J Mol Evol (2003) 56:597–606
DOI: 10.1007/s00239-002-2428-8

JOURNAL OF MOLECULAR EVOLUTION

Mytilus edulis Core Histone Genes Are Organized in Two Clusters Devoid of Linker Histone Genes

Werner Albig,[1] Ursula Warthorst,[1] Birgit Drabent,[1] Eva Prats,[2] Luis Cornudella,[2] Detlef Doenecke[1]
[1] Institut für Biochemie und Molekulare Zellbiologie, Universität Göttingen, Humboldtallee 23, D-37073 Göttingen, Germany
[2] Department of Molecular and Cell Biology, Institut de Biologia Molecular de Barcelona, CSIC, Jordi Girona 18-26, E-08034 Barcelona, Spain

Received: 15 July 2002 / Accepted: 12 December 2002

Abstract. Comparison of histone gene cluster arrangements in several species has revealed a broad spectrum of histone gene patterns. To elucidate the core histone gene organization in a mollusk, we have analyzed a *Mytilus edulis* genomic library and have isolated eight phage clones containing core histone genes. Analysis of insert DNA revealed that the core histone genes are arranged as regular gene repeats of all four core histones. The repeats do not contain linker histone genes. The clones are distributed into two groups of dissimilar repeated units with a common size of about 5.6 kb. The genes of each core histone class in the distinct repeats encode identical histone proteins and have comparable gene arrangements in the two repeat units. However, the intergenic sequences differ significantly. The core histone genes are organized as large clusters of about 100 repeats each. Previously, we have shown that the linker histone genes in *M. edulis* are also organized in a cluster of repeats of solitary H1 genes. Hence, this is the first case of a separate, clustered organization of both core and linker histone genes, repectively.

Key words: *Mytilus edulis* — Core histone genes — Gene organization — Nucleotide sequence

Introduction

Histones are the major protein constituents of eukaryotic chromatin. The family of histone proteins has been subdivided into the core histones (H2A, H2B, H3, and H4), which form the core particle of the nucleosome, and the linker histones (H1), which are involved in the generation of the higher-order chromatin structure (Thoma et al. 1979). The H3 and H4 histone proteins are among the most conserved eukaryotic proteins, and their genes as well as their arrangement have been conserved over a long period in evolution (Miller et al. 1993). For example, the H4 amino acid sequences of slime molds, plants, and mammals are nearly identical (Wells and Brown 1991). In humans, 14 genes code for H4 proteins with identical amino acid sequences (Albig and Doenecke 1997). In contrast to the highly conserved family of core histones, H1 histones are less evolutionary conserved. Seven H1 subtypes have been characterized and isolated from the genomes of human and mouse (Eick et al. 1989; Albig et al. 1991, 1997a; Drabent et al. 1991, 1993, 1995; Wang et al. 1997). In numerous organisms, core and linker histone genes are organized in clusters containing several copies of all five histone genes (Hentschel and Birnstiel 1981). Such gene clusters can be arranged in tandem repeats of genes encoding all five histones as in several invertebrate genomes (Maxson et al. 1983a). In the genome of the sea urchins *Strongylocentrotus purpuratus* and *Psammechinus miliaris*, several hundred quintets of histone genes are tandemly linked and are expressed at early embryonic stages (Hentschel and

The sequences for the repeat units of histone genes described in this paper have been deposited in the EMBL Nucleotide Sequence Database under the following accession numbers: UW06, AJ492923; UW15, AJ492924; and UW27, AJ492925.
Correspondence to: Dr. D. Doenecke; *email:* ddoenec@gwdg.de

Birnstiel 1981; Kedes 1979; Sures et al. 1978). The sea urchin histone genes that are expressed later in embryogenesis are clustered with no regular organization, while some histone genes appear dispersed throughout the genome (Maxson et al. 1983b). In the chicken, murine, and human genomes, histone genes are clustered but not tandemly repeated, and the H1 histone genes are located within these large clusters together with core histone genes (Heintz et al. 1981; D'Andrea et al. 1985; Albig et al. 1997a, b; Wang et al. 1997). In the annelids *Chaetopterus variopedatus* and *Platynereis dumerilii*, histone gene clusters of all five histones and additional clusters of only core histones without H1 genes have been reported (Sellos et al. 1990; del Gaudio et al. 1998). Gene clusters containing exclusively core histones, with no H1 genes, have also been described in the genomes of the newt *Notophthalmus viridescens* (Stephenson et al. 1981), the sea stars *Pisaster ochraceus*, *P. brevispinus*, and *Dermasterias imbricata* (Cool et al. 1988), and the coral *Acropora formosa* (Miller et al. 1993), albeit with variable histone gene arrangements.

In the fruit fly *Drosophila virilis*, core histone gene clusters without H1 genes have been found in addition to tandemly arranged quintets of all five histone genes (Domier et al. 1986). In contrast, histone gene clusters in *Drosophila melanogaster* consist exclusively of quintets containing H1 and core histone genes (Lifton et al. 1977). In the genome of the nematode *Caenorhabditis elegans* dispersed clusters of only core histone genes have been found (Roberts et al. 1987). Moreover, in *Volvox carteri* (Lindauer et al. 1993), *Caenorhabditis elegans* (Sanicola et al. 1990), and *Arabidopsis thaliana* (Gantt and Lenvik 1991; Ascenzi and Gantt 1997), single-copy genes of H1 separated from core histone genes have been described. In human (Albig et al. 1993), mouse (Brannan et al. 1992), and rat (Walter et al. 1996), the genes coding for the histone H1° are located solitarily within the genome. Sellos et al. (1990) have described core histone genes tandemly repeated in the annelid *Platynereis dumerilii* and have compared the cluster arrangements in several species. Summarizing the evolutionary picture emerging from these data, the authors conclude that regularly repeated histone gene quintets are characteristic of echinoderms and arthropods.

In a previous report, we studied the arrangement of the linker histones in the genome of the marine mollusk *Mytilus edulis*, which has been the subject of several studies in chromatin research. This work on linker histone genes has shown that all H1 genes are arranged in tandem with no neighboring core histone genes (Drabent et al. 1999). Here, the organization of the core histone genes within the genome of this bivalve mollusk is described. The core histone genes are organized in clusters, each containing all four core histone genes repeated in tandem. There are two clusters, with a similar arrangement of the coding sequences but with distinct intergenic sequences.

Materials and Methods

Isolation and Sequence Analysis of Mytilus edulis *Histone Genes*

A *Mytilus edulis* genomic library in the λGEM-11 bacteriophage (Promega, Madison, WI) was prepared from sperm DNA and screened with a human H4 histone gene probe (Albig et al. 1991). Hybridization was performed under low-stringency conditions (60°C, 5 × SSPE [1 × SSPE: 0.18 *M* NaCl, 0.01 *M* sodium phosphate, 0.001 *M* EDTA]). Eight phage clones hybridizing with the probe were isolated and further analyzed by restriction enzyme digestion and Southern blotting. The fragments hybridizing with core histone gene probes were isolated and subcloned into the pUC19 plasmid vector. Sequence analysis was carried out by Dye Terminator Cycle sequencing with an ABI 373A DNA Sequencer (Applied Biosystems, Foster City, CA). Alignments of the sequences were obtained using the DNA Star and BLAST software.

Restriction Mapping of the Phage Inserts

Restriction mapping of the phage inserts was performed following Smith and Birnstiel (1976) with some modifications. Briefly, 10–15 μg of phage DNA was completely digested with *Sfi*I to release inserts from the phage arms. These were then partially digested with *Sac*I, *Pst*I, *Hin*dIII, and *Eco*RI, respectively. The resulting restriction fragments were separated by gel electrophoresis and blotted onto a nylon membrane. The Southern blot was hybridized with radioactively labeled oligonucleotide probes for the Sp6 and T7 promoter sites, delimiting each end of the inserts. Size analysis of the labeled restriction fragments revealed the arrangement of the *Sac*I fragments within phage inserts.

Determination of the Copy Number of the Repeat Unit within the Genome

Four, 2, 1, 0.5, and 0.2 μg *of Mytilus edulis* genomic DNA and, for comparison purposes, 0.5, 0.2, 0.1, 0.05, 0.02, and 0.01 ng of plasmid DNA from pUW06 and pUW15, containing one copy of each of the *Mytilus edulis* core histone genes on a 4.8 or 5.6-kb *Sac*I fragment, were digested with *Sac*I and blotted onto a nylon membrane. Considering a haploid DNA complement for *Mytilus edulis* of 1.5×10^9 bp (Ruiz-Lara et al. 1992), 1 μg of genomic DNA corresponds to approximately 6×10^5 *Mytilus edulis* genome equivalents. Because the size of pUW06 and pUW15 is 8.2 kb each, 1 ng of plasmid DNA corresponds to 1.2×10^8 copies of each of the core histone genes. The blot was subsequently hybridized with the radioactively labeled *Mytilus edulis* H4 probe and the hybridization intensity was measured with a Phosphor Imager and evaluated with the software Image-Quant (Amersham Biosciences, Freiburg). The quantitation was done in three independent experiments. The linearity of the intensities of the hybridization signals was verified by a dose–response curve.

Results

Isolation of Clones Containing Core Histone Gene Repeats from a Mytilus edulis *Genomic Library*

For the isolation of the core histone genes from *Mytilus edulis*, a genomic library was analyzed with a

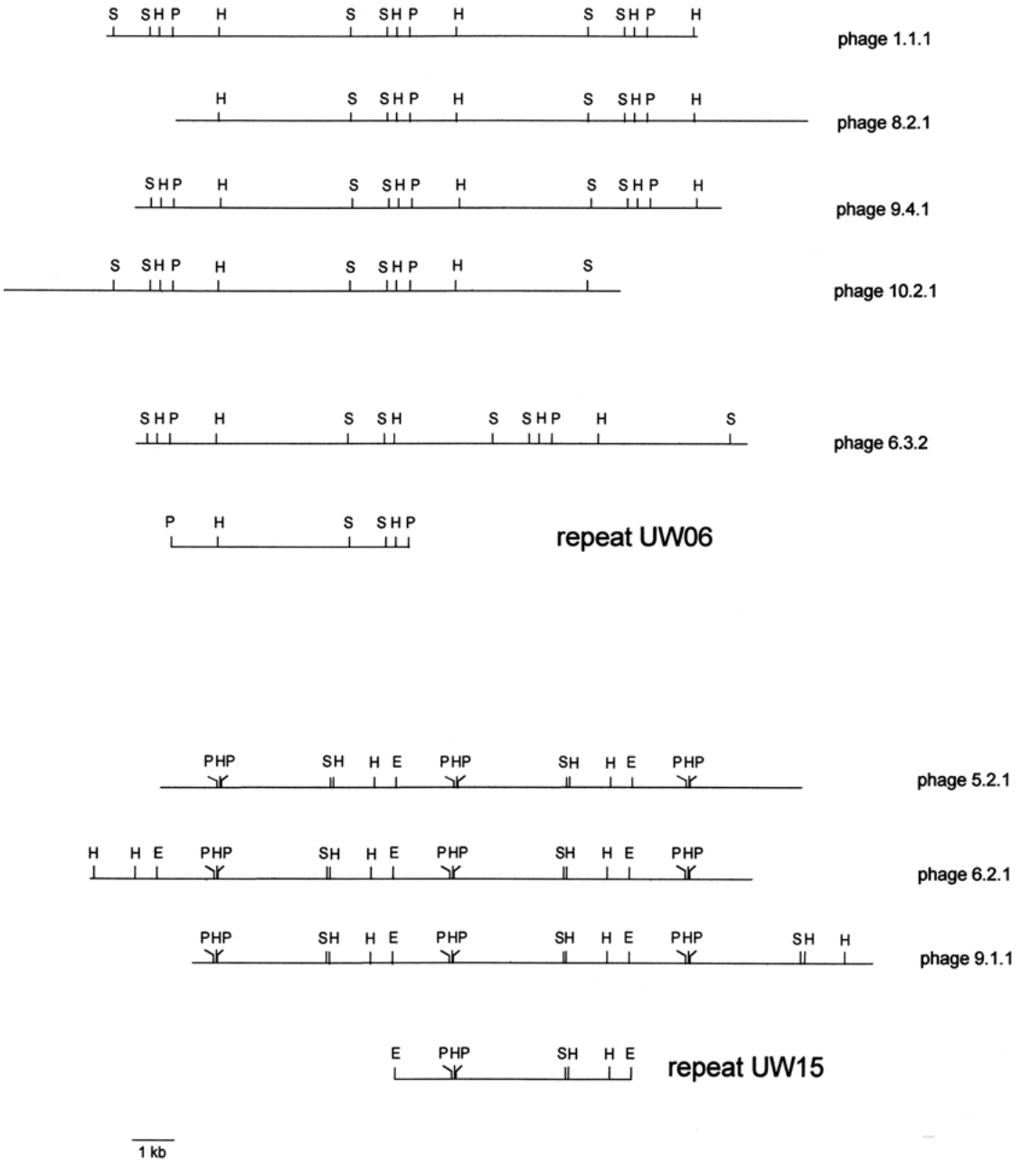

Fig. 1. Restriction maps of the inserts of the phage clones containing core histone genes. Recombinant phages were isolated from a *Mytilus edulis* genomic library. The phage inserts were mapped according to the method described by Smith and Birnstiel (1976), with the restriction enzymes *Sac*I (S), *Pst*I (P), *Hin*dIII (H), and *Eco*RI (E). The repeat units of the two types (UW06 and UW15) are shown below the corresponding restriction maps.

human H4 histone gene probe (Albig et al. 1991). The screening yielded eight phage clones. Restriction enzyme analysis with *Sac*I, *Pst*I, and *Sac*I/*Pst*I revealed that the clones could be distributed into two groups according to size. The inserts of one group contained identical 4.8 and 0.8-kb *Sac*I fragments, whereas the other group contained only 5.6-kb *Sac*I fragments. The first group comprised five clones, and the second three. All clones were further analyzed by restriction enzyme mapping and Southern blot hybridization, which was performed with recombinant phage DNA digested with the restriction enzymes *Sac*I and *Pst*I and the combination thereof (*Sac*I/*Pst*I). Hybridizations were carried out with the human H1.3 probe (Albig et al. 1991) and with probes for the human core histones H2A, H2B, H3, and H4 (Albig et al. 1997b). None of the eight phage DNA fragment patterns hybridized with the H1 histone gene probe, whereas all phage clones yielded insert fragments that hybridized with all four core histone gene probes. The size pattern of the hybridizing fragments confirmed the subdivision into two groups.

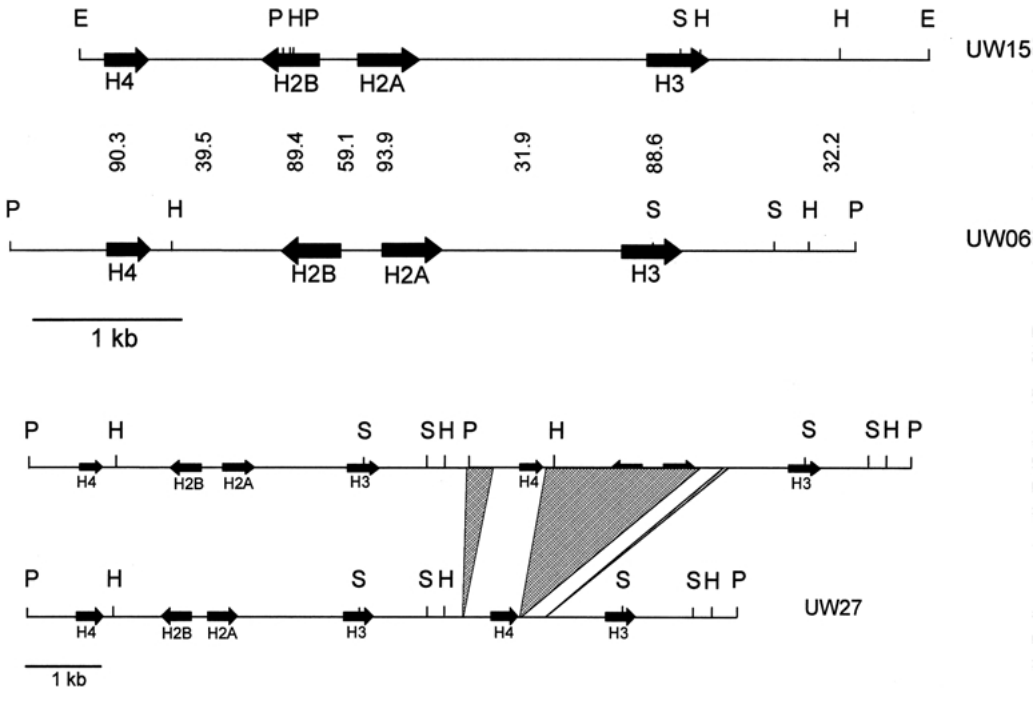

Fig. 2. Histone gene arrangement in different repeat units. Restriction enzyme cleavage sites are indicated as in Fig. 1. *Bold arrows* denote the coding regions of the core histone genes. **Top:** Comparison of the two repeat units. *Vertical numbers* indicate the nucleotide sequence similarity of the corresponding genes and the intergenic regions between the two repeat units (UW06 and UW15), expressed as a percentage. **Bottom:** Interpretation of the origin of the aberrant *Sac*I fragment UW27 of the recombinant phage 6.3.2. UW27 is the result of three deletions from two adjacent UW06 repeats. Deleted regions are indicated as *hatched triangles*. Note: The size scale for the bottom is half that for the top.

Phage inserts were mapped in detail by partial digestion with *Sac*I and hybridization with probes for each of the phage arms, following Smith and Birnstiel (1976). Partial *Sac*I digestion clearly revealed a repeated structure in the insert, with several *Sac*I fragments within the DNA of each phage (electrophoresis data not shown). Almost all phage inserts contained *Sac*I fragments of the same size as shown in Fig. 1. The phage inserts 1.1.1, 8.2.1, 9.4.1, and 10.2.1 contained alternating *Sac*I fragments of 4.8 and 0.8 kb, whereas the phage inserts 5.2.1, 6.2.1, and 9.1.1 contained multiple 5.6-kb fragments. One of the phage inserts (6.3.2) yielded a restriction fragment pattern different from that of the two prototypes (Fig. 1). It displayed an additional *Sac*I fragment of 2.5 kb interspersed within the regular arrangement of the *Sac*I fragments of 4.8 and 0.8 kb found in the corresponding cluster type. Sequence comparison indicates that this is the result of three deletions from the regular 4.8-kb *Sac*I fragment as shown in Fig. 1.

Characterization of the Core Histone Genes

Various *Sac*I fragments hybridizing with the core histone probe were isolated, subcloned, and mapped with several restriction enzymes. Fragments of the same size revealed an identical restriction enzyme pattern, whereas fragments of different sizes revealed distinct but similar restriction patterns (not shown).

The observed differences in fragment lengths suggest that phage inserts contain genomic DNA tracts from core histone clusters with distinct repeating units or distinct parts of the same gene cluster. It was difficult to sort out the fragments of various lengths in phages with several *Sac*I fragments of the same size. Since end-sequencing of a set of subcloned 5.6-kb *Sac*I fragments from several clones revealed identical sequences, we further analyzed the repeat unit for each group: UW06 for the *Sac*I 4.8- + 0.8-kb repeat group and UW15 for the *Sac*I 5.6-kb group, together with the interspersed 2.5-kb *Sac*I fragment of the 4.8-kb repeat unit of phage 6.3.2. The sequence analysis revealed a similar arrangement of the histone genes within the two repeat units. The coding sequences of the corresponding core histone genes were also similar in the two repeats, but there was little resemblance as regards intergenic regions (Fig. 2). The open reading frames (ORFs) of the corresponding core histone genes encoded the same primary structures. All core histone genes show the characteristic regulatory elements found in all replication-expressed histone genes (see Fig. 5): (1) a CAAT box, (2) a TATA box (exception, H3), and (3) the palindromic consensus sequence that forms the

A

Histone H2A

```
                    N-terminal domain                          globular domain
                   1----------------------#-------------------------------------------------------80
human              MSGRGKQGGKARAKAKTRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYLAAVLEYLTAEILELAGNAARDNKKTRII
mouse              MSGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYMAAVLEYLTAEILELAGNAARDNKKTRII
sea urchin         MSGRGKGAGKARAKAKSRSARAGLQFPVGRVHRFLRKGNYAQRVGAGAPVYLAAVLEYLAAEILELAGNAARDNKKTRII
starfish           MSGRGK-GGKARAKAKSRSSRAGLQFPVGRVHRFLRKGNYAERVGAGAPVYLAAVMEYLAAEILELAGNAARDNKKTRIN
spoon worm         MSGRGK-GGKAKGKSKSRSSRAGLQFPVGRIHRLLRKGNYAERIGAGAPVYLAAVMEYLAAEVLELAGNAARDNKKTRII
polychaete worm    MSGRGK-GGKAKGKSKTRSSRAGLQFPVGRIHRLLRKGNYAERARAGAPAYLAAVMEYLAAEVLELAGNAARDNKKSRII
cuttlefish         MSGRGK-GGKVKGKSKTRSSRAGLQFPVGRIHRLLRKGNYAQRVGAGAPVYLAAVMEYLAAEVLELAGNAARDNKKSRII
Mytilus edulis     MSGRGK-GGKAKAKAKSRSSRAGLQFPVGRIHRLLRKGNYAERVGAGAPVYLAAVLEYLAAEVLELAGNAARDNKKSRII

                              C-terminal domain
                   81--------#-----------------------------------130          variances
human              PRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK                17
mouse              PRHLQLAVRNDEELNKLLGGVTIAQGGVLPNIQAVLLPKKTESHKPGKNK                17
sea urchin         PRHLQLAVRNDEELNKLLSGVTIAQGGVLPNIQAVLLPKKT-S---KASK                13
starfish           PRHLQLAIRNDEELNKLLSGVTIAQGGVLPNIQAVLLPKKTA----KAAK                11
spoon worm         PRHLQLAIRNDEELNKLLSGVTIAQGGVLPNIQAVLLPKKS-SQ--K-TK                 9
polychaete worm    PRHLQLAIRNDEELNKLLSGVTIAQGGVLPNIQAVLLPKKT--Q--KPAK                 8
cuttlefish         PRHLQLAIRNDEELNKLLSGVTIAQGGVLPNIQAVLLPKKT--Q--KAAK                 6
Mytilus edulis     PRHLQLAIRNDEELNKLLSGVTIAQGGVLPNIQAVLLPKKT--Q--KAAK
```

B

Histone H2B

```
                         N-terminal domain                          globular domain
                   1-------------------------------------#-----------------------------------------80
human              M-PEPAKSAPAPKKGSKKAVTKAQKKDGKKRKRSRKESYSVYVYKVLKQVHPDTGISSKAMGIMNSFVNDIFERIAGEAS
rat                M-PEPAKSAPAPKKGSKKAVTKAQKKDGKKRKRSRKESYSVYVYKVLKQVHPDTGISSKAMGIMNSFVNDIFERIAGEAS
chicken            M-PEPAKSAPAPKKGSKKAVTKTQKKGDKKRRRTRKESYSIYVYKVLKQVHPDTGISSKAMGIMNSFVNDIFERIAGEAS
sea urchin         M-P-AKAQA-AGKKGSKKA-KAPKPSGDKKRRRKRKESYGIYIYKVLKQVHPDTGISSRAMSIMNSFVNDVFERIAAEAS
Chironomus         MAP-PKTSGKAAKKAGKAQ-KAISKDDKKRRRHKRKESYAIYIFKVLKQVHPDTGVSSKAMSIMNSFVNDIFERIAAEAS
copepod            M-P-PKVSGKAAKKAGKAQ-KNISKGDKKKNR-KRKESYAIYIYKVLKQVHPDTGISSKAMSIMNSFVNDIFERIASEAS
polychaete worm    M-P-PKVSSKGAKKAVSKS-KAPRSG-DKKKRRKRRESYSIYVYKVMKQVHPDTGISSKAMSIMNSFVNDLFERIASEAS
Mytilus edulis     M-P-PKVGTKGAKKAVTKA-KTARPGGDKKRRRKRRESYAIYIYKVLRQVHPDTGVSSKAMSIMNSFVNDIFERIAAEAS

                                    C-terminal domain
                   81--------------------#---------------------127          variances
human              RLAHYNKRSTITSREIQTAVRLLLPGELAKHAVSEGTKAVTKYTSSK                28
rat                RLAHYNKRSTITSREIQTAVRLLLPGELAKHAVSEGTKAVTKYTSSK                28
chicken            RLAHYNKRSTITSREIQTAVRLLLPGELAKHAVSEGTKAVTKYTSSK                26
sea urchin         RLAHYNKKSTITSREVQTAVRLLLPGELAKHAVSEGTKAVTKYTTSK                22
Chironomus         RLAHYNKRSTITSREIQTAVRLLLPGELAKHAVSEGTKAVTKYTSSK                18
copepod            RLAHYNKRSTITSREIQTAVRLLLPGELAKHAVSEGTKAVTKYTSSK                17
polychaete worm    RLAHYNKRSTITSREIQTAVRLLLPGELAKHAVSEGTKAVTKYTSSK                14
Mytilus edulis     RLAHYNKRSTITSREIQTAVRLLLPGELAKHAVSEGTKAVTKYTSSK
```

Fig. 3. Alignment of the amino acid sequences of histones H2A (**A**) and H2B (**B**) from *Mytilus edulis* with a selected set of homologues from other eukaryotes. Residues other than those of the *M. edulis* histones are shown by *italicized letters*; identical residues, by *boldface letters*; deletions are indicated by *light dashes*; *wild cards*, by *heavy dashes*. The variability is expressed as the number of variant amino acids together with insertions and/or deletions with respect to the *M. edulis* histone proteins. Boundaries between N termini and globular domains and between the latter and C termini are indicated by *pound signs* above the alignments. Sequences are from the NCBI Protein database and are identified by their respective accession numbers. **H2A histones:** human, *Homo sapiens*, XP_004457; mouse, *Mus musculus*, AAB04767; sea urchin, *Strongylocentrotus purpuratus*, P16886; starfish, *Asterias rubens*, P02269; spoon worm, *Urechis caupo*, P27325; polychaete worm, *Chaetopterus variopedatus*, AF007904; and cuttlefish, *Sepia officinalis*, P02268. **H2B histones:** human, *Homo sapiens*, XP_011484; rat, *Ratus norvegicus*, 223096; chicken, *Gallus gallus*, A30221; sea urchin, *Strongylocentrotus purpuratus*, P16889; chironomus, *Chironomus thummi thummi*, P21897; copepod, *Tigriopus californicus*, P35069; and polychaete worm, *Chaetopterus variopedatus*, AAC15915.

stem–loop motif responsible for proper 3′ histone mRNA processing (Birnstiel et al. 1985).

The coding regions of the core histone genes consist of ORFs for a 124-amino acid (aa) H2A histone, a 123-aa H2B histone, a 135-aa H3 histone, and a 102-aa H4 histone. Sequence comparison of the amino acid sequences with the histone sequence database revealed that the H4 and H3 histones are

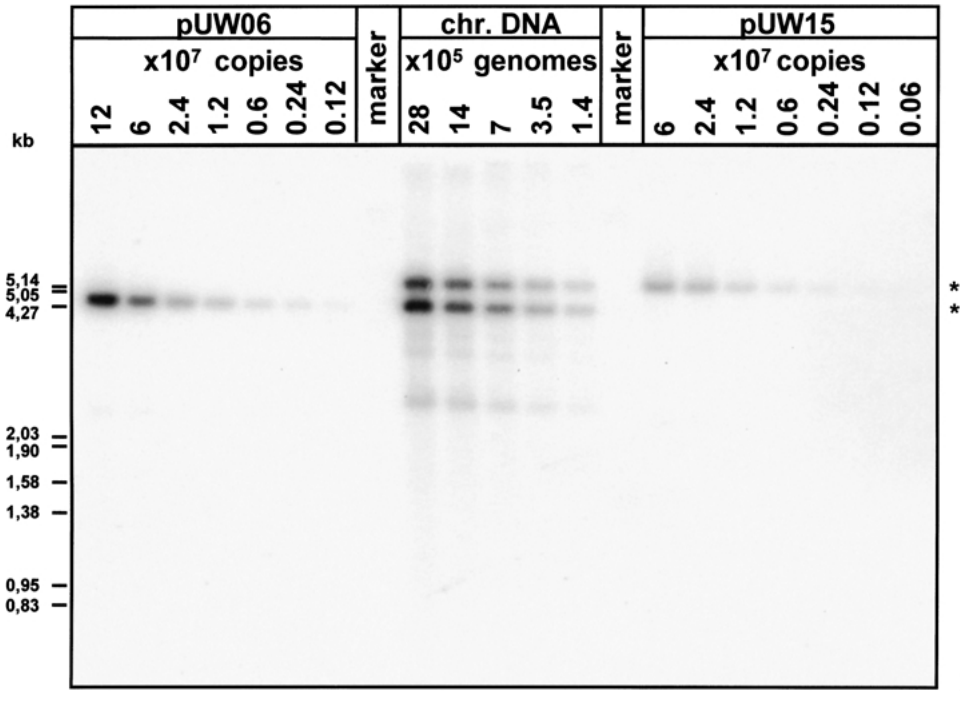

Fig. 4. Determination of the copy number of the core histone genes within the *Mytilus edulis* genome. DNA samples, ranked in the order of the decreasing amounts shown at the *top*, of *Mytilus edulis* genomic DNA and, for calibration, plasmid DNA pUW15 and pUW06, each containing one copy of the corresponding *Mytilus edulis* repeat unit, respectively, were digested with *Sac*I and blotted onto a nylon membrane. Considering that the haploid DNA complement of *Mytilus edulis* has been estimated as 1.5×10^9 bp (Ruiz-Lara et al. 1992) and the size of the plasmids is 8.2×10^3 bp, the copy numbers of genomes and plasmids, respectively, are calculated from the amount of DNA applied on the gel and are indicated at the top. The Southern blot was hybridized with the radioactively labeled 5.6-kb *Sac*I fragment of the *Mytilus edulis* core histone repeat unit (pUW15). Hybridization intensity was measured with a Phosphor Imager and quantified with the software Image-Quant (Molecular Dynamics). The asterisks indicate the bands used for calculation of the copy number.

identical to the H4 and H3 proteins of several species as divergent as mammals and nematodes. The H3 histone sequence is identical to the H3.1 variant of mammals.

The H2A and H2B histones differ from other known H2A and H2B histone proteins. The H2A histone shows six aa exchanges compared with the most closely related H2A from *Sepia officinalis* (Fig. 3A). The variations are clustered within the N-terminal domain of the H2A histone. The similarity of the *Mytilus* H2A to the human H2A sequences is even slighter, since there are 17 exchanges and the largest difference is found within the distal C terminus, where the mammalian H2A contains an octapeptide that is totally absent in the mollusk as well as in the other invertebarates used for comparison. The aa sequence conservation was reasonably high in the globular domains of histone H2A and histone H2B. The *Mytilus* H2B histone shows the greatest similarity to the H2B from the annelid *Chaeopterus variopedatus*, a polychaete worm (Fig. 3B). Again, the variability concentrates within the N-terminal domain, with 8 of the 14 overall changes. The N terminus of the mollusk H2B differs substantially from that of the vertebrate H2B, as 22 of the 36 aa comprising the N-terminal domain differ in these two species.

Copy Number of the Core Histone Gene Repeat Units in the Mytilus edulis *Genome*

During the screening about 20 copies of each repeat unit of the two cluster types according to size were isolated. To assess the total copy number of the core histone genes in the *Mytilus edulis* genome, increasing amounts of genomic DNA and, for comparison purposes, graded amounts of plasmid DNA from each cluster group containing one copy of the corresponding repeat units were digested with *Sac*I and transferred to a nylon membrane (Fig. 4). The hybridization intensity of the chromosomal DNA was assessed and compared with that of the isolated core histone gene repeat units. Using this method the copy number of each core histone gene repeat was estimated to be about 100 copies per haploid genome, with an estimated uncertainty of about 20%. Therefore the gene copy number for each core histone is about 200 copies (about 100 of each repeat).

The hybridization of probes from each core histone gene unit to genomic DNA cleaved with several

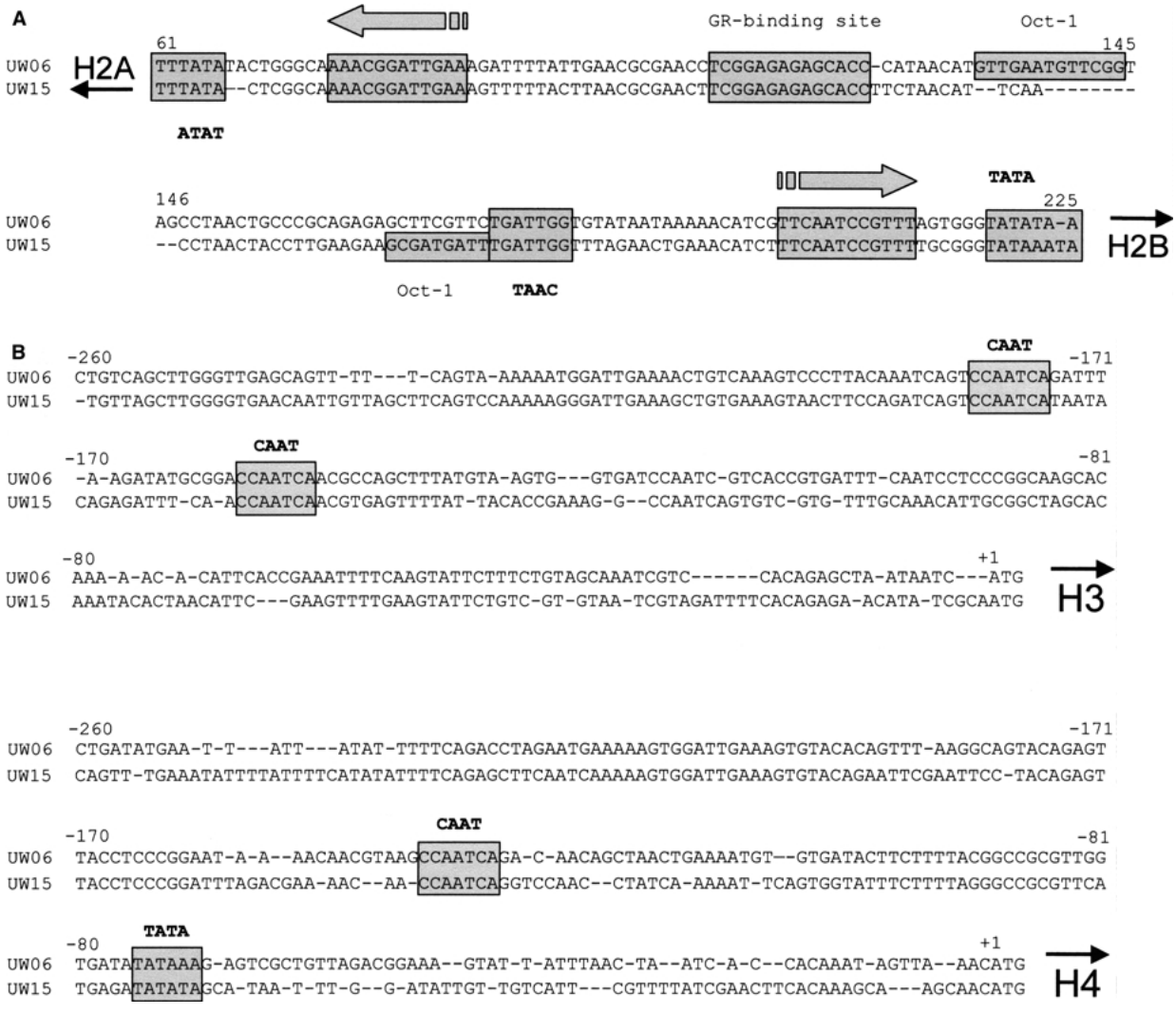

Fig. 5. Nucleotide sequence comparison of the core histone gene promoters from the different repeat units. **A** Comparison of the divergent promoter region of the H2A/H2B gene pair. The nucleotide sequences of the two TATA motifs are *boxed*. Consensus binding sequence motifs of transcription factors are identified by the TRANSFAC 4.0 software and indicated by *hatched boxes*. GR-binding site, glucocorticoid binding site; Oct-1, consensus binding site for the octamer transcription factor 1. The two dodecanucleotides with opposite orientations are also *boxed* and the orientation is indicated by the *arrows* above. The nucleotides are numbered in relation to the start codons of the H2A genes. **B** Comparison of the H3 and H4 promoter regions. Only the consensus binding sequences of two factors were identified by the software. The TATA and CAAT boxes are indicated as *gray boxes*. The nucleotides are numbered in relation to the start codons of the respective genes.

restriction endonucleases revealed no signal besides those of the two repeat units. However, hybridization with the H3 and H4 probe generated a third signal resulting from the additional 2.5-kb *Sac*I repeat (clone UW27) interspersed in the UW06 repeat type. In summary, we conclude that the vast majority of core histone genes from *M. edulis* are organized into the two repeat units UW06 and UW15.

As reported elsewhere upon analysis of the chromosomal organization of the H1 histone gene cluster in *Mytilus edulis* by Southern blot hybridization of large genomic restriction fragments (Drabent et al. 1999), the H4 histone genes are located in two chromosome regions unrelated to H1 genes. This, together with the results presented here, indicates that core histone gene repeats do not colocalize with H1 genes and are not randomly dispersed but clustered within the *Mytilus* genome. However, it cannot be safely assessed whether the core histone genes in the mussel genome are organized in one or two clusters or whether the two repeat units colocalize or are located on distinct genomic regions.

Analysis of Promoter Regions

The major difference between the *Sac*I repeats of the two core histone gene units was their length (Fig. 2), since the genes within the repeat units encoded almost-identical primary structures. This prompted a

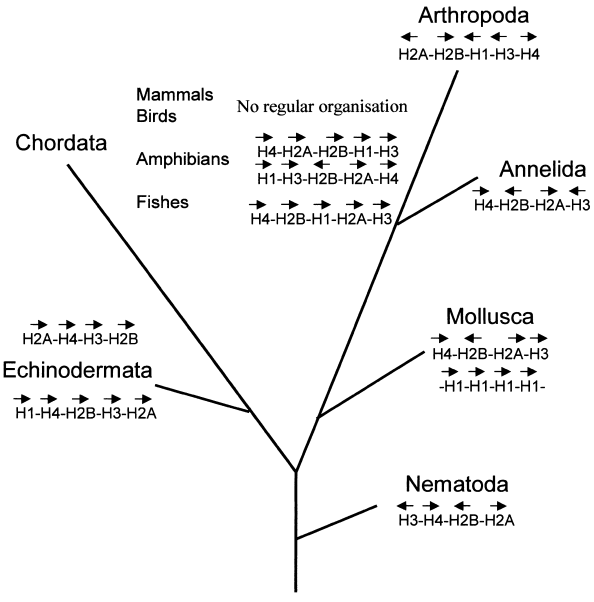

Fig. 6. Organization of the histone genes in several species. Genomic organization of clustered histone genes in the genomes of a representative member of each phylum of this simplified phylogenetic tree (following Sellos et al. 1990; supplemented by data from this paper and Drabent et al. [1999]). Trivial names are shown in *light letters*. *Arrows* indicate the direction of transcription.

detailed analysis of the DNA tracts immediately upstream of the coding sequences. The H2A and H2B genes appear to be transcribed in opposite orientations and regulated by a divergent promoter as in most eukaryotic genomes (Trappe et al. 1999). Comparison of the consensus binding sequences within this promoter with the elements found in the H2A–H2B human promoter revealed sharp differences. We only identified a CAAT box, potential binding sites for the Oct-1 transcription factor, and a sequence homologous to a corticoid receptor binding motif (Fig. 5A). Strikingly, a conserved dodecanucleotide motif immediately upstream of the TATA boxes, identical in both the H2A and the H2B genes, was also detected. Motif searches by computer, however, failed to identify this sequence as a binding motif for a known transcription factor. Comparison with binding motifs of transcription factors identified in the promoters of histone genes indicates a distant resemblance to the conserved HiNF-P/H4TF-2 binding site found in the promoter of vertebrate H4 genes (van Wijnen et al. 1992). In the promoter region of the H4 genes, only the consensus sequences of the TATA box and the CAAT box could be identified (Fig. 5B). We could not identify a real TATA box in the promoters of the H3 genes. These promoters comprise only two classical CAAT boxes with the extension CA, which is typical for CAAT boxes in histone gene promoters (Meergans et al. 1998). The CAAT box is the binding site for NF-Y (Mantovani et al. 1999) and the phylogenetically conserved histone gene HiNF-D/CDP-cut (van Wijnen et al. 1996).

Discussion

In this study, we have isolated and analyzed several phage clones from a *Mytilus edulis* genomic library containing complete sets of core histone genes. The analysis has revealed that the inserts encompass two different core histone gene repeats, containing each of the four core histone genes. These repeats are organized in clusters as in most eukaryotes. The arrangement of histone genes within the clusters can vary from one species to another (see Fig. 6 and Maxson et al. [1983a]). In birds (D'Andrea 1985; Tönjes and Doenecke 1987), human (Albig et al. 1997b; Albig and Doenecke 1997), and mouse (Sittman et al. 1981; Wang et al. 1997), histone gene clusters are irregularly distributed, whereas in amphibians, fish, and invertebrates histone genes are found in clusters of tandemly repeated units of all five histone subtypes (Maxson et al. 1983a). In the sea urchin, most histone genes are clustered in tandem repeats and contain one copy of each of the five histone genes (Schaffner et al. 1978; Sures et al. 1978; Kedes 1979; Hentschel and Birnstiel 1981; Maxson et al. 1983b). The trout (*Salmo gairdnerii*) genome also contains tandem repeats of clustered histone genes (Connor et al. 1984), although they are arranged differently. In both organisms, histone mRNAs are transcribed from the same DNA strand. In contrast, the fruit fly *Drosophila melanogaster* displays clusters with all five histone genes, where the H3, H2A, and H1 genes are transcribed from one strand and the H2B and H4 genes are transcribed from the other (Lifton et al. 1977). Several types of histone gene organization have been reported in amphibians. In the newt *Notophthalmus viridescens*, histone genes are organized in a major cluster containing all five histone genes and the genes are transcribed from both DNA strands (Stephenson et al. 1981). In *Xenopus laevis*, some clusters consist of all five histone genes, while other histone gene clusters lack H1 genes (van Dongen et al. 1981; Turner and Woodland 1983). In some clusters, all genes are located on the same strand, while in others the genes are distributed on both strands.

Core histone gene clusters with no neighboring H1 genes have been described in several organisms. Various types of core histone gene clusters have been found in the nematode *Caenorhabditis elegans* (Roberts et al. 1987). H1 coding sequences have not been detected in any of these clusters and no repeated organization of the clusters has been observed. In addition, the histone genes are transcribed from both DNA strands. Tandem repeats of clusters with only

core histone genes at high copy numbers have been reported in several sea star species, as detailed in the Introduction. In the marine annelid *Platynereis dumerilii*, core histone genes are clustered, showing a highly conserved gene arrangement within the clusters. These genes are tandemly arrayed and transcribed from both strands (Sellos et al. 1990). In contrast, the annelid *Chaetopterus variopedatus* shows histone gene clusters with all five histone genes including H1 genes (del Gaudio et al. 1998). Furthermore, clusters of core histone genes devoid of H1 genes have been identified in the coral *Acropora formosa* (Miller et al. 1993). In none of these cases have tandem repeats of H1 genes been described. However, the genomes of *Volvox carteri* (Lindauer et al. 1993), *Caenorhabditis elegans* (Sanicola et al. 1990), and *Arabidopsis thaliana* (Gantt and Lenvik 1991; Ascenzi and Gantt 1997) contain H1 genes that are separated from core histone genes and individually dispersed within the genome. None of these genes are arranged in tandemly repeated units. In this respect, the mollusk *Mytilus edulis* is an exception. Its genome, as reported here, shows two separate groups of repeats, one containing the four core histone genes and the other solitary H1 linker histone genes.

The genomic content of the *Mytilus edulis* core histone genes has been calculated to be about 200 copies (about 100 copies of each type of repeat) per haploid genome. This figure is in excellent agreement with the 100 copies of H1 histone genes reported previously (Drabent et al. 1999), in accordance with the 2:1 stoichiometry of the core and linker histones in the chromatin. Moreover, hybridization of genomic DNA digested with rare cutting restriction enzymes with a histone H4 probe revealed two hybridizing fragments (Drabent et al. 1999). The total size of these two fragments was calculated to be about 1000 kb. Since the lengths of the core histone gene repeat units are 5.6 kb, the maximum number of core histone genes cannot exceed 200 copies.

The finding of two clusters of core histone gene repeat units reported here is also consistent with the results of the hybridization of restricted genomic DNA with the histone H4 probe. However, the data do not allow us to conclude whether the two tandemly repeated units are associated in a large cluster or whether, on the contrary, they arise from distinct clusters, each comprising one type of repeat unit.

The strong conservation of the structural elements CAAT and TATA boxes and the 3′-end palindromic sequence, together with the arrangement of the core histone gene promoters and the opposite orientation of the H2A and H2B genes, indicates that this is the optimal organization for the regulation of expression of the histone gene repeats. Likewise, the conservation of the core histone gene sequences can be considered a requirement of their key role in the organization of nucleosomal chromatin and the regulation of eukaryotic genes. The particular conservation of the amino acid sequences of the histone globular domains supports their function in chromatin organization, whereas the sequence differences found in the N- and C-terminal regions of the proteins suggest the modulation of chromatin-mediated gene regulation in the mussel compared with other species. The structural similarities of histone genes and the homology between the individual histone proteins clearly support a shared ancestry.

The results shown in this report, in combination with those reported in our previous paper, describe for the first time the organization of the histone genes of a bivalve mollusk. This fills a gap in the knowledge of histone gene organization in the different lineages of the phyla (see Fig. 6). To our knowledge, it is also the first case in which core histone genes and H1 linker histone genes appear to be organized in separate tandem repeats.

Acknowledgments. This work was supported by EC Grant SC1-CT91-0693 to D.D. and L.C. Support from the Spanish DGES (PB97-1136) and the Catalan DGR (SGR01-354) within the framework of the Centre de Referència en Biotecnologia de la Generalitat de Catalunya is also acknowledged. The technical assistance of Ms Diana Otto is gratefully acknowledged.

References

Albig W, Doenecke D (1997) The human histone gene cluster at the D6S105 locus. Hum Genet 101:284–294

Albig W, Kardalinou E, Drabent B, Zimmer A, Doenecke D (1991) Isolation and characterization of two human H1 histone genes within clusters of core histone genes. Genomics 10:940–948

Albig W, Drabent B, Kunz J, Kalff-Suske M, Grzeschik K-H, Doenecke D (1993) All known human H1 histone genes except the H1° gene are clustered on chromosome 6. Genomics 16:649–654

Albig W, Meergans T, Doenecke D (1997a) Characterisation of the H1.5 gene completes the set of human H1 subtype genes. Gene 184:141–148

Albig W, Kioschis P, Poustka A, Meergans K, Doenecke D (1997b) Human histone gene organization: Nonregular arrangement within a large cluster. Genomics 40:314–322

Ascenzi R, Gantt J (1997) A drought-stress-inducible histone gene in *Arabidopsis thaliana* is a member of a distinct class of plant linker histone variants. Plant Mol Biol 34:629–641

Birnstiel M, Busslinger M, Strub K (1985) Transcription termination and 3′ processing: The end is in site! Cell 41:349–359

Brannan CI, Gilbert DJ, Ceci JD, Matsuda Y, Chapman VM, Mercer JA, Eisen H, Johnston LA, Copeland NG, Jenkins NA (1992) An interspecific linkage map of mouse chromosome 15 positioned with respect to the centromere. Genomics 13:1075–1081

Connor W, Mezquita J, Winkfein RJ, States JC, Dixon GH (1984) Organization of the histone genes in the rainbow trout (*Salmo gairdnerii*). J Mol Evol 20:272–235

Cool D, Banfield D, Honda BM, Smith MJ (1988) Histone genes in three sea star species: Cluster arrangement, transcriptional

polarity and analysis of the flanking regions of H3 and H4 genes. J Mol Evol 27:36–44

D'Andrea R, Coles LS, Lesnikowski C, Tabe L, Wells JRE (1985) Chromosomal organization of chicken histone genes: Preferred association and inverted duplications. Mol Cell Biol 5:3108–3115

del Gaudio R, Potenza N, Stefanoni P, Chiusano ML, Geraci G (1998) Organization and nucleotide sequence of the cluster of five histone genes in the polichaete worm *Chaetopterus variopedatus:* First record of a H1 histone gene in the phylum annelida. J Mol Evol 46:64–73

Domier LL, Rivard JJ, Sabatini LM, Blumenfeld M (1986) *Drosophila virilis* histone gene clusters lacking H1 coding segments. J Mol Evol 23:149–158

Drabent B, Kardalinou E, Doenecke D (1991) Structure and expression of the human gene encoding testicular H1 histone (H1t). Gene 103:263–268

Drabent B, Bode C, Doenecke D (1993) Structure and expression of the mouse testicular H1 histone gene (H1t). Biochim Biophys Acta 1216:311–313

Drabent B, Franke K, Bode C, Kosciessa U, Bouterfa H, Hameister H, Doenecke D (1995) Isolation of two murine H1 histone genes and chromosomal mapping of the H1 gene complement. Mammal Genome 6:505–511

Drabent B, Kim JS, Albig W, Prats E, Cornudella L, Doenecke D (1999) *Mytilus edulis* histone gene clusters containing only H1 genes. J Mol Evol 49:645–655

Eick S, Nicolai M, Mumberg D, Doenecke D (1989) Human H1 histones: Conserved and varied sequence elements in two H1 subtype genes. Eur J Cell Biol 49:110–115

Gantt JS, Lenvik TR (1991) *Arabidopsis thaliana* H1 histones. Analysis of two members of a small gene family. Eur J Biochem 202:1029–1039

Heintz N, Zernik M, Roeder RG (1981) The structure of the human histone genes: Clustered but not tandemly repeated. Cell 24:661–668

Hentschel CC, Birnstiel ML (1981) The organization and expression of histone gene families. Cell 25:301–313

Kedes LH (1979) Histone genes and histone messengers. Annu Rev Biochem 48:837–870

Lieber T, Angerer LM, Angerer RC, Childs G (1988) A histone H1 protein in sea urchin is encoded by a poly(A)$^+$ mRNA. Proc Natl Acad Sci USA 85:4123–4127

Lifton RP, Goldberg ML, Karp RW, Hogness DS (1977) The organization of the histone genes in *Drosophila melanogaster*: Functions and evolutionary implication. Cold Spring Harbour Symp Quant Biol 42:1047–1051

Lindauer A, Müller K, Schmitt R (1993) Two histone H1-encoding genes of the green alga *Volvox carteri* with features intermediate between plant and animal genes. Gene 129:59–68

Mantovani R (1999) The molecular biology of the CCAAT-binding factor NF-Y. Gene 239:15–27

Maxson R, Cohn R, Kedes L, Mohun T (1983a) Expression and organization of histone genes. Annu Rev Genet 17:239–277

Maxson R, Mohun T, Gormezano G, Childs G, Kedes L (1983b) Distinct organizations and patterns of expression of early and late histone gene sets in the sea urchin. Nature 301:120–125

Meergans T, Albig W, Doenecke D (1998) Conserved sequence elements in human main type-H1 histone gene promoters: Their role in H1 gene expression. Eur J Biochem 256:436–446

Miller DJ, Harrison PL, Mahony TJ, McMillan JP, Miles A, Odorico DM, ten Lohuis MR (1993) Nucleotide sequence of the histone gene cluster in the coral *Acropora formosa* (*Cnidaria. Scleractinia*). Features of histone gene structure and organization are common to diploblastic and triploblastic metazoans. J Mol Evol 37:245–253

Roberts SB, Sanicola M, Emmons SW, Childs G (1987) Molecular characterization of the histone gene family of *Caenorhabditis elegans*. J Mol Biol 196:27–38

Ruiz-Lara S, Prats E, Sainz J, Cornudella L (1992) Cloning and characterization of a highly conserved satellite DNA from the mollusc *Mytilus edulis*. Gene 117:237–242

Sanicola M, Ward S, Childs G, Emmons SW (1990) Identification of *a Caenorhabditis elegans* histone H1 gene family. Characterization of a family member containing an intron and encoding a poly(A)+ mRNA. J Mol Biol 212:259–268

Schaffner W, Kunz G, Daetwyler H, Telford J, Smith HO, Birnstiel ML (1978) Genes and spacers of cloned sea urchin histone DNA analysed by sequencing. Cell 14:655–671

Sellos D, Krawetz SA, Dixon GH (1990) Organization and complete nucleotide sequence of the core-histone-gene cluster of the annelid *Platynereis dumerilii*. Eur J Biochem 190:21–29

Sittman DB, Chiu IM, Pan CJ, Cohn RH, Kedes LH, Marzluff WF (1981) Isolation of two clusters of mouse histone genes. Proc Natl Acad Sci USA 78:4078–4082

Smith HO, Birnstiel ML (1976) A simple method for DNA restriction site mapping. Nucleic Acids Res 3:2387–2398

Stephenson E, Erba H, Gall J (1981) Characterization of a cloned histone gene cluster of the newt *Notophthalmus viridescens*. Nucleic Acids Res 9:2281–2295

Sures I, Lowry J, Kedes LH (1978) The DNA sequence of sea urchin (*S. purpuratus*) H2A, H2B and H3 histone coding and spacer regions. Cell 15:1033–1044

Thoma F, Koller T, Klug A (1979) Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. J Cell Biol 83:403–427

Tönjes R, Doenecke D (1987) A highly conserved sequence in H1 histone genes as an oligonucleotide hybridization probe: isolation and sequence of a duck H1 gene. J Mol Evol 25:361–370

Trappe R, Doenecke D, Albig W (1999) The expression of human H2A-H2B histone gene pairs is regulated by multiple sequence elements in their joint promoters. Biochim Biophys Acta 1446:341–351

Turner PC, Woodland HR (1983) Histone gene number and organisation in Xenopus: *Xenopus borealis* has a homogeneous major cluster. Nucleic Acids Res 11:971–986

van Dongen W, de Laaf L, Zaal R, Moorman A, Destree O (1981) The organization of the histone genes in the genome of *Xenopus leavis*. Nucleic Acids Res 9:2297–2311

van Wijnen AJ, van den Ent FM, Lian JB, Stein JL, Stein GS (1992) Overlapping and CpG methylation-sensitive protein-DNA interactions at the histone H4 transcriptional cell cycle domain: Distinctions between two human H4 gene promoters. Mol Cell Biol 12:3273–3287

van Wijnen AJ, van Gurp MF, de Ridder MC, Tufarelli C, Last TJ, Birnbaum M, Vaughan PS, Giordano A, Krek W, Neufeld EJ, Stein JL, Stein GS (1996) CDP/cut is the DNA-binding subunit of histone gene transcription factor HiNF-D: A mechanism for gene regulation at the G1/S phase cell cycle transition point independent of transcription factor E2F. Proc Natl Acad Sci USA 93:11516–11521

Walter L, Klinga-Levan K, Helou K, Albig W, Drabent B, Doenecke D, Günther E, Levan G (1996) Chromosome mapping of rat histone genes H1fv, H1d, H1t, Th2a and Th2b. Cytogenet Cell Genet 75:136–139

Wang ZF, Sirotkin AM, Buchold GM, Skoultchi AI, Marzluff WF (1997) The mouse histone H1 genes: gene organization and differential regulation. J Mol Biol 271:124–138

Wells D, Brown D (1991) Histone and histone gene compilation and alignment update. Nucleic Acids Res 19(Suppl):2173–2188