MOLECULAR AND CELLULAR BIOLOGY, Mar. 1997, p. 1469–1475 Vol. 17, No. 3
0270-7306/97/$04.00+0

Methylation of Genomes and Genes at the Invertebrate-Vertebrate Boundary

SUSAN TWEEDIE,* JILLIAN CHARLTON, VICTORIA CLARK, AND ADRIAN BIRD

Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, United Kingdom

Received 10 October 1996/Returned for modification 26 November 1996/Accepted 13 December 1996

Patterns of DNA methylation in animal genomes are known to vary from an apparent absence of modified bases, via methylation of a minor fraction of the genome, to genome-wide methylation. Representative genomes from 10 invertebrate phyla comprise predominantly nonmethylated DNA and (usually but not always) a minor fraction of methylated DNA. In contrast, all 27 vertebrate genomes that have been examined display genome-wide methylation. Our studies of chordate genomes suggest that the transition from fractional to global methylation occurred close to the origin of vertebrates, as amphioxus has a typically invertebrate methylation pattern whereas primitive vertebrates (hagfish and lamprey) have patterns that are typical of vertebrates. Surprisingly, methylation of genes preceded this transition, as many invertebrate genes have turned out to be heavily methylated. Methylation does not preferentially affect genes whose expression is highly regulated, as several housekeeping genes are found in the heavily methylated fraction whereas several genes expressed in a tissue-specific manner are in the nonmethylated fraction.

Patterns of genomic DNA methylation have been studied previously with methylation-sensitive restriction endonucleases in a selection of animal species. Vertebrate genomes (including representatives of fish, amphibia, reptiles, birds, and mammals) are globally methylated in the vast majority of cell types (see, for example, reference 33), with 60 to 90% of CpG dinucleotides containing 5-methylcytosine (m^5C). In contrast, most invertebrate genomes that have been examined so far are predominantly nonmethylated (10, 11), and in a few cases (e.g., *Drosophila melanogaster* and *Caenorhabditis elegans*) they appear to be free from methylation (28). Invertebrate genomes that do have methylated DNA (including representatives of cnidarians, molluscs, and echinoderms) conform to a common pattern. A minor portion of the genome (10 to 40%) belongs to a methylated fraction (or compartment) made up of long stretches of heavily methylated DNA, while the major genomic fraction comprises long domains of nonmethylated DNA (10, 11).

A notable feature of methylation patterns generally is their stability. In the sea urchin, for example, the methylated fraction is apparent at all tested stages of early embryonic and larval development, as well as in a variety of adult tissues and in male and female germ cells (11). Reassociation analysis shows that methylated and nonmethylated fractions retain broadly the same sequence composition throughout development, and this view is reinforced by the finding that specific sequences are always found in the same fraction. The embryo-specific histone genes, for example, which are transcriptionally active in one cell type but inactive in another, do not alter their methylation status but remain nonmethylated regardless of expression. Relative constancy of methylation patterns is also a feature of vertebrate genomes. Here the nonmethylated fraction is reduced to a small percentage of the genome, the so-called CpG island fraction (6, 16). The great majority of CpG islands are nonmethylated at all stages of development. Although the general pattern is stable, some vertebrate sequences have been found to alter their methylation status during development. CpG islands on the inactivated X chromosome of eutherian mammals, for example, become methylated in somatic cells (see, for example, references 26 and 30).

Since overall DNA methylation patterns are a relatively stable attribute of an organism's genome, it is of interest to study when and how major transitions in the patterns arose during evolution. We have focused particularly on organisms whose lineages diverged from a common ancestor within the chordate phylum in order to determine when the transition from fractional to global methylation may have occurred. We have also addressed the methylation status of genes. Previous data indicated that invertebrate genes are almost invariably in the nonmethylated fraction of the genome (11), and on this basis it was suggested that methylation of genes might be exclusively a vertebrate phenomenon (see, for example, reference 8). The present study demands a revision of this idea, as significant numbers of genes are in fact found in the methylated fraction of the invertebrate genome.

MATERIALS AND METHODS

DNA and RNA preparation. Unless otherwise specified in the text, the species of sea urchin, sea squirt, amphioxus, and lamprey used in this study were *Strongylocentrotus purpuratus*, *Ciona intestinalis*, *Branchiostoma lanceolatum*, and *Lampetra planeri*, respectively. DNA was extracted from sea urchin sperm as described previously (11) and from other animal tissues by standard methods (25). Total RNA was isolated from adult amphioxus and sea urchin (*Psammechinus miliaris*) eggs by acid guanidinium thiocyanate-phenol-chloroform extraction (15).

Identification of m^5C in DNA immobilized on nitrocellulose membranes. Slot blots on nitrocellulose were prepared with 50 ng of undigested, denatured DNA in each slot. *Hpa*II-digested DNA was separated on 1.2% agarose gels and transferred to nitrocellulose. Filters were blocked in phosphate-buffered saline containing 0.1% Tween 20 (PBS-T) and 5% dried milk powder and incubated with a mouse monoclonal antibody against m^5C (diluted 1 in 60 in PBS-T) (29), and methylated DNA was visualized with the Amersham ECL detection system.

Two-dimensional TLC. Methylation of CpG in recognition sites for *Taq*I (TCGA) and *Msp*I (CCGG) was assayed by the method of Cedar et al. (13) as modified by Antequera et al. (2). *Taq*I or *Msp*I fragments (500 ng) were end labelled and digested to deoxyribonucleoside 5′-monophosphates by incubation in 10 μl of 50 mM Tris-HCl (pH 7.8)–5 mM $MgCl_2$ containing 5 μg of DNase I (Worthington) and 60 ng of snake venom phosphodiesterase (Boehringer) at 37°C for 1 h. After hydrolysis, one-half of each of the products was spotted on a

* Corresponding author. Mailing address: Institute of Cell and Molecular Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JR, United Kingdom. Phone: 44-131-650-8695. Fax: 44-131-650-5379. E-mail: stweedie@srv0.bio.ed.ac.uk.

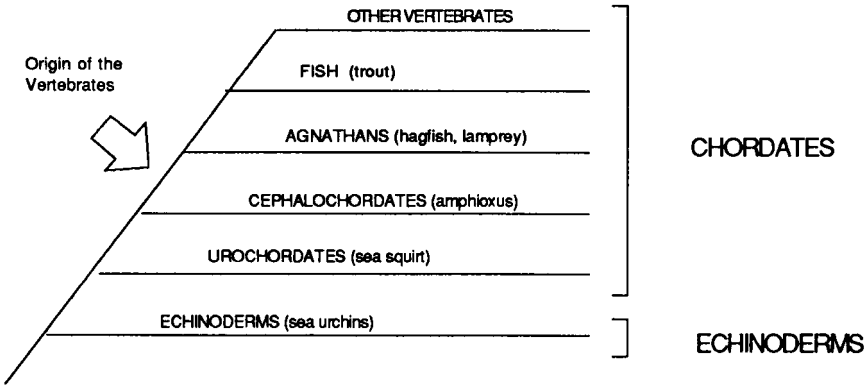

FIG. 1. Accepted phylogenetic relationships among the organisms used in this study. The two metazoan phyla concerned are indicated at the right of the figure. Common names for the types of organisms from which DNA or RNA was isolated are in parentheses above each branch.

cellulose thin-layer chromatography (TLC) plate (20 by 20 cm; Kodak), and the plates were developed as described previously (13). The spots were quantified by using ImageQuant software after the plates were scanned on a PhosphorImager (Molecular Dynamics).

Southern and Northern blotting and hybridization. For Southern blots, samples of DNA (10 μg) were digested with *Msp*I and *Hpa*II; an aliquot of each digest (usually 4 μl from a total volume of 40 μl) was added to 200 ng of (dry) plasmid DNA and assayed for complete digestion. The remainder of the digest was separated on 0.9% agarose gels and transferred to Hybond-N+ (Amersham). In most cases the blots were also hybridized to a probe which recognizes nonmethylated sequences (e.g., genes coding for rRNA [rDNA]) as a further control for digestion. For Northern blots, total RNA (10 μg) was fractionated on 1.5% agarose gels containing 2.2 M formaldehyde and transferred to Hybond-N (Amersham).

Isolation of random cDNA probes. Random cDNA clones were isolated from a λ-zap library prepared from a 2- to 4-day-old amphioxus (*Branchiostoma floridae*) (22) (a gift from Linda Holland, Scripps Institution of Oceanography, San Diego, Calif.) and a λgt10 library prepared from unfertilized sea urchin (*P. miliaris*) eggs (a gift from Meinrad Busslinger, IMP, Vienna, Austria).

The partial DNA sequence was obtained with an Applied Biosystems 373A DNA sequencer by using the universal forward and reverse primers. The cDNA sequences were used as queries in BLAST searches (1) of the databases maintained by the National Center for Biotechnology Information, Bethesda, Md., and several significant matches were found (see Table 1). Details of the BLAST results can be obtained from the authors.

Gene-specific probes. Additional sea urchin probes for hydroxy-3-methylglutaryl coenzyme A reductase (HMG-CoA reductase), protein disulfide isomerase (PDI), glutamine synthetase, chicken ovalbumin upstream promoter transcription factor (COUP) exon 1, spicule matrix protein SM30-α exon 2, and ribosomal protein S6 were generated by PCR from *S. purpuratus* genomic DNA or an *S. purpuratus* pGAD424 cDNA library (prepared by F. Calzone). A genomic fragment from the *B. lanceolatum* insulin-like peptide gene was amplified by PCR with primers derived from the *Branchiostoma californiensis* insulin-like peptide sequence (14), and a partial alkali myosin light chain cDNA was amplified from the *B. floridae* cDNA library with primers from the published sequence (22). In each case the PCR products were subcloned into the pTAg vector (R&D Systems) and partially sequenced to check the authenticity of the product. Two overlapping regions of the *B. lanceolatum* S6 ribosomal protein gene were amplified from genomic DNA. Each PCR product was subcloned into pTAg, and six subclones of the 5′ fragments and four subclones of the 3′ fragments were sequenced. All of the intron-exon boundaries contained within these fragments showed conservation with the human gene (4). Details of the primers used to generate these probes and conditions for the PCR are available from the authors.

Nucleotide sequence accession numbers. The GenBank-EMBL accession numbers for the new sequences used in this paper are Z83259 to Z83283.

RESULTS

Global patterns of genomic methylation. Figure 1 shows the principal organisms used in this study and their accepted phylogenetic relationships (31). The majority of selected species were chordates, as the invertebrate-vertebrate transition occurred within this phylum. The organisms thought to be most closely related to the ancestral transitional forms are amphioxus on the invertebrate side and the agnathans, or jawless fishes (hagfish and lamprey), on the vertebrate side.

We analyzed the distribution of methylated sites in DNA from these species by comparing the patterns of restriction fragments generated by methylation-sensitive enzymes (Fig. 2). First, we compared amphioxus and sea squirt DNAs with DNA from the sea urchin. *Hpa*II digestion of all these invertebrate DNAs gave similar patterns (Fig. 2A). Most of the DNA was cleaved to a heterogeneous mixture of fragments that entered the gel, while a fraction, apparently undigested, remained at the top of the gel. The high-molecular-weight component is not devoid of *Hpa*II recognition sites (CCGG) since it was largely removed by digestion with the methylation-insensitive isoschizomer of *Hpa*II, *Msp*I. This fraction most likely was not

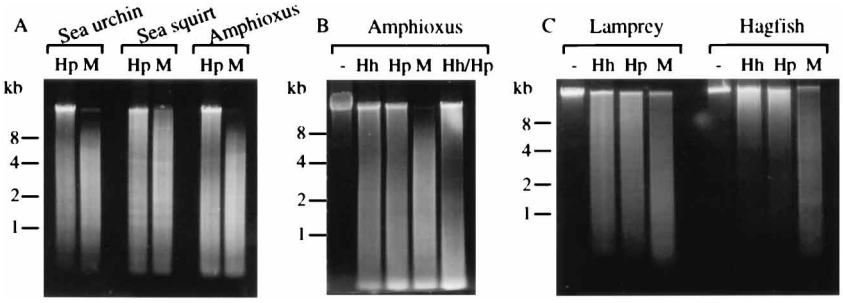

FIG. 2. Fragment patterns obtained after digestion of total genomic DNA from various organisms with *Msp*I (M), *Hpa*II (Hp), and *Hha*I (Hh). DNA samples were fractionated on agarose gels and stained with ethidium bromide. Sizes of fragments, in kilobase pairs (kb), are indicated to the left of each panel. (A) Fractional methylation of the genomes of three invertebrates: sea urchin, sea squirt, and amphioxus. (B) Fragment patterns obtained with amphioxus DNA after treatment with no enzyme (−) or with the enzymes listed above. (C) Fragment patterns obtained with DNA from lamprey and hagfish.

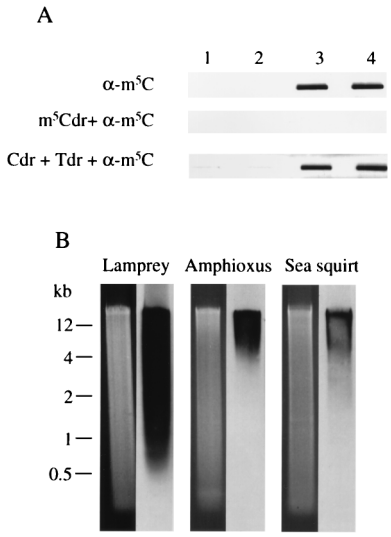

FIG. 3. An antibody to m^5C (α-m^5C) specifically recognizes the methylated compartment in fractionally methylated genomes. (A) Denatured DNA from nonmethylated bacteriophage λ (lane 1), *Drosophila melanogaster* (lane 2), sea urchin (lane 3), and mouse (C57BL/6J) (lane 4) (50 ng each) was immobilized on nitrocellulose filters. The filters were incubated with a monoclonal antibody to m^5C in the presence or absence of 0.2 mM 5-methylcytidine (m^5Cdr) or 0.1 mM each cytidine (Cdr) and thymidine (Tdr). (B) *Hpa*II-digested DNA from lamprey, amphioxus, and sea squirt was separated on a 1.2% agarose gel, stained with ethidium bromide (left panel of each pair), blotted onto nitrocellulose, and incubated with the anti-m^5C antibody in the absence of competitors (right panel of each pair).

cut by *Hpa*II because of CpG methylation. The fragment pattern obtained when amphioxus DNA was digested with another methylation-sensitive enzyme, *Hha*I (GCGC), was similar to that obtained with *Hpa*II (Fig. 2B). Furthermore, when *Hpa*II and *Hha*I were used together, the amount of undigested DNA was the same as that resulting from use of either enzyme alone but the cleaved fragments were reduced in size. This confirms that there are two fractions of the genome in amphioxus: one resistant to both *Hpa*II and *Hha*I digestion and the other susceptible to digestion by both enzymes.

As an independent test for the presence of methylated and nonmethylated DNA fractions in these genomes, blots of the digested DNAs were challenged with a monoclonal antibody against m^5C (29). Slot blot analysis confirmed the specificity of the antibody, as methylated genomes (sea urchin and mouse) were readily detected but nonmethylated genomes (*Drosophila* and bacteriophage λ) gave background signal (Fig. 3A). The signal was competed out by addition of 5-methylcytidine but not by addition of Cdr (cytidine) and Tdr (thymidine). When blots of invertebrate genomic DNAs were tested with the antibody, it was clear that digestion with *Hpa*II separated the genome into methylated and nonmethylated fractions. The *Hpa*II-resistant fragments were strongly stained, whereas smaller fragments were not (Fig. 3B). *Msp*I led to the appearance of antibody staining in the lower regions of the lane (data not shown).

Equivalent experiments with vertebrate DNAs produced different results. Hagfish DNA gave a typical vertebrate digestion pattern (see reference 10), as it was largely resistant to digestion by both *Hpa*II and *Hha*I (Fig. 2C). Thus, the hagfish genome exhibits global methylation, unlike its nonvertebrate relatives amphioxus and sea squirt. The methylation pattern of lamprey was initially less obvious. As in other vertebrate genomes, *Hpa*II fragments were significantly larger on average than the *Msp*I fragments, but lamprey DNA was significantly more digestible by *Hpa*II than hagfish DNA or that of other vertebrates (Fig. 2C). The gel pattern suggested that methylation is spread throughout the genome but at a comparatively low proportion of all CpG sites. This interpretation was supported when blots of *Hpa*II-digested lamprey DNA were challenged with anti-m^5C antibody. Methylated DNA was apparent throughout most of the *Hpa*II lane (Fig. 3B). Thus, there was no evidence of a nonmethylated fraction like that of invertebrates. The absence of antibody signal below ~500 bp probably reflects both the limited ability of nitrocellulose to retain small DNA fragments and the reduced probability that a short DNA fragment will contain a methylated CpG. Further analysis indicated that the proportion of CpGs that are methylated is lower in lamprey DNA (about 25%) than in other vertebrates (over 50%) (Fig. 4B)—hence, the comparatively extensive digestion by *Hpa*II.

Methylation status of genes. To study the methylation of chordate genes, we initially took advantage of the conservation of 18S and 28S rRNA gene (rDNA) sequences, which allows a single probe (from *Xenopus* rDNA) to be used against Southern blots of a variety of DNA samples. Previous studies have established that the rDNAs of many invertebrates are nonmethylated (10). Figure 5 shows that the rDNA of amphioxus, like that of the sea urchin, is not methylated at *Hpa*II sites. The rDNA of sea squirts is also nonmethylated (data not shown). Agnathans, on the other hand, have extensively methylated rDNA, as do other fishes (Fig. 5 and data not shown). These studies of a specific gene, like those of whole-genome methylation described above, show a sharp distinction between invertebrate and vertebrate methylation patterns.

To assess the methylation status of protein-coding genes, cDNAs were selected from an amphioxus (*B. floridae*) cDNA library. Each cDNA was partially sequenced to allow database searches and to eliminate contaminating mitochondrial sequences. In addition to the random cDNAs, a genomic sequence corresponding to the insulin-like gene of *B. californiensis* (14) and a partial cDNA for the *B. floridae* alkali myosin light chain (22) were generated by PCR. All the probes were

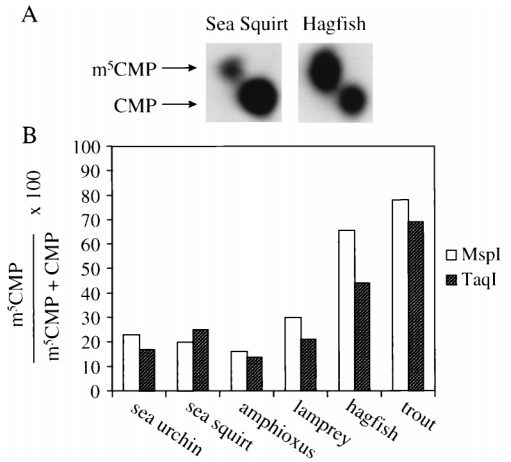

FIG. 4. Levels of CpG methylation within the sequences CCGG and TCGA of various chordate and echinoderm genomes. (A) Sea squirt and hagfish DNA was cleaved with *Msp*I (CCGG), and the end-labelled fragments were digested to nucleotides and separated by two-dimensional TLC. Cytosine (CMP) and m^5CMP spots are indicated with arrows. (B) The intensities of the spots from panel A and other, similar experiments were quantified with a PhosphorImager, and average values from up to three experiments were expressed as percentages of m^5C.

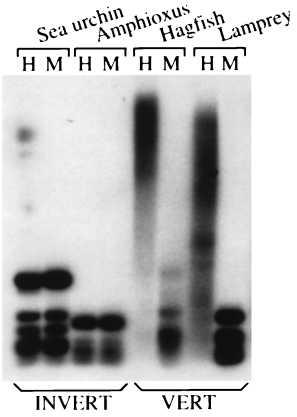

FIG. 5. Methylation status of 18S and 28S rDNA in invertebrate (INVERT) and vertebrate (VERT) genomes. DNA from the invertebrates sea urchin and amphioxus and the vertebrates hagfish and lamprey was digested with *Hpa*II (H) and *Msp*I (M), blotted, and hybridized to an rDNA probe derived from *Xenopus laevis.*

hybridized against Southern blots of DNA from amphioxus (*B. lanceolatum*) that had been digested with *Msp*I and *Hpa*II and Northern blots of total amphioxus RNA (Fig. 6). Only cDNAs that detected a discrete transcript on Northern blots were included in the analysis (Fig. 6B). Of 17 probes, 10 hybridized to nonmethylated genes as judged by the identity of *Hpa*II and *Msp*I band patterns (Fig. 6A, lanes 1 and 2; Fig. 6C, lane 3). These included the genomic sequence of an insulin-like protein, alkali myosin light chain (data not shown), and three cDNAs which showed a high degree of sequence similarity to known genes from other species (see Table 1). Three probes hybridized to multiple bands, and two of these were identified as β-actin (Fig. 6A, lane 3) and α-tubulin (data not shown). In these cases, the majority of the bands detected were common to both the *Hpa*II and *Msp*I digests and were therefore nonmethylated, but a minority of weakly hybridizing bands were methylated. Since actin and tubulin are known to be encoded by gene families in most organisms, the results suggest that nearly all family members are nonmethylated. There may be a few methylated genes present as well, or the methylated sequences may be pseudogenes or spurious related sequences.

For seven of the cDNAs, there was clear evidence of gene methylation. Six probes hybridized to completely methylated sequences located in the undigested high-molecular-weight component of amphioxus DNA (see, e.g., Fig. 6A, lane 5). Probe Brf6, which was homologous to ribosomal protein P2, was also clearly methylated but had a single band in common between the *Msp*I and *Hpa*II digests (Fig. 6A, lane 4). The results, summarized in Table 1, demonstrate conclusively that amphioxus genes are not confined to the nonmethylated fraction of the genome.

Methylation was mapped in more detail across the amphioxus S6 ribosomal protein gene. A cDNA containing the entire S6 coding region was used to obtain genomic sequence from amphioxus (*B. lanceolatum*) DNA (Fig. 6D) by PCR. The sequences of different cloned PCR products varied by small insertions and deletions that were confined mainly to the intervening sequences. This probably reflects the high degree of polymorphism that is observed within the amphioxus population (22a). The positions of the *Hpa*II sites are illustrated in Fig. 6D; potentially polymorphic sites are indicated. Southern blots of *Msp*I-digested genomic DNA showed a doublet corresponding to the predominant internal fragments of 0.74 and 0.66 kbp, together with two larger bands representing flanking sequences (Fig. 6C, lane 2M). The presence of signal at a high molecular weight in the *Hpa*II lanes indicates that all the tested sites are methylated. The possibility that either digest was partial was eliminated by the presence of identical band patterns in the *Hpa*II and *Msp*I lanes when the same blot was reprobed with the insulin-like gene (Fig. 6C, lane 3).

Is the presence of methylated genes in amphioxus a reflection of its evolutionary proximity to the vertebrates (whose genes are predominantly methylated), or is gene methylation widespread among the invertebrates? To answer this question, we extended the study to the sea urchin by selecting probes

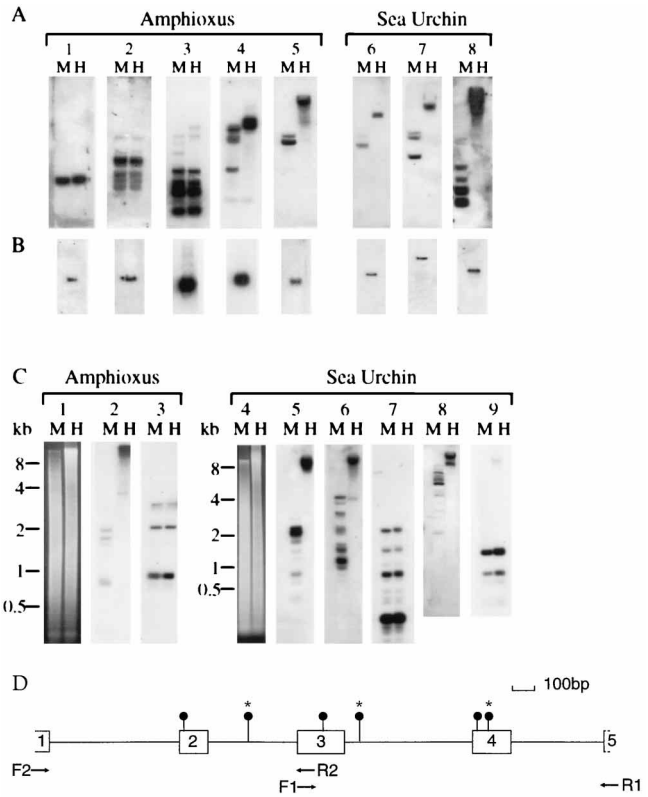

FIG. 6. Methylation status of genes in amphioxus and sea urchin. (A) Southern blots of genomic DNA digested with *Msp*I (M) and *Hpa*II (H) were hybridized to randomly selected cDNA probes. Amphioxus probes have a Brf prefix, as they were isolated from a *B. floridae* library, and sea urchin probes have a Psm prefix, as they were derived from *P. miliaris*. Lanes 1 to 5, amphioxus; lanes 6 to 8, sea urchin (*P. miliaris*). Probes: lanes 1, Brf5; lanes 2, Brf21; lanes 3, Brf51; lanes 4, Brf6; lanes 5, Brf18; lanes 6, Psm115; lanes 7, Psm139; lanes 8, Psm109. (B) Northern analysis with the same probes as used for panel A hybridized to adult amphioxus RNA (lanes 1 to 5) or sea urchin egg RNA (lanes 6 to 8). (C) Methylation of specific genes in the amphioxus and sea urchin genomes. Lanes 1 show amphioxus DNA stained with ethidium bromide after digestion with *Hpa*II (H) and *Msp*I (M) and fractionation on an agarose gel. This DNA was blotted and hybridized to a probe for the amphioxus S6 ribosomal protein gene (lanes 2) and the insulin-like protein gene of amphioxus (lanes 3). Lanes 4 show *S. purpuratus* DNA stained with ethidium bromide after digestion with *Hpa*II (H) or *Msp*I (M) and fractionation on an agarose gel. Blots of these digests were hybridized to *S. purpuratus* probes for HMG-CoA reductase cDNA (lanes 5), glutamine synthetase cDNA (lanes 6), spicule matrix protein SM30-α exon 2 (lanes 7), S6 ribosomal protein (lanes 8), and human rDNA (lanes 9). Lanes 7 to 9 are not from the same gel as lanes 4 to 6. (D) Map of *Hpa*II/*Msp*I sites in the genomic DNA sequence of the amphioxus S6 ribosomal protein gene. The positions of primers (F1, F2, R1, and R2) that were used to generate the fragments are shown by the arrows. The open boxes and single lines represent exons and introns, respectively. The exons are numbered relative to those in the human gene (17). Asterisks denote sites that are polymorphic in the population. Filled circles show sites that are methylated in genomic DNA.

TABLE 1. Methylation status of invertebrate genes

Methylation status	Gene	
	Amphioxus	Sea urchin
Methylated	40S ribosomal protein S6 (Brf61)[a] 60S ribosomal protein P2 (Brf6)[a] Mitochondrial proton carrier-like protein (Brf38)[a] NADH dehydrogenase subunit (Brf16)[a] snRNP-like protein (Brf43)[a] Unidentified cDNAs (Brf18 and -32)	40S ribosomal protein S6 HMG-CoA reductase Sterol regulatory binding protein 2 (Psm153)[a] PDI Glutamine synthetase Mitochondrial carrier protein (Psm139)[a] Unidentified cDNAs (Psm108, -109, -115, and -155)
Nonmethylated	Insulin-like protein Myosin light chain Triacylglycerol lipase (Brf21)[a] α-Tubulin (Brf37)[a,b] β-Actin (Brf51)[a,b] Unidentified cDNAs (Brf5, -41, -52, -54, and -57) rDNA	Spicule matrix protein SM30-α SpCOUP transcription factor Putative anion-exchange protein (Psm12)[a] Unidentified cDNAs (Psm131 and -151) rDNA

[a] Randomly selected cDNAs with significant database matches.
[b] Gene families that are predominantly nonmethylated but contain a minor methylated component (see Fig. 6A, lane 3).

from a cDNA library derived from *P. miliaris* egg RNA. Again, only probes that detected discrete bands on Northern blots were chosen. We found that six of nine of these genes were heavily methylated (Fig. 6A and Table 1). Additional Southern analysis with probes that were derived from sea urchin cDNA by PCR revealed that the genes for HMG-CoA reductase, glutamine synthetase, and PDI are also methylated (Fig. 6C and data not shown).

DISCUSSION

Genomic methylation patterns can be classified into three distinct types: nonmethylated, fractionally methylated, and globally methylated (Fig. 7). The data presented here suggest that there is a transition between the fractionally methylated genomes (type B) found in many invertebrates and the globally methylated genomes (type C) that are characteristic of vertebrates. The genomes of sea squirts and amphioxus, the most vertebrate-like invertebrate, exhibit a minor fraction of methylated DNA, as do those of representative echinoderms and annelids, a nemertean, a priapulid, a bryozoan, a mollusc, and a cnidarian (Fig. 7). In fact, compartmental methylation of this kind is probably the most widespread among multicellular eukaryotes, as it is also found in the plant *Arabidopsis thaliana* (32), the slime mold *Physarum polycephalum* (35), and a variety of fungi (3, 19). Vertebrates, on the other hand, consistently display global methylation patterns. This applies also to the earliest vertebrates, exemplified by the jawless fishes or agnathans. Based on these studies of present-day forms, it is reasonable to propose that globally methylated genomes arose within the phylum Chordata, near the time when vertebrates originated.

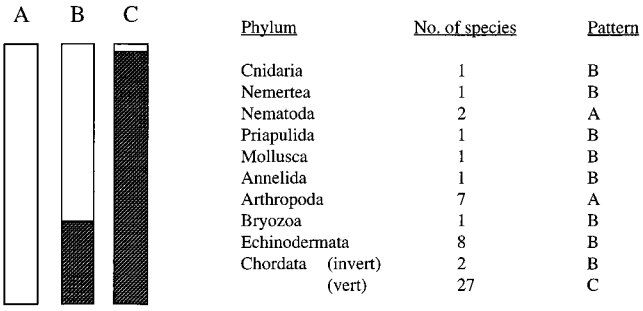

FIG. 7. Schematic representations of genomic methylation patterns and a summary of the patterns seen in various animal phyla. This list refers only to animal genomes that have been examined by restriction enzyme analysis. Crosshatching indicates the approximate proportion of the genome which is methylated in each genome type. Pattern A represents genomes in which no fraction that is resistant to methylation-sensitive enzymes can be detected. DNA methylation has been detected biochemically in some arthropods, but overall genomic methylation patterns of these genomes have not been reported. Pattern B represents fractionally methylated genomes; the proportion of methylated DNA in this group is variable but generally less than 50%. Pattern C represents globally methylated genomes; the small unshaded region in the C-type pattern denotes the CpG island fraction, which constitutes less than 3% of the genome. Data for the table are taken from references 10, 17, and 28 and include our unpublished data (34) on the bryozoan *Membranipora membranacea*, the annelid *Aphrodita aculeata*, the priapulid *Priapulus caudatus*, the nematode *Nippostrongylus brasiliensis*, and the nemertean worm *Lineus longissimus*. Abbreviations: invert, invertebrates; vert, vertebrates.

A significant fraction of invertebrate genes were found to be heavily methylated. This was surprising, as previous studies had shown that histone genes, 28S and 18S rRNA genes, and 5S rRNA genes of the sea urchin all resided in the nonmethylated fraction of the genome. There are two reports of methylated sites associated with invertebrate genes. First, in the sea urchin *S. purpuratus*, actin, bindin, and histone gene probes detected strong bands on Southern blots that were identical for the *Hpa*II and *Msp*I lanes, implying a lack of methylation. A subset of weaker bands, however, showed differences that were due to methylation (20). The patterns resemble those obtained in this study with actin and tubulin genes as probes (Fig. 6A, lane 3) in that complementary sequences are predominantly nonmethylated but a few members contain methylated sites. It is not known whether the methylated bands correspond to functional genes. A second report describes the only characterized example of gene methylation in an invertebrate (18). The esterase genes that are amplified in insecticide-resistant aphids show clear evidence of methylation on Southern blots. The remainder of the aphid genome is apparently free from methylation. Whether methylation of genes is part of the normal development of the aphid or is related to the mechanism of drug resistance is not known.

A striking feature of the extensive methylation of invertebrate genes uncovered here is that several unequivocal housekeeping genes, such as those for ribosomal proteins P2 and S6, a mitochondrial carrier protein, and HMG-CoA reductase, are

found in the methylated fraction of the genome whereas genes that may be expressed in a tissue-specific manner, such as those for lipase, insulin-like protein, spCOUP transcription factor, alkali myosin light chain, and spicule matrix protein SM30-α, belong to the nonmethylated compartment. We saw no evidence for variable methylation of genes in DNA derived from whole bodies of amphioxus. Sequences were either fully methylated or totally nonmethylated. The alkali myosin light chain gene, for example, has been shown to be expressed in a subset of cell types (22) but is uniformly nonmethylated in DNA from whole amphioxus. Mosaic methylation would be apparent if the gene was differentially methylated according to cell type. Similarly, transcripts for SM30-α are confined to the primary mesenchyme cells of the sea urchin embryo (21), but this gene is completely nonmethylated in DNA from sperm. These findings agree with those from previous work showing that embryonic histone genes belong to the nonmethylated fraction of the genome both when they are expressed and when they are silent (11). No evidence was found for switching of this or other sequences between the methylated and nonmethylated compartments of the genome in different sea urchin tissues or during development. A conventional view of the role of DNA methylation might predict that tightly regulated genes would make use of DNA methylation as a transcriptional control while housekeeping genes that are active in all cells would not benefit. Our findings contradict this expectation.

What is it that earmarks a fraction of the invertebrate genome for methylation but leaves the remainder unmodified? Speculation is hampered by our ignorance of the parameters that determine methylation patterns in any system (see reference 5). Triggering of de novo methylation by the presence of repetitive sequences cannot account for the presence of repeated genes (such as histone and 5S, 18S, and 28S rRNA genes) in the nonmethylated fraction of the genome. Neither does it explain why many single-copy housekeeping genes fall within the methylated fraction. The observation that the highly conserved rRNA genes are nonmethylated in all invertebrates that were tested but methylated in fish and amphibia compounds the uncertainty. One possible explanation that was raised previously involves replication timing (11). It has been argued that the vertebrate genome replicates in two phases, early and late (24). The banding of mammalian chromosomes may be a physical manifestation of this two-phase replication. By analogy, methylated and nonmethylated fractions of invertebrate genomes may represent late-replicating and early-replicating DNA sequences, respectively. According to this hypothesis, DNA methyltransferase would be present in the replication complex only late in S phase. Thus, only late-replicating DNA would become methylated.

It has been speculated that the total number of genes (or, more strictly, transcription units) is significantly higher in the vertebrates than in their invertebrate relatives (7, 9). In specific cases, we know that invertebrates, including amphioxus, have one copy of a gene that occurs in several distinct versions in vertebrates (14, 22, 23). The expansion is thought to have occurred predominantly via duplications of genes in the protovertebrate genome followed by divergence to generate novel family members (27). The capacity for more genes in the genome might have been made possible by improved mechanisms for damping transcriptional noise (7, 9). Low-density CpG methylation (about 1 methyl-CpG per 100 bp) of the kind found throughout globally methylated genomes has the characteristics that are expected of a noise reduction system. In model experiments, weak promoters were more severely affected than strong promoters by methylation at this density (12). The present results have implications for these ideas. First, the spread of methylation throughout the genome does, as far as we can tell, occur at the vertebrate-invertebrate boundary. Thus, it remains possible that vertebrate evolution was facilitated by innovations at the level of genome modification. A second relevant finding is that gene methylation was already well established at the time vertebrates arose and was not, therefore, a vertebrate invention. Several questions deserve further study. When did genes first become methylated? Since DNA methylation can repress transcription, how do methylated genes remain active? Are their promoters associated with methylation-free CpG islands? Future work is needed to answer these questions and perhaps illuminate the relationship between genomic DNA methylation patterns and the evolution of metazoan complexity.

ACKNOWLEDGMENTS

We thank Richard Meehan, Paco Antequera, and Donald Macleod for critical reading of the manuscript. We are grateful to L. Holland, P. Holland, M. Busslinger, P. Robson, D. Finnegan, F. Calzone, and M. Grunstein for providing DNA and cDNA libraries and to E. Viegas-Pequignot for the anti-m^5C monoclonal antibody. We are also grateful to Aileen Greig and Joan Davidson for technical assistance.

This work was supported by the Wellcome Trust and the Howard Hughes Medical Institute.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Antequera, F., D. Macleod, and A. P. Bird.** 1989. Specific protection of methylated CpGs in mammalian nuclei. Cell **85:**509–517.
3. **Antequera, F., M. Tamame, J. R. Villanueva, and T. Santos.** 1984. DNA methylation in the fungi. J. Biol. Chem. **259:**8033–8036.
4. **Antoine, M., and M. Fried.** 1992. The organisation of the intron-containing human S6 ribosomal protein (rpS6) gene and determination of its location at chromosome 9p21. Hum. Mol. Genet. **1:**565–570.
5. **Bestor, T., and B. Tycko.** 1996. Creation of genomic methylation patterns. Nat. Genet. **12:**363–367.
6. **Bird, A., M. Taggart, M. Frommer, O. J. Miller, and D. Macleod.** 1985. A fraction of the mouse genome that is derived from islands of non-methylated, CpG-rich DNA. Cell **40:**91–99.
7. **Bird, A., and S. Tweedie.** 1995. Transcriptional noise and the evolution of gene number. Philos. Trans. R. Soc. Lond. B **349:**249–253.
8. **Bird, A. P.** 1993. Functions for DNA methylation in vertebrates. Cold Spring Harbor Symp. Quant. Biol. **58:**281–285.
9. **Bird, A. P.** 1995. Gene number, noise reduction and biological complexity. Trends Genet. **11:**94–100.
10. **Bird, A. P., and M. H. Taggart.** 1980. Variable patterns of total DNA and rDNA methylation in animals. Nucleic Acids Res. **8:**1485–1497.
11. **Bird, A. P., M. H. Taggart, and B. A. Smith.** 1979. Methylated and unmethylated DNA compartments in the sea urchin genome. Cell **17:**889–901.
12. **Boyes, J., and A. Bird.** 1992. Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. EMBO J. **11:**327–333.
13. **Cedar, H., A. Solage, G. Glaser, and A. Razin.** 1979. Direct detection of methylated cytosine in DNA by use of the restriction enzyme Msp1. Nucleic Acids Res. **6:**2125–2132.
14. **Chan, S. J., Q.-P. Cao, and D. F. Steiner.** 1990. Evolution of the insulin superfamily: cloning of a hybrid insulin/insulin-like growth factor cDNA from amphioxus. Proc. Natl. Acad. Sci. USA **87:**9319–9323.
15. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:**156–159.
16. **Cooper, D. N., M. H. Taggart, and A. P. Bird.** 1983. Unmethylated domains in vertebrate DNA. Nucleic Acids Res. **11:**647–658.
17. **Cross, S., P. Kovarik, J. Schmidtke, and A. P. Bird.** 1991. Non-methylated islands in fish genomes are GC-poor. Nucleic Acids Res. **19:**1469–1474.
18. **Field, L. M., A. L. Devonshire, R. H. French-Constant, and B. G. Forde.** 1989. Changes in DNA methylation are associated with loss of insecticide resistance in the peach-potato aphid *Myzus persicae* (Sulz.). FEBS Lett. **243:**323–327.
19. **Freedman, T., and P. J. Pukkila.** 1993. De novo methylation of repeated sequences in *Coprinus cinereus*. Genetics **135:**357–366.
20. **Fronk, J., G. A. Tank, and J. P. Langmore.** 1992. DNA methylation pattern changes during development of a sea urchin. Biochem. J. **283:**751–753.
21. **George, N. C., C. E. Killian, and F. H. Wilt.** 1991. Characterisation and expression of a gene encoding a 30.6-kDa *Strongylocentrotus purpuratus* spi-

cule matrix protein. Dev. Biol. **147:**334–342.
22. **Holland, L. Z., D. A. Pace, M. L. Blink, M. Kene, and N. D. Holland.** 1995. Sequence and expression of amphioxus alkali myosin light chain (AmphiMLC-alk) throughout development: implications for vertebrate myogenesis. Dev. Biol. **171:**665–676.
22a. **Holland, P., and L. Holland.** Personal communication.
23. **Holland, P. W. H., J. Garcia-Fernandez, N. A. Williams, and A. Sidow.** 1994. Gene duplication and the origins of vertebrate development, p. 125–133. *In* M. E. Akam, P. W. H. Holland, P. W. Ingham, and G. Wray (ed.), The evolution of developmental mechanisms. The Company of Biologists Ltd., Cambridge, United Kingdom.
24. **Holmquist, G. P.** 1989. Evolution of chromosome bands: molecular ecology of noncoding DNA. J. Mol. Evol. **28:**469–486.
25. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 196–198. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Mohandas, T., R. S. Sparkes, and L. J. Shapiro.** 1981. Reactivation of an inactive human X-chromosome: evidence for X-inactivation by DNA methylation. Science **211:**393–396.
27. **Ohno, S.** 1970. Evolution by gene duplication. Springer-Verlag, Berlin, Germany.
28. **Rae, P. M. M., and R. E. Steele.** 1979. Absence of cytosine methylation at CCGG and GCGC sites in the rDNA coding regions and intervening sequences of *Drosophila* and the rDNA of other higher insects. Nucleic Acids Res. **6:**2987–2995.
29. **Reynaud, D., C. Bruno, P. Boullanger, J. Grange, S. Baresti, and A. Niveleau.** 1991. Monitoring of urinary excretion of modified nucleosides in cancer patients using a set of six monoclonal antibodies. Cancer Lett. **61:**255–262.
30. **Riggs, A. D., and G. P. Pfeifer.** 1992. X-chromosome inactivation and cell memory. Trends Genet. **8:**169–174.
31. **Satoh, N., and W. R. Jeffrey.** 1995. Chasing tails in ascidians: developmental insights into the origin and evolution of chordates. Trends Genet. **11:**354–359.
32. **Simoens, C. R., J. Gielen, M. Van Montagu, and D. Inze.** 1988. Characterization of highly repetitive sequences of *Arabidopsis thaliana*. Nucleic Acids Res. **16:**6753–6767.
33. **Singer, J., J. Roberts-Ems, F. W. Luthardt, and A. D. Riggs.** 1979. Methylation of DNA in mouse early embryos, teratocarcinoma cells and adult tissues of mouse and rabbit. Nucleic Acids Res. **7:**2369–2385.
34. **Tweedie, S., S. Cross, and A. Bird.** Unpublished data.
35. **Whittaker, P. A., A. McLachlan, and N. Hardman.** 1981. Sequence organization in nuclear DNA from *Physarum polycephalum*: methylation of repetitive sequences. Nucleic Acids Res. **9:**801–814.