Bioinform.* 2013; 10: 645–656. <https://doi.org/10.1109/TCBB.2013.68> PMID: 24091398
47. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* 1986; 3: 418–426. <https://doi.org/10.1093/oxfordjournals.molbev.a040410> PMID: 3444411
48. Jukes TH, Cantor CR, Munro H. Evolution of protein molecules. In: Munro HN, editor. *Mammalian protein metabolism.* Academic Press, New York. 1969. 21–132.
49. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 2009; 25: 1451–1452. <https://doi.org/10.1093/bioinformatics/btp187> PMID: 19346325

50. Meredith RW, Westerman M, Case JA, Springer MS. A phylogeny and timescale for marsupial evolution based on sequences for five nuclear genes. *J Mammal Evol.* 2008; 15: 1–36.
51. Meredith RW, Westerman M, Springer MS. A phylogeny of Diprotodontia (Marsupialia) based on sequences for five nuclear genes. *Mol Phylogenet Evol.* 2009; 51: 554–571. <https://doi.org/10.1016/j.ympev.2009.02.009> PMID: 19249373
52. Duchêne DA, Bragg JG, Duchêne S, Neaves LE, Potter S, Moritz C, et al. Analysis of Phylogenomic Tree Space Resolves Relationships Among Marsupial Families. *Syst Biol.* 2017; syx076.
53. Ding F, Patel C, Ratnam S, McCarrey JR, Chaillet JR. Conservation of Dnmt1 to cytosine methyltransferase in the marsupial *Monodelphis domestica*. *Genesis.* 2003; 36: 209–213. <https://doi.org/10.1002/gene.10215> PMID: 12929092
54. Hobbs M, Pavasovic A, King AG, Prentis PJ, Eldridge MD, Chen Z, et al. A transcriptome resource for the koala (*Phascolarctos cinereus*): insights into koala retrovirus transcription and sequence diversity. *BMC genomics.* 2014; 15: 786. <https://doi.org/10.1186/1471-2164-15-786> PMID: 25214207
55. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007; 24: 1586–1591. <https://doi.org/10.1093/molbev/msm088> PMID: 17483113
56. Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* 2005; 22: 1107–1118. <https://doi.org/10.1093/molbev/msi097> PMID: 15689528
57. Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol.* 2005; 22: 2472–2479. <https://doi.org/10.1093/molbev/msi237> PMID: 16107592
58. Richardson B. Impact of aging on DNA methylation. *Ageing Res Rev.* 2003; 2: 245–261. PMID: 12726774
59. Mugal CF, Arndt PF, Holm L, Ellegren H. Evolutionary consequences of DNA methylation on the GC content in vertebrate genomes. *G3 (Bethesda).* 2015; 5: 441–447.
60. Romiguier J, Ranwez V, Douzery EJ, Galtier N. Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome sizes. *Genome Res.* 2010; 20: 1001–1009. <https://doi.org/10.1101/gr.104372.109> PMID: 20530252
61. Do CB, Mahabhashyam MS, Brudno M, Batzoglou S. ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Res.* 2005; 15: 330–340. <https://doi.org/10.1101/gr.2821705> PMID: 15687296
62. Hall TA. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser.* 1999; 41: 95–98.
63. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016; 33: 1870–1874. <https://doi.org/10.1093/molbev/msw054> PMID: 27004904
64. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol.* 1993; 10: 512–526. <https://doi.org/10.1093/oxfordjournals.molbev.a040023> PMID: 8336541
65. Whelan S, Goldman N. Distributions of Statistics Used for the Comparison of Models of Sequence Evolution in Phylogenetics. *Mol Biol Evol.* 1999; 16: 1292–1299.