Molecular evolutionary characterization of the mussel *Mytilus* histone multigene family: first record of a tandemly repeated unit of five histone genes containing an H1 subtype with "orphon" features*

José M. Eirín-López[a], M. Fernanda Ruiz[b], Ana M. González-Tizón[a], Andrés Martínez[a], Lucas Sánchez[b], Josefina Méndez[a†]

[a] Departamento de Biología Celular y Molecular, Universidade da Coruña, Campus de A Zapateira s/n, E-15071, A Coruña, Spain
[b] Departamento de Biología Celular y del Desarrollo, Centro de Investigaciones Biológicas - CSIC, E-28006, Madrid, Spain

Journal of Molecular Evolution, volume 58**,** issue 2, pages 131–144, february 2004
Received 15 may 2003, accepted 21 july 2003

How to cite:

Molecular evolutionary characterization of the mussel *Mytilus* histone multigene family: first record of a tandemly repeated unit of five histone genes containing an H1 subtype with "orphon" features. José M. Eirín-López, M. Fernanda Ruiz, Ana M. González-Tizón, Andrés Martínez, Lucas Sánchez, Josefina Méndez, *J Mol Evol.*, (2004), 58(2): 131-144. https://doi.org/10.1007/s00239-003-2531-5

Abstract

The present work represents the first characterization of a clustered histone repetitive unit containing an H1 gene in a bivalve mollusk. To complete the knowledge on the evolutionary history of the histone multigene family in invertebrates, we undertake its characterization in five mussel *Mytilus* species, as an extension of our previous work on the H1 gene family. We report the quintet H4–H2B–H2A–H3–H1 as the major organization unit in the genome of *Mytilus galloprovincialis* with two 5S rRNA genes with interspersed nontranscribed spacer segments linked to the unit, which is not justified by their cotranscription with histone genes. Surprisingly, 3′ UTR regions of histone genes show two different mRNA termination signals, a stem-loop and a polyadenylation signal, both related to the evolution of histone gene expression patterns throughout the cell cycle. The clustered H1 histones characterized share essential features with "orphon" H1 genes, suggesting a common evolutionary origin for both histone subtypes which is supported by the reconstructed phylogeny for H1 genes. The characterization of histone genes in four additional *Mytilus* species revealed the presence of strong purifying selection acting among the members of the family. The chromosomal location of most of the core histone genes studied was identified by FISH close to telomeric regions in *M. galloprovincialis*. Further analysis on nucleotide variation would be necessary to assess if H1 proteins evolve according to the birth-and-death model of evolution and if the effect of the strong purifying selection maintaining protein homogeneity could account for the homologies detected between clustered and "orphon" variants.

Keywords: mussel *Mytilus*, histone gene quintets, "orphon" genes, phylogeny, copy number, FISH

The nucleotide sequences for histone genes described in the present work have been deposited in the GenBank Database under the following accession numbers: *Mytilus galloprovincialis* histone repetitive unit (AY267739); *M. californianus* H2A (AY267759), H2B (AY267741), H3 (AY267745), H4 (AY267752); *M. chilensis* H2A (AY267756), H2B (AY267744), H3 (AY267746), H4 (AY267751); *M. edulis* H2A (AY267757), H2B (AY267742), H3 (AY267749), H4 (AY267754); *M. galloprovincialis* H2A (AY267755), H2B (AY267740), H3 (AY267748), H4 (AY267750); *M. trossulus* H2A (AY267758), H2B (AY267743), H3 (AY267747), H4 (AY267753).

Introduction

Histones are a small set of basic proteins found in all eukaryotic organisms, involved in DNA packaging in nucleosomes and also in the regulation of gene expression. There are five histone classes which can be classified in two major groups according to structural and functional criteria: core histones (H2A, H2B, H3, H4), which form multiprotein complexes and interact with DNA to constitute the nucleosome structure, and linker histones (H1), which interact with DNA stretches between nucleosomes giving rise to the chromatosome structure and participate in nucleosome positioning (Simpson 1978).

Core histones show a typical tripartite structure with a central globular domain containing a fold motif involved in protein–protein interactions forming H2A–H2B heterodimers and (H3–H4)2 tetramers. The central domain is flanked by two terminal tails (N and C terminal) without tertiary structure, where the N-terminal domains play a critical role in regulation of gene expression by interacting with different transcription factors (Arents and Moudrianakis 1995). The typical structure of H1 proteins consists of a central trypsin-resistant globular domain, which contains a winged-helix domain, flanked by two nonorganized terminal tails (Ramakrishnan et al. 1993). The long C-terminal tail is the most important domain, playing a role in nucleosome positioning and regulation of gene expression (Khochbin and Wolffe 1994; Wolffe et al. 1997).

With the exception of the H4 histone, for which variants have not been described, these proteins can also be classified on the basis of their genomic organization and expression patterns, such as replication dependent, replication independent, stage specific, and tissue specific (Isenberg 1979; Maxson et al. 1983; Doenecke et al. 1997). Replication-dependent histones are expressed only in the S-phase of the cell cycle during DNA replication, whereas replication-independent types are expressed at low levels but continuously throughout the cell cycle. Stage specific histones are generally and specifically expressed during early embryogenesis and tissue specific types are detected only in particular cell types such as testis and avian nucleate erythrocytes (Hentschel and Birnstiel 1981; D'Andrea et al. 1985; Ohsumi and Katagiri 1991).

In many organisms core and linker histones are organized in clusters containing several copies of the five histone gene classes (Hentschel and Birnstiel 1981). These show a typical tandem arrangement in invertebrate genomes (Maxson et al. 1983), whereas this tandem configuration is generally lost in vertebrate genomes (D'Andrea et al. 1985; Albig et al. 1997; Wang et al. 1997). Additional organizations of histone genes are usually observed. Independent clusters containing only H1 genes have been described in annelids (Sellos et al. 1990; del Gaudio et al. 1998) and mussels (Drabent et al. 1999), which coexist in the genome with clustered repetitions of the five histone classes. The presence of histone quartets (without H1) has also been reported for newts (Stephenson et al. 1981), sea stars (Cool et al. 1988), corals (Miller et al. 1993), recently for the mussel *Mytilus edulis* (Albig et al. 2003), and also in the genome of Drosophila virilis coexisting with histone quintets (Domier et al. 1986). Different reports have proposed an "orphon" hypothesis to explain the evolutionary origin of the independently organized H1 genes. It is generally accepted that the exclusion of a gene from a gene family, and consequently from the homogenization events acting over the members, could give rise to an "orphon" gene, which hereafter would be under totally

different evolutionary constraints to those guiding the evolution of the gene family (Childs et al. 1981; Schulze and Schulze 1995). Such an "orphon" origin has been proposed for the independent H1 clusters isolated in the mussel *M. edulis* (Drabent et al. 1999) and for a group of H1 genes in four additional *Mytilus* species (Eirín-López et al. 2002).

So far, H1 clusters and core histone clusters have been independently characterized in one mussel species. The purpose of this paper is to undertake the molecular characterization of the histone repetitive unit in the mussel *M. galloprovincialis* and to study its evolution by analyzing the core histone genes in four additional *Mytilus* species. These studies are an extension of our previous work on H1 genes. Our results show that clustered H1 genes share common features with "orphon" variants at their promoter regions and at the protein central conserved domain, suggesting a common evolutionary origin for both histone subtypes. Additionally, nucleotide substitution numbers and codon bias values were estimated for histone genes and discussed in relation to copy number and chromosomal location analyses.

Materials and methods

General methods

Mussel specimens 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 Balcobo (Atlantic coast of Spain), and *M. trossulus* from Esquimalt Lagoon (Pacific coast of Canada). Genomic DNA was purified in CTAB buffer (2% hexaclecytrimethylammonium bromide, 1.4 M NaCl, 1 00 mM Tris–HCl, pH 8.0, 20 mM EDTA, 0.2% β-mercaptoethanol) following the protocols described by Rice and Bird (1990) and Winnepenninckx et al. (1993).

Screening of a *Mytilus galloprovincialis* genomic library

A genomic library from the mussel *M. galloprovincialis* synthesized in EcoRI/λ DASH II vector (Stratagene) was screened for histone genes using as a probe the coding region for the H1 gene (1100 bp) of this species (Eirín-López et al. 2002). The probe was labeled with [^{32}P]dCTP using the Rediprime II Kit (Amersham Pharmacia Biotech) and subsequently used in the hybridization procedure as follows: filters were prehybridized under high stringency conditions for 12 h at 68°C, in a solution of 5× SSC, 0.1% SDS, 2% nonfat powdered milk, 100 µg/mL denatured herring sperm DNA. The labeled probe was added to the solution and the hybridization was carried for 24 h at 68°C. Posthybridization washes of the filters were successively performed in a solution of 2× SSC, 1% SDS for 30 min at room temperature. Positive clones were verified by Southern blotting experiments, again using the H1 probe.

DNA sequencing of the repetitive unit

PCR primers were designed from consensus of coding regions for histone genes in invertebrates to amplify core histone genes in the mussel *M. galloprovincialis*. Primers annealing in coding regions were defined as: 5′-H2A-partial (AAG AGG TAA AAG TGG AAA GGC CCG) and 3′-H2A-partial (TAG CTT GAT TTG CCG GTC TTC TTG), with a resulting fragment of 366 bp; 5′-H2B-partial (CAA AGT CAA CGG CAC CCC GTG) and 3′-H2B-partial (TTT GGC GAG TTC ACC TGG CAG), amplifying a 278-bp-long fragment; 5′-H3-partial (TCG CAA ATC TAC AGG AGG GAA GGC) and 3′-H3-partial (CAT GAT GGT AAC CCT CTT GGC GTG), giving a fragment of 340 bp; and 5′-H4-partial (AAA GGA GGA AAG GGA CTG GGA) and 3′-H4-partial (CTG GCG TTT CAA GGC GTA CAC), with a resulting fragment of 267 bp. PCR amplifications from template DNA (25 ng) were performed in a final volume of 25 µL, where primers were used at 10 µM in PCR reactions, with 1 U/µL of *Taq* DNA polymerase (Roche Molecular

Biochemicals). Reaction conditions were the same for the four histones, with a first denaturation step of 4 min and 30 s at 95°C, followed by 35 cycles consisting of a 30 s denaturation step at 95°C, 30 s of annealing at 51°C, and 30 s of extension at 72°C. A final extension step of 5 min was performed at 72°C.

The resulting products were digoxigenin-labeled (Roche Molecular Biochemicals) by PCR and subsequently used in Southern blot to verify the presence of core histone genes in the positive clones of the screening. Additionally, the primers annealing in coding regions were also used for PCR amplifications to corroborate the Southern blot results. The sequence of the histone cluster was determined using a primer walking strategy where intermediate primers were designed (Table 1) until determining the complete DNA sequence of the unit, through automatic DNA sequencing directly from PCR products in an ABI PRISM 377 sequencer (Applied Biosystems, Perkin–Elmer). Coding regions for histone genes were identified by means of BLAST software by comparing the obtained sequences with those of the ORFs in the databases. The complete nucleotide sequence of the unit has been deposited in the GenBank Database under accession number AY267739.

Table 1. Primers used for DNA sequencing of the histone repetitive unit in *M. galloprovincialis*

Primer	Sequence (5′→3′)	Position
H3.15	GGATTCGGAGGTTAAACAGC	461–480
H3.13	ATTCTTCATCTGATTAGTCCG	1023–1043
H3.11	TACATTTCAGGAGTATACATC	1640–1661
H3.9	CATCCAGGCTGTACTTCTGCC	2244–2264
H3.7	TTTAAGAGAATGTGACAGCAG	2881–2901
H3.5	TCCGATTGGTATAGACATGC	3256–3275
H3.3	TGCAAACATTGCGGCTAGC	3671–3689
H3.1	CTATGTGGCTTCTTAACTCC	3866–3885
H3.2	TTCGTTTCCAGAGCTCTGC	4011–4029
H3.4	GATGTACAACATATTCAGACTC	4234–4255
H3.6	CAATCCATCTTCTAATTGCAG	4617–4637
H3.8	TGTCAATTTCCCGGGACATC	5224–5243
H3.10	ATCATTAGTTATTTGGTCCATC	5593–5614
H3.20	GACTTCCTTGACCGAGATG	5987–6005
H3.19	GATGTGTGCGAATGAACGAC	6507–6526
H3.18	TTTGCAGATTTGGCAGCAGG	7050–7069
H3.17	AAGAAGACAAAGGCTGCAGC	7089–7108
H3.22	GTGTCTACGACCATATCACG	7484–7503
H3.21	GAATACCGGGTGTTGTAGAC	7584–7603

PCR amplification and DNA sequencing of core histone genes in *Mytilus* species

Additional sets of primers were defined from DNA sequences of the repetitive unit characterized in *M. galloprovincialis* to efficiently amplify core histone genes coding and noncoding flanking regions in mussel species belonging to the genus *Mytilus*. These primers were defined as: 5′-H2A-full (ACT ACC TGG AAG AAG CGA T) and 3′-H2A-full (ACA GAG AAA TGG AGG GAG T), 5′-H2B-full (GTC ATT TTG GGG TGG GAC ACA G) and 3′-H2B-full (CAA AAC ATC GCT TCT TCC AGG TAG), 5′-H3-full (TGT GTG CCA AAT GTT AGC TTG G), and 3′-H3-full (CAG TAA CCT GAC TGT CTT GGT CT), and 5′-H4-full (ATT CCT ACA GAG TTA CCT CCC GGA T) and 3′-H4-full (AAG TTG GAC AAG TTG GAC AGG

AGA). Histone genes were amplified in four other species (*M. californianus*, *M. chilensis*, *M. edulis*, and *M. trossulus*) by PCR under the same conditions as described above for *M. galloprovincialis* except for the annealing temperature, which was 52°C. A 661-bp-long fragment was obtained in the case of H2A, 663 bp for H2B, 878 bp for H3, and 601 bp for H4. Automatic DNA sequencing was performed directly from the PCR products in a CEQ 8000 sequencer (Beckman Coulter). The GenBank accession numbers of these sequences are: *M. californianus* H2A (AY267759), H2B (AY267741), H3 (AY267745), H4 (AY267752); *M. chilensis* H2A (AY267756), H2B (AY267744), H3 (AY267746), H4 (AY267751); *M. edulis* H2A (AY267757), H2B (AY267742), H3 (AY267749), H4 (AY267754); *M. galloprovincialis* H2A (AY267755), H2B (AY267740), H3 (AY267748), H4 (AY267750); and *M. trossulus* H2A (AY267758), H2B (AY267743), H3 (AY267747), H4 (AY267753).

Nucleotide substitution numbers in *Mytilus* histone genes

Multiple alignments of amino acid and nucleotide sequences were conducted using Clustal X software (Thompson et al. 1997) with the default parameters given by the program. The proportions of synonymous (p_S) and nonsynonymous (p_N) substitution differences per site were computed for all sequences by the modified Nei–Gojobori method (Zhang et al. 1998). Phylogenetic trees were reconstructed from sequence alignments by means of the neighbor-joining algorithm (Saitou and Nei 1987), testing the inferred topology with 1000 bootstrap replicates. All the steps in the analysis were conducted using the MEGA package, version 2.1 (Kumar et al. 2001). The amount of codon bias presented by *Mytilus* histone 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]), where the highest value (61) indicates that all synonymous codons are used equally (no bias), and the lowest (20) that only one codon is used in each synonymous class (extreme bias).

Copy number of histone genes in mussel genomes

Different amounts of genomic DNA (200, 100, 50, 25 ng) from each of the five species were transferred onto a nylon membrane together with 0.8, 0.4, 0.2, and 0 ng of PCR product from the amplification of histone genes (coding regions) as a reference. The haploid DNA complement consists of 1.605 pg in *M. californianus* and 1.510 pg in *M. trossulus* (González-Tizón et al. 2000) and of 1.710 pg in *M. chilensis* and *M. edulis* and 1.920 pg in *M. galloprovincialis* (Rodríguez-Juiz et al. 1996). The blot was hybridized with 100 ng/mL of each of the digoxigenin-labeled probes (whole coding regions of H2B and H3 from *M. galloprovincialis* lacking noncoding regions to avoid nonspecific signals), in a final volume of 25 mL of solution, where probes were not at limiting conditions. The resulting signals were detected by chemiluminescence and hybridization intensity was quantified using Q-Win Image Analysis Software (Leica Imaging System).

Fluorescent in situ hybridization on *Mytilus galloprovincialis* chromosomes

M. galloprovincialis specimens were collected from intertidal rocky shores and, once in the laboratory, were placed in tanks with filtered seawater and fed with a microalgae suspension of *Isochrysis* sp. and *Tetraselmys* sp. for at least 10 days. Metaphases were obtained as described by González-Tizón et al. (2000). The DNA probe used was a digoxigenin-labeled fragment of approximately 4.2 Kbp containing only core histone genes that were amplified by PCR using the clone λ2[5,1]1a DNA as a template and the primers 5′-H4-full/3′-H3-full at 10 µM. Hybridization was performed using 100 ng of labeled probe and the signal was detected by immunocytochemical incubations in mouse anti-digoxigenin, rabbit anti-mouse FITC (fluorescein isothiocyante), and goat anti-rabbit FITC antibodies, with a final chromosome counterstaining using propidium iodide (50 ng/mL antifade). Chromosome preparations were visualized and photographed using a Leica DM RXA fluorescence microscope on Sensia-Fujichrome 400 ASA film (Fujifilm).

Results

Isolation and organization of histone genes

Five positive clones, λ2[5,1]1a, λ2[5,2]1a, λ4[13,7]1a, λ7[25,13]1a, and λ8[29,21]1a, were isolated from the genomic library of *Mytilus galloprovincialis* following the procedure described in Materials and methods. The presence of H1 genes was corroborated by means of Southern blot analyses on phage DNA digested with *Eco*RI (Fig. 1A). The presence of the remaining histone genes (H2A, H2B, H3, H4) was evidenced in the clones λ2[5,1]1a and λ2[5,2]1a through PCR amplifications using specific primers annealing in histone coding regions. The unit was subsequently sequenced by primer walking (GenBank accession No. AY267739). The sequence analysis displayed the presence of five histone genes and two 5S rRNA genes in a fragment of approximately 8 Kbp with the following organization: H4>, <H2B, H2A>, H3>, H1>, 5S>, 5S>. Intergenic spacers separate the genes from one another (Fig. 1B). A final round of DNA sequencing of the positive clone λ2[5,1] revealed the first nucleotides of the open reading frame for the H4 gene near the 5S rDNA genes in the 3′ region of the histone unit (the beginning of a new unit) and the last nucleotides of the open reading frame for the H1 gene flanking the H4 gene in the 5′ region of the unit (the final segment of the previous unit). We understand these results as evidence for the presence of histone genes organized in repetitive units which are grouped in clusters.

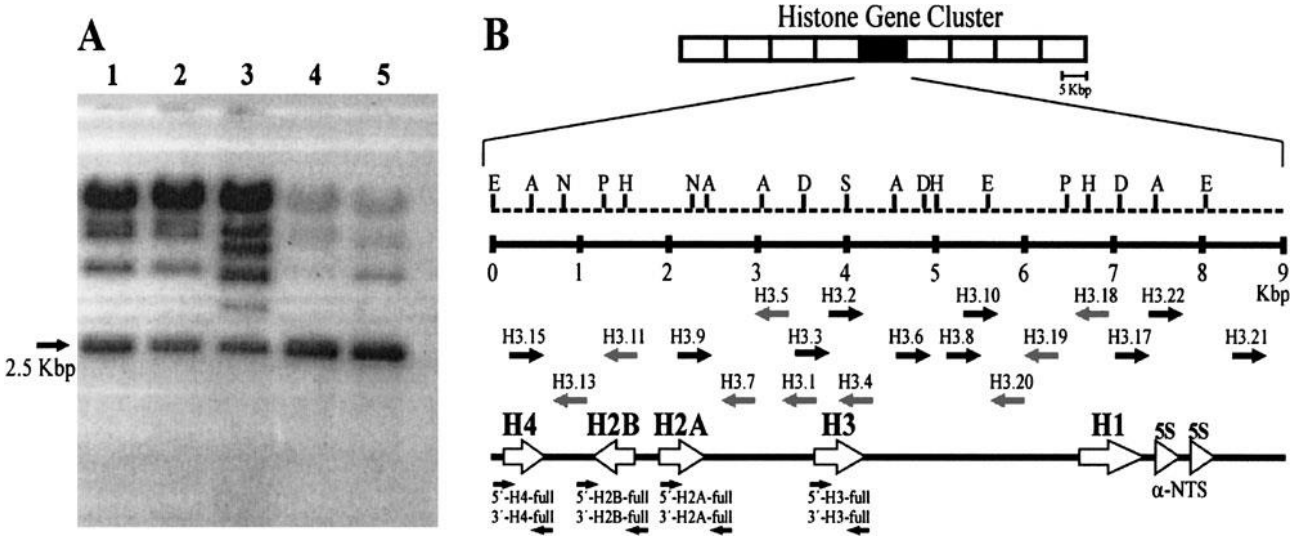

Figure 1. Southern blot of phages isolated from the genomic library of *M. galloprovincialis* (**A**) and histone gene cluster in clone λ2[5,1]1a. (**B**). **A** Lanes are as follows: 1, λ2[5,1]1a; 2, λ2[5,2]1a; 3, λ4[13,7]1a; 4, λ7[25,13]1a; 5, λ8[29,21]1a. All clones showed a common hybridization signal of 2.5 Kbp. **B** Restriction map of the characterized repetitive unit in the clone λ2[5,1]1a. E, *Eco*RI; A, *Acc*I; H, *Hin*dIII; P, *Pst*I; N, *Hin*dII; D, *Dra*I; S, *Sac*I. Position and orientation of the intermediate and specific primers used in the amplification of the repetitive unit, organization, and polarity of histone genes and 5S rRNA genes in the unit are indicated below.

The units characterized in the present work (quintets) were not under the ancestral exclusion events which gave rise to clusters containing only tandemly arranged “orphon” H1 histone genes. Taking into account that these quintets gave rise to the two additional types of clusters reported in *Mytilus* genomes (Drabent et al. 1999; Albig et al. 2003) and also regarding the coordinate expression of histone genes (Hentschel and Birnstiel 1981), we define major units as those gathering the five histone classes.

Analysis of histone sequences in the repetitive unit

The sequence of the unit is composed of a first ORF of 309 bp coding for an H4 protein of 103 residues, located 163 bp downstream of the 5′ *Eco*RI site (Fig. 2A). The next ORF is inverted and spaced by 782 bp from the H4 gene, and consists of 372 bp encoding an H2B protein of 124 amino acids (Fig. 2B). A short spacer segment of 282 bp separates this gene from a 375-bp-long ORF, which encodes for an H2A histone type of 125 residues (Fig. 2C). After a long spacer region of 1470 bp, the ORF for histone H3 is found, which is 408 bp long and encodes for 136 amino acids (Fig. 2D). The next ORF comprises 594 bp encoding an H1 protein of 198 amino acids, which is separated by a 2374 bp-long spacer region from H3 (Fig. 2E). Both 5S rRNA genes are 300 bp downstream from the end of the stop codon of H1 and encoded by the same DNA strand. A short segment of 41 bp corresponding to an α-NTS region is first identified, followed by a 5S rRNA gene of 121 bp separated by another α-NTS region of 132 bp from the last 5S rRNA gene (Fig. 2F).

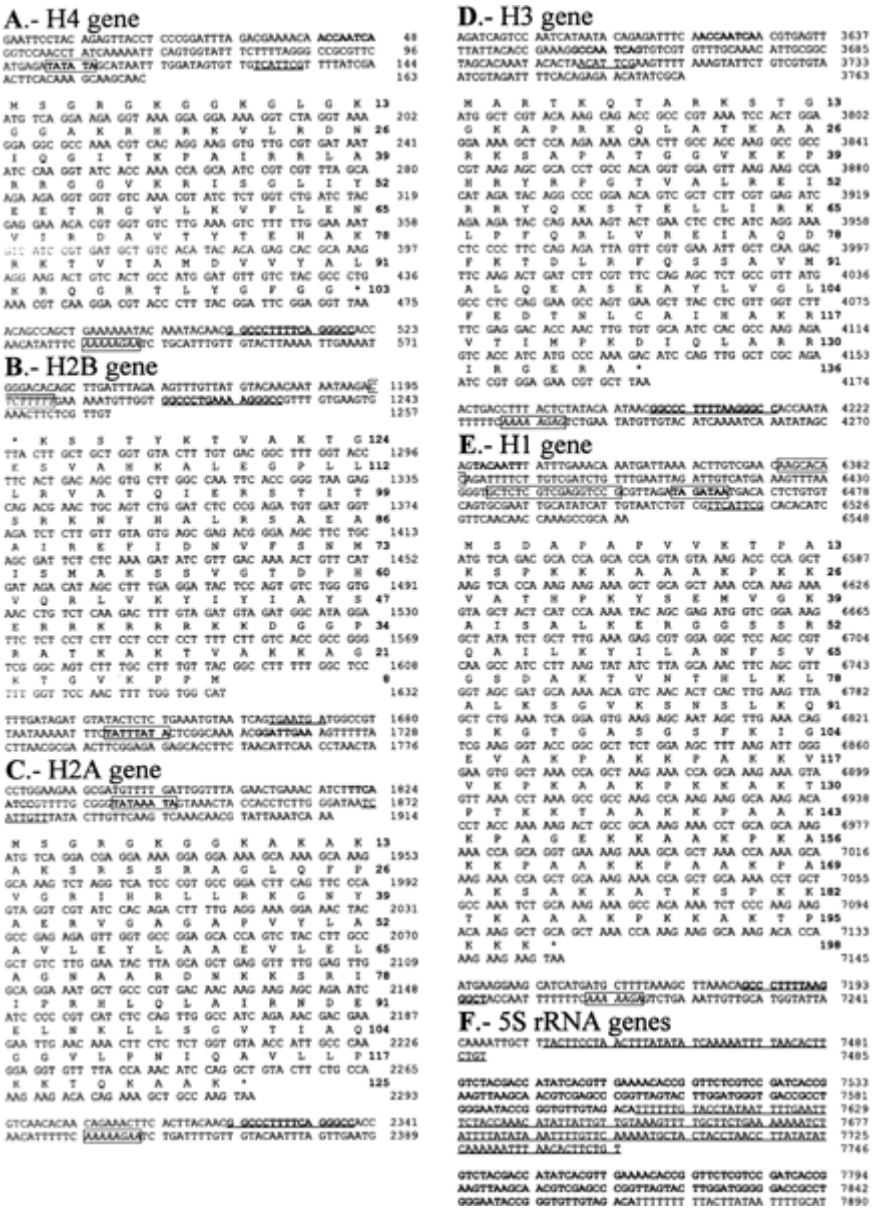

Figure 2. Nucleotide and amino acid sequences of *M. galloprovincialis* histone genes encoded by the repetitive unit isolated from clone λ2[5,1]1a with flanking 5′/3′ untranslated regions. Numbering on the right refers to the nucleotide sequence and numbering in boldface refers to amino acid residues. Translated amino acids are placed above the correspondingcodons. The conserved promoter motifs are indicated as follows: TATA boxes are in boldface and boxed, putative CAP sites are underlined, CAAT boxes are in boldface, and the H1 box-like (positions 6376 to 6383) and the

H4 box (positions 6435 to 6451) elements are boxed in the promoter region of the H1 gene. In the 3′ UTR, regions of conserved elements are indicated as follows: stem-loop orhairpin terminal structures are in boldface and underlined, and purine-rich elements are in italics and boxed. Note that the amino acid sequence of H2B appears inverted because it is encoded by the complementary DNA strand. 5S rRNA genes linked to the histone repetitive unit are reflected in **F**, where 5S genes are indicated in boldface and the interspersed α-NTS spacers are underlined.

Specific subtype features were absent, with the exception of the H4 gene, where the presence of an alanine residue at position 74, typical of stage-specific subtypes (Drabent et al. 1995), has been identified. No typical features of vertebrate-specific histones were identified in mussel histone genes, except for H1 promoter regions, although the high K:R ratio presented by these genes (58:2) is typical of somatic subtypes (del Gaudio et al. 1998).

The nucleotide sequence for the H1 histone is clearly divergent from the H1 genes characterized in other *Mytilus* species, with several indel (insertion/deletion) events at the C-terminal domain of the molecule. A deletion of two residues is observed at position 114 (Ala–Lys), together with an insertion of three, five, and one residue at positions 137 (Thr–Ala–Ala), 166 (Pro–Ala–Ala–Lys–Pro), and 185 (Thr), respectively. A high density of amino acid motifs enriched in basic residues is also observed in this domain, possibly representing phosphorylation sites (Mezquita et al. 1995). Comparison of the protein central conserved domain reveals that, at the amino acid level, clustered H1 proteins of *M. galloprovincialis* show the highest degree of homology with the consensus for H1 proteins in mussels, followed by that of H1D proteins in sea urchin and that of H5/H1^0 in vertebrates (Fig. 3).

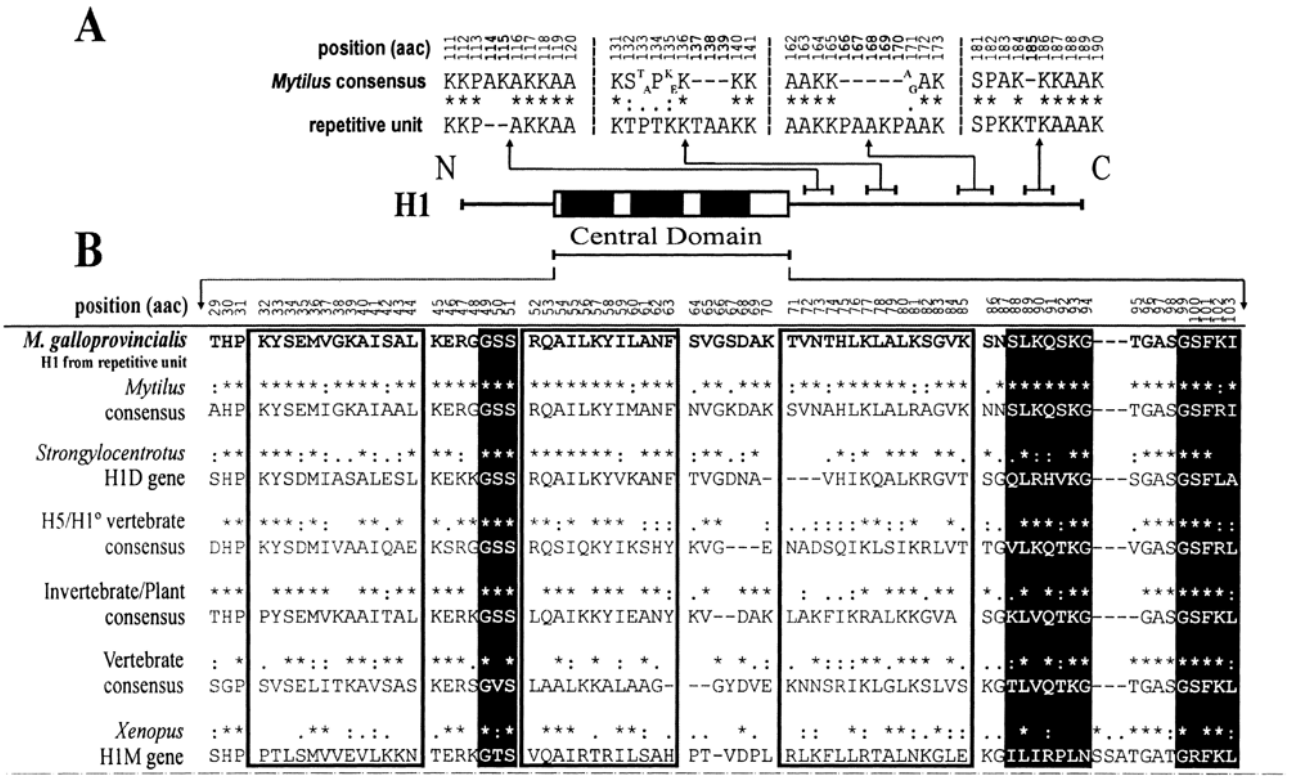

Figure 3. Length variation and analysis of the central protein domain in H1 histones. **A** Location and characterization of amino acid indel events (in boldface) in the H1 C-terminal domain, compared with the consensus sequence in the genus *Mytilus.* Asterisks indicate perfect matches, colons indicate a high degree of homology, and dots indicate a low degree of homology between residues. **B** Comparison of the central conserved domain between clustered H1 proteins and consensus sequences for other organisms. Homology is indicatedas in **A**. The three α-helix regions in the winged-helix motif are emphasized by open boxes and β-sheet structures are indicated with black boxes.

Noncoding flanking regions of the histone genes

The promoter regions contain several elements common to other genes transcribed by RNA polymerase II. Perfect TATA signals were identified for H2A, H2B, and H4 genes at positions −62 to −69, −69 to −76, and −56 to −61, respectively. H1 had a slight modification, showing the sequence TAGATAA. The typical sequences 5′-GATCC-3′ and 5′-CCTAATTTGCATATG-3′ (Maxson et al. 1983) were not detected. The short sequence 5′-MCATTCP-3′, which represents putative CAP sites, was also present in all genes, generally 50 to 70 bp upstream of the start codon (Sures et al. 1980). From our results, the consensus CAP sequence for *M. galloprovincialis* histone genes was defined as 5′-$^{Pu}{}_{T}{}^{T}{}_{A}$CATTCPu-3′. These genes do not contain either the short sequence CCCTCT/G, which is present upstream from the CAP sites in *Drosophila* histone genes, or ATTTGCAT, which is specific for the H2B promoter region and involved in replication-dependent expression (Sturm et al. 1988). The CAAT box signal was identified in all cases and was present twice in the H3 gene, a characteristic feature of vertebrate histone genes (Connor et al. 1984). Histone H1 promoter regions showed the presence of typical elements of linker genes such as an H1 box-like element (−166 to −173), followed by an H4 box element (−98 to −114). The latter element occupies the same position that the CAAT box occupies in somatic subtypes, and is typical of H4 genes and linker "orphon" variants (Peretti and Khochbin 1997). Only the presence of one enhancer sequence was identified for the H4 gene at positions −41 to −47, matching perfectly with the consensus sequence defined for histone genes (TG$^{A}{}_{T}{}^{A}{}_{T}{}^{A}{}_{T}$G [Connor et al. 1984]), TGGATAG being for *M. galloprovincialis*. Additionally, a high degree of homology between promoter regions of H2A and H4 genes was detected, where the first nine residues of both proteins match perfectly. This is consistent with that previously reported for sea urchins (Sures et al. 1978), suggesting a common evolutionary origin for both proteins.

Each of the histone genes analyzed show the typical palindrome sequence forming the stem-loop structure at noncoding 3′ terminal regions, followed by a purine-rich element 14 bp downstream. The stem-loop consensus sequence for the genes analyzed was defined as 5′-$^{G}{}_{A}$GCCCTTTT$^{C}{}_{A}$AGGGC$^{T}{}_{C}$-3′ (Table 2). Surprisingly, all genes show at least one additional mRNA termination signal downstream to the palindrome sequence, a polyadenylation signal which is typical of replication-independent histone genes expressed at constant but low levels throughout the cell cycle and in quiescent differenciated cells (Hentschel and Birnstiel 1981; Hankeln and Schmidt 1991; Akhmanova et al. 1997; del Gaudio et al. 1998; Barzotti et al. 2000).

The spacer regions between histone genes showed a great density of repetitions of the sequences TACA, TA, and TGAT. The simple sequence $(AAAG)_{17}$ was detected between H4 and H2B genes, and 12 repetitions of the motif TTCG together with 4 reiterations of the heptanucleotide $A(GA)_2(AT)_2$ were observed in the spacer segment downstream of the H2A histone. Series of three repetitions forming the sequence 5′-$AC(TA)_2(GA)_2$ATACAGAG-3′ and two repetitions of 5′-$TA(GA)_2$ATACAAAGAA$(TA)_2$GC-3′ were also present in the intergenic region between H2A and H3.

Table 2. Transcription termination signals in *M. galloprovincialis* genes

***M. galloprovincialis* histone genes**	**Stem-loop signal**	**Purine-rich motif**	**Poly(A) signal**
H1	+37 AGCCCTTTTAAGGGCT	+65 AAAAAGAG	Y
H2A	+29 GGCCCTTTTCAGGGCC	+58 AAAAAGAA	Y
H2B	+27 GGCCCTTTTCAGGGCC	+55 AAAAAGAG	Y
H3	+25 GGCCCTTTTAAGGGCC	+54 AAAAAGAG	Y
H4	+29 GGCCCTTTTCAGGGCC	+58 AAAAAGAA	Y
M. galloprovincialis	$^{C}{}_{A}$GCCCTTTT$^{C}{}_{A}$AGGGC$^{T}{}_{C}$	AAAAAGA$^{G}{}_{A}$	Y
S. purpuratus	GGC$^{C}{}_{T}$CTTTTCAG$^{G}{}_{A}$GCC	CAAGAAAGA	N
P. dumerilii	GGCC$^{T}{}_{A}$TTTTAA$^{T}{}_{A}$GGCC	CAAAAGA	N
C. variopedatus	GG$^{C}{}_{T}$CCTT^{T}A$_{C}$T$^{T}{}_{C}$AGG$^{G}{}_{A}$CC	C$^{C}{}_{A}$ $^{G}{}_{A}$ $^{G}{}_{A}$GAAA	Y
C. thummi	^{C}G$_{A}$GTC$^{T}{}_{C}$TTTT$^{C}{}_{T}$A$^{A}{}_{G}$G$^{A}{}_{G}$C^{C}G$_{T}$	AA$^{G}{}_{A}$A$^{G}{}_{A}$A$^{G}{}_{A}$	Y
A. aquaticus	GG^{G}C$_{T}$ $^{T}{}_{C}$C$^{C}{}_{T}$ATT$^{C}{}_{T}$ $^{G}{}_{A}$G$^{T}{}_{C}$ ^{C}G$_{A}$CC	A$^{A}{}_{C}$AA$^{A}{}_{G}$AGA	Y
D. hydei	$^{G}{}_{T}$ ^{G}T$_{C}$CCTTTTCAGG$^{A}{}_{G}$C^{T}C$_{G}$	$^{C}{}_{A}$ $^{C}{}_{A}$A$^{A}{}_{G}$GA$^{G}{}_{A}$ ^{A}C$_{T}$ $^{A}{}_{T}$	N
O. mykiss	GGCTCTTTTAAGAGCC	$^{A}{}_{T}$ $^{G}{}_{C}$AAA$^{G}{}_{A}$	N

Nucleotide substitution numbers in *Mytilus* histone genes

To complete the molecular and the evolutionary characterization of the histone multigene family in the genus *Mytilus*, core histone genes were isolated and sequenced in four additional mussel species. The numbers of synonymous (p_S) and nonsynonymous (p_N) nucleotide differences per site were estimated by means of the modified Nei–Gojobori method (Zhang et al. 1998). Nucleotide variability in histone genes was essentially synonymous and there were no particular species clearly divergent from the others with respect to all histone genes. The highest values of synonymous and nonsynonymous divergence were observed in the case of the H4 histone, with 0.265 ± 0.034 and 0.013 ± 0.007 substitutions per site, respectively (Table 3). *M. californianus* and *M. edulis* were the most divergent species at the synonymous level in the case of H2A and H4 genes. On the other hand, *M. chilensis* and *M. edulis* showed the highest values of synonymous divergence in the case of H2B and H3 genes, and they were also the most divergent pair of species at nonsynonymous level in the case of H3. The minimum synonymous divergence was presented by the pair *M. galloprovincialis/M. trossulus* in all genes but H2A, which showed the lowest value in *M. chilensis/M. galloprovincialis.*

Nonsynonymously, *M. trossulus* was the most divergent species in the case of histones H2A and H2B, and *M. chilensis* in the case of H3 and H4.

Table 3. Average numbers of synonymous (p_S) and nonsynonymous (p_N) substitutions in mussel histone genes[a]

	H2A		H2B		H3		H4	
	$p_S \pm SE$	$p_N \pm SE$	$p_S \pm SE$	$p_N \pm SE$	$p_S \pm SE$	$p_N \pm SE$	$p_S \pm SE$	$p_N \pm SE$
Mc/Mch	0.063 ± 0.023	0.000 ± 0.000	0.046 ± 0.020	0.000 ± 0.000	0.229 ± 0.036	0.010 ± 0.006	0.371 ± 0.054	0.023 ± 0.012
Mc/Me	0.090 ± 0.027	0.000 ± 0.000	0.028 ± 0.015	0.000 ± 0.000	0.134 ± 0.031	0.007 ± 0.005	0.403 ± 0.054	0.023 ± 0.012
Mc/Mg	0.072 ± 0.024	0.000 ± 0.000	0.009 ± 0.009	0.000 ± 0.000	0.202 ± 0.034	0.003 ± 0.003	0.194 ± 0.043	0.009 ± 0.006
Mc/Mt	0.063 ± 0.023	0.004 ± 0.004	0.019 ± 0.013	0.004 ± 0.004	0.202 ± 0.034	0.003 ± 0.003	0.215 ± 0.044	0.009 ± 0.006
Mch/Me	0.027 ± 0.016	0.000 ± 0.000	0.055 ± 0.022	0.000 ± 0.000	0.263 ± 0.039	0.014 ± 0.007	0.032 ± 0.017	0.000 ± 0.000
Mch/Mg	0.009 ± 0.009	0.000 ± 0.000	0.037 ± 0.018	0.000 ± 0.000	0.034 ± 0.016	0.007 ± 0.005	0.338 ± 0.053	0.020 ± 0.013
Mch/Mt	0.018 ± 0.013	0.004 ± 0.004	0.046 ± 0.020	0.004 ± 0.004	0.034 ± 0.016	0.007 ± 0.005	0.349 ± 0.053	0.015 ± 0.012
Me/Mg	0.036 ± 0.018	0.000 ± 0.000	0.018 ± 0.013	0.000 ± 0.000	0.245 ± 0.038	0.007 ± 0.005	0.370 ± 0.054	0.020 ± 0.013
Me/Mt	0.045 ± 0.019	0.004 ± 0.004	0.028 ± 0.016	0.004 ± 0.004	0.245 ± 0.038	0.007 ± 0.005	0.359 ± 0.054	0.015 ± 0.012
Mt/Mg	0.027 ± 0.015	0.004 ± 0.004	0.009 ± 0.009	0.004 ± 0.004	0.000 ± 0.000	0.000 ± 0.000	0.022 ± 0.015	0.000 ± 0.000
Average	0.045 ± 0.012	0.002 ± 0.002	0.025 ± 0.008	0.002 ± 0.002	0.159 ± 0.024	0.007 ± 0.003	0.265 ± 0.034	0.013 ± 0.007

[a] Species name abbreviations: Mc, *M. californianus*; Mch, *M. chilensis*; Me, *M. edulis*; Mg, *M. galloprovincialis*; Mt, *M. trossulus.* Estimations are in units of substitutions per site ± standard deviation. The average is the arithmetic mean.

Histone proteins show a tripartite structure with a central conserved domain flanked by two terminal tails without tertiary structure. The nucleotide substitution numbers have also been determined for the three protein segments in the four core histone classes in order to analyze the presence of specific constraints acting on each domain. Comparisons of p_S and p_N values between the different protein domains are shown in Fig. 4. Regarding the central folded domain, H3 and H4 histones showed the highest levels of synonymous divergence, with average values of 0.189 ± 0.036 and 0.333 ± 0.042 substitutions per site, respectively, followed by the N-terminal and C-terminal tails. In the case of the H2A gene, the N-terminal domain was the most divergent at the synonymous level (0.055 ± 0.029) substitutions per site on average), whereas the C-terminal domain was the most divergent in the H2B gene (0.037 ± 0.034 synonymous substitutions per site on average). Additionally, the central domain showed the highest values of nonsynonymous divergence in all cases, especially in the H4 gene (0.016 ± 0.011 substitutions per site on average), followed by H2B (0.014 ± 0.006), H3 (0.011 ± 0.005), and H2A (0.003 ± 0.003).

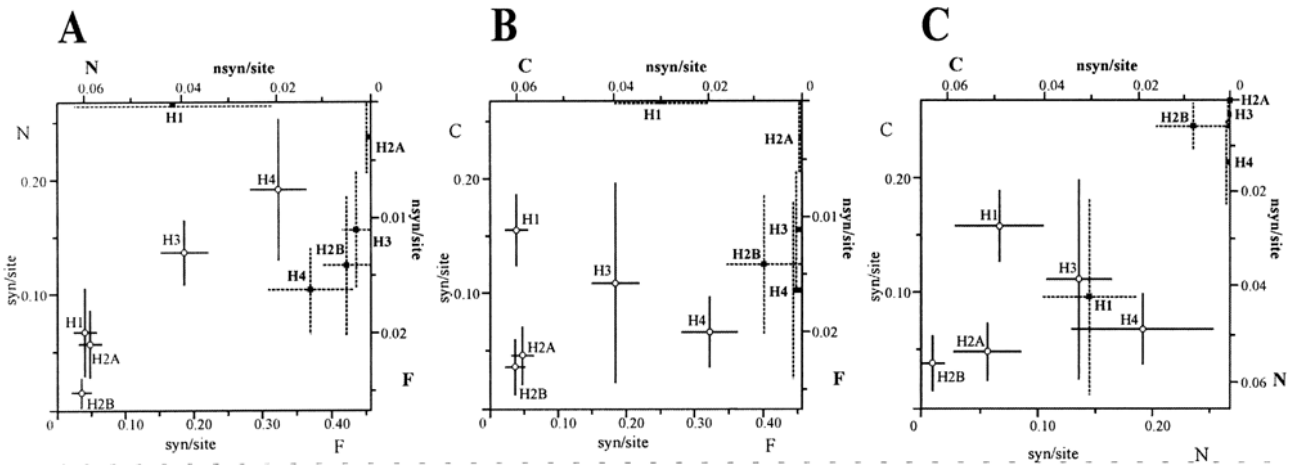

Figure 4. Synonymous and nonsynonymous average substitution numbers per site among the analyzed *Mytilus* species in the three functional domains of the five histone classes. N, N-terminaldomain; F, central folded domain; C, C-terminal domain. **A** N-terminal domain versus central folded domain. **B** C-terminal domain versus central folded domain. **C** C-terminal domain versus N-terminal domain. Open circles represent synonymous substitution numbers per site (lower axis) and filled boxes represent nonsynonymous substitution numbers per site (upper axis). Standard deviations are represented by solid lines (synonymous substitutions) and dashed lines (nonsynonymous substitutions).

Although a high degree of nucleotide divergence was presented by H3 and H4 genes, the phylogeny reconstructed from the complete nucleotide coding regions clearly discriminates between the five different histone classes and reflects a monophyletic origin for each histone class (Fig. 5). It is important to note that the clustered H1 gene from *M. galloprovincialis* is placed in an independent branch from the remaining *Mytilus* H1 genes previously described by Eirín-López et al. (2002). The characterization of H1 histone genes was completed by including mussel clustered H1 genes in the phylogenetic analysis of H1s from a broad number of species initially reported by Eirín-López et al. (2002). The bootstrapped topology (1000 replicates) was rooted with the H1-like gene from the trypanosomatidae *Leishmania braziliensis*. Results show that clustered H1 genes are placed in the monophyletic subgroup of the vertebrate differentiation-specific subtypes, where "orphon" H1 genes from *Mytilus* and the H1D gene from the sea urchin *Strongylocentrotus purpuratus* are also included. The amount of codon bias in *Mytilus* histone genes was estimated as the "effective number of codons" (ENC) index (Wright 1990). H4 is the most biased gene (35.015 ± 6.975, averaged for all species), followed by H3 (43.429 ± 2.339), H2A (43.507 ± 1.299), and H2B (50.522 ± 1.642). A comparison of the three protein functional domains of each histone revealed that the central folded domain including the fold-motif is the most biased segment in all histone genes but H2B (Fig. 6).

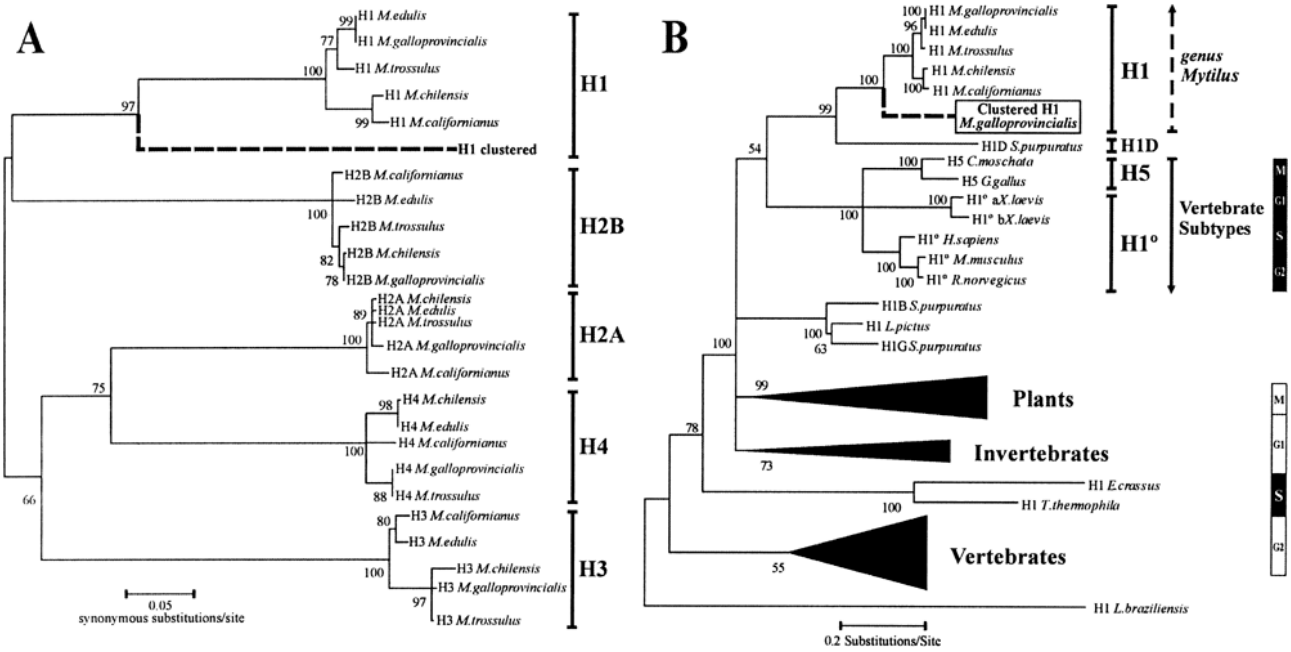

Figure 5. Phylogenetic relationships among histone genes. **A** Phylogeny reconstructed from complete histone nucleotide coding regions in the genus *Mytilus* discriminating between the five major eukaryotic histone classes, marked by brackets. Bootstrap values (1000 replicates) are placed in the corresponding nodes. The sequences for histone genes are available under the following GenBank accession numbers: *M. californianus* H1 (AJ416421), H2A (AY267759), H2B (AY267741), H3 (AY267745), H4 (AY267752); *M. chilensis* H1 (AJ416422), H2A (AY267756), H2B (AY267744), H3 (AY267746), H4 (AY267751); *M. edulis* H1 (AJ416423), H2A (AY267757), H2B (AY267742), H3 (AY267749), H4 (AY267754); *M. galloprovincialis* H1 (AJ416425), H2A (AY267755), H2B (AY267740), H3 (AY267748), H4 (AY267750); and *M. trossulus* H1 (AJ416425), H2A (AY267758), H2B (AY267743), H3 (AY267747), H4 (AY267753). **B** Relationships among H1 genes from several species, grouped in different phyla and in one group of differentiation-specific subtypes. Gene expression patterns along the cell cycle are indicated in the right margin, where black boxes indicate active gene expression in the highlighted cell phase.

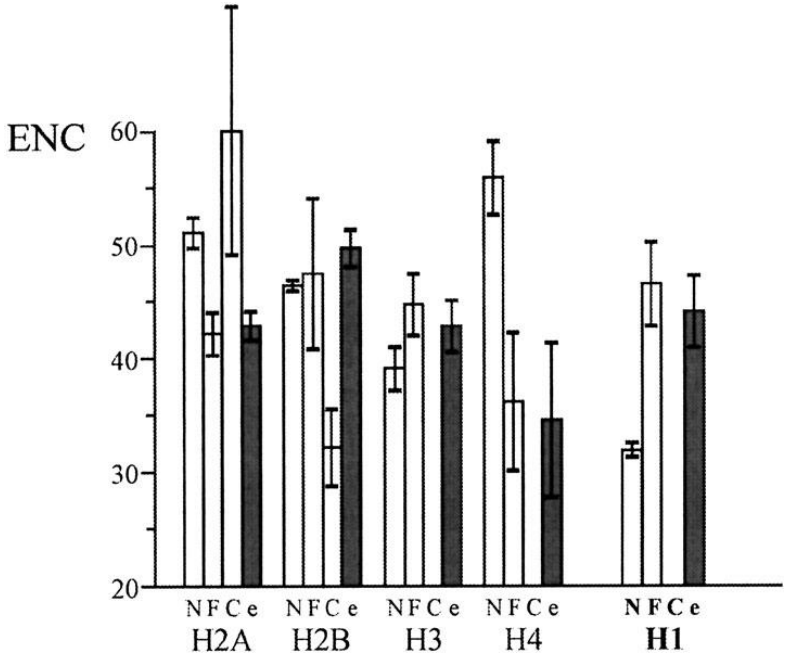

Figure 6. Codon bias values in the structural domains of mussel histone genes. Abbreviations are as in Fig. 4. Values are given as an average, and in the case of the entire protein (e) they are represented by shaded bars. Thin lines indicate the standard deviations. Absence of bars indicates that the ENC index is not applicable.

Copy number determination and chromosomal location of histone genes

The copy number of histone genes in each mussel species was estimated as indicated in the Materials and methods. The average copy number of H2A/H2B genes was 212 copies per haploid genome (*M. trossulus*, 152 copies; *M. californianus*, 195 copies; *M. galloprovincialis*, 224 copies; *M. edulis*, 239 copies; and *M. chilensis*, 251 copies) and about 201 copies per haploid genome for H3/H4 (*M. californianus*, 133 copies; *M. trossulus*, 178 copies; *M. chilensis* and *M. galloprovincialis*, 224 copies; *M. edulis*, 248 copies) (Fig. 7A).

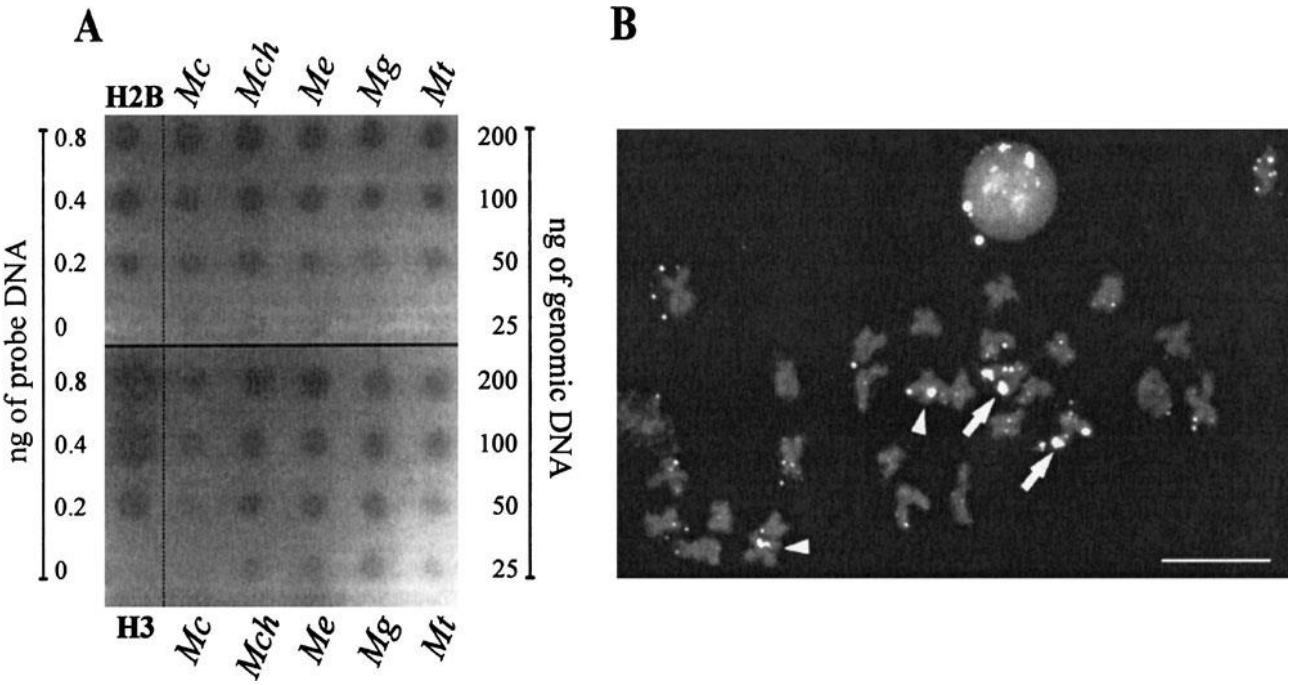

Figure 7. Copy number and chromosomal location of mussel histone genes. **A** Dot-blot results for core histone genes in the five *Mytilus* species analyzed (abbreviations are as in Table 3). The intensity of one histone copy was estimated from the signal pattern given by different amounts of histone probe. The copy number was resolved from this pattern. **B** FISH microphotograph after hybridization with a digoxigenin-labeled histone probe in metaphase chromosomes of *M. galloprovincialis* counterstained with propidium iodide. Arrows indicate terminal hybridization signals, and arrowheads indicate interstitial signals. Bar =10 μm.

In situ hybridization of the repetitive unit without H1 and 5S rRNA genes to *M. galloprovincialis* chromosomes revealed the presence of two pairs of signals in two chromosome pairs (Fig. 7B), revealing the presence of two *loci* containing core histone genes in two different sites in the genome. One *locus*, which presumably gathers most of the histone genes due to the high intensity of its hybridization signal, is located at a telomeric position close to heterochromatic regions.

Discussion

Histone genes in the mussel *Mytilus galloprovincialis*

The present work is the first characterization of a histone unit containing a H1 gene together with core histone genes in a bivalve mollusk. This unit is composed of one copy of each of the five histone classes, showing the organization H4–H2B–H2A–H3–H1, where all but H2B are transcribed from the same DNA strand. DNA sequencing of regions flanking the histone unit, organization and polarity of the histone genes, and fluorescent in situ hybridization results suggest that histone genes are organized in repetitive units which are clustered in two *loci* in the genome.

The isolated unit is approximately 8 Kbp long and corresponds to a single PCR band, suggesting the presence of a single type of unit with respect to their length. Results contrast with those recently reported by

Albig et al. (2003) for *M. edulis*, where no H1 histone genes were identified in the repetitive units and two size classes were described.

Generally, there is great variability in the organization and polarity of histone genes in the clustered repetitions, even between closely related species. Previous works reveal the absence of H1 genes from the main repetition units in several invertebrate genomes. These genes are arranged in independent clusters as in the case of the mussel *M. edulis* (Drabent et al. 1999), with an "orphon" evolutionary origin which is probably common to all *Mytilus* species (Eirín-López et al. 2002). Our results show the presence of at least an additional number of H1 copies located together with the remaining histone classes forming quintets. It is not possible, however, to discard the presence of a second type of clusters comprising only core histone genes as recently described for *M. edulis* (Albig et al. 2003). The histone repetitive unit also contains two 5S rRNA genes, with an interspersed nontranscribed spacer (α-NTS) segment, which are linked to the histone unit and are encoded by the same DNA strand. Such an association has been previously reported for the crustaceans *Artemia salina* (Drouin and Moniz de Sá 1995) and *Asellus aquaticus* (Barzotti et al. 2000), although a different polarity of 5S genes was observed. Transcriptional control elements of the 5S rRNA genes reside within their coding region so that they are expected to be functionally autonomous. Given that 5S rRNA and histone genes are transcribed by different RNA polymerases (type III and type II, respectively), their linkage would not be justified by their cotranscription. A possible hypothesis could involve the invasion of the histone units by 5S genes through transposition events, and this linkage would not provide any selective advantage (Drouin and Moniz de Sá 1995).

A very interesting feature of all histone genes in the repetitive unit is the presence of two different mRNA termination signals in their 3′ UTR region, a unique characteristic of histone genes in a cluster. The typical stem-loop or hairpin-loop signal (Birnstiel et al. 1985) is followed by a purine-rich element, both signals involved in mRNA processing of replication-dependent genes (Hentschel and Birnstiel 1981). Additionally, a polyadenylation signal AATAAA, typical of replication-independent histone genes (Hentschel and Birnstiel 1981; Akhmanova et al. 1997), is located downstream to the purine-rich element. The presence of a double mRNA termination signal is unique of histone genes and common for other invertebrates such as *Chironomus thummi* (Hankeln and Schmidt 1991), *Drosophila melanogaster* (Akhmanova et al. 1997), annelids (del Gaudio et al. 1998), and crustaceans (Barzotti et al. 2000). A possible speculation about the meaning of both signals involves the progressive replacement of one of them by the other, more efficient and specific (del Gaudio et al. 1998). The presence of polyadenylation signals must be subjected to a more detailed analysis, however, because of their location with respect to the stop codons of the histone ORFs and regarding the abundance of the AT dinucleotide in several spacer regions, which can lead to confusion in the identification of polyadenylation signals (Hankeln and Schmidt 1991).

Nucleotide substitution numbers in *Mytilus* spp. histone genes

Most of the nucleotide variation presented by histone genes is essentially synonymous and H4 is the most variable gene, followed by H3, H2A, and H2B. Average synonymous divergence between *Mytilus* species for H1 genes was estimated as being 0.127 ± 0.025 substitutions per site (Eirín-López et al. 2002). The high degree of conservation of histone proteins has been maintained by purifying selection. However, the high degree of similarity at the DNA level between the gene members in the family has been classically explained as a consequence of concerted evolution, involving mechanisms such as selection, gene conversion, and unequal crossing-over. Notwithstanding, synonymous variation is still very high even between virtually identical proteins and there is great similarity between different copies of genes among distantly related species. This agrees with the birth-and-death model of evolution under strong purifying selection without concerted evolution (Nei and Hughes 1992), which has been recently proposed as the major mechanism guiding the evolution of the histone multigene families H3 and H4 (Piontkivska et al. 2002; Rooney et al. 2002).

Our results reveal that, in the genus *Mytilus*, the central folded domains of core histone genes are the most variable regions in comparison with terminal tails. This might reflect specific constraints operating on the individual domains of the molecule. Considering that core histones are critical in gene expression regulation through interactions with specific transcription factors (Wolffe et al. 1997), the higher conservation of N-terminal domains would indicate that these are the main protein segments responsible for regulation of gene expression (Ponte et al. 1998). Estimation of the synonymous (p_S) and nonsynonymous (p_N) substitution numbers in the three protein domains show that, as for entire proteins, p_S is substantially higher than p_N in the three protein segments. These results follow the assumptions made by the birth-and-death model of evolution.

The reconstructed neighbor-joining phylogenetic tree shows the relationships between the five histone classes in the five *Mytilus* species regarding synonymous substitution numbers. The tree topology suggests a monophyletic origin for histone genes, supported by high bootstrap values, being possible that all histone classes have evolved nearly simultaneously, with H1 genes arising later, as proposed by Piontkivska et al. (2002). A common evolutionary origin for H2A and H4 genes is also supported by the reconstructed topology, where both histone classes are clustered together. The phylogenetic analysis of H1 genes from several species grouped in different phyla suggests that clustered and "orphon" H1 genes from mussels share a common evolutionary origin. Both histone subtypes are included in a monophyletic group together with the differentiation-specific subtypes from vertebrates and also with the H1D gene from the sea urchin *Strongylocentrotus purpuratus*, the unique differentiation-specific subtype identified until now in invertebrates (Lieber et al. 1988). These relationships are supported by the presence of a polyadenylation signal in 3′ UTR segments of clustered H1s, typical from replication-independent histones such as vertebrate differentiation-specific subtypes. Although clustered H1 genes present "orphon" features, they are located in a branch different from that occupied by the "orphon" variants inside the *Mytilus* subtree. Further analysis will be necessary to clearly determine and characterize the different expression patterns of histone genes, a determinant process in the divergence of the "orphon" H1 histone group.

Evolution of the clustered H1 genes

The comparison between the H1 protein central conserved domain among different organisms also reveals that clustered H1 proteins share essential characteristics with "orphon" H1 genes from mussels and sea urchins (Lieber et al. 1988) and with vertebrate differentiation-specific subtypes H5 and $H1^0$. It is likely that the homologies between clustered and "orphon" H1 proteins observed in the present work are a consequence of the common origin of both molecules (Schulze and Schulze 1995). An ancestral exclusion event of several H1 copies from the main units followed by nonconservative changes, probably made during the rise of the phylum Vertebrata, would be specifically responsible for the rise of the "orphon" variants (Childs et al. 1981). Additionally, it must be taken into account that purifying selection without concerted evolution is actually assumed to be the major evolutionary force maintaining protein homogeneity in multigene families such as ubiquitins and H3 and H4 histones (Nei et al. 2000; Piontkivska et al. 2002; Rooney et al. 2002). In this case, genes can evolve independently or according to the birth-and-death model of evolution (Nei and Hughes 1992). The latter model assumes that new genes are created by repeated gene duplication and that some of them are maintained in the genome for a long time, whereas others are deleted or become nonfunctional. This model assumes that (a) replication-dependent and replication-independent histone variants are divergent, (b) p_S is substantially higher than p_N, (c) genes are clustered by type in the phylogenies (i.e., H1t, H5, $H1^0$), not by species, and (d) pseudogenes are generated. Results presented in this work and additional data from a manuscript in preparation are consistent with these assumptions, suggesting the possibility that also the H1 family is evolving following the birth-and-death model, as previously reported for the H3 and H4 families (Piontkivska et al. 2002; Rooney et al. 2002). Otherwise, it is unlikely that clustered H1s observed in *M. galloprovincialis* are pseudogenes because of the integrity of their

promoter regions (presence of H1 box, H4 box, and TATA box elements) and the detected indels do not alter the expected reading frame.

Chromosomal location and copy number of histone genes

FISH results on *M. galloprovincialis* chromosomes locate core histone genes at two *loci* in two different chromosome pairs, which support and complete previous results from Southern blot analyses in *M. edulis* (Albig et al. 2003). One of the *loci*, presumably gathering most of the genes, is located fairly close to heterochromatic regions of telomeres. The chromosomal position of H1 clusters in this species was located at three *loci* in three different chromosome pairs, two of them also fairly close to telomeres (Eirín-López et al. 2002). Taking into account chromosome morphology data, it is likely that both *loci* containing core histones described here correspond to two of the three *loci* containing H1 genes, so that the remaining *locus* would only contain H1 genes.

The proximity of core histones to heterochromatic regions in telomeres could be critical in two aspects. First, close proximity to heterochromatic regions is likely to reduce gene activity with respect to their euchromatic counterparts, as in the case of *Drosophila* histone genes (Fitch et al. 1990). Thus, a high copy number would be necessary to obtain similar amounts of histone proteins, taking into account that products of all four core histone classes are demanded nearly stoichiometrically. Indeed, the average copy number for core histone genes has been estimated to be 212 and 201 copies per haploid genome for H2A/H2B and for H3/H4, respectively. These results roughly duplicate the estimations of about 100–110 copies per haploid genome for H1 genes in the genus *Mytilus* (Eirín-López et al. 2002) and are in accordance with results obtained for *M. edulis* by Albig et al. (2003), agreeing with the 2:1 stoichiometry of core and linker histone genes.

Second, proximity to heterochromatic regions can also be conditioning codon usage, as predicted by the hitchhiking and background selections models, diminishing the amount of codon bias due to the fixation of slightly deleterious mutations (Kaplan et al. 1989; Charlesworth et al. 1993). Nevertheless, the mussel histone genes are highly biased, as expected for very actively expressed genes under strong selective constraints, suggesting that their proximity to heterochromatic regions might not be close enough to significantly affect codon usage. From our results it seems that chromosomal location of core histone genes does not affect copy number and codon usage bias in histone genes. A possible explanation involves that such a proximity is not enough to modify these parameters, but we must take into account that mussel chromosomes show relatively short heterochromatic segments at centromeres and telomeres when compared with most of invertebrate chromosomes (Martínez-Lage et al. 1994).

To our knowledge, the results described in this report are of relevance to the field in two main sections. We first report the presence of units containing H1 genes together with the remaining core histone genes in the genome of bivalve mollusks, and we also report the organization H4–H2B–H2A–H3–H1 as the major histone gene arrangement in mussel *Mytilus* repetitive units, where all but H2B genes are transcribed from the same DNA strand. On the other hand, the characterization of the five histone classes in several *Mytilus* species represents a very important contribution to improve the knowledge of this multigene family in invertebrates. Far from the classical notion of homogeneity among the family members, our results raise new questions as those concerning the evolutionary origin of clusters located at the two different chromosomal *loci*, about the homologies detected between clustered and “orphon” H1 genes, and also regarding the presence of two different mRNA termination signals in 3′ UTR regions. We are just beginning to fill the gap in the knowledge of histone genes in bivalve mollusks, and further studies will be necessary to determine clearly the evolutionary meaning of such outstanding features.

Acknowledgements

This work was supported by a PGIDT Grant (10PX110304PR) awarded to J. Méndez and by a predoctoral FPU fellowship from the Spanish Government given to J.M. Eirín- López. Thanks are due to ANFACO-CECOPESCA, Dr. H. Hummel, and Dr. J. Ausió for kindly supplying mussel specimens. We also thank two anonymous reviewers for helpful discussions and comments.

References

Akhmanova, A, Miedema, K, Kremer, H, Henning, W (1997). Two types of polyadenilated mRNAs are synthesized from *Drosophila* replication-dependent histones. *Eur J Biochem, 244*:294-300.

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.

Albig, W, Warthorst, U, Drabent, B, Prats, E, Cornudella, L, Doenecke, D (2003). *Mytilus edulis* core histone genes are organized in two clusters devoid of linker histone genes. *J Mol Evol, 56*:597-606.

Arents, G, Moudrianakis, EN (1995). The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc Natl Acad Sci USA, 92*:11170-11174.

Barzotti, R, Pelliccia, F, Bucciarelli, E, Rocchi, A (2000). Organization, nucleotide sequence, and chromosomal mapping of a tandemly repeated unit containing the four core histones genes and a 5S rRNA gene in an isopod crustacean species. *Genome, 43*:341-345.

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

Brown, D, Cook, A, Wagner, M, Wells, D (1992). Closely linked H2B genes in the marine copepod *Tigriopus californicus* indicate a recent gene duplication or gene conversion event. *DNA Sequence, 2*:387-396.

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.

Connor, W, States, JC, Mezquita, J, Dixon, GH (1984). Organization and nucleotide sequence of rainbow trout histone H2A and H3 genes. *J Mol Evol, 20*:236-250.

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 polychaete worm *Chaetopterus variopedatus*: first record of a H1 histone gene in the phylum Annelida. *J Mol Evol, 46*:64-73.

Doenecke, D, Albig, W, Bode, C, Drabent, B, Franke, K, Gavenis, K, Witt, O (1997). Histones: genetic diversity and tissue-specific gene expression. *Histochem Cell Biol, 107*:1-10.

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, 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.

Drabent, B, Louroutziatis, A, Prats, E, Cornudella, L, Doenecke, D (1995). Structure of histone H2B and H4 genes of the sea cucumber *Holothuria tubulosa*. *DNA Sequence, 6*:41-45.

Eirín-López, JM, González-Tizón, AM, Martínez, A, Méndez, J (2002). Molecular and evolutionary analysis of mussel histone genes (*Mytilus* spp.): possible evidence of an "orphon" origin for H1 histone genes. *J Mol Evol, 55*:272-283.

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.

Hankeln, T, Schmidt, ER (1991). The organization, localization and nucleotide sequence of the histone genes of the midge *Chironomus thummi*. *Chromosoma, 101*:25-31.

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

Isenberg, I (1979). Histones. *Annu Rev Genet, 48*:159-191.

Kaplan, NL, Hudson, RR, Langley, CH (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.

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.

Martínez-Lage, A, Gónzalez-Tizón, AM, Méndez, J (1994). Characterization of different chromatin types in *Mytilus galloprovincialis* L. after C-banding, fluorochrome and restriction endonuclease treatments. *Heredity, 72*:242-249.

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

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

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 tripoblastic metazoans. *J Mol Evol, 37*:245-253.

Nei, M, Gu, X, Sitnikova, T (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci USA, 94*:7799-7806.

Nei, M, Hughes, AL (1992). Balanced polymorphism and evolution by the birth-and-death process in the MHC *loci*. In: Tsuji, K, Aizawa, M, Sasazuki, T eds. 11th. *Histocompatibility workshop and conference*. Oxford University Press. Oxford, pp. 27-38.

Nei, M, Rogozin, IB, Piontkivska, H (2000). Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc Natl Acad Sci USA, 97*:10866-10871.

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.

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

Piontkivska, H, Rooney, AP, Nei, M (2002). Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol Biol Evol, 19*:689-697.

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.

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

Rooney, AP, Piontkivska, H, Nei, M (2002). Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. *Mol Biol Evol, 19*:68-75.

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.

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.

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.

Sturm, RA, Dalton, S, Wells, JRE (1988). Conservation of histone H2A/2B intergene regions: a role for the H2B specific element in divergent transcription. *Nucleic Acids Res, 16*:8571-8586.

Sures, I, Levy, S, Kedes, L (1980). Leader sequences of *Strongylocentrotus purpuratus* histone mRNAs start at a unique heptanucleotide common to all five histone genes. Proc Natl Acad Sci USA, 77:1265-1269.

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

Thompson, JD, Gibson, TJ, Plewniak, F, Jeanmougin, F, Higgins, DG (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res, 25*:4876-4882.

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.

Winnepenninckx, B, Backeljau, T, De Watcher, R (1993). Extraction of high molecular weigh DNA from molluscs. *Trends Genet, 9*:407.

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.

Zhang, J, Rosenberg, HF, Nei, M (1998). Positive darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci USA 95*:3708-3713.

* This is a post-peer-review, pre-copyedit version of an article published in [Journal of Molecular Evolution]. The final authenticated version is available online at: [https://doi.org/10.1007/s00239-003-2531-5].

† fina@udc.es