DNA methylation in amphioxus: from ancestral functions to new roles in vertebrates

Ricard Albalat, Josep Martí-Solans and Cristian Cañestro

Advance Access publication date 2 March 2012

Abstract

In vertebrates, DNA methylation is an epigenetic mechanism that modulates gene transcription, and plays crucial roles during development, cell fate maintenance, germ cell pluripotency and inheritable genome imprinting. DNA methylation might also play a role as a genome defense mechanism against the mutational activity derived from transposon mobility. In contrast to the heavily methylated genomes in vertebrates, most genomes in invertebrates are poorly or just moderately methylated, and the function of DNA methylation remains unclear. Here, we review the DNA methylation system in the cephalochordate amphioxus, which belongs to the most basally divergent group of our own phylum, the chordates. First, surveys of the amphioxus genome database reveal the presence of the DNA methylation machinery, DNA methyltransferases and methyl-CpG-binding domain proteins. Second, comparative genomics and analyses of conserved synteny between amphioxus and vertebrates provide robust evidence that the DNA methylation machinery of amphioxus represents the ancestral toolkit of chordates, and that its expansion in vertebrates was originated by the two rounds of whole-genome duplication that occurred in stem vertebrates. Third, in silico analysis of CpGo/e ratios throughout the amphioxus genome suggests a bimodal distribution of DNA methylation, consistent with a mosaic pattern comprising domains of methylated DNA interspersed with domains of unmethylated DNA, similar to the situation described in ascidians, but radically different to the globally methylated vertebrate genomes. Finally, we discuss potential roles of the DNA methylation system in amphioxus in the context of chordate genome evolution and the origin of vertebrates.

Keywords: DNA methylation; amphioxus; chordate and vertebrate genome evolution; transposable elements; conserved genomic synteny; Dnmt and Mbd

INTRODUCTION

DNA methylation is an epigenetic mechanism that has been mainly associated with regulation of gene expression. DNA methylation marks can be inherited cell to cell, and play important roles maintaining differential gene expression profiles between different cell lineages during embryo development, states of cell differentiation in adult tissues or
pluripotency in germ cells, as well as preserving chromosomal imprinting [1–4]. DNA methylation consists in the addition of a methyl group covalently bound to the 5-carbon of cytosine of a CpG dinucleotide. In animals, DNA methylation is mediated by DNA methyltransferases (Dnmt), which are evolutionarily conserved enzymes that have been classified into three groups, Dnmt1, Dnmt2 and Dnmt3 [5]. Dnmt3 enzymes are responsible for de novo methylation during development [6, 7]. Dnmt1 enzymes are important in maintaining methylation patterns after DNA replication on hemimethylated symmetric motifs [8]. Dnmt2 (also termed Trdmt1) enzymes show a robust methyltransferase activity of tRNA molecules, and only present low activity against DNA [9, 10].

The addition of methyl groups to the DNA regulates gene expression in two ways, either by direct interference with the binding of proteins that interact with DNA elements [11–13], or indirectly, through recruitment of proteins that contain a methyl-CpG-binding domain (MBD), which are called Mbd1, Mbd2, Mbd3, Mbd3L, Mbd4 and MeCP2 (methyl CpG-binding protein 2) [14, 15]. Most Mbd proteins specifically recognize and bind to methylated DNA, and it is thought that they can associate with histone deacetylases (HDACs) and other chromatin remodeling proteins, contributing to the transcriptional repression of promoters ([16, 17] reviewed in [18]). Each Mbd protein has, however, its own features. MeCP2 was the first discovered Mbd protein with methyl-CpG DNA-binding activity, and it has been related to the HDAC complex in the maintenance of transcriptional silencing [17]. Mbd1 is the only methyl-CpG-binding protein that is capable of repressing transcription from both methylated and unmethylated promoters, at least in cell culture [19–21]. The DNA binding activity of Mbd2 and Mbd3 proteins is mediated by their MBD, and although such activity has been lost in the mammalian Mbd3, it is still preserved in amphibians and fishes [22]. In mammals, Mbd2 represses transcription in a methylation density-dependent fashion [23]. Mbd3L1 is a protein with significant homology to Mbd2 and Mbd3. It lacks, however, the MBD, although it might interact with Mbd2 as a methylation-dependent transcriptional repressor, likely competing with Mbd3 [24]. Finally, Mbd4 is the only member of the Mbd family that does not seem to be involved in transcriptional repression, but it plays a role in decreasing the mutability of methyl-CpG in the genome [25]. Additional Mbd proteins have also been described (e.g. Mbd5 and Mbd6), and in contrast to the rest of the MBD proteins, they do not bind methylated DNA, but might play a role in the formation of heterochromatin and epigenetic reprogramming after fertilization [26].

Despite that the DNA methylation is present throughout the phylogenetic ladder, substantial differences in the multiplicity of Dnmt and Mbd genes, and in the amount and genomic distribution of methylated cytosines are found among different taxa. Thus, vertebrates have more Dnmt and Mbd genes, and higher methylated genomes than non-vertebrates. Vertebrates have at least one Dnmt, one Dnmt2, one to five Dnmt3 genes and at least six Mbd genes in their DNA methylation toolkit, whereas invertebrates have a variable lower number of Dnmt and Mbd genes ([5, 22, 27–32] and Figure 1). Caenorhabditis elegans, for instance, lacks Dnmt genes, whereas related nematodes preserve one Dnmt2-related gene [33]. Among insects, fruit flies and mosquitoes only have Dnmt2, the silk moth has both Dnmt1 and Dnmt2, and the honeybee possesses the full set of Dnmt genes [34, 35]. The protostome Mbd complement also varies depending on the species analyzed. Most lophotrochozoan and ecdysozoan species have both Mbd1/Mbd2/3 (formerly Mbd2/3) and Mbd4/MeCP2 (formerly Mbd4) genes, although some ecdysozoans have a truncated or even no Mbd4/MeCP2 genes ([28, 31, 33, 36] and Figure 1). Outside bilaterians, a full complement of Dnmt is present in cnidarians and sponges, but so far a sole Dnmt2 has been found in placozoans [31, 37, 38]. Mbd1/2/3 and Mbd4/MeCP2 have been also identified in cnidarians as well as in placozoans [31]. In addition to the diversity of Dnmt and Mbd genes, the extent of CpG methylation and its genomic distribution also differ among non-vertebrate species, from undetectable or very low levels to mosaic patterns of substantial methylation in different non-vertebrate genomes, although never reaching the high levels of global methylation of vertebrates [39–43].

Amphioxus belongs to the subphylum of the cephalochordates, the most basally divergent group in our own phylum, the chordates. The privileged position of amphioxus permits the comparison of its methylation system with that of vertebrates, thereby allowing the inference of the methylation condition of the last chordate common ancestor. Such evolutionary framework helps to understand how changes
in the DNA methylation system might have led to functional innovations in the stem vertebrate genome. Here, we first review the main components of the DNA methylation machinery (i.e. \textit{Dnmt} and \textit{Mbd} genes) in amphioxus, and perform a comparative analysis with those in vertebrates. Second, we describe the pattern of distribution of the DNA methylation throughout the amphioxus genome, comparing it with that reported in other chordates. Finally, we discuss the evolution of the DNA methylation machinery in chordates, in the context of the major genomic changes related to the two rounds of whole-genome duplication (2R-WGD) that were concomitant with the origin and diversification of vertebrate innovations.

\section*{Evolution of the Chordate DNA Methylation Toolkit: \textit{Dnmt} and \textit{Mbd} Genes}

The amphioxus DNA methylation toolkit

Genomic surveys revealed that amphioxus has three \textit{Dnmt} genes, namely \textit{Bf\_Dnmt1}, \textit{Bf\_Dnmt2} and \textit{Bf\_Dnmt3}, each one ortholog to one of the three \textit{Dnmt} groups \cite{31,37}. Structurally, the C-terminal
region of Dnmt enzymes contains the cytosine methyltransferase active site, organized into 10 motifs, of which 6 of them (I, IV, VI, VIII, IX and X) are evolutionarily conserved [45–47]. These six motifs are recognizable in the three Bf_Dnmt sequences (Figure 2), suggesting the enzymatic conservation of the amphioxus Dnmt proteins. Each amphioxus Dnmt contains additional domains specific for each Dnmt group (Figure 2), which agrees with the tree topology obtained by phylogenetic analysis [31]. Hence, Bf_Dnmt1 contains the typical Dmap1-binding region, the nuclear localization signal (NLS), the Cys-rich zinc-binding domain and the two bromo-adjacent homology (BAH) motifs. Bf_Dnmt2 presents the characteristic Dnmt2 CFTXXXXXY (CFT) sequence between motifs VIII and IX. Bf_Dnmt3 contains the typical PWPP domain and the Cys-rich region in the N-terminal region of the protein.

Genomic surveys also revealed that amphioxus has two Mbd genes. Considering phylogenetic reconstructions [22, 31] and analyses of conserved synteny (discussed in the next section), here we suggest a new nomenclature that reflects their evolutionary relationships with vertebrate genes: Bf_Mbd1/2/3 (formerly Bf_Mbd2/3) because it is the co-ortholog of vertebrate Mbd1, Mbd2 and Mbd3 paralogs, and Bf_Mbd4/MeCP2 (formerly Bf_Mbd4) because it is the co-ortholog of vertebrate Mbd4 and MeCP2 paralogs. Structurally, Bf_Mbd1/2/3 contains, in addition to the MBD, the coiled-coil domain at the C-terminal region characteristic of vertebrate Mbd2 and Mbd3 [48], but lacks the CXXC motifs and the hydrophobic transcripional repression domain (hTRD) of Mbd1 (Figure 2). Bf_Mbd4/MeCP2 concurs with the structural features of Mbd4 proteins, including the thymine DNA glycosylase domain and the Fas-associated death-domain (FADD) protein binding region in its C-terminal half [31], but lacks the highly basic transcripional repression domain (bTRD) present in MeCP2 (Figure 2). Domain analysis reveals, therefore, that Bf_Mbd1/2/3 is more similar to vertebrate Mbd2 and 3 than to Mbd1, and Bf_Mbd4/MeCP2 is more similar to Mbd4 than to MeCP2. These similarities suggest that Bf_Mbd1/2/3, Bf_Mbd4/MeCP2 and vertebrate Mbd2, Mbd3 and Mbd4 might have maintained ancestral roles, whereas Mbd1 and MeCP2 might have acquired vertebrate-specific novel functions.

Most Dnmt and Mbd protein domains are involved in protein–protein interactions. Vertebrate Dnmt1 is able to bind different proteins, including Dnmt3 [49], histone deacetylase 2 (Hdac2), DNA methyltransferase 1 associated protein 1 (Dmap1) [50] and proliferating cell nuclear antigen (PcnA) [51]. Dnmt3 binds zinc fingers and homeoboxes 1 protein (Zhxl) [52], and Mbd2 and Mbd3 interact with many proteins in the NuRD-Mi-2 complex, including Hdac2, retinoblastoma binding protein 4 (Rbbp4), metastasis associated 1 (Mta1) and GATA zinc finger domain containing 2A (Gatad2a) [53, 54]. Homologs of these Dnmt- and Mbd partners have been identified in amphioxus (e.g. Dmap1, Hdac2, PcnA, Rbbp4, Mta1 and Gatad2a) [31], and in silico reconstruction of a putative interaction network in amphioxus is therefore feasible (Figure 3). Although this in silico approximation is only an initial step that requires experimental validation, the identification in amphioxus of bona fide Dnmt and Mbd proteins and their plausible partners suggests the presence of an interaction network similar to that in vertebrates.

Vertebrate expansion of the DNA methylation toolkit

In contrast to non-vertebrate chordates, vertebrate Dnmt and Mbd gene families have suffered distinct degrees of expansions. Dnmt3, for instance, has been duplicated in vertebrates, and analysis of the human genome reveals that DNMT3A and DNMT3B paralogs are in conserved syntenic chromosomal segments in Hsa2 and Hsa20, arguing in favor of the hypothesis that both genes likely originated during the 2R-WGD, rather than via local duplications (Figure 4A). Genomic analysis of conserved synteny also reveals that Mbd genes have lost most of their components [31, 37].
Figure 2: Structural protein domains of the DNA methylation toolkit. (A) Structural comparison of Dnmt enzymes in amphioxus (Bf) and human (Hs). Dnmt enzymes share six highly conserved methyltransferase motifs (I, IV, VI, VIII, IX and X) at the C-terminal region, and several group-specific domains, mostly found at the N-terminal region. A Dmap1-binding region, a nuclear localization signal (NLS), a region involved in targeting to replication foci, a Cys-rich zinc-binding domain, and two bromo-adjacent homology (BAH) motifs are distinctive of the N-terminal region of Dnmt1 enzymes, which is joined to the C-terminal catalytic region by a tract of glycine and lysine residues (GK). A CFTXXYXXY (CFT) sequence between motifs VIII and IX is characteristic of Dnmt2. A NLS, a PWVP domain and a Cys-rich region are typically found in the N-terminal region of Dnmt3. (B) Structural comparison of amphioxus and human Mbd proteins. The methyl-CpG binding domain (MBD), distinctively found at the N-terminal half of all these proteins, is capable to recognize and bind specifically to methylated DNA. In addition to the MBD, amphioxus Mbd1/2/3 contains a coiled-coil domain (CC) at the C-terminal region characteristic of vertebrate Mbd2 and Mbd3, but lacks the CXXC motifs and the hydrophobic transcription repression domains (hTRD) of the human MBD1. Amphioxus Mbd4/MeCP2 includes a thymine DNA glycosylase domain and a Fas-associated death-domain protein binding region (FADD) in its C-terminal half, distinctive of vertebrate Mbd4, but lacks the basic transcription repression domain (bTRD) of MECP2.

Conservation is observed between the MBD1/MBD2-bearirng paralogon in Hsa18 and the MBD3/MBD3L-bearing paralogon in Hsa19, as well as between the MBD4-bearing paralogon in Hsa3 and the MECP2-bearing paralogon in HsaX (Figure 4B and C). Notably, genomic analysis reveals conserved synteny between the amphioxus Bf Mbd1/2/3-bearing paralogon within a 2-Mb fragment in
scaffold *Bf* V2_32 and the *MBD1*/*MBD2* and *MBD3*/*MBD3L* paralogons in 10-Mb fragment in *Hsa*18 and a 20-Mb fragment in *Hsa*19, respectively (Figure 4D). It is tempting to speculate that there might be still unknown evolutionary constraints maintaining such high degree of synteny conservation during >500 Mya, the time of divergence between cephalochordates and vertebrates. Currently, different projects are being developed to uncover the causes that could explain the maintenance of macro- and micro-synteny conservation across such distantly related taxa [57].

*Dnmt1* and *Dnmt2*, however, have been maintained as single copy genes despite the vertebrate 2R-WGD events. Although genomic analysis reveals conserved synteny between the *DNMT1* region on human chromosome 9 (*Hsa9*) and chromosome segments in *Hsa*1 and *Hsa*19, as well as between the *DNMT2* region on chromosome *Hsa*10 and chromosome segments in *Hsa*2 and *Hsa*17 (CC, unpublished data), none of the *DNMT1* or *DNMT2* ohnologs (i.e. paralogs originated by genome duplication) survived in their respective paralogous chromosomal segments. *Dnmt1* and *Dnmt2* have remained as single copy genes in chordates, suggesting that these genes might be refractory to copy number variation, a phenomenon observed in ohnologs that have dosage-balance evolutionary constraints [60].

**DNA METHYLATION DISTRIBUTION PATTERN IN AMPHIOXUS**

The extent of CpG methylation in the amphioxus genome was initially studied by comparing electrophoretic patterns of genomic DNA digested with methylation-sensitive and methylation-insensitive restriction endonucleases [39]. Southern blots of these digestions were hybridized with gene-specific probes to assess the methylation status of certain amphioxus genes, providing an approximate idea of the genomic distribution of this epigenetic mark. These early studies led to the conclusion that there are two roughly comparable fractions of methylated and unmethylated DNA in the amphioxus genome, with genes distributed in both fractions [39].

Sequencing of the whole genome of *Branchiostoma floridae* has allowed an estimate of the DNA methylation pattern throughout the amphioxus genome. The distribution of DNA methylation in

![Figure 3: Amphioxus network of protein interactions linked to Dnmt and Mbd inferred by STRING v9.0 [55]. Spheres represent amphioxus proteins connected to interacting partners by solid lines (confidence score > 0.3). Stronger interactions are represented by thicker lines. Amphioxus protein IDs from JGI and their respective human partners with the highest bit-score value in sequence comparisons are shown.](https://academic.oup.com/bfg/article-abstract/11/2/142/213567)
Figure 4: Conserved syntenies between DNMT (A) and MBD (B and C) bearing paralogs in human chromosomes provide evidence that the expansion of the DNA methylation toolkit was originated by the two rounds of whole-genome duplication (2R-WGD) that occurred in stem vertebrates. Interestingly, synteny for the MBD bearing paralogs has been preserved between human and amphioxus chromosomes for >500 Mya (D), which may suggest that selective evolutionary constrains might preserved the structural organization of this genomic region because of a potential functional role. HsaN: human chromosome N; Bfl_v2N, B. floridae scaffold N. The analysis of synteny conservation is inferred from the Synteny Database [58] using a 100-gene sliding window as previously described [59].
the genome can be inferred from the ratio between the observed and expected CpG content, a metric termed CpGo/e ratio or CpG score [41, 61]. In general, unmethylated regions show a CpGo/e ratio close to 1, whereas hyper-methylated regions tend to contain few CpG sites due to the mutability of 5-methylcytosine, therefore showing CpGo/e ratios far smaller than 1 (e.g. $<0.5$ as in humans). Using this indirect approach, Okamura et al. [42] analyzed one genomic region of 2 Mb, and concluded that amphioxus genome has a mosaic pattern of DNA methylation. The mean length of methylated and unmethylated regions was estimated 4.8 kb (SD = 5.0 kb) and 7.7 kb (SD = 11.1 kb), respectively, comparable to those of other non-vertebrate deuterostome genomes (e.g. 3.6 and 6.6 kb for sea urchin, and 5.3 and 6.8 kb for ascidians, respectively), but distinct from the global methylated genomes of mammals (e.g. 25.5 and 2.6 kb for human, respectively). This indirect approach has been experimentally validated in ascidian and human genomes [42], but no experimental analysis has been reported in amphioxus. To validate whether or not CpGo/e ratio is a good estimate of the DNA methylation status in amphioxus, we have examined here the CpGo/e ratios of 18 amphioxus genes that had been experimentally verified to be either methylated (8 genes) or unmethylated (10 genes). Our analysis shows that the CpGo/e ratios of 17 out of the 18 genes analyzed correlate with their methylation status (Figure 5A), supporting Okamura’s analysis [42].

To test if the inferences about the methylation pattern made by Okamura and collaborators from the analysis of 2 Mb could be extended to other genomic regions, here we have analyzed the CpGo/e ratio across 17 Mb from three arbitrarily selected scaffolds of the amphioxus genome (Figure 5B–D). According to the CpGo/e values, the distribution of DNA methylation across different scaffolds shows a mosaic pattern, with methylated segments interspersed with similar segments of unmethylated DNA. Consistent with Okamura’s analysis, the plot of the distribution of the frequencies of the CpGo/e ratios reveals statistically significant bimodal distributions in all regions analyzed ($G^2$ over 5% significance level $>5.99$; left and right curves in Figure 5E–G). The bimodal distributions have very similar means among different scaffolds. One of the distributions might represent the methylated fraction (average CpGo/e ratio $\sim 0.4$, left curve), whereas the other distribution may represent the hypo-methylated fraction (average CpGo/e ratio $\sim 0.8$, right curve). Interestingly, the relative weight of each fraction shows small differences among scaffolds, but globally the methylated fraction tends to be slightly smaller than the unmethylated one (Figure 5H). Future experimental analyses are needed to test these CpGo/e-based inferences that suggests that amphioxus follows an overall methylation pattern similar to that in ascidians, in which about half of the genome is heavily methylated, whereas the other half is poorly methylated [41, 42]. These non-vertebrate patterns are in sharp contrast with those of globally methylated vertebrate genomes, in which 80% of the CpGs are methylated [4].

In mammals and plants, DNA methylation has been associated not only to gene silencing, but also to transcriptional repression of transposable elements (TEs), decreasing their mobility, and thereby functioning as a mechanism of genome defense against their mutational activity [66–68]. The methylation status of the amphioxus transposons remains mostly unknown. Experimental data have been provided only for one type of TE, in which analysis of methylation-sensitive DNA digestion revealed that the non-LTR retrotransposon Bf-CR1 was predominantly free of methylation [69, 70]. Important caveats need to be considered if CpGo/e ratios are used to infer the methylation status of TEs due to their peculiar behavior. For instance, populations of TEs may suffer sequence bias due to bottlenecks derived from the burst of mobilization of a reduced number of TEs, and thereby the CpGo/e ratios of the new TEs could reflect the CpG ratio of the original genomic hosting region, rather than a biased ratio due to TE-specific targeted methylation. Another aspect to be considered is the possibility of horizontal transfers of TEs between different phyla, in which CpG ratios of recently acquired TEs might still reflect the methylation status of the donor genome rather than that of the host genome. However, considering that TE horizontal transfers between different phyla are rare, and that the TEs of amphioxus do not appear to have suffered significant recent expansions [71], it is plausible to assume that most of the current amphioxus TEs might be ancient and vertically inherited. We argue, therefore, that if TEs had been major targets for DNA methylation over a long evolutionary time in the amphioxus lineage, the CpGo/e ratios of most TEs should be biased toward low values.
Figure 5: Mosaic pattern of DNA methylation in the amphioxus genome. (A) Histograms ordered by increasing values of the CpGo/e ratios of 18 amphioxus genes of known methylation status. Methylation genes include
DNA methylation in amphioxus

To test this possibility, we have analyzed the CpGo/e ratios of all the 623 TE families provided in the amphioxus genome database (http://genome.jgi-psf.org/Brafl1/Brafl1.home.html), and the results do not show any obvious bias toward low CpGo/e ratios. In fact, only ~7% of TE sequences showed CpGo/e ratios <0.5 (0.4 is the average of the methylated genomic fraction in amphioxus), whereas >62% of TEs have CpGo/e ratios >0.7 (0.8 is the average of the hypo-methylated genomic fraction in amphioxus) (Figure 5I). These theoretical inferences suggest that, globally, TEs are not major targets of methylation, a similar condition to that described in ascidians, in which experimental evidence has proved that most TEs are poorly methylated [41, 70, 72]. If future experimental analysis (e.g. bisulfite sequencing) of amphioxus TEs corroborates this inference, we could conclude that the methylation of TEs as a defense mechanism might be an evolutionary innovation acquired in vertebrates, not present in cephalochordates and urochordates.

DNA METHYLATION, FROM ANCESTRAL FUNCTIONS TO NEW ROLES

What functions might DNA methylation be playing in amphioxus? In ascidians, which similarly to amphioxus also have a mosaic methylation pattern, DNA methylation is targeted to 'gene bodies' rather than to intergenic regions or TEs [41]. It has been proposed that DNA methylation in gene bodies might prevent aberrant transcription from cryptic initiation sites, which may interfere with the normal transcription machinery [73]. Zilberman et al. [73] have hypothesized that a large number of elongating RNA polymerases might prevent the use of cryptic sites in highly transcribed genes, whereas in moderately transcribed genes, methylation of gene bodies is needed to avoid aberrant transcripts. This hypothesis predicts that the methylated fraction of a given genome will be enriched in genes moderately transcribed, and that this enrichment will be perceptible in genomes in which DNA methylation has not been recruited for other functions such as silencing of TEs. The analyses of DNA methylation of ascidians and other protostome genomes agree with these predictions, and genes with weak–moderate expression (e.g. housekeeping genes) belong to the heavily methylated fraction, while genes with high tissue-specific expression belong to the poorly methylated fraction [41, 74–76]. Analysis of a small set of genes in amphioxus shows a tempting correlation between housekeeping genes in the methylated fraction [e.g. ribosomal proteins P2 (Brf6) and S6 (Brf61), mitochondrial carrier protein (Brf38)], and tissue-specific genes in the unmethylated fraction [e.g. Lipase.

Figure 5: Continued

Brf6 (60S ribosomal protein P2, NW.003101422), Brf66 (NADH dehydrogenase subunit, NW.003101225), Brf32 (unidentified cDNA, NW.003101419), Brf38 (mitochondrial proton carrier-like protein, NW.003101433), Brf43 (snRNP-like protein, NW.003101422), Brf61 (4OS ribosomal protein S6, NW.003101524) [39], Adh3 (alcohol dehydrogenase class III, from exons 2 to 6, NW.003101484) [62], and presinilin (AF69890) [63], and unmethylated genes are AmphimLC-alk (alkali myosin light chain, NW.003101500), Brf5 (unidentified cDNA, NW.003101462), Brf21 (triacylglycerol lipase, NW.003101520), Brf37 (α-tubulin, NW.003101522), Brf41 (unidentified cDNA, NW.003101534), Brf51 (β-actin, NW.003101548), Brf52 (unidentified cDNA, NW.003101505), Brf54 (unidentified cDNA, NW.003101377), Brf57 (unidentified cDNA, NW.003101328) and rDNA (AF061796) [39]. (B–D) Pattern of DNA methylation inferred by CpGo/e ratios across 17 Mb from three arbitrarily selected genomic regions (scaffold1: 7.2 Mb, scaffold2: 6.5 Mb and scaffold 40: 3.1 Mb) splitted in 2 kb segments. This plot reveals a mosaic pattern of interspersed regions of heavily methylated and poorly methylated fractions across the amphioxus genome. Estimations of the CpGo/e ratio were performed with CpGproD algorithm [64]. (E–G) Analysis of the frequency distribution of CpGo/e ratios in 0.1 intervals revealed statistically significant (G2 over 5% significance level >5.99) bimodal distributions of methylated (left curve) and unmethylated (right curve) fractions in the three regions analyzed. Test for uni- or bi-modal distributions was performed with the NOCOM software [65]. (H) Global analysis of the three previously selected scaffolds. (I) Analysis of the frequency distribution of CpGo/e ratios of all the 623 TE sequences provided in the amphioxus genome database (http://genome.jgi-psf.org/Brafl1/Brafl1.home.html), representing a total of 18 Mb (Supplementary Table S1). The distribution of the frequencies of the CpGo/e ratios of all amphioxus TE families does not show a bimodal distribution as does the rest of the genome in amphioxus [dashed lines in (I)], and importantly, the distribution is not biased toward low values, suggesting that most TE might not be main targets of methylation in amphioxus.
(Brf21), insulin-like protein, alkali myosin light chain] [39]. A definitive answer about the function of DNA methylation in amphioxus, however, awaits future analysis of large-scale DNA methylation footprinting in order to test if the mosaic-methylated amphioxus genome consists of poorly methylated segments enriched in highly expressed genes, whereas the significantly methylated segments mostly span throughout gene-bodies of weak–moderately expressed genes.

In summary, the mosaic pattern of DNA methylation in amphioxus and ascidians likely represents the ancestral condition of the chordate genome, while the global pattern of DNA methylation might be a vertebrate innovation (Figure 6). This radical change of DNA methylation patterns correlates with the occurrence of the 2R-WGD, and with the amplification of the Dnmt and Mbd gene families that originated during this event (Figure 6). It is plausible to postulate that after the 2R-WGD, positive selection might have facilitated the expansion of the DNA methylation toolkit, leading to the acquisition of new roles, improving, for instance, the silencing of spurious transcription of the increased number of vertebrate genes [77], counteracting deleterious gene dosage effects [78] or decreasing the mutational activity of TEs [66] in the newly polyploidized genome of stem vertebrates.

SUPPLEMENTARY DATA
Supplementary data are available online at http://bfg.oxfordjournals.org/.

FUNDING
This work has been supported by grant 2009SGR.336 from Generalitat de Catalunya and
grant BFU2010-14875 from the Ministerio de Ciencia e Innovación (Spain).

**Key Points**

- Phylogenetic inferences and analysis of conserved synteny suggest that amphioxus has a DNA methylation toolkit that represents the ancestral chordate condition.
- The expansion of the DNA methylation toolkit in vertebrates was originated by the 2R-WGD that occurred in stem vertebrates.
- Genomic sequence analysis suggests that amphioxus has a mosaic pattern of distribution of DNA methylation, similar to ascidians, but radically different to the globally methylated genome of vertebrates.
- The methylation status of amphioxus TE is remains unknown, but no evidence has been found that suggests that TEs are major targets of DNA methylation.
- The methylation of gene bodies, which might prevent transcription from cryptic initiation sites of moderately expressed genes, appears to be a potential role of DNA methylation in ancestral chordates.

**Acknowledgements**

We would like to thank to Jerzy Jurka for kindly sharing valuable TE information on the amphioxus genome. We appreciate the valuable comments of three anonymous reviewers for helping to improve the manuscript.

**References**


