J Mol Evol (2002) 55:272–283
DOI: 10.1007/s00239-002-2325-1

JOURNAL OF MOLECULAR EVOLUTION

Molecular and Evolutionary Analysis of Mussel Histone Genes (*Mytilus* spp.): Possible Evidence of an "Orphon Origin" for H1 Histone Genes

José M. Eirín-López, Ana M. González-Tizón, Andrés Martinez, Josefina Méndez

Departamento de Biología Celular y Molecular, Universidade da Coruña, Campus de A Zapateira s/n, E-15071 A Coruña, Spain

Received: 19 December 2001 / Accepted: 18 February 2002

Abstract. Linker histones are a divergent group of histone proteins with an independent evolutionary history in which, besides somatic subtypes, tissue- and differentiation-specific subtypes are included. In the present work H1 histone coding and noncoding segments from five *Mytilus* mussel species (Mollusca: Bivalvia) widely distributed throughout the world have been determined and characterized. Analysis of promoter regions shows clear homologies among *Mytilus* H1 genes, sea urchin H1 genes, and vertebrate differentiation-specific H1 subtypes (H5 and $H1^0$), all having an H4 box motif in common. The amino acid sequence of the H1 protein central conserved domain is also closely related to that previously defined for the vertebrate divergent subtypes. A phylogenetic tree reconstructed from different H1 genes from several species strengthens the hypothesis of an "orphon" origin for the *Mytilus* H1 genes, as well as for the $H1^0$/H5 genes from vertebrates and the H1D gene from the sea urchin *Strongylocentrotus purpuratus*, is suggested. As additional data, the average copy number of the H1 genes in the species analyzed was estimated as being 100 to 110 copies per haploid genome, where FISH revealed telomeric chromosomal location for several H1 copies in *M. galloprovincialis*. The contribution of such proximity to heterochromatic regions over the amount of codon bias detected for H1 genes is discussed.

Key words: *Mytilus* — H1 genes — "Orphon" genes — Copy number — FISH — Phylogeny

Introduction

Histone proteins, subdivided into core histones (H2A, H2B, H3, H4) and linker histones (H1 and its variants), are the major constituents of the eukaryotic chromatin. The typical structure of H1 proteins consists in a central trypsin-resistant globular region, with a winged-helix domain, flanked by two terminal arms (N and C terminal) without tertiary structure (Ramakrishnan et al. 1993). Besides the structural role of linker proteins constituting the chromatosomal structure (Simpson 1978), preferential binding to scaffold-associated regions and participation in nucleosome positioning are currently accepted mechanisms by which H1 could have a regulatory role in transcription, through modulation of chromatin higher-order structure (Zlatanova and Van Holde 1992; Khochbin and Wolffe 1994; Wolffe et al. 1997). Linker histones can also be subdivided into distinct types according to the stage of development and to the tissue where they are expressed (Khochbin and Wolffe 1994). During early embryogenesis, the expression of specific H1 histones has been described in amphibians (Ohsumi and Katagiri 1991). Later in development and during the activation of zygotic genes, there is an accumulation of somatic H1

The sequences for histone H1 described in Fig. 1 have been deposited in the EMBL Nucleotide Sequence Database under the following accession numbers: *Mytilus californianus*, AJ416421; *M. chilensis*, AJ416422; *M. edulis*, AJ416423; *M. galloprovincialis*, AJ416424; and *M. trossulus*, AJ416425.

Correspondence to: Josefina Méndez Felpeto; *email:* fina@udc.es

histones in cells, and finally, the formation of different tissues is accompanied by the accumulation of differentiation-specific subtype H1 histones (Grunwald et al. 1995). Three vertebrate-specific H1 histones ($H1^0$, H5, H1M), which are encoded by polyadenylated mRNAs and expressed independently of the S- phase in the cell cycle (Pandey and Marzluff 1987), stand out from all the others because of their high degree of structural and functional divergence. Otherwise, vertebrate somatic H1 histones are encoded by mRNAs with a 3′ stem–loop structure, correlated with S- phase-dependent expression. The only known example of a differentiation-specific subtype in invertebrates is the H1D histone in the sea urchin *Strongylocentrotus purpuratus* (Lieber et al. 1988), which shows great homology with H5 and $H1^0$ histones from vertebrates and also with H1 histone from the mussel *Mytilus edulis* (Drabent et al. 1999).

In many organisms core and linker histone genes are grouped together in clusters with several repetitions of the five histone classes (Hentschel and Birnstiel 1981). The tandem organization of histone genes within the repetitive units is characteristic of invertebrate genomes (Maxson et al. 1983) and generally lost in vertebrate genomes (D'Andrea et al. 1985; Albig et al. 1997; Wang et al. 1997). Many times along the evolutionary scale, an independent organization of linker histone genes from the remaining core histone genes has been observed. Such an arrangement was reported for invertebrates such as *Drosophila virilis* (Domier et al. 1986), *Caenorhabditis elegans* (Roberts et al. 1987), sea stars (Cool et al. 1988), corals (Miller et al. 1993), annelids (del Gaudio et al. 1998), and *M. edulis* (Drabent et al. 1999). This arrangement has also been observed for vertebrates such as newts (Stephenson et al. 1981) and for vertebrate specific linker subtypes as H5 (Krieg et al. 1983) and $H1^0$ (Albig et al. 1993; Walter et al. 1996).

In general, previous works have revealed that the exclusion of a gene from a gene family could be more precisely defined as an exclusion from the events of sequence homogenization operating over the family members, and such exclusion gives rise to an "orphon" gene (Childs et al. 1981; Sculze and Schulze 1995). Till now, studies on "orphon" genes have involved mostly vertebrate genomes (Coles and Wells 1985; Schulze and Schulze 1995; Brocard et al. 1997; Peretti and Khochbin 1997), only two groups of invertebrates being analyzed: the sea urchin *S. purpuratus* (Lieber et al. 1988) and the mussel *M. edulis* (Drabent et al. 1999). To complete and clarify several aspects of the evolutionary history of linker histone genes in invertebrates, we have sequenced and characterized H1 histone genes in five mussel species, distributed worldwide, and belonging to the genus *Mytilus* (*M. californianus*, *M. chilensis*, *M. edulis*, *M. galloprovincialis*, and *M. trossulus*). Results obtained from molecular and phylogenetic analyses support the hypothesis that H1 histone genes in the genus *Mytilus* share a common evolutionary origin with a group of "orphons" composed of the differentiation-specific subtypes from vertebrates and also the H1D histone gene from the sea urchin *S. purpuratus*. Additionally, the copy number and the chromosomal location of the H1 genes are reported and discussed in the present work with regard to the estimated codon bias values.

Materials and Methods

PCR Amplification and DNA Sequencing of Mytilus *H1 Histone Genes*

Mussels analyzed were collected from different localities along the European and American coasts as follows: *Mytilus californianus* from Point no Point, Pacific coast of Canada; *M. chilensis* from Puerto Aguirre, Chile; *M. edulis* from Yerseke, Holland; *M. galloprovincialis* from A Coruña, Atlantic coast of Spain; and *M. trossulus* from Esquimalt Lagoon, Pacific coast of Canada.

Genomic DNA from muscle tissue was purified in CTAB buffer in the five mussel species analyzed, following the protocol described by Rice and Bird (1990). PCR amplifications from template genomic DNA were performed in a final volume of 25 µl (10 ng/µl), where primers 5′-H1-full (TAC CTG CGA AGA CAA TTC AG), and 3′-H1-full (AGA AAG GGT AGG GCT CAG) were used at 10 µ*M*, with 1 U/µl of *Taq* DNA polymerase (Roche Molecular Biochemicals). The reactions were performed with a first denaturation step of 4 min 30 s at 95°C, followed by 35 cycles consisting of a 30 s of denaturation at 95°C, 30 s of annealing at 55.5°C, and 30 s of extension at 72°C. A final extension step of 5 min was performed at 72°C. In the resulting 1100-bp band, the whole H1 coding region together with the noncoding flanking segments is included. Automatic DNA sequencing was performed directly from the PCR products using the 5′-H1-full/3′-H1-full primer set.

Generation of Nucleotide and Amino Acid Alignments

Multiple alignments of nucleotide and amino acid sequences were conducted using the Clustal X program (Thompson et al. 1997), with the default parameters given by the program. Alignments of the amino acid central conserved domains were performed following the criteria previously established by Ramakrishnan et al. (1993) and by Schulze and Schulze (1995) to define the borders of the H1 central domain. Further alignments with different gap penalizations were performed to estimate the stability and validity of the assignments between the sequences of the final alignment.

Nucleotide Substitutions and Phylogenetic Analysis

The numbers of synonymous (d_S) and nonsynonymous (d_N) substitutions were estimated by the method of Nei and Gojobori (1986) with the Jukes–Cantor correction. The reconstruction of a phylogenetic tree from the aligned nucleotide sequences was achieved using the neighbor-joining method (Saitou and Nei 1987). Evolutionary distances were estimated by means of Jukes–Cantor's one parameter model and the inferred phylogeny was tested with 1000 bootstrap replicates. All the steps in the analysis were carried out using the MEGA 2 package (Kumar et al. 2001).

The nucleotide substitution patterns were determined by the maximum-likelihood method, using the baseml program for the 5′ region (positions −300 to −1), and the codeml program for the coding region (positions 1 to 576), both included in the PAML package (Yang 2000). Ancestral sites were discarded if their probability was less than 0.75 and the probability of multiple hits in the branches from the ancestral species to closely related species was considered to be low and not significantly affected (Petrov and Hartl 1999). The amount of codon bias presented by *Mytilus* H1 genes was estimated by means of the DnaSP 3 program (Rozas and Rozas 1999) and is referred to as the effective number of codons (ENC) (Wright 1990). The highest value (61) indicates that all synonymous codons are used equally, and the lowest (20) that only one codon is used in each synonymous class.

Analysis of the H1 Copy Number in the Mytilus *Genomes*

Varying amounts (200, 100, 50, and 25 ng) of genomic DNA from each *Mytilus* species analyzed were transferred onto a nylon membrane together with 0.8, 0.4, 0.2, and 0 ng of H1 PCR product (complete coding region) for comparison. The haploid DNA complement consists of 1.605 pg in *M. californianus*, and 1.510 pg in *M. trossulus* (González-Tizón et al. 2000), 1.710 pg in *M. chilensis* and *M. edulis*, and 1.920 pg in *M. galloprovincialis* (Rodriguez-Juiz et al. 1996). The blot was hybridized with 100 ng/ml of the digoxigenin-labeled whole H1 coding region probe from *M. galloprovincialis* (in a final volume of 25 ml of hybridization solution, the probe was not under limitant conditions) and the resulting signal was detected by a chemiluminescent reaction using CSPD (Roche Molecular Biochemicals) as substrate for alkaline phosphatase. Hybridization intensity was quantified and evaluated using the Q-Win image analysis software (Leica Imaging System).

Obtaining Chromosomes and Fluorescent In Situ Hybridization

Once in the laboratory, collected *M. galloprovincialis* specimens were placed in tanks containing filtered seawater and fed continuously on a suspension of *Isochrysis* sp. and *Tetraselmis* sp. microalgae for 10–15 days. Metaphases were obtained following the protocol described by González-Tizón et al. (2000). The DNA probe used in fluorescent in situ hybridization (FISH) was an H1-positive recombinant phage from a *M. galloprovincialis* genomic library (λDASH II/*Eco*RI; Stratagene), labeled with digoxigenin-11-dUTP using a Nick Translation Kit (Roche Molecular Biochemicals). The slide preparation procedure was performed as described previously by González-Tizón et al. (2000), using 100 ng of labeled DNA probe. Hybridization sites were detected by immunocytochemical incubations in mouse antidigoxigenin, rabbit anti-mouse FITC (fluorescein isothiocyanate), and goat anti-rabbit FITC. Chromosomes were counterstained with propidium iodide (50 ng/ml antifade) and visualized and photographed, using a Leica DM RXA fluorescence microscope, on Kodak Ektachrome 400 ASA film.

Results

Characterization of the HI Histone Sequence in the Genus Mytilus

PCR reactions with the 5′-H1-full/3′-H1-full primer set yielded a 1100-bp fragment containing the coding and noncoding flanking regions of the H1 gene. The resulting fragments for each species were subsequently purified and sequenced. Coding regions, translated amino acid sequences, and noncoding flanking regions are shown in Fig. 1. Promoter regions (Fig. 1A) are highly conserved, especially the major regulatory elements involved in the transcriptional activity of H1. We have identified an H1 box-like element, an H4 box, and a TATA box. The gaps in the alignment of the promoter regions are due mainly to the presence of indels in the sequence of *M. californianus*. The consensus obtained for the promoter region from the five *Mytilus* species was compared with other H1 promoter regions from different taxa (Fig. 2), where the presence of an H4 box element (instead of a CAAT box) clearly discriminates between vertebrate somatic H1s and the rest of the linker subtypes.

The coding sequences (Fig. 1B) reveal the presence of a 190-amino acid (aa)-long H1 protein in all cases, with the exception of *M. chilensis* (189 aa) and *M. californianus* (188 aa). These sequences possess a high density of amino acid motifs that are enriched in basic residues. Length variation involves deletion of two amino acid residues at the C-terminal arm of the protein: a proline (P) in position 138 in both cases and an alanine (A) in position 170 only in *M. chilensis*. The nucleotide variation in the coding region is essentially synonymous, although 13 of the 15 amino acid replacements are represented by nonsynonymous substitutions between *M. californianus* and/or *M. chilensis* and the remaining species. At the 3′-terminal region (Fig. 1C), once again, nucleotide variation is practically absent, especially in the region corresponding to the stem–loop or hairpin structure of the mRNA.

Analysis of the Protein Central Conserved Domain

An amino acid alignment was performed from 58 H1 protein central domains (including *Mytilus* H1s), in which the whole folded region is represented as described for histone H5 by Ramakrishnan et al. (1993). Given that many positions show shared residues between invertebrates and plants, and finding that they are clearly differentiated from the residues observed in vertebrate sequences, consensus sequences were separately defined for invertebrates/plants, vertebrates, $H5/H1^0$, and *Mytilus*. These consensus sequences agreed with those previously reported by Schulze and Schulze (1995). A high degree of divergence between the central domain of the H1M protein from *Xenopus* and the remaining sequences was detected, and that is why it was independently analyzed. By individually aligning each consensus sequence against that obtained for the central domain of *Mytilus* H1 proteins (Fig. 3),

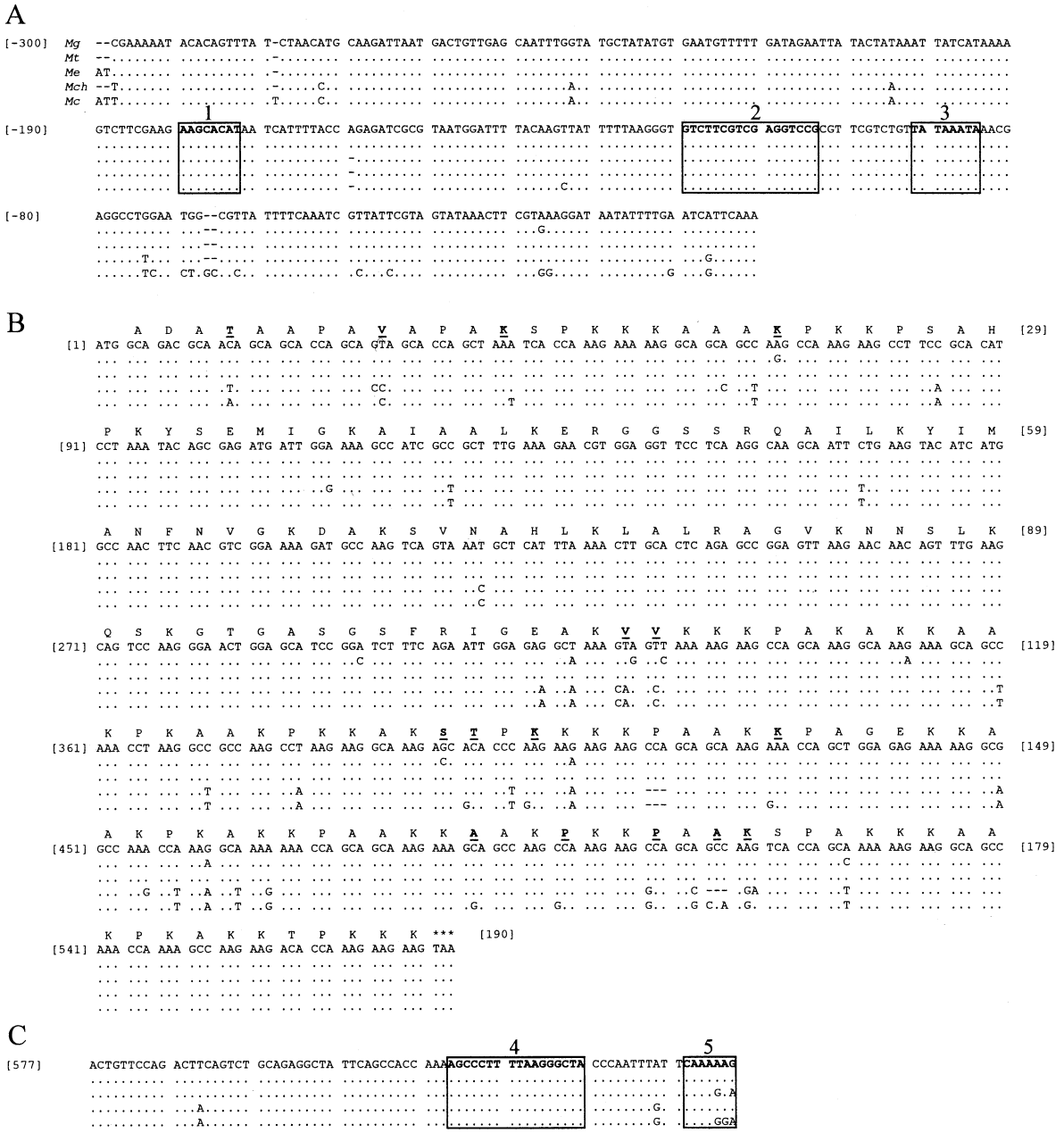

Fig. 1. Nucleotide sequences determined for the five mussel species analyzed and the corresponding EMBL Nucleotide Sequence Database accession numbers: *M. galloprovincialis* (*Mg*) AJ416424, *M. trossulus* (*Mt*) AJ416425, *M. edulis* (*Me*) AJ416423, *M. chilensis* (*Mch*) AJ416422, and *M. californianus* (*Mc*) AJ416421. **(A)** Sequence for *Mytilus* H1 promoter regions (positions −1 to −300). Numbering on the *left* refers to the nucleotide sequence, *dots* represent complete matches between sequences, and gaps are indicated by *dashes*. Conserved regulatory elements (positions −80 to −180) are indicated by *open boxes*: (1) H1 box-like element (positions −173 to −180); (2) H4 box (positions −104 to −120); (3) TATA box (positions −85 to −92). **(B)** H1 coding regions. Numbering on the *right* refers to amino acid residues. Translated amino acids are placed above the corresponding codon, with positions where replacements were detected in *boldface* and *underlined*. **(C)** H1 noncoding 3′ flanking regions (positions 577 to 655). Conserved elements are again in *open boxes*: (4) stem–loop or hairpin terminal structure (positions 620 to 636); (5) part of a purine-rich element (positions 648 to 654).

our results show that the highest degree of homology is presented by the consensus obtained for H5/H1^0 central domains, followed by the consensus for invertebrates/plants, vertebrates, and H1M from *Xenopus*, which shows the lowest degree of homology.

Nucleotide Substitution Numbers and Patterns in Mytilus *H1 Histone Genes*

Synonymous and nonsynonymous substitution numbers were estimated using the method of Nei and Gojobori (1986). The analysis of the individual

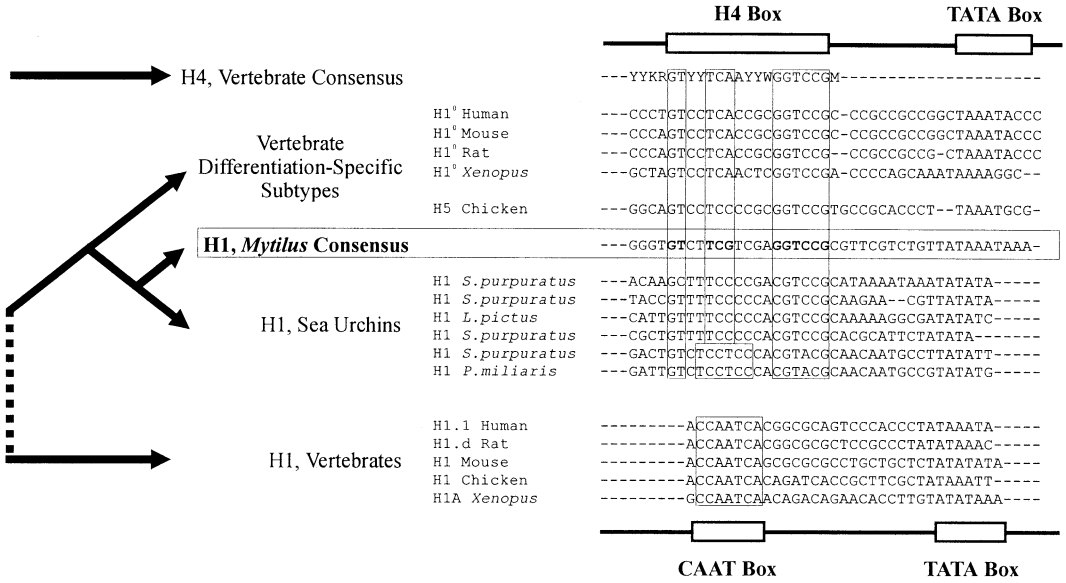

Fig. 2. The structure of the proximal promoter region of the H1 gene is compared between the consensus for the genus *Mytilus* and its homologous region in vertebrates, sea urchins, and vertebrate differentiation-specific subtypes and the consensus of vertebrate H4 histone gene (Peretti and Khochbin 1997). The TATA box and the upstream sequence including the CAAT box (in vertebrate common-type somatic H1s) and the H4 box (in the remaining H1 genes) are indicated. *Black arrows* on the left indicate evolutionary relationships among the H1 sequences analyzed.

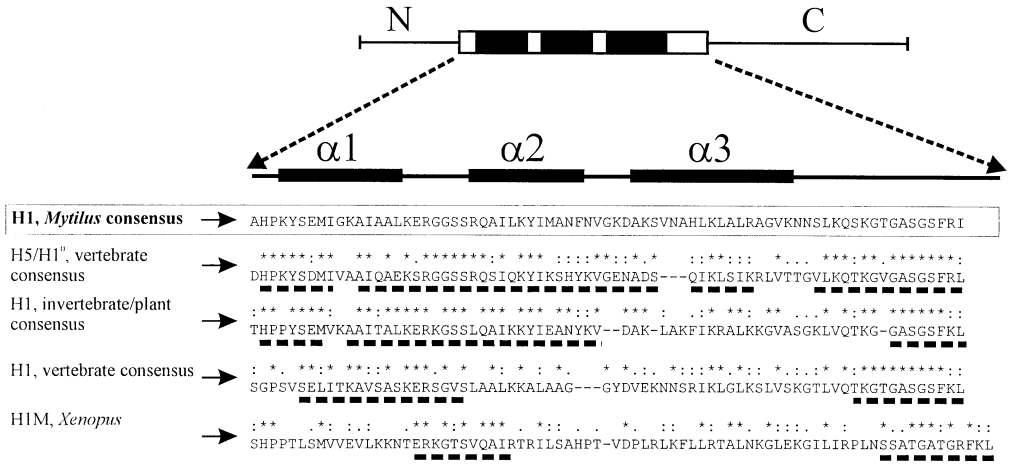

Fig. 3. Analysis of the H1 protein central conserved domain, represented above the alignment as an *open box* flanked by the two terminal arms (carboxy and amino) of the protein. *Black boxes* in the central domain indicate positions in which the three α-helical regions of the winged-helix motif are placed. Regions of homology with the consensus defined for *Mytilus* H1 protein are emphasized with *discontinuous bars* below the amino acid sequence compared, where *asterisks* indicate totally conserved residues, *colons* indicate a high degree of conservation, and *dots* indicate a low degree of conservation.

domains shows a total absence of amino acid replacements among the *Mytilus* species in the central segment of the protein (see Table 1). The highest synonymous divergence was observed at the C-terminal arm, followed by the central domain and the N-terminal arm of the protein. On the other hand, the highest levels of nonsynonymous divergence were present at the N-terminal domain. Nucleotide sequences corresponding to the coding regions of *M. galloprovincialis* and *M. edulis* matched perfectly, and the most divergent species in the genus *Mytilus* were *M. chilensis* and *M. trossulus*. On average, synonymous and nonsynonymous divergence among the five mussel species analyzed was 0.127 ± 0.025 and 0.018 ± 0.005 substitution/site, respectively. In the differentiation-specific subtypes (H5 and $H1^0$) and in the tissue-specific subtype H1t from vertebrates, the central domain, again, is more conserved than the terminal regions, at the amino acid level (Fig. 4). Once more, nucleotide divergence is essen-

Table 1. Numbers of synonymous (d_S) and nonsynonymous (d_N) substitutions in H1 histone genes[a]

Histone H1 (*Mytilus*)	Complete protein		N-Terminal domain		Central domain		C-Terminal domain	
	d_S	d_N	d_S	d_N	d_S	d_N	d_S	d_N
Mg/Me	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
Mg/Mt	0.063 ± 0.023	0.005 ± 0.003	0.000 ± 0.000	0.017 ± 0.015	0.036 ± 0.026	0.000 ± 0.000	0.119 ± 0.055	0.005 ± 0.005
Mg/Mch	0.201 ± 0.044	0.019 ± 0.008	0.105 ± 0.078	0.051 ± 0.039	0.137 ± 0.052	0.000 ± 0.000	0.319 ± 0.098	0.026 ± 0.014
Mg/Mc	0.162 ± 0.039	0.030 ± 0.009	0.051 ± 0.054	0.051 ± 0.029	0.116 ± 0.048	0.000 ± 0.000	0.265 ± 0.082	0.054 ± 0.019
Me/Mt	0.063 ± 0.023	0.005 ± 0.003	0.000 ± 0.000	0.017 ± 0.015	0.036 ± 0.026	0.000 ± 0.000	0.119 ± 0.055	0.005 ± 0.005
Mc/Mch	0.201 ±0.044	0.019 ± 0.008	0.105 ± 0.078	0.051 ± 0.039	0.137 ± 0.052	0.000 ± 0.000	0.319 ± 0.098	0.026 ± 0.014
Me/Me	0.162 ± 0.039	0.030 ± 0.009	0.051 ± 0.054	0.051 ± 0.029	0.116 ± 0.048	0.000 ± 0.000	0.265 ± 0.082	0.054 ± 0.019
Mt/Mch	0.210 ± 0.046	0.023 ± 0.008	0.104 ± 0.077	0.069 ± 0.043	0.137 ± 0.051	0.000 ± 0.000	0.345 ± 0.102	0.032 ± 0.015
Mt/Mc	0.171 ± 0.040	0.035 ± 0.009	0.050 ± 0.054	0.069 ± 0.033	0.116 ± 0.048	0.000 ± 0.000	0.289 ± 0.087	0.060 ± 0.020
Mch/Mc	0.040 ± 0.018	0.019 ± 0.006	0.051 ± 0.054	0.051 ± 0.029	0.018 ± 0.018	0.000 ± 0.000	0.058 ± 0.035	0.026 ± 0.012
Average	0.127 ± 0.025	0.018 ± 0.005	0.052 ± 0.038	0.043 ± 0.023	0.085 ± 0.031	0.000 ± 0.000	0.210 ± 0.055	0.029 ± 0.010

[a] Estimations are in units of substitutions per site ± standard deviation. The average is the arithmetic mean.

Fig. 4. Numbers of nucleotide substitutions in vertebrate differentiation- and tissue-specific histone genes and in *Mytilus* H1 genes. Results are given as an average, discriminating between the whole coding region (*shaded bars*) and the three protein domains (*open bars*). *Thin bars* indicate the standard deviation. **(A)** Synonymous substitutions per site. **(B)** Nonsynonymous substitutions per site. T, complete protein; N, amino-terminal region; G, central domain; C, carboxy-terminal region.

tially synonymous in all cases and located mainly in the N-terminal segment of histones H5 and $H1^0$ (0.553 ± 0.258 and 0.877 ± 0.285 substitution/site, respectively, on average) and in the C-terminal segment of the H1t histone (1.746 ± 0.415 substitutions/site, on average). Histone H1t is the most divergent subtype at all levels, reaching synonymous and nonsynonymous substitution numbers of almost 1 and 0.2 substitution/site, respectively, for the whole protein.

To determine the nature of nucleotide substitutions in the H1 gene, ancestral sequences were obtained using the maximum-likelihood method. Although no functional constraints are thought to be present in flanking regions (except for promoter motifs) in the 5′ region, transitions from G to A (four) and from C to T (five) occurred more frequently than transitions from A to G (two) and T to C (one). A significantly higher frequency of change to A or T than to G or C was observed here (9:3, $p > 0.05$), indicating a bias toward A or T, with high transition numbers and relative frequencies of A and T ($A = 0.33$, $T = 0.35$). In the protein coding region, significantly higher frequencies of changes to A or T than to G or C were not observed (8:9; $p > 0.05$), indicating an equilibrium state for the base composition.

The amount of codon bias in *Mytilus* H1 histone genes was estimated as the "effective number of codons" (ENC) index (Wright 1990), which ranges between 20 (extreme bias) and 61 (no bias). For *Mytilus* H1 genes, the codon bias measured as ENC

Table 2. Codon bias values in *Mytilus* histone genes[a]

Species	Effective number of codons (ENC)		
	H1	H3	H4
Mc	43.896	37.181	35.845
Mch	45.611	38.793	30.238
Me	47.186	39.182	34.887
Mg	47.186	40.378	39.785
Mt	47.991	39.008	39.902
Average	46.898	38.908	36.131

[a] Species abbreviations are as in Fig. 2.

is an average of 46.898 (Table 2). This value is very low compared with those obtained for core histone genes such as the H3 and H4 histones in the same species.

Phylogenetic Analysis

A neighbor-joining phylogenetic tree (Saitou and Nei 1987) was reconstructed from 58 complete nucleotide H1 coding regions in several organisms and rooted with the H1-like gene from the trypanosomatidae *Leishmania braziliensis* (Fig. 5). The bootstrapped topology (1000 replicates) obtained clearly discriminates among the taxonomic groups of plants, invertebrates, and vertebrates. The monophyletic subgroup of the differentiation-specific subtypes from vertebrates is also well defined, where H1 histone genes from *Mytilus* and the H1D gene from the sea urchin *S. purpuratus* are included. Three H1 genes from sea urchins (H1B and H1G from *S. purpuratus* and H1 from *Lytechinus pictus*) are uncertainly placed between plants and vertebrate differentiation-specific subtypes. *Xenopus* H1M histone is grouped together with the vertebrate somatic types in the present topology.

Number of H1 Copies in the Genomes Analyzed and Chromosomal Location in M. galloprovincialis

Increasing amounts of genomic DNA from each species were transferred onto a nylon membrane and, for comparison, corresponding amounts of H1 PCR product DNA (Fig. 6A). The probe lacking the noncoding flanking regions (to avoid nonspecific signals) was digoxigenin-labeled, and after hybridization, the signal was subsequently detected by chemiluminescence. Hybridization intensity was measured and the copy number of the H1 genes was calculated at about 84 copies per haploid genome in *M. californianus*, 110 in *M. chilensis* and *M. edulis*, 115 in *M. galloprovincialis*, and 94 in *M. trossulus*. FISH experiments on *M. galloprovincialis* chromosomes, using an H1-recombinant phage from a *M. galloprovincialis* genomic library as a probe, revealed the presence of three pairs of signals in three chromosome pairs (Fig. 6B). With the exception of one pair located at an interstitial position, all the hybridization signals map H1 genes at telomeric chromosomal positions.

Discussion

H1 Histone Genes in the Genus Mytilus

In the present work, coding and noncoding flanking regions of the H1 histone gene were amplified by PCR using specific primers in five species of mussels belonging to the genus *Mytilus*. Within the promoter region (Fig. 1A) typical regulatory elements of H1 genes are represented (Peretti and Khochbin 1997). Between the H1 box-like element (Dalton and Wells 1988) and the TATA box, it is characteristic to find a region that is homologous to the promoter sequence of the H4 histone gene (H4 box). This region is located at the same position as that occupied by the CAAT box in somatic H1s from vertebrates (Martinelli and Heintz 1994). This H4 box motif strongly accounts for the high degree of homology detected at promoter regions among *Mytilus* H1 genes and vertebrate differentiation-specific H1s, sea urchin H1s, and vertebrate H4 genes (Fig. 2). In this sense, the evolutionary vinculation established between H1 histones from the mussel *M. edulis* and the differentiation-specific subtypes from vertebrates by Drabent et al. (1999) extends to all the species analyzed in the present work.

The coding sequences (Fig. 1B) reveal polymorphism in protein length caused by indel events at the C-terminal arm of the H1 protein. A high density of amino acid motifs highly enriched in basic residues was identified at the C-terminal arm, possibly representing phosphorylation sites (Mezquita et al. 1985), although the presence of the H1-specific motif T(S)PKKAKKP from vertebrates (Tönjes and Doenecke 1987) was not found. The presence of a stem–loop or hairpin termination signal (Birnstiel et al. 1985) in the 3′ flanking region (Fig. 1C) is correlated with the appearance of these mRNAs in the S phase of the cell cycle, which contrasts with the polyadenylation signal presented by the "orphon" subtypes and also by the H1D gene from *S. purpuratus*, expressed independently of the S phase.

The comparisons between the protein central conserved domains in the organisms analyzed show that H1 proteins in *Mytilus* share essential characteristics with H1 proteins from invertebrates and plants and also with H5 and $H1^0$ proteins from vertebrates (Fig. 3). The residues constituting the first two α-helical regions and also the terminal β-sheet structure of the protein-folded domain show the highest degree of conservation among the groups analyzed. As reported previously by Schulze and

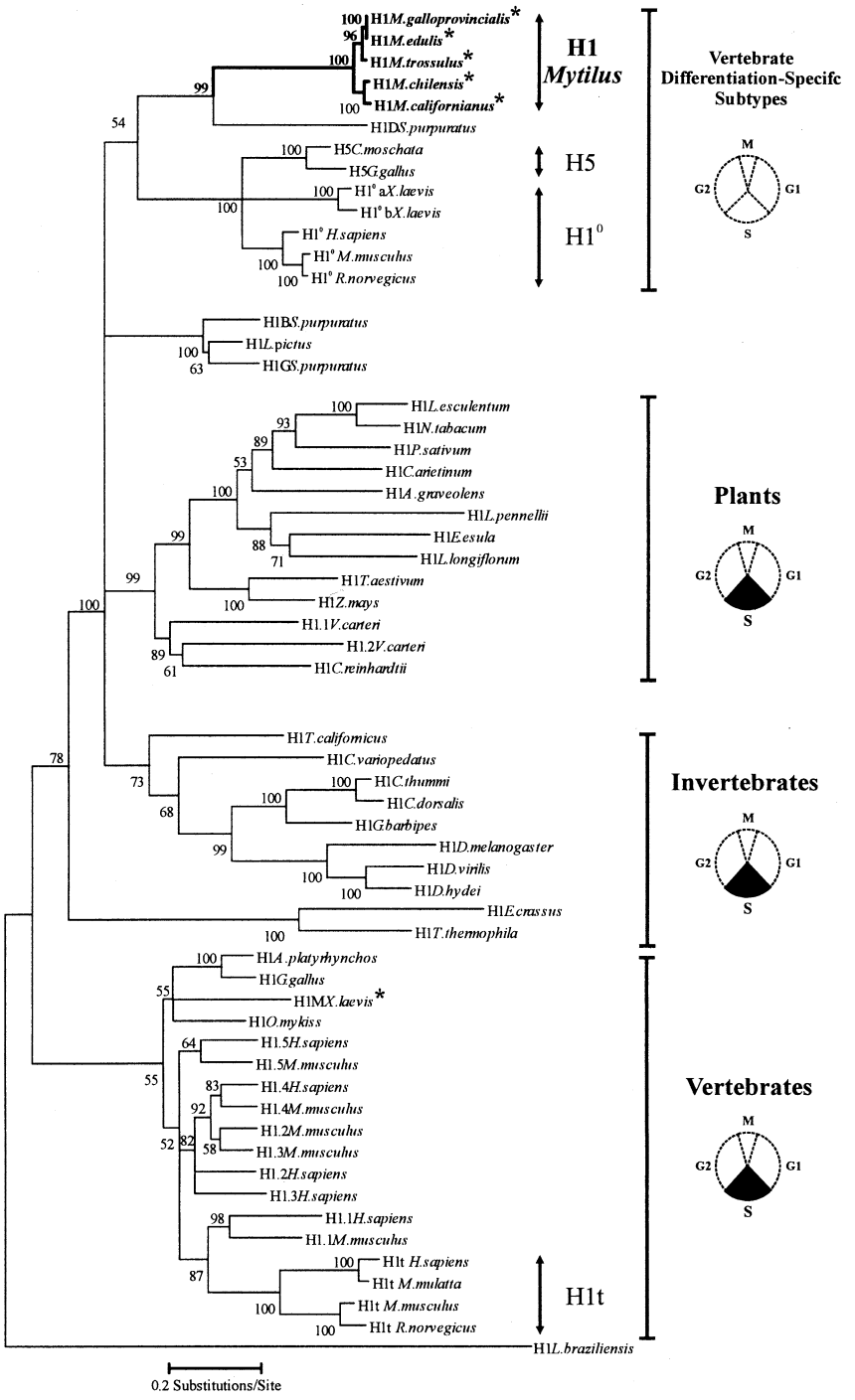

Fig. 5. Phylogenetic relationships among H1 histone genes from several species grouped in different phyla and in one group of differentiation-specific subtypes. Bootstrap values are placed in the corresponding nodes. Patterns of gene expression related to the cell cycle are indicated as follows: the *open diagram* of the cell cycle indicates gene expression independent from the S phase; S phase-dependent expression is indicated by *shading* of the correspondent segment in the cell cycle diagrams (*asterisks* indicate exceptions). The five *Mytilus* species analyzed in the present work are in *boldface*. Nucleotide coding regions for H1 genes were obtained from the NHGRI/NCBI database under the following accesion numbers: *Lycoperscion esculentum* (AJ224933), *Nicotiana tabacum* (AB029614), *Pisum sativum* (X05636), *Cicer arietinum* (AJ006767), *Apium graveolens* (Y12599), *L. pennellii* (U01890), *Euphorbia esula* (AF222804), *Lilium longiflorum* (AB012694), *Triticum aestivum* (D87064), *Zea mays* (X57077), *Volvox carteri* H1.1 (L07946), *V. carteri* H1.2 (L07947), *Chlamydomonas reinhardtii* (U16726), *Strongylocentrotus purpuratus* H1B (H1β; M20314), *Lytechinus pictus* (X04488), *S. purpuratus* H1G (H1γ; M16033), *Mytilus galloprovincialis, M. edulis*, *M. trossulus*, *M. chilensis*, *M. californianus*, and *S. purpuratus* H1D (H1δ; J03807), *Cairina moschata* H5 (X01065), *Gallus gallus* H5 (J00870), *Xenopus laevis* $H1^0$a (Z71502), *X. laevis*, $H1^0$b (Z71503), *Homo sapiens* $H1^0$ (X03473), *Mus musculus* $H1^0$ (U18295), *Rattus norvegicus* $H1^0$ (X72624), *Tigriopus californicus* (M84797), *Chaetopterus variopedatus* (U96764), *Chironomus thummi* (L28724), *C. dorsalis* (U21211), *Glyptotendipes barbipes* (L29102), *Drosophila melanogaster* (X14215), *D. virilis* (L76558), *D. hydei* (X17072), *Euplotes crassus* (AF127331), *Tetrahymena thermophila* (M14854), *Anas platyrhynchos* (X06128), *G. gallus* (M17021), *X. laevis* H1M (S69089), *Oncorhynchus mykiss* (X02624), *H. sapiens* H1.5 (Z98744), *M. musculus* H1.5 (Z46227), *H. sapiens* H1.4 (NM_005321), *M. musculus* H1.4 (Y12292), *M. musculus* H1.2 (Y12291), *M. musculus* H1.3 (Y12291), *H. sapiens* H1.2 (X57129), *H. sapiens* H1.3 (M60747), *H. sapiens* H1.1 (X57130), *M. musculus* H1.1 (L26164), *H. sapiens* H1t (M60094), *Macaca mulatta* H1t (M97756), *M. musculus* H1.t (X72805), *R. norvegicus* H1t (M28409), *Leishmania braziliensis* (AF131910).

Schulze (1995) for vertebrate differentiation-specific subtypes, it is possible that the central domain is also an evolutionarily old structure in *Mytilus* H1 proteins, and nonconservative changes made during the rise of the phylum Vertebrata represent evolutionary innovations. This could be due to an ancestral exclusion event of a gene from the clustered gene family, and subsequently from the operating events of nucleotide homogenization, which could account for the rise of an "orphon" group with an independent evolutionary history with respect to the somatic types.

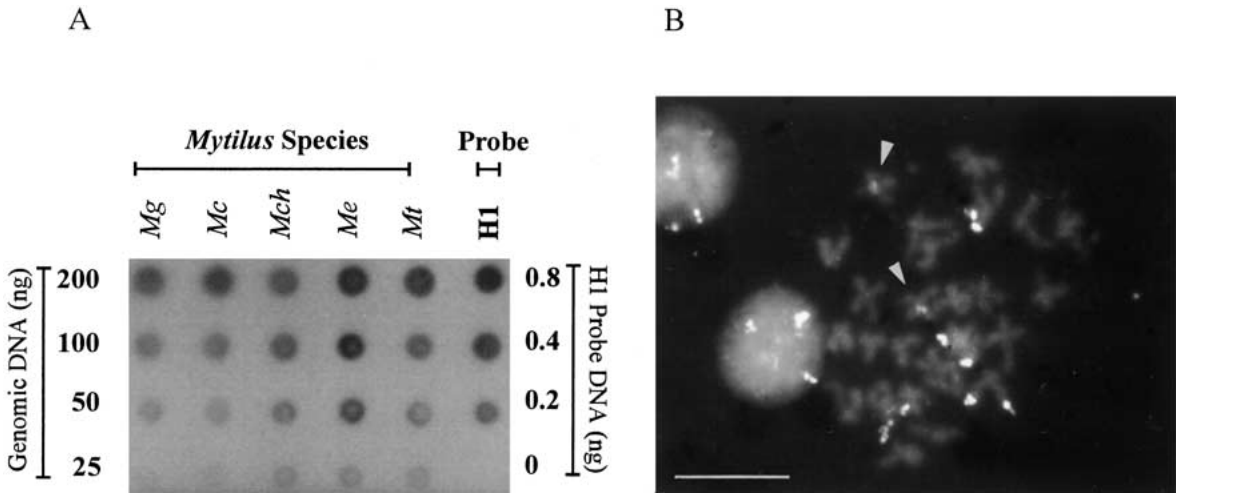

Fig. 6. Copy number and chromosomal location of the H1 genes. **(A)** Dot-blot analysis of the H1 histone copy number in the five species analyzed. The signal intensity of one H1 copy was estimated from the intensity pattern given by different amounts of H1 probe. The H1 copy number in genomic DNA from *Mytilus* species was subsequently resolved from this pattern. Species abbreviations are as in Fig. 2. **(B)** FISH with a digoxigenin-labeled H1 histone probe in metaphase chromosomes of *M. galloprovincialis* counterstained with propidium iodide. *Arrows* indicate interstitial hybridization signals. Bar = 10 μm.

Nucleotide Substitutions and Phylogenetic Analysis of H1 Genes

The group of linker histones is the fastest-evolving histone class. However, the H1 subtypes, except for H1t and H1a, can still be considered highly conserved proteins (Ponte et al. 1998). The vast majority of the nucleotide variation detected among *Mytilus* species is focused on the nonorganized terminal segments of the protein (Table 1), where synonymous divergence greatly exceeds nonsynonymous divergence, indicating the presence of purifying selection acting on H1 genes. The H1 histone C-terminal arm comprises nearly 50% of the protein and its function is associated with chromatin condensation and gene expression modulation, whereas the short N-terminal arm has little or no critical function (Widom 1998). Probably, the absence of tertiary structure in these regions contributes to the higher tolerance of amino acid substitutions (Ponte et al. 1998). The vertebrate tissue-specific H1t gene is the most variable linker histone, followed by the differentiation-specific subtypes H5 and H1° (Fig. 4). It is possible that the absence of sequence homogenization events, due to an independent genomic location of these divergent variants with respect to the clusters of somatic types (Krieg et al. 1983; Albig et al. 1993; Walter et al. 1996), accounts for the high nucleotide variation and also for the fast evolutionary rates observed. Finally, the central domain of H1 histones is generally the most conserved region, due to the presence of tertiary structure (Ramakrishnan et al. 1993) and also to its functional importance in the process of binding to DNA.

It is likely that bias toward AT in the nucleotide substitution patterns observed for the 5′ flanking region of *Mytilus* H1 genes may be due to a need to maintain the structure of the characteristic promoter motifs involved in the correct expression of the "orphon" genes, avoiding the presence of transversions which could drastically modify these signals. The same results have been recently reported for the H3 histone gene in eight species of the *D. melanogaster* species subgroup (Matsuo 2000), indicating a recent change in the selective pressure acting over this region to maintain the structure of the regulatory elements. It is not surprising that histone genes (highly expressed during certain stages of the cell cycle) have their promoter regions under strong selective constraints to assess the effective expression of the gene. Meanwhile, the same effect would be present in the coding regions through a homogeneous synonymous codon usage (high codon bias).

The exclusion event of a group of linker histones from the main group during the short period of divergent evolution which gave rise to the phylum Vertebrata is reflected in the inferred tree topology (Fig. 5). This group would constitute the direct predecessor of the present divergent linker histones from invertebrates, plants, and vertebrate differentiation-specific subtypes. Further evolutionary changes involving major structural changes would have arisen related mainly to gene expression, where one group underwent changes oriented toward S phase-dependent gene expression, while other linker histone genes, present at the same time and encoded by "orphon" genes (Childs et al. 1981), were not subject to these particular changes. In the tree topology,

Mytilus H1 histone genes are included in a monophyletic group together with the differentiation-specific subtypes from vertebrates, supporting the data previously reported by Drabent et al. (1999) for the mussel *M. edulis* H1 genes. The H1D histone gene from the sea urchin *S. purpuratus* is also included in this group, strengthening the hypothesis that this gene is defined as a differentiation-specific subtype from invertebrates (Lieber et al. 1988). An additional group that includes the *S. purpuratus* H1B and H1G histone genes is also placed in a basal position, close to the differentiation-specific subtypes group.

In a previous work by Schulze and Schulze (1995) the H1M histone gene from *Xenopus laevis* was defined as a vertebrate differentiation-specific subtype, although clear dissimilarities from H5/H1^0 genes were reported. This result contrasts with those obtained in the present work, in which the H1M gene is placed in the vertebrate somatic H1s monophyletic group. Possibly, the restriction of the data analyzed by Schulze and Schulze to amino acid comparisons, referring only to the H1 protein central domain, constitutes a serious bias in their results. In any case, further analysis will be necessary to clarify the evolutionary status of this gene, given that "orphon" characteristics such as polyadenylated mRNAs and independent organization from multiple histone gene clusters are presented by H1M.

Copy Number and Chromosomal Location of H1 Genes

The copy number of the H1 genes in the genomes of the species analyzed has been estimated to be about 100 to 110 copies per haploid genome on average (Fig. 6A), supporting previous results obtained in *M. edulis* (Drabent et al. 1999). The high copy number of H1 genes could be related to the telomeric chromosomal position of several H1 copies in *M. galloprovincialis* (Fig. 6B), as stated previously for *Drosophila* histone genes (Fitch et al. 1990). Genes located near heterochromatic regions are generally less active than their euchromatic counterparts, and it is probable that the former may have come to exist in higher copy numbers to achieve similar levels of expression.

In this sense, the estimated amount of codon bias in *Mytilus* H1 histone genes (Table 2) differs greatly from the codon bias values obtained for two core histone genes (H3 and H4) in the same species. The easiest and classical explanation involves differences in selective constraints between core and linker histone genes, but the observed codon bias is still low for the high expression levels presented by H1 histone genes. Taking into account the telomeric chromosomal position of H1 genes in *M. galloprovincialis* (Fig. 6B), the low recombination rates presented by such heterochromatic regions are likely to be a critical factor in determining the low codon bias values presented by genes located near these chromosomal positions, as predicted by the hitchhiking and background selection models, together with the Hill–Robertson effect (Hill and Robertson 1969; Kaplan et al. 1989; Charlesworth et al. 1993). Low codon bias has been empirically reported for histone genes in *D. melanogaster*, which are located fairly close to the centromere of chromosome 2 (Fitch and Strausbaugh 1993), and also for highly expressed genes close to telomeric regions in *D. melanogaster* (Munté et al. 2001). It is probable that differences in selective constraints between core and linker histone genes account for the codon bias and high synonymous divergence values in *Mytilus* H1 genes, but is not possible to discard the additional effect of the telomeric chromosomal location of these genes.

The results reported in the present work show the presence of significant homologies between the H1 histones characterized in the genus *Mytilus* and the vertebrate differentiation-specific linker histones (H5 and H1^0) and sea urchin H1D histone. These histone genes would have their direct evolutionary predecessor in an "orphon" group, which would have diverged from the main group during the rise of the phylum Vertebrata. Thus, it is not possible from our results to discard the effect of the telomeric chromosomal position of several H1 repetitions on the copy number or on the amount of codon bias detected for H1 histone genes. Further analysis of H1 histone genes from invertebrates will be necessary to trace and place the evolutionary events which gave rise to the presence of sequence elements such as the poly (A) tail, related to gene expression patterns in an independent way from DNA replication, a key proccess in the divergence of this "orphon" group of histone genes.

Acknowledgments. This work was supported by a FEDER grant awarded to Dr. J. Méndez (IFD97-1295). We are grateful to ANFACO–CECOPESCA, Dr. H. Hummel, and Dr. J. Ausió for kindly supplying the mussel specimens and to Dr. L. Cornudella for sharing an H1 histone gene clone from *M. edulis*. Thanks are also due to Dr. L. Sánchez and Dr. M.F. Ruiz for their critical discussions and for reading the English manuscript. J.M. Eirín-López was presented a predoctoral FPU fellowship award from the Spanish Government.

References

Albig W, Drabent B, Kunz J, Kalff-Suske M, Grzeschik KH, Doenecke D (1993) All known human H1 histone genes except the H1^0 gene are clustered on chromosome 6. Genomics 16:649–654

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

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

Brocard MP, Triebe S, Peretti M, Doenecke D, Khochbin S (1997) Characterization of the two $H1^0$-encoding genes from *Xenopus laevis*. Gene 189:127–134

Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular evolution. Genetics 134:1289–1303

Childs G, Maxson R, Kedes L (1981) Orphons: Dispersed genetic elements derived from tandem repetitive genes of eucaryotes. Cell 23:651–663

Coles LS, Wells JRE (1985) An H1 histone gene-specific 5′ element and evolution of H1 and H5 genes. Nucleic Acids Res 13:585–594

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

Dalton S, Wells JR (1988) A gene-specific promoter element is required for optimal expression of the histone H1 gene in S-phase. EMBO J 7:49–56

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 histone gene in the phylum annelida. J Mol Evol 46:64–73

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

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

Fitch DHA, Strausbaugh LD (1993) Low codon bias and high rates of synonymous substitution in *Drosophila hydei* and *D. melanogaster* histone genes. Mol Biol Evol 10:397–413

Fitch DHA, Strausbaugh LD, Barrett V (1990) On the origins of tandemly repeated genes: Does histone gene copy number in *Drosophila* reflect chromosomal location? Chromosoma 99:118–124

González-Tizón AM, Martínez-Lage A, Rego I, Ausió J, Méndez J (2000) DNA content, karyotypes, and chromosomal location of 18S–5.8S–28S ribosomal loci in some species of bivalve molluscs from the Pacific Canadian coast. Genome 43:1065–1072

Grunwald D, Lawrence JJ, Khochbin S (1995) Accumulation of histone $H1^0$ during early *Xenopus laevis* development. Exp Cell Res 218:586–595

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

Hill WG, Robertson A (1969) The effect of linkage on limits to artificial selection. Genet Res 8:269–294

Kaplan NL, Hudson RR, Langley C (1989) The ''hitchhiking'' effect revisited. Genetics 123:887–899

Khochbin S, Wolffe AP (1994) Developmentally regulated expression of linker-histone variants in vertebrates. Eur J Biochem 225:501–510

Krieg PA, Robins AJ, D'Andrea R, Wells JRE (1983) The chicken H5 gene is unlinked to core and H1 histones. Nucleic Acids Res 11:619–627

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular Evolutionary Genetic Analysis software. Bioinformatics 17:1244–1245

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

Martinclli R, Heintz N (1994) H1TF2A, the large subunit of a heterodimeric, glutamine-rich CCAAT-binding transcription factor involved in histone H1 cell cycle regulation. Mol Cell Biol 14:8322–8332

Matsuo Y (2000) Molecular evolution of the histone 3 multigene family in the *Drosophila melanogaster* species subgroup. Mol Phylogenet Evol 16:339–343

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

Mezquita J, Connor W, Winkfein RJ, Dixon GH (1985) An H1 histone gene from rainbow trout (*Salmo gairdnerii*). J Mol Evol 21:209–219

Miller DJ, Harrison PL, Mahony TJ, et al. (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

Munté A, Aguadé M, Segarra C (2001) Changes in the recombinational environment affect divergence in the *yellow* gene of Drosophila. Mol Biol Evol 18:1045–1056

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous snd nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Ohsumi K, Katagiri C (1991) Occurrence of H1-subtypes specific to pronuclei and cleavage stage cell nuclei of anuran amphibians. Dev Biol 147:110–120

Pandey NB, Marzluff WF (1987) The stem loop structure at the 3′ end of histone mRNA is necessary and sufficient for regulation of histone mRNA stability. Mol Cell Biol 7:4557–4559

Peretti M, Khochbin S (1997) The evolution of the differentiation-specific histone H1 gene basal promoter. J Mol Evol 44:128–134

Petrov DA, Hartl DL (1999) Patterns of nucleotide substitutions in *Drosophila* and mammalian genomes. Proc Natl Acad Sci USA 96:1475–1749

Ponte I, Vidal-Taboada JM, Suau P (1998) Evolution of the vertebrate H1 histone class: Evidence for the functional differentiation of the subtypes. Mol Biol Evol 15:702–708

Ramakrishnan V, Finch JT, Graziano V, Lee PL, Sweet RM (1993) Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. Nature 362:219–223

Rice EL, Bird CJ (1990) Relationships among geographically distant population of *Gracilaria verrucosa* (Gracilariales, Rhodophyta) and related species. Phycologia 29:501–510

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

Rodríguez-Juiz AM, Torrado M, Méndez J (1996) Genome-size variation in bivalve molluscs determined by flow cytometry. Mar Biol 126:489–497

Rozas J, Rozas R (1999) DnaSP version 3: An integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15:174–175

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Schulze E, Schulze B (1995) The vertebrate linker histones $H1^0$, H5, and H1M are descendants of invertebrate ''orphon'' histone H1 genes. J Mol Evol 41:833–840

Simpson RT (1978) Structure of chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. Biochemistry 17:5524–5531

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

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface:Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

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

Walter L, Klinga-Levan K, Helou K, Albig W, Drabent B, Doenecke D, Gunther 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

Widom J (1998) Chromatin structure: Linking structure to function with histone H1. Curr Biol 8:R788–R791

Wolffe AP, Khochbin S, Dimitrov S (1997) What do linker histones do in chromatin? BioEssays 19:249–255

Wright F (1990) The "effective number of codons" used in a gene. Gene 87:23–29

Yang Z (2000) Phylogenetic analysis by maximum likelihood (PAML), version 3.0. University College London, London

Zlatanova J, Van Holde K (1992) Histone H1 and transcription: Still an enigma? J Cell Sci 103:889–895