

# Histone genes of the razor clam *Solen marginatus* unveil new aspects of linker histone evolution in protostomes

Rodrigo González-Romero, Juan Ausió, Josefina Méndez, and José M. Eirín-López

**Abstract:** The association of DNA with histones results in a nucleoprotein complex called chromatin that consists of repetitive nucleosomal subunits. Nucleosomes are joined together in the chromatin fiber by short stretches of linker DNA that interact with a wide diversity of linker H1 histones involved in chromatin compaction and dynamics. Although the long-term evolution of the H1 family has been the subject of different studies during the last 5 years, the lack of molecular data on replication-independent (RI) H1 variants from protostomes has been hampering attempts to complete the evolutionary picture of this histone family in eukaryotes, especially as it pertains to the functional specialization they impart to the chromatin structure in members of this bilaterian lineage. In an attempt to fill this gap, the present work characterizes the histone gene complement from the razor clam *Solen marginatus*. Molecular evolutionary analyses reveal that the H1 gene from this organism represents one of the few protostome RI H1 genes known to date, a notion which is further supported by its location within the monophyletic group encompassing the RI H1 variants in the overall phylogeny of eukaryotic H1 proteins. Although the detailed characterization of the nucleotide substitution patterns in RI H1 variants agrees with the model of birth-and-death evolution under strong purifying selection, maximum-likelihood approaches unveil the presence of adaptive selection during at least part of the evolutionary differentiation between protostomes and deuterostomes. The presence of increased levels of specialization in RI H1 proteins from deuterostomes as well as the significant differences observed in electrostatic properties between protostome and deuterostome RI H1s represent novel and important preliminary results for future studies of the functional differentiation of this histone H1 lineage across bilaterians.

*Key words:* histones, protostomes, chromatin, molecular evolution, razor clam.

**Résumé :** L'association de l'ADN avec les histones produit un complexe nucléoprotéique appelé chromatine, laquelle est composée de sous-unités nucléosomiques répétées. Les nucléosomes sont attachés les uns aux autres dans la fibre de chromatine grâce à de courtes séquences d'ADN de liaison qui interagissent avec une grande diversité d'histones H1 de liaison, lesquelles sont impliquées dans la compaction et la dynamique de la chromatine. Bien que l'évolution à long terme de la famille des histones H1 ait fait l'objet de diverses études au cours des cinq dernières années, l'absence de données moléculaires sur les variants H1 indépendants de la réplication (RI) chez les protostomes ont nuï aux efforts visant à compléter le tableau évolutif de cette famille d'histones chez les eucaryotes. Cela est particulièrement vrai en ce qui a trait à la spécialisation fonctionnelle qu'elles confèrent à la structure de la chromatine chez les membres de ces bilatériens. Afin de combler ce vide, ce travail présente la caractérisation des gènes codant pour les histones chez le couteau gaine, *Solen marginatus*. Des analyses moléculaires évolutives ont révélé que le gène H1 chez cet organisme est un des rares cas de gènes H1 RI connus à ce jour chez les protostomes, une conclusion qui est appuyée par sa position au sein d'un groupe monophylétique réunissant les variants H1 RI parmi l'ensemble de la phylogénie des protéines H1 eucaryotes. Bien que la caractérisation détaillée des substitutions nucléotidiques au sein des variants H1 RI soit en accord avec une évolution de type naissance-mort sous l'effet d'une forte sélection purificatrice, les approches de vraisemblance maximale révèlent la présence de sélection adaptative durant au moins une partie de la différenciation évolutive entre les protostomes et les deutérostomes. La présence d'une plus grande spécialisation des protéines H1 RI chez les deutérostomes ainsi que les différences significatives observées entre les H1 RI des protostomes et des deutérostomes quant aux propriétés électrostatiques constituent des résultats préliminaires inédits et importants en vue de futures études sur la différenciation fonctionnelle au sein de ce groupe d'histones H1 chez les bilatériens.

*Mots-clés :* histones, protostomes, chromatine, évolution moléculaire, couteau gaine.

[Traduit par la Rédaction]

Received 7 January 2009. Accepted 8 April 2009. Published on the NRC Research Press Web site at genome.nrc.ca on 5 June 2009.

Corresponding Editor: A. Civetta.

**R. González-Romero, J. Méndez, and J.M. Eirín-López.**<sup>1</sup> Departamento de Biología Celular y Molecular, Universidade da Coruña, E15071 A Coruña, Spain.

**J. Ausió.** Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC V8W 3P6, Canada.

<sup>1</sup>Corresponding author (e-mail: jeirin@udc.es).

## Introduction

In eukaryotes and some archaeobacteria, DNA is found associated with histones in a nucleoprotein complex called chromatin that allows for the high extent of compaction of genomic DNA within the limited space of the cell nucleus. Chromatin also provides the support on which most DNA metabolic processes (i.e., replication, repair, transcription) take place. At the structural level, a fundamental repetitive subunit, the nucleosome, results from the association of 2 left-handed superhelical turns of DNA wrapping about a protein core (consisting of 2 copies of each of H2A, H2B, H3, and H4 core histones) (van Holde 1988). Nucleosomes are joined together in the chromatin fiber by short stretches of linker DNA that interact with linker H1 histones, providing an additional folding to the chromatin fiber.

Although the different functional domains of eukaryotic chromatin all share a common nucleosomal structure, the dynamic processes responsible for the local heterogeneity observed across the genome are potentially regulated by 3 main mechanisms: replacement of canonical (replication-dependent [RD]) histones with specialized (replication-independent [RI]) variants that have dedicated functions (Malik and Henikoff 2003), the occurrence of post-translational histone modifications (Jenuwein and Allis 2001), and the association with remodeling complexes responsible for nucleosome mobilization (Owen-Hughes 2003). The wide range of possible configurations that facilitate different chromatin metabolic needs are the result of the synergistic action of the aforementioned mechanisms, through a recognition mechanism that has been referred to as the “histone code” (Strahl and Allis 2000).

The histone H1 family stands out among histones for being the most diverse (Ausió 2006). It encompasses canonical (RD) subtypes largely coupled to the S phase of the cell cycle, and replacement histone variants often encoded by solitary genes which are expressed independently of replication at basal but constant levels throughout the cell cycle (Eirín-López et al. 2009). Contrary to the classical concept of homogenization of these proteins through a process of concerted evolution, we have demonstrated that the long-term evolution of the H1 family is subject to a birth-and-death process under strong purifying selection which promotes genetic diversity (Eirín-López et al. 2004a).

In addition to its physiological relevance to chromatin of somatic tissues, histone H1 shares common features with a group of chromosomal sperm proteins referred to as sperm nuclear basic proteins (SNBPs) (Ausió 1999). During spermiogenesis, these proteins replace histones to different extents depending on the organism, and provide the tight packing of the genetic material which is characteristically found in the mature sperm nucleus (Ausió 1999; Eirín-López et al. 2006a). We have recently provided evidence that histone H1 and SNBPs are in fact descendants of a common RI histone precursor whose diversification process early in metazoan evolution led to the differentiation of canonical RD and variant RI lineages (Eirín-López et al. 2006b). Concomitantly, the functional compartmentalization of the somatic and germinal lines allowed further differentiation between RI histone H1 proteins and RI SNBPs. This led to the vertical parallel evolution of histone H1 and

SNBPs which is observed across protostomes and deuterostomes (Eirín-López et al. 2008).

Although the long-term evolution of the H1 family has been thoroughly studied in deuterostomes during the last 5 years (Eirín-López et al. 2005, 2004a, 2006a, 2006b; Nei and Rooney 2006), the lack of molecular data on RI H1 variants from protostomes has been hampering attempts to complete the evolutionary picture of this histone family in eukaryotes, especially as it pertains to the functional specialization they impart to the chromatin structure of this bilaterian lineage. Indeed, RI H1 proteins were believed to be exclusive to deuterostomes until “orphan” H1 genes with RI features were described in molluscs (Eirín-López et al. 2002, 2004b; González-Romero et al. 2008). Mollusca is of critical interest for the study of histone H1 and SNBP evolution not only because it represents the only protostome phylum where RI H1 proteins have been described so far but also because it encompasses different species representative of the 3 main types of SNBPs (histones, protamine-like proteins, and protamines) (Ausió 1999).

The present work represents an attempt to fill the gap in the knowledge of the protostome RI H1 lineage by characterizing the histone gene complement from the razor clam *Solen marginatus* and performing molecular evolutionary analyses to investigate the long-term evolution of these H1 genes. Our results reveal that the H1 gene from *S. marginatus* represents a protostome RI H1 gene subject to birth-and-death evolution under strong purifying selection. Comparisons between protostome and deuterostome RI H1 genes show increased levels of specialization in the latter case, as well as significant differences in electrostatic properties between RI H1s from the two groups of organisms. These findings represent novel and important preliminary results for future studies of the functional differentiation of this histone H1 lineage across bilaterians.

## Materials and methods

### PCR amplification and DNA sequencing of *Solen marginatus* histone genes

Razor clam specimens of the species *Solen marginatus* were collected in the locality of Boiro on the Galician coast (northwest Spain) and identified by taxonomists at the Center of Marine Investigations from Vilaxoán (Pontevedra, Spain). Genomic DNA from muscle tissue was purified in CTAB buffer following standard protocols (Rice and Bird 1990; Winnepenninckx et al. 1993). PCR primers for the 5 histone genes were designed using the repetitive unit of histones from the mussel *Mytilus galloprovincialis* (Eirín-López et al. 2004b), as follows: H1 fwd (5'-ACATATTCTG AAAGAAAAT TC-3'), H1 rev (5'-AGCAAGTACA CATGGACTTT A-3'), H2A fwd (5'-ACATTCAACC TAAC-TACCTG-3'), H2A rev (5'-TTCATTTTTT TCCCACCAAC TATT-3'), H3 fwd (5'-GAACAATTGT TAGCTTCAA-3'), H3 rev (5'-TTTCTTCTTC TTTCAATACA-3'), H4 fwd (5'-GAATTCCTAC AGAGTTACC-3'), and H4 rev (5'-TGT-ATCCACA GACTTGCTTG CC-3'). Amplification reactions from template genomic DNA were performed in a final volume of 25  $\mu$ L (10 ng/ $\mu$ L of template DNA) including 10  $\mu$ mol/L primers and 25 units of *Taq* DNA polymerase (Roche Molecular Biochemicals). The reactions consisted of

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 annealing step at 53.5 °C, and 30 s extension step at 72 °C. A final extension step of 5 min at 72 °C was performed. Automatic DNA sequencing was performed directly from the isolated PCR products, using the specific primers mentioned above. The DNA sequences of the 5 histone genes from *S. marginatus* were deposited in the GenBank Database with the following accession numbers: histone H1, FJ595834; histone H2A, FJ595835; histone H2B, FJ595836; histone H3, FJ595837; histone H4, FJ595838.

### Molecular evolutionary analyses of replication-independent histone H1 genes

A total of 209 histone H1 and related SNBP sequences (see Table S1 for details<sup>2</sup>) were included in the overall phylogenetic analysis of the razor clam H1 gene within the histone H1 family. There are no less than 12 different nomenclatures for H1 genes; in the present work the nomenclature from Doenecke's lab was followed (Albig et al. 1997). Multiple sequence alignments were conducted and edited for potential errors using the CLUSTAL\_X (Thompson et al. 1997) and BIOEDIT (Hall 1999) programs, on the basis of the translated amino acid sequences. Alignment of amino acid sequences corresponding to the core domain of H1 histones was carried out using histone H1 sequence fragments defined by previously established criteria for determining the boundaries of this domain (Ramakrishnan et al. 1993; Schulze and Schulze 1995).

All molecular evolutionary analyses were conducted using the program MEGA version 4 (Tamura et al. 2007). The extent of amino acid sequence divergence among H1 proteins and SNBPs in the global phylogeny was estimated by means of the uncorrected differences (*p*-distances), as this approach is known to give better results owing to its smaller variance (Nei and Kumar 2000). Estimations of protein and nucleotide sequence divergence within the RI H1 lineage were performed using the Poisson correction and the Kimura 2-parameter method, respectively. The numbers of synonymous (*p<sub>S</sub>*) and non-synonymous (*p<sub>N</sub>*) differences per site were also computed using the modified method of Nei and Gojobori (Zhang et al. 1998), providing the transition/transversion ratio (*R*). Evolutionary distances were calculated using the complete deletion option in all cases except for the overall H1 and SNBP protein phylogeny shown in Fig. 2A, where the pairwise deletion option was used; standard errors of the estimations were calculated using the bootstrap method (1000 replicates). The presence and nature of selection in H1 genes was tested using the codon-based Z-test for selection, defining *H*<sub>0</sub> as *p<sub>S</sub>* = *p<sub>N</sub>* and *H*<sub>1</sub> as *p<sub>S</sub>* > *p<sub>N</sub>* (Nei and Kumar 2000).

The neighbor-joining tree-building method (Saitou and Nei 1987) was used to reconstruct the phylogenetic trees in this work. To confirm that our results are not dependent on this choice, phylogenetic analyses were completed by reconstructing maximum parsimony trees. The reliability of the

resulting topologies was tested using both the bootstrap (Felsenstein 1985) and the interior branch-test (Sitnikova 1996) methods, producing the bootstrap probability (BP) and confidence probability (CP) values, respectively, for each interior node in the trees after 1000 replicates. Given the known conservative nature of the bootstrap method, BP > 80% was interpreted as high statistical support for groups, whereas CP ≥ 95% was considered statistically significant (Sitnikova et al. 1995). The amount of codon bias in RI H1 genes was referred to as the effective number of codons (Wright 1990) and was estimated using DnaSP version 4 (Rozas et al. 2003).

### Reconstruction of ancestral sequences and electrostatic potentials of H1 family members

Ancestral sequences corresponding to the internal nodes of the phylogeny of RI histone H1 genes were reconstructed by maximum likelihood using the codeml program within the PAML 4 package (Yang 2007). This allows estimation of the nucleotide substitutions involved in the differentiation between protostome and deuterostome RI lineages and the diversification of H1 members, as well as their nature (synonymous or replacement). The three-dimensional structure of the H1 protein from *S. marginatus* as well as that of all RI H1 histones used in the estimation of electrostatic distances was modeled using the coordinates determined for the crystal structure of histone H5 from chicken (Protein Data Bank accession code 1HST) as a reference in the context of the SWISS-MODEL workspace (Arnold et al. 2005); the obtained structures were rendered with the MacPyMOL program (DeLano 2007). Comparisons between the electrostatic properties of protostome and deuterostome RI H1 histones were conducted in the webPIPSA pipeline (Richter et al. 2008). Electrostatic potentials were determined using the University of Houston Brownian Dynamics program (Madura et al. 1995), and the absolute distances calculated from the similarity indices for the electrostatic potentials were represented in a colorized matrix and in an epogram (tree representation of the relationships among potentials). The representation of the electrostatic potentials in the modeled structures was implemented with the VMD program ( Humphrey et al. 1996).

## Results and discussion

### Characterization of the *Solen marginatus* histone gene sequences

PCR using primers specific for *Mytilus galloprovincialis* histone genes yielded DNA fragments encompassing the coding regions of core and linker histones as well as their 5' and 3' untranslated regions (UTRs), allowing for the analysis of some of the motifs involved in the regulation of the expression of these genes. The sequences thus obtained are shown in Fig. 1. The histone genes of *S. marginatus* are sequentially arranged in the following way: an open reading frame (ORF) of 573 bp encoding a linker H1 protein of 190 residues, a 378 bp ORF encoding an H2A protein of 125

<sup>2</sup>Supplementary data for this article are available on the journal Web site (<http://genome.nrc.ca>) or may be purchased from the Depository of Unpublished Data, Document Delivery, CISTI, National Research Council Canada, Building M-55, 1200 Montreal Road, Ottawa, ON K1A 0R6, Canada. DUD 3953. For more information on obtaining material refer to <http://cisti-icist.nrc-cnrc.gc.ca/eng/ibp/cisti/collection/unpublished-data.html>.

**Fig. 1.** Nucleotide and amino acid sequences of *Solen marginatus* histone genes. Numbering on the right refers to the nucleotide sequences and numbering in boldface refers to amino acid residues. Translated amino acids are placed above the corresponding codons. Conserved promoter elements are indicated as follows: CAAT boxes in boldface, putative CAP sites underlined, TATA boxes in boldface and underlined, the H1 box-like element in a black box, and the H4 box element in an open box (the last two both within the histone H1 promoter region). Conserved elements at 3' UTRs are indicated as follows: stem-loop structure underlined, purine-rich element in boldface.

residues, an ORF of 375 bp encoding an H2B protein of 124 amino acids, an ORF of 411 bp encoding an H3 protein of 136 residues, and a 312 bp ORF encoding an H4 protein of 103 residues. Analyses of the promoter regions revealed the presence of several *cis*-acting elements that are common to many other genes transcribed by RNA polymerase II. These include TATA signals (region -84 to -91 for histone H1, region -63 to -68 for histone H2A, region -62 to -69 for histone H2B, and region -56 to -61 for histone H4). However, the typical elements 5'-GATCC-3' and 5'-CCTAATTTGCA-TATG-3' (Maxson et al. 1983) could not be identified. Putative CAP sequences are usually present in all genes; they have the consensus sequence 5'-MCATTCAP-3' and are generally located -40 to -100 bp upstream of the initiation codon (Sures et al. 1980). The CAAT box signal was also identified in all cases, although consensus sequences upstream of CAP sites, such as CCCTCT/G (typical from *Drosophila* histone genes) or ATTTGCAT (typical from H2B promoter regions), were not detected in *S. marginatus*.

The promoter region of the *S. marginatus* H1 gene contains typical linker histone gene elements such as an H1 box-like element (-171 to -178) (Dalton and Wells 1988) followed by an H4 box element (-136 to -158), which replaces the CAAT box found in canonical RD H1 genes. The presence of an H4 box, which is typical of H4 genes, allows for clear discrimination between variant RI H1 genes and canonical RD H1 genes (Peretti and Khochbin 1997; Eirín-López et al. 2002, 2005). Therefore, the presence of an H4 box element in the H1 gene of *S. marginatus* suggests an evolutionary link with protostome "orphan" RI H1 genes of other clams and mussels (Eirín-López et al. 2002; González-Romero et al. 2008). The promoter regions of H2A and H4 exhibit a high degree of homology that extends to the conservation of the first 9 residues of the coding regions. This is in agreement with previous data obtained from sea urchins and mussels (Sures et al. 1978; Eirín-López et al. 2004b) and is consistent with the notion of a common evolutionary origin for both genes (Eirín-López et al. 2009).

Analyses of the 3' UTRs revealed in all instances the presence of the typical palindromic sequence that results in the formation of a stem-loop structure which is typical of RD histone genes (Marzluff 1992). This was followed by a purine-rich element approximately 15 bp downstream. The stem-loop sequences in the histone genes of *S. marginatus* are shown in Table 1 in comparison with those of other representative species of protostomes and deuterostomes. The consensus sequence in *S. marginatus* is 5'-G<sub>A</sub>GCCCTTTT - C<sub>A</sub>AGGC<sub>T</sub>-3'. Despite this, the H1 gene promoter elements are strongly indicative of an RI expression pattern typical of "orphan" genes (Schulze and Schulze 1995; González-Romero et al. 2008) and it is not possible to rule out the presence of a dual mechanism of gene expression that also involves polyadenylated transcripts, as was previously re-

ported for mussel and other clam H1 genes (Eirín-López et al. 2005; González-Romero et al. 2008).

### Phylogenetic location of protostome RI H1 variants within the histone H1 family

Given the presence of typical RI linker histone regulatory elements in the H1 gene of *S. marginatus*, we decided to analyze the evolution of its sequence within a broader phylogenetic context that includes all H1 proteins (both RI and RD) described until now. Furthermore, considering the close relationship between H1 proteins and SNBPs, these germinal chromatin components were also included in the phylogeny (Fig. 2A; see Fig. S1 for details on the alignment<sup>2</sup>). The topology thus obtained points to the common evolutionary origin shared by H1 and SNBPs early in metazoan evolution (Eirín-López et al. 2006a, 2006b). This encompasses a process of vertical parallel evolution across protostomes and deuterostomes that leads to differentiation between the protamine and protamine-like components of SNBPs (Ausió 1999; Eirín-López et al. 2008) as well as between RI and RD H1 proteins (Eirín-López et al. 2004a). In the latter instance, RI H1 proteins share a common monophyletic origin that consists of protostome ("orphan" H1) and deuterostome (H5/H1<sup>o</sup>) representatives including the H1 gene of *S. marginatus* (Figs. 2A and 2B).

The nucleotide-based phylogeny corresponding exclusively to RI H1 genes is shown in Fig. 2B and depicts a well-defined differentiation between the protostome and deuterostome RI lineages. The topology reflects a functional clustering of deuterostome RI variants, as would be predicted from the long-term birth-and-death evolution of these proteins (Eirín-López et al. 2005). As expected, a close relationship is observed between the H1 gene of *S. marginatus* and RI H1 genes from other bivalve molluscs. In this regard, the RI "orphan" status of the razor clam H1 gene is supported by its position in the overall H1 phylogeny and in the tree specific for RI genes. In terms of linker histone evolution, such an observation is of critical interest considering that the characterization of protostome RI H1 proteins has remained elusive for so long.

The study of the protein and nucleotide variation among H1 lineages reveals significantly more synonymous substitutions than non-synonymous substitutions ( $P < 0.001$  in all Z-test comparisons). This provides conclusive evidence for a mechanism of purifying selection guiding the long-term evolution of H1 genes, in agreement with the birth-and-death model (Eirín-López et al. 2004a). Protostome H1s seem to display amino acid and nucleotide variation levels slightly higher than those of deuterostome H1s. However, differences in the patterns of variation are also detected between different H1 lineages and taxonomic groups. Although no significant differences in variation between RD and RI lineages have been reported in deuterostomes, the results in Table 2 indicate the presence of a significantly higher degree

### Histone H1

accaatttga tatgctatat gtgaatgttt ttgatagaat tataactaaaa -250  
 attatcataa aagtcttcga ac**aaagcacat** aatcatttta cc**gagatcgc** -200  
**tctcgtcgcg gtcocg** agcta tttttaaggg tgtcttcgct gaggtccgcg -150  
 ttcgtctgtt **ataaata**aac gaggccctca actggccgct attttcaaac -100  
 gctattccat tctgtatactt cgtggaggat aatattttgg atcgttccaaa -50  
**M A D A K A A P A A A P A 13**  
 ATG GCA GAC GCA AAA GCA GCA CCA GCA GCA GCA CCA GCT 39  
**N S P K K K A A A K R K K K 26**  
 AAT TCA CCA AAG AAA AAG GCA GCA GCT AAG CGG AAG AAG 78  
**P S A H P K Y S E M I G K 39**  
 CCT TCA GCA CAT CCT AAA TAC AGC GAG ATG ATT GGA AAA 117  
**A I A A L K E R G G S S R 52**  
 GCC ATT GCT GCT TTG AAA GAA CGT GGA GGT TCC TCA AGG 156  
**Q A I L K Y I M A N F N V 65**  
 CAA GCA ATT TTG AAG TAC ATC ATG GCC AAC TFC AAC GTC 195  
**G K D A K S V N A H L K L 78**  
 GGA AAA GAT GCC AAG TCT GTA AAC GCT CAT TTA AAA CTT 234  
**A L R A G V K N N R L K Q 91**  
 GCA CTC AGA GCC GGA GTT AAG AAC AAC AGA TTG AAG CAG 273  
**S K G T G A S G S F R I G 104**  
 AGC AAG GGA ACT GGA GCA TCC GGA TCT TTC AGA ATT GGA 312  
**Q A K Q A K K K P A K A K 117**  
 CAA GCA AAG CAA GCT AAA AAG AAG CCA GCA AAG GCA AAG 351  
**A A A K P K A A A K P K E A 130**  
 TCA GCA GCT AAA CCT AAG GCA GCC AAG CCA AAG GAG GCA 390  
**K S A P E K K R A A K E P 143**  
 AAG AGC GCA CCT GAG AAA AAG AGG GCA GCA AAG GAA CCA 429  
**A R E K K A A K P K A L K 156**  
 GCT AGA GAG AAA AAA GCG GCC AAA CCT AAA GCT TTG AAA 468  
**P A A K K V A K A K K A A 169**  
 CCA GCA GCA AAG AAA GTA GCG AAG GCA AAG AAG GCA GCG 507  
**P R S P A A K K A A A K P K 182**  
 CCT AGG TCA CCA GCT AAA AAG AAG GCA GCC AAA CCA AAA 546  
**A K K T P K N K \*** 190  
 GCC AAG AAG ACA CCA AAG AAT AAG TAA 573  
 actgttcag actacagtct gcagaggcta ttcagccacc **aaaagccctt** +623  
**tttaagggct**a cccaatttgt **tcaaaaggaag** tctcaggtt atgctgtagg +673  
 tgtcagtagta gcctcagatca tagctcatgt ttacactgta acacttcgta +723

### Histone H2A

aggtcgaatc cttgcgagtc actttagcgtt at**ttcaatcc** gaattatgcy -150  
 acccttgaat gtattatatg ctggccgcgg ta**taaatatt** gacctcttgg -100  
 ggtaa**tcatt** gttttatactt gttcagtgct aaacaacgta ttaaatcaaaa -50  
**M S G R G K G G K A K A K 13**  
 ATG TCA GGG CGA GGA AAA GGA GGT AAA GCA AAG GCA AAG 39  
**A K S R S S R A G L Q F P 26**  
 GCA AAG TCT AGG TCA TCC CGT GCC GGG CTT CAG TTC CCA 78  
**V G R I H R A L R K G N Y 39**  
 GTA GGT CGT ATC CAC AGA CTT TTG AGG AAA GGA AAC TAC 117  
**A E R V G A G A P V Y L A 52**  
 GCC GAG AGA GTA GGT GCC GGA GCT CCA GTC TAC CTT GCC 156  
**A V L E Y L A V E V L E L 65**  
 GCT GTC TTG GAA TAT TTG GCA GTT GAG GTT TTG GAG TTG 195  
**A G N A V R D N K K S R I 78**  
 GCA GGA AAT GCT GTT CGT GAC AAC AAG AAG AGC AGA ATC 234  
**I P R H L Q L A I R N D E 91**  
 ATC CCC CGT CAT CTT CAG TTG GCC ATC AGA AAC GAC GAA 273  
**E L N K L L S G V T I A Q 104**  
 GAA TTG AAC AAA CTT CTC TCT GGT GTA ACC ATT GCC CAA 312  
**G G V F P N I Q A V L L P 117**  
 GGT GGT GTT TTC CCA AAC ATC CAG GCT GTA CTT CTG CCA 351  
**K K S Q K A A K \*** 125  
 AAA AAA TCA CAG AAA GCT GCC AAG TAA 378  
 agtgtccaat accctatcca atttc**ggccc** **ttttaagggc** ctcgaatatt +428  
**tttcaaaaag** agtctcgcgc agtagtatgc caatgaagat ctactcgtaa +478  
 tctaactcgt caatctctct ctctctctctg agttatcoac atctttcttt +528  
 attctctctgt ctcttttaggt cttcatttta gcacacggga tgcaccatta +578

### Histone H2B

tccaggtagt taggttgaat gttagaaggt gctctctccg aagtctcgcgt -150  
 taagtaaaaa **ctttcaatcc** gttttgccga **gtataaata**g aaatttttat -100  
 taacggccat **cattcactga** ttacatttca gagagtatac atctatcaag -50  
**M P Q K V G T K G A K K A 13**  
 ATG CCA CAA AAA GTC GGA ACC AAA GGA GCC AAA AAA GCC 39  
**V T K A K T A R P G S D K 26**  
 GTA ACA AAG GCA AAG ACT GCC CGA CCC GGC AGT GAC AAG 78  
**K R R R R K R R E S Y A I Y 39**  
 AAA AGG AGG AGG AAG AGG AGA GAA TCC TAC GCC ATC TAC 117  
**I Y K V L K Q V H P D T G 52**  
 ATC TAC AAA GTC CTG AAA CAG GTT CAC CCA GAC ACT GGA 156  
**V S S K A I S I M Y S F V 65**  
 GTA TCC TCA AAG GCT ATC TCT ATC ATG TAC AGT TTT GTC 195  
**N D I F E R I T A G A S R 78**  
 AAC GAC ATC TTT GAG AGA ATC ACT GCA GGA GCT TCC CGT 234  
**L A H Y N K R S T I T S R 91**  
 CTC GCT CAC TAC AAC AAG AGA TCT ACC ATC ACA TCT CGG 273  
**D V Q T A V R L L L P G E 104**  
 GAT GTA CAG ACT GCA GTT CGT CTG CTC TTA CCC GGT GAA 312  
**L P K H A V S E G T K A V 117**  
 TTG CCC AAG CAC GCT GTC AGT GAA GGT ACC AAA GCA GTC 351  
**T K Y T T S S R \*** 124  
 ACA AAG TAT ACC AGC AGC AGA TAA 375  
 agtcaataca acagaaactt cacttacaac **ggcccttttc** **agggcc**acca +425  
 acatttttcca **aaaagaat**ct gactttgttg tacaagtgt tagatgaaga +475  
 atacaagatt gtaactagta cactccctcc agttaagtgt actggggagga +525  
 gactgcgagat atttcccttc agataagcgt ctcgtgctat acatcccta +575

### Histone H3

ggattgaaaa ctgtgaaagt aacttcccga tcagtccaat cataatacag -200  
 agatttcaac **aatcaac**gctg agtttttatta caccgaaagg ccaatcagtg -150  
 tctgtctcgc aaacattgcy gctagcacia **atacactaac** **attcga**agtt -100  
 ttttaagtatt cctgtcgtgta atcgtagatt **ttca**cagaga acatatcgca -50  
**M A R T K Q T A R K S T G 13**  
 ATG GCT CGT ACA AAG CAG ACC GCC CGT AAA TCC ACT GGA 39  
**R K A P R K Q L A T K A A 26**  
 AGA AAA GCT CCA AGA AAA CAA CTT GCC ACC AAG GCC GCC 78  
**R K S A P A T G G V K K P 39**  
 CGT AAG AGC CCA CCT GCC ACT GGT GGA GTT AAG AAC CCA 117  
**Y R Y R P G T V A L R E I 52**  
 TAT AGA TAC AGG CCA GAC ACA GTC GCT CTT CGT GAG ATC 156  
**R R Y Q S S T E L L I R K 65**  
 AGA AGA TAC CAG AGT AGT ACT GAA TTA CTT ATC AGG AAG 195  
**L P F Q R L V R E I A Q D 78**  
 CTC CCC TTC CAG AGG TTA GTT CGT GAA ATT GCT CAG GAC 234  
**F K T D L R F Q S S A V M 91**  
 TTC AAG ACC GAT CTT CGT TTC CAG AGC TCT GCA GTG ATG 273  
**A L Q E A S E A Y L V G L 104**  
 GCC CTC CAG GAG GCC AGT GAG GCT TAC CTC GTT GGT CTT 312  
**F E D T N L R A I H A K R 117**  
 TTC GAG GAC ACA AAC TTG CGT GCA ATC CAC GCC AAG CGG 351  
**V P I M P K D I Q L A R R 130**  
 GTC CCC ATC ATG CCC AAA GAC ATC CAG TTG GCT CGC AGA 390  
**I R G E R A \*** 136  
 ATC CGT GGG GAA CGT GCT TAA 411  
 agtgtccaat accctattca **atttccggccc** **ttttaagggc** ctcgaatatt +461  
**tttcaaaaag** agtctcgcgc agtagtatgc caatgaagat ctactcgtaa +511  
 tctaactcgt caatctctct ctctctctctg agttatccac atctttcttt +561  
 attctctctgt ctcttttaggt cttcatttta gcacacggga tgcaccatta +611

### Histone H4

gttacctccc ggatttagac gaaaacaacc **aatc**aggtcc aacctatcaa -150  
 aaattcagtg gtatttcttt tagggccgcg **ttca**tgagat **atata**gcata -100  
 atttggatat **tgttgcatt** cgtttttatcg aacctcaaaa agcaacaacc -50  
**M S G R G K G G K G L S K 13**  
 ATG TCA GGA AGA GGT AAA GGA GGA AAA GGT CTA AGT AAA 39  
**G A K R H R K V V L R D N 26**  
 GGA GGC GCC AAA CGA CAC AGG AAG GTG TTG CGT GAT AAT 78  
**I Q G I T K P A I R R L A 39**  
 ATC CAA GGT ATA ACC AAG CCA GCA ATC CGT CGT TTA GCA 117  
**R R S G V K R I S G L I Y 52**  
 AGA AGA AGT GGT GTC AAA CGT ATC TCG GGT CTT ATC TAC 156  
**E E T R G V L K V F L E N 65**  
 GAA GAA ACA CGT GGT GTC TTA AAA GTC TTT TTG GAA AAC 195  
**V I R D A V T Y T E H A K 78**  
 GTG ATC CGT GAT GCT GTC ACA TAC ACT GAG CAC GCA AAG 234  
**R K T V T A M D V I Y A L 91**  
 CGC AAG ACT GTC ACT GCC ATG GAC GTT ATC TAC GCC CTG 273  
**K R Q G R T L Y G F G G \*** 103  
 AAG CGT CAA GGA CGT ACC CTT TAC GGA TTC GGA GGA TAA 312  
 actcaagcgt gtatttaaca **taaacgccc** **cccttttcag** **ggggc**accta +362  
 caaattt**aaa** **aaagaat**cag cttaattgat acgtacaaat gacaacataa +412  
 gcccttctgc aatgatcgcg gtaacgcaata tacgagattc **ccgc**ggtccc +462

**Table 1.** Transcription termination signals in *Solen marginatus* histone genes.

Histone gene	Stem-loop signal	Purine-rich motif
H1	+44 AGCCCTTTTAAGGGCT	+73 AAAGGAAG
H2A	+26 GGCCCTTTTCAGGGCC	+55 AAAAAAGAG
H2B	+31 GGCCCTTTTCAGGGCC	+60 AAAAAAGAA
H3	+26 GGCCCTTTTAAGGGCC	+55 AAAAAAG
H4	+29 GGCCCTTTTCAGGGGCC	+58 AAAAAAGAA
Consensus		
<i>Solen marginatus</i>	<sup>G</sup> <sub>A</sub> GCCCTTTT <sup>C</sup> <sub>A</sub> AGGC <sup>C</sup> <sub>T</sub>	AAA <sup>G</sup> <sub>A</sub> <sup>G</sup> <sub>A</sub> <sup>G</sup> <sub>A</sub> <sup>G</sup> <sub>A</sub> A
Veneridae	<sup>G</sup> <sub>A</sub> GCCCTTTT <sup>C</sup> <sub>A</sub> AGGGC <sup>C</sup> <sub>T</sub>	AAA <sup>A</sup> <sub>T</sub> AGA <sup>A</sup> <sub>G</sub>
Mytilidae	<sup>G</sup> <sub>A</sub> GCCCTTTT <sup>C</sup> <sub>A</sub> AGGGC <sup>C</sup> <sub>T</sub>	AAAAAGA <sup>G</sup> <sub>A</sub>
<i>Strongylocentrotus purpuratus</i>	GGC <sup>C</sup> <sub>T</sub> CTTTTCAG <sup>G</sup> <sub>A</sub> GCC	CAAGAAAGA
<i>Platynereis dumerilii</i>	GGCC <sup>T</sup> <sub>A</sub> TTTTAA <sup>T</sup> <sub>A</sub> GGCC	CAAAAAGA
<i>Chaetopterus variopedatus</i>	GG <sup>C</sup> <sub>T</sub> CCTT <sup>T</sup> <sub>A</sub> C <sup>T</sup> <sub>T</sub> AGG <sup>G</sup> <sub>A</sub> CC	C <sup>C</sup> <sub>A</sub> <sup>G</sup> <sub>A</sub> <sup>G</sup> <sub>A</sub> GAAA
<i>Chironomus thummi</i>	<sup>C</sup> <sub>G</sub> A <sup>G</sup> <sub>T</sub> C <sup>T</sup> <sub>T</sub> TTTT <sup>C</sup> <sub>T</sub> A <sup>A</sup> <sub>G</sub> <sup>G</sup> <sub>A</sub> <sup>G</sup> <sub>C<sup>C</sup><sub>G</sub>T</sub>	AA <sup>G</sup> <sub>A</sub> <sup>G</sup> <sub>A</sub> <sup>G</sup> <sub>A</sub> A
<i>Asellus aquaticus</i>	GG <sup>G</sup> <sub>C</sub> T <sup>T</sup> <sub>C</sub> C <sup>C</sup> <sub>T</sub> ATT <sup>C</sup> <sub>T</sub> <sup>G</sup> <sub>A</sub> G <sup>T</sup> <sub>C</sub> <sup>C</sup> <sub>G</sub> ACC	A <sup>A</sup> <sub>C</sub> AAA <sup>G</sup> <sub>A</sub> GAGA
<i>Drosophila hydei</i>	<sup>G</sup> <sub>T</sub> <sup>G</sup> <sub>T</sub> C <sup>C</sup> <sub>CTTTTCAGG<sub>A</sub><sup>C</sup><sub>T</sub><sup>C</sup><sub>G</sub></sub>	C <sup>A</sup> <sub>A</sub> <sup>A</sup> <sub>G</sub> <sup>G</sup> <sub>A</sub> <sup>G</sup> <sub>A</sub> <sup>A</sup> <sub>C</sub> <sup>T</sup> <sub>T</sub>
<i>Onchorynchus mykiss</i>	GGCTCTTTTAAGAGCC	A <sup>T</sup> <sub>G</sub> <sup>C</sup> <sub>AAA<sup>G</sup><sub>A</sub></sub>

of protein ( $0.373 \pm 0.030$ ) and nucleotide ( $0.355 \pm 0.014$ ) variation in protostome RD H1 variants when compared with their RI counterparts ( $0.066 \pm 0.006$  and  $0.078 \pm 0.006$ , respectively). The lower variation exhibited by protostome RI variants is in agreement with the absence of functionally specialized RI H1 variants in these organisms, likely as a result of the lower complexity of these organisms. This is in contrast to deuterostomes, which have highly specialized RI variants such as histone H5 and histone H1<sup>o</sup> (Eirín-López et al. 2008).

The promoter regions of RI histone H1 genes contain characteristic and specific regulatory elements that differ from those of RD H1 genes. The razor clam H1 gene described in this work was aligned with other RI and RD H1 genes as well as SNBPs to identify any shared conserved elements. Figure 2C shows the ubiquitous presence of the H4 box element across promoters of RI H1 genes (van Wijnen et al. 1992; Peretti and Khochbin 1997), which is replaced by the CAAT box in RD H1 genes and by a conserved control element in SNBPs. Although there is no apparent similarity between the nucleotide sequences of the H4 box and the SNBP control element, the lack of the CAAT box supports a closer evolutionary relationship between SNBPs and RI H1 histones (Eirín-López et al. 2006b), as observed in Fig. 2C. Overall, the phylogenetic analysis shown in Fig. 2C suggests that eukaryotic histone H1 arose from a replication-independent precursor gene with polyadenylated transcripts that subsequently evolved into the RD H1 lineage (Eirín-López et al. 2008, 2009).

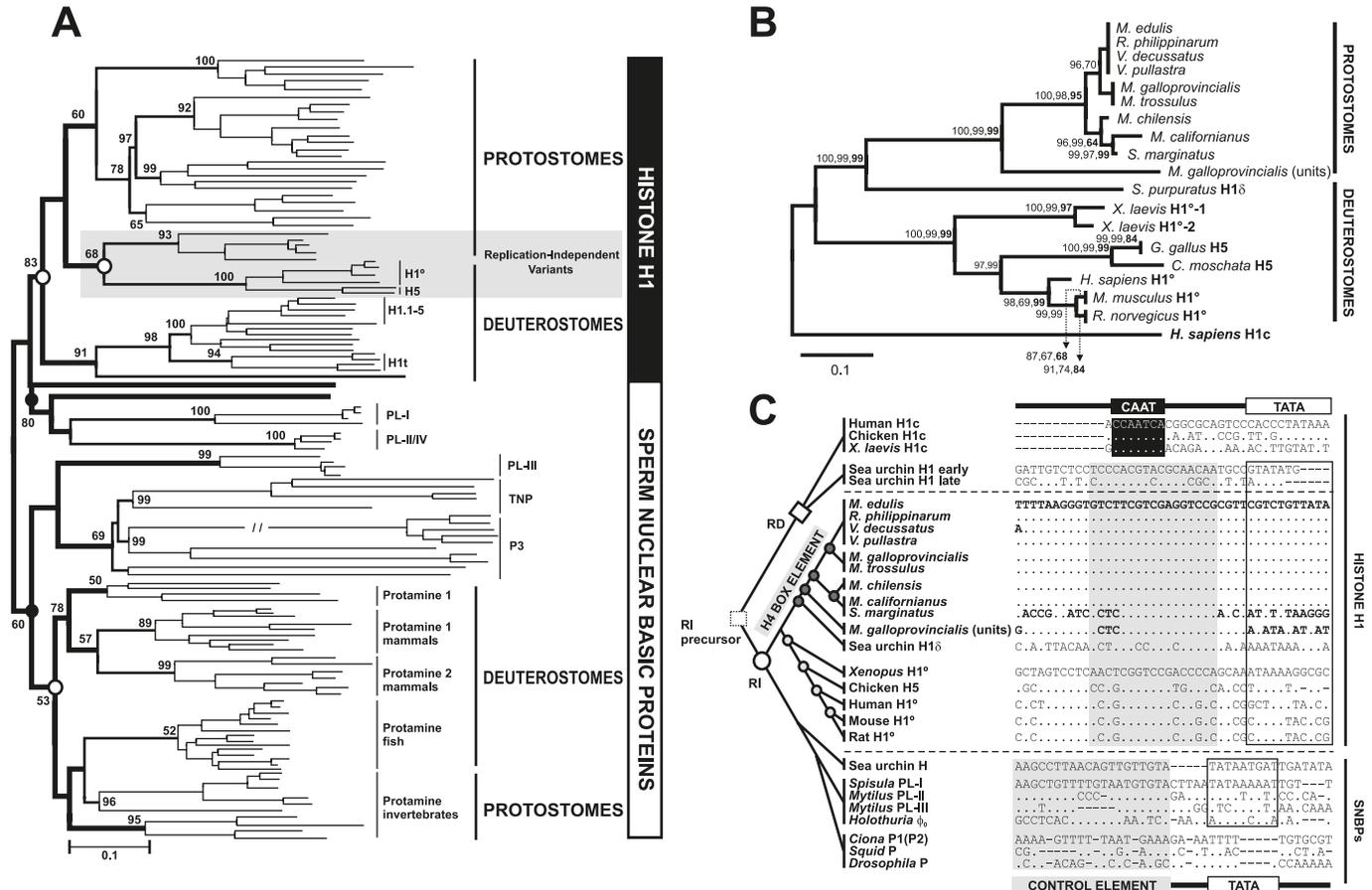
#### Long-term evolution of RI variants across protostomes and deuterostomes

Histone H1 RI variants have shorter amino acid sequences than their canonical RD counterparts. They contain 190 residues in the mussel (Eirín-López et al. 2002) and the razor clam analysed here and 185 amino acid residues in sea urchin (Lieber et al. 1988), which is smaller than the H1s encoded by the RD H1 genes of the same organisms, which

range from 211 to 217 residues. We have previously shown that the highly characteristic winged-helix core domain of metazoan H1 histones provides a “footprint” for the different H1 subtypes (Eirín-López et al. 2006b). The alignment of protostome and deuterostome RI H1 histones shown in Fig. 3A using this domain supports the phylogenetic and promoter considerations described earlier for the razor clam histone H1. The high extent of similarity with the “orphan” protostome H1 RI variants of other bivalve molluscs points towards an RI status for the razor clam histone H1 identified here, representing an “orphan” H1 protein.

To investigate the nature of the nucleotide substitution patterns that led to the diversification of the RI lineages from protostomes and deuterostomes in the course of evolution, the ancestral sequences for the internal nodes in the topology shown in Fig. 2B were reconstructed and the nucleotide changes subsequently analyzed. The results shown in Fig. 3B indicate high confidence levels for the groups of sequences defined by the internal nodes of the phylogeny generated in this way. Our maximum likelihood estimates indicate that a total of 1078 nucleotide substitutions are necessary for the current differentiation among the extant protostome RI H1 variants (nodes I–IV), including 411.9 synonymous substitutions and 644.7 non-synonymous substitutions. Conversely, 1230 nucleotide changes are involved in the differentiation of deuterostome RI H1 variants (nodes 1–5), probably as a result of the higher level of specialization of this group of organisms. Of these changes, 474.8 correspond to synonymous substitutions and 744.5 are replacements. Importantly, the non-synonymous changes outnumber the synonymous substitutions during the differentiation of RI variants, suggesting a process of adaptative evolution during at least part of the evolutionary history of the genes encoding these proteins. This evolutionary pattern is probably related to the exclusion of these RI histone H1 variant genes from the main repetitive RD histone gene units as well as to their solitary locations in the genome (Schulze and Schulze 1995; Eirín-López et al. 2002).

**Fig. 2.** (A) Phylogenetic relationships reconstructed between histone H1 and SNBPs based on complete amino acid sequences using uncorrected *p*-distances. Numbers for interior branches represent bootstrap probability (BP) values based on 1000 replicates. The monophyletic origin of RI H1 proteins and that of protamines is indicated by open circles, while the polyphyletic origin of protamine-like proteins is indicated by solid circles. (B) Nucleotide phylogeny encompassing protostome and deuterostome RI H1 genes. Numbers for interior nodes represent bootstrap and confidence probabilities based on 1000 replicates, followed by the BP corresponding to the maximum parsimony tree topology (shown only when greater than 50%). The topology was rooted with the canonical histone H1 (H1c) from human. (C) Proximal promoter regions from different members of the histone H1 family and SNBPs. Characteristic elements defining H1 lineages are indicated, including CAAT elements (RD lineage), H4 box elements (RI lineage), and control elements (SNBP lineage). Evolutionary relationships among H1 and SNBP representatives are summarized by the branching pattern to the left of the comparisons.



**Table 2.** Average numbers of amino acid ( $p_{AA}$ ), total nucleotide ( $p_{NT}$ ), and synonymous ( $p_S$ ) and non-synonymous ( $p_N$ ) nucleotide differences per site and Z-test of selection in the H1 variants analyzed and within different taxonomic groups and H1 lineages.

Histone H1	$p_{AA}$ (mean±SE)	$p_{NT}$ (mean±SE)	$p_S$ (mean±SE)	$p_N$ (mean±SE)	R	Z-test <sup>a</sup>
Protostomes	0.412±0.033	0.381±0.014	0.634±0.023	0.299±0.024	0.7	10.773**
RD	0.373±0.030	0.355±0.014	0.631±0.020	0.272±0.023	0.7	12.534**
RI	0.066±0.006	0.078±0.006	0.174±0.026	0.050±0.003	0.8	5.829**
Deuterostomes	0.341±0.028	0.334±0.015	0.557±0.021	0.261±0.023	0.7	9.126**
RD	0.151±0.024	0.203±0.013	0.492±0.015	0.090±0.016	0.8	13.879**
RI	0.132±0.026	0.165±0.014	0.468±0.027	0.071±0.017	1.1	9.859**
Overall RD	0.420±0.033	0.417±0.015	0.684±0.012	0.327±0.026	0.7	11.909**
Overall RI	0.305±0.031	0.286±0.015	0.560±0.023	0.202±0.022	0.7	11.024**

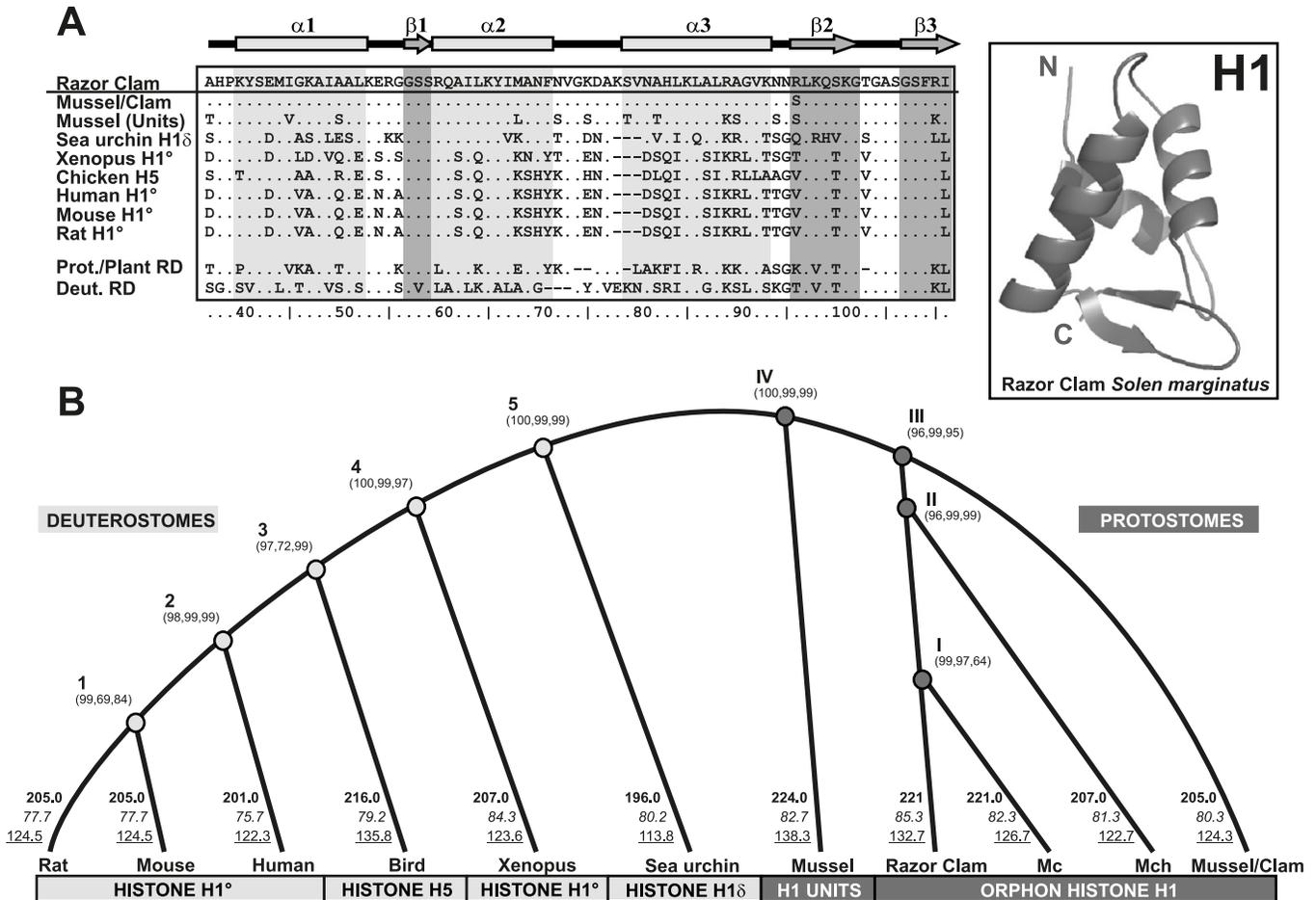
Note: SE, standard error; R, average transition/transversion ratio; \*\*,  $P < 0.001$ .

<sup>a</sup>H<sub>0</sub>:  $p_N = p_S$ ; H<sub>1</sub>:  $p_N < p_S$ .

Further inference regarding the evolutionary mechanisms leading to the differentiation of RI H1 variants across protostomes and deuterostomes can be obtained from study of the codon usage bias of their encoding genes. The results shown

in Table 3 indicate that, except for the divergent H1° genes from *Xenopus*, RI H1 genes from deuterostomes are significantly more biased than their protostome counterparts (*t* test, 4.349,  $P < 0.001$ ). Such results can be interpreted in light of

**Fig. 3.** (A) Comparison of the conserved central region from histone H1 in different members of the RI lineage (including the razor clam H1 sequence as a reference) as well as with consensus RD H1 representatives from protostomes and plants (Prot./Plant) and deuterostomes (Deut.). The three-dimensional structure modeled for histone H1 in the razor clam is indicated in the right margin. (B) Evolutionary pathways leading to differentiation among RI histone H1 variants. The numbers of total nucleotide (boldface), synonymous (italics), and non-synonymous (underlined) changes from ancestral sequences reconstructed for nodes 1–5 (deuterostomes) and I–IV (protostomes) are indicated. Confidence levels for the internal nodes are indicated as in Fig. 2B.



**Table 3.** Estimation of the amount of codon usage bias in protostome and deuterostome RI histone H1 genes.

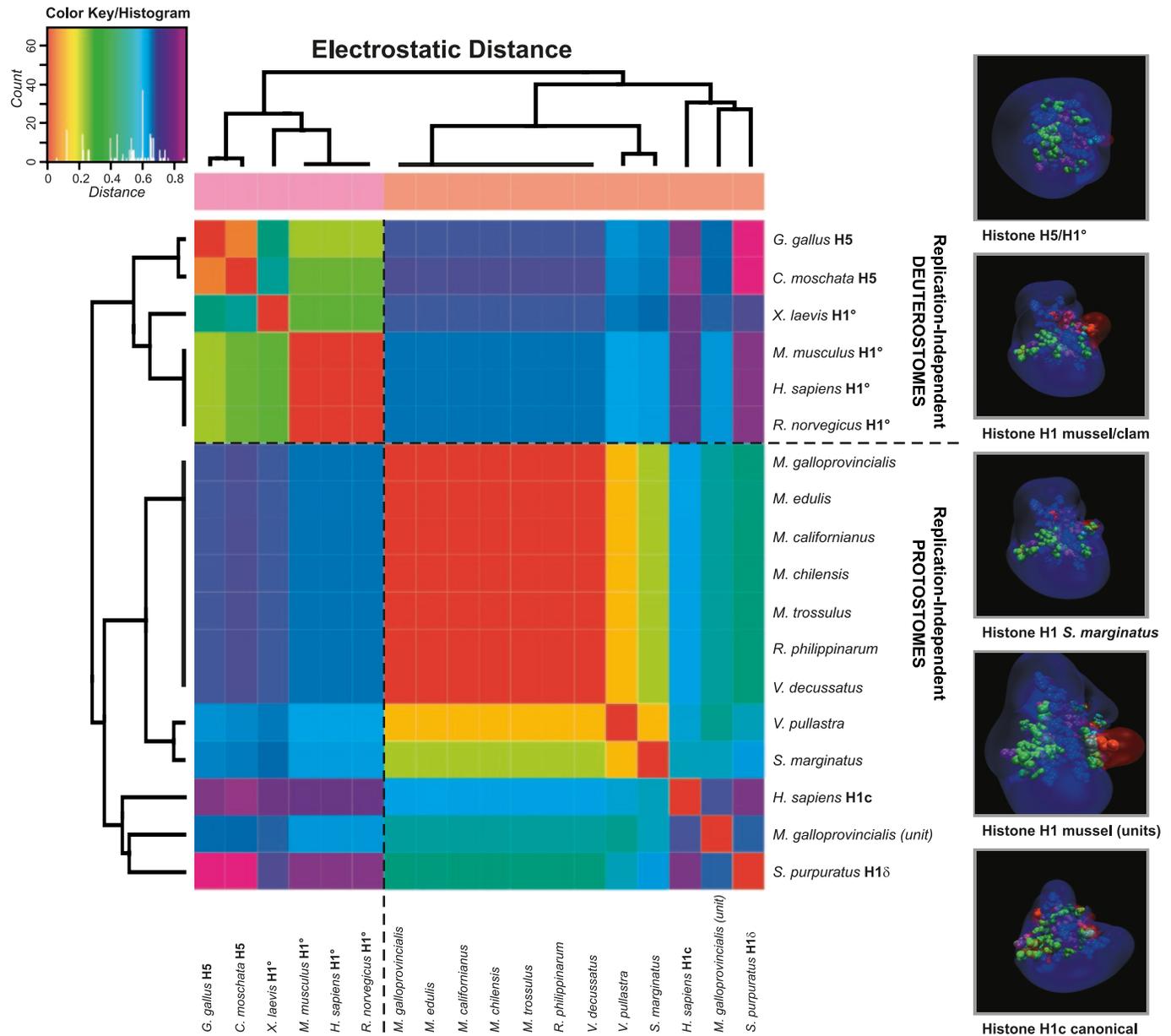
Group of organisms	Effective no. of codons
Mussel	51.562±1.445
Clam	50.810±0.742
Razor clam	37.639
Mussel (units)	53.778
Sea urchin H1δ	53.477
H5 birds	35.274±3.089
H1° <i>Xenopus</i>	59.316±2.381
H1° mammals	41.208±1.306
Protostomes RI	49.520±4.572
Deuterostomes RI	40.655±5.585

a higher degree of functional specialization of deuterostome RI H1 genes in comparison with the apparently less differentiated RI H1s from protostomes. The razor clam H1 gene characterized in this work exhibits an exceptionally high level of codon bias (37.639) compared with H1 genes from

other mussels and clams, thus providing a notable exception to this trend.

Given that ionic interactions play an important role in the interaction of histone H1 with the nucleosome and linker DNA segments that modulate chromatin dynamics, the electrostatic interaction properties of protostome and deuterostome RI histone H1 lineages were analyzed to investigate the potential effects of the observed variation on binding abilities in both RI H1 lineages. The epogram shown in Fig. 4 distinctively points to the presence of 2 different groups based on electrostatic potentials, corresponding to deuterostome and protostome RI histone H1s. Furthermore, sea urchin H1δ and H1 from the repetitive units of mussels are somewhat more closely related to canonical RD histone H1, indicating a slightly divergent status within the RI lineage. The comparisons made on the basis of the electrostatic potential variation provide support to the other molecular evolutionary and phylogenetic analyses revealing a differentiation between RI H1 proteins from protostomes and deuterostomes. Such results suggest the presence of different constraints acting upon protostome and deuterostome RI H1 proteins and leading to their differentiation during evolution.

**Fig. 4.** Electrostatic distances calculated from the similarity indices for the electrostatic potentials of histone H1 RI variants, represented in a color-coded matrix (heat map). The distance between similarity indices of every pair of molecules ( $a$  and  $b$ ) is defined as  $D_{a,b} = \sqrt{2 - 2SI_{a,b}}$ . The color code and the number of comparisons for each distance interval are indicated in the key/histogram. The tree along the left side of the image assembles the proteins into groups with similar electrostatic potentials (epogram), with discontinuous black lines delimiting two different groups: RI and RD histone H1 variants. Representations of electrostatic potentials for four representative H1 molecules belonging to the different groups defined in the epogram are indicated in the right margin of the figure. Negatively charged surfaces are red and positively charged surfaces are blue; colors were assigned to amino acids according to their physical and chemical structural characteristics (red, acidic; blue, basic; green, polar uncharged; purple, nonpolar hydrophobic).



## Conclusions

There is now very little doubt about the RI nature of the ancestral histone genes in early eukaryotes that led to the differentiation of canonical RD lineages later on during evolution (Malik and Henikoff 2003; Eirín-López et al. 2009). However, our understanding of the evolutionary processes responsible for the differentiation of RI H1 histones has been hindered by the lack of information on protostome RI H1 variants. The information provided in this work shows

that RI variants seem to be the rule rather than the exception among mollusc H1 histones. Such prevalence is most likely also connected to the origin and presence of several different types of SNBPs in this group of organisms (Eirín-López et al. 2006a, 2008). Although the RI variants share a common long-term evolutionary mechanism of birth-and-death with their RD counterparts (Eirín-López et al. 2005), our results show that RI H1 histones from protostomes and deuterostomes exhibit distinctive structural differences. While the

general trend in protostomes appears to be the presence of a single functional RI H1 type, at least 2 highly differentiated RI H1 variants (H5 and H1<sup>o</sup>) have been described in deuterostomes. Such an increase in heterogeneity during the specialization process was most likely determined by the higher functional complexity of deuterostomes, which might imply the existence of differences in chromatin organization with respect to protostomes. This is supported by the higher RI H1 sequence diversity and higher codon bias observed within this group. Some preliminary hints about the functional significance of the RI histone H1 diversification can be drawn from the electrostatic potentials analyzed in the present work, as ionic interactions play an important role in the interaction of histone H1 with the nucleosome and linker DNA segments. Despite all this, questions still remain, especially pertaining to the expression of linker histones in protostomes and their involvement in different nuclear metabolic processes. Functional studies of protostome linker histones will be required to further decipher these issues.

## Acknowledgments

We are grateful to Dr. J. Fernández-Tajes for kindly supplying razor clam specimens and for fruitful comments on an earlier version of the manuscript. This work was supported by a Marie Curie Outgoing International Fellowship (MOIF-CT-2005-021900) within the 6th Framework Programme (European Union) and by a contract within the Isidro Parga Pondal Program (Xunta de Galicia) (to J.M.E.-L.), and by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC-OGP-0046399-02) (to J.A.). R.G.-R. is the recipient of a fellowship from the Diputación da Coruña (Spain) and a predoctoral fellowship from the Universidade da Coruña.

## References

- Albig, W., Meergans, T., and Doenecke, D. 1997. Characterization of the H1.5 gene completes the set of human H1 subtype genes. *Gene*, **184**(2): 141–148. doi:10.1016/S0378-1119(96)00582-3. PMID:9031620.
- Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, **22**(2): 195–201. doi:10.1093/bioinformatics/bti770. PMID:16301204.
- Ausió, J. 1999. Histone H1 and evolution of sperm nuclear basic proteins. *J. Biol. Chem.* **274**(44): 31115–31118. doi:10.1074/jbc.274.44.31115. PMID:10531297.
- Ausió, J. 2006. Histone variants: the structure behind the function. *Brief. Funct. Genomics Proteomics*, **5**(3): 228–243. doi:10.1093/bfgp/ell020.
- Dalton, S., and Wells, J.R.E. 1988. A gene-specific promoter element is required for optimal expression of the histone H1 gene in S-phase. *EMBO J.* **7**(1): 49–56. PMID:2896124.
- DeLano, W.L. 2007. MacPyMOL: a PyMOL-based molecular graphics application for MacOS X. DeLano Scientific LLC, Palo Alto, Calif.
- Eirín-López, J.M., González-Tizón, A.M., Martínez, A., and 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**(3): 272–283. doi:10.1007/s00239-002-2325-1. PMID:12187381.
- Eirín-López, J.M., González-Tizón, A.M., Martínez, A., and Méndez, J. 2004a. Birth-and-death evolution with strong purifying selection in the histone H1 multigene family and the origin of *orphon* H1 genes. *Mol. Biol. Evol.* **21**(10): 1992–2003. doi:10.1093/molbev/msh213. PMID:15254261.
- Eirín-López, J.M., Ruiz, M.F., González-Tizón, A.M., Martínez, A., Sánchez, L., and Méndez, J. 2004b. 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. *J. Mol. Evol.* **58**(2): 131–144. doi:10.1007/s00239-003-2531-5. PMID:15042333.
- Eirín-López, J.M., Ruiz, M.F., González-Tizón, A.M., Martínez, A., Ausió, J., Sánchez, L., and Méndez, J. 2005. Common evolutionary origin and birth-and-death process in the replication-independent histone H1 isoforms from vertebrate and invertebrate genomes. *J. Mol. Evol.* **61**(3): 398–407. doi:10.1007/s00239-004-0328-9. PMID:16082565.
- Eirín-López, J.M., Frehlick, L.J., and Ausió, J. 2006a. Protamines, in the footsteps of linker histone evolution. *J. Biol. Chem.* **281**(1): 1–4. doi:10.1074/jbc.R500018200. PMID:16243843.
- Eirín-López, J.M., Lewis, J.D., Howe, A., and Ausió, J. 2006b. Common phylogenetic origin of protamine-like (PL) proteins and histone H1: evidence from bivalve PL genes. *Mol. Biol. Evol.* **23**(6): 1304–1317. doi:10.1093/molbev/msk021. PMID:16613862.
- Eirín-López, J.M., Frehlick, L.J., Chiva, M., Saperas, N., and Ausió, J. 2008. The sperm proteins from amphioxus mirror its basal position among chordates and redefine the origin of vertebrate protamines. *Mol. Biol. Evol.* **25**(8): 1705–1713. doi:10.1093/molbev/msn121. PMID:18503046.
- Eirín-López, J.M., González-Romero, R., Dryhurst, D.D., Méndez, J., and Ausió, J. 2009. Long-term evolution of histone families: old notions and new insights into their diversification mechanisms across eukaryotes. *In* *Evolutionary biology from concept to application II*. Edited by P. Pontarotti. Springer-Verlag, Berlin Heidelberg. In press.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**(4): 783–791. doi:10.2307/2408678.
- González-Romero, R., Ausió, J., Méndez, J., and Eirín-López, J.M. 2008. Early evolution of histone genes: prevalence of an ‘orphon’ H1 lineage in protostomes and birth-and-death process in the H2A family. *J. Mol. Evol.* **66**(5): 505–518. doi:10.1007/s00239-008-9109-1. PMID:18443735.
- Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**: 95–98.
- Humphrey, W., Dalke, A., and Schulten, K. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**(1): 33–38, 27–28. doi:10.1016/0263-7855(96)00018-5. PMID:8744570.
- Jenuwein, T., and Allis, C.D. 2001. Translating the histone code. *Science*, **293**(5532): 1074–1080. doi:10.1126/science.1063127. PMID:11498575.
- Lieber, T., Angerer, L.M., Angerer, R.C., and Childs, G. 1988. A histone H1 protein in sea urchins is encoded by a poly(A)+ mRNA. *Proc. Natl. Acad. Sci. U.S.A.* **85**(12): 4123–4127. doi:10.1073/pnas.85.12.4123. PMID:2898141.
- Madura, J.D., Briggs, J.M., Wade, R.C., Davis, M.E., Luty, B.A., Ilin, A., et al. 1995. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics Program. *Comput. Phys. Commun.* **91**(1–3): 57–95. doi:10.1016/0010-4655(95)00043-F.
- Malik, H.S., and Henikoff, S. 2003. Phylogenomics of the nucleosome. *Nat. Struct. Biol.* **10**(11): 882–891. doi:10.1038/nsb996. PMID:14583738.

- Marzluff, W.F. 1992. Histone 3' ends: essential and regulatory functions. *Gene Expr.* **2**(2): 93–97. PMID:1633440.
- Maxson, R., Cohn, R., Kedes, L., and Mohun, T. 1983. Expression and organization of histone genes. *Annu. Rev. Genet.* **17**(1): 239–277. doi:10.1146/annurev.ge.17.120183.001323. PMID:6421226.
- Nei, M., and Kumar, S. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, New York.
- Nei, M., and Rooney, A.P. 2006. Concerted and birth-and-death evolution in multigene families. *Annu. Rev. Genet.* **39**(1): 121–152. doi:10.1146/annurev.genet.39.073003.112240.
- Owen-Hughes, T. 2003. Colworth memorial lecture. Pathways for remodelling chromatin. *Biochem. Soc. Trans.* **31**(Pt 5): 893–905. PMID:14505445.
- Peretti, M., and Khochbin, S. 1997. The evolution of the differentiation-specific histone H1 gene basal promoter. *J. Mol. Evol.* **44**(2): 128–134. doi:10.1007/PL00006129. PMID:9069173.
- Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L., and Sweet, R.M. 1993. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature*, **362**(6417): 219–223. doi:10.1038/362219a0. PMID:8384699.
- Rice, E.L., and Bird, C.J. 1990. Relationships among geographically distant populations of *Gracilaria verrucosa* (Gracilariales, Rhodophyta) and related species. *Phycologia*, **29**: 501–510.
- Richter, S., Wenzel, A., Stein, M., Gabdoulline, R.R., and Wade, R.C. 2008. webPIPSA: a web server for the comparison of protein interaction properties. *Nucleic Acids Res.* **36**(Web Server): W276–W280. doi:10.1093/nar/gkn181.
- Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X., and Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**(18): 2496–2497. doi:10.1093/bioinformatics/btg359. PMID:14668244.
- Saitou, N., and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**(4): 406–425. PMID:3447015.
- Schulze, E., and Schulze, B. 1995. The vertebrate linker histones H1<sup>0</sup>, H5, and H1M are descendants of invertebrate “orphan” histone H1 genes. *J. Mol. Evol.* **41**(6): 833–840. doi:10.1007/BF00173162. PMID:8587127.
- Sitnikova, T. 1996. Bootstrap method of interior-branch test for phylogenetic trees. *Mol. Biol. Evol.* **13**(4): 605–611. PMID:8882503.
- Sitnikova, T., Rzhetsky, A., and Nei, M. 1995. Interior-branch and bootstrap tests of phylogenetic trees. *Mol. Biol. Evol.* **12**(2): 319–333. PMID:7700156.
- Strahl, B.D., and Allis, C.D. 2000. The language of covalent histone modifications. *Nature*, **403**(6765): 41–45. doi:10.1038/47412. PMID:10638745.
- Sures, I., Lowry, J., and Kedes, L.H. 1978. The DNA sequence of sea urchin (*S. purpuratus*) H2A, H2B and H3 histone coding and spacer regions. *Cell*, **15**(3): 1033–1044. doi:10.1016/0092-8674(78)90287-8. PMID:728984.
- Sures, I., Levy, S., and Kedes, L.H. 1980. Leader sequences of *Strongylocentrotus purpuratus* histone mRNAs start at a unique heptanucleotide common to all five histone genes. *Proc. Natl. Acad. Sci. U.S.A.* **77**(3): 1265–1269. doi:10.1073/pnas.77.3.1265. PMID:6154927.
- Tamura, K., Dudley, J., Nei, M., and Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**(8): 1596–1599. doi:10.1093/molbev/msm092. PMID:17488738.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL\_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**(24): 4876–4882. doi:10.1093/nar/25.24.4876. PMID:9396791.
- van Holde, K.E. 1988. *Chromatin*. Springer-Verlag, New York.
- van Wijnen, A.J., van den Ent, F.M.I., Lian, J.B., Stein, J.L., and Stein, G.S. 1992. Overlapping and CpG methylation-sensitive protein–DNA interactions at the histone H4 transcriptional cell cycle domain: distinctions between two human H4 gene promoters. *Mol. Cell. Biol.* **12**(7): 3273–3287. PMID:1620129.
- Winnepenninckx, B., Backeljau, T., and De Wachter, R. 1993. Extraction of high molecular weight DNA from molluscs. *Trends Genet.* **9**(12): 407. doi:10.1016/0168-9525(93)90102-N. PMID:8122306.
- Wright, F. 1990. The ‘effective number of codons’ used in a gene. *Gene*, **87**(1): 23–29. doi:10.1016/0378-1119(90)90491-9. PMID:2110097.
- Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**(8): 1586–1591. doi:10.1093/molbev/msm088. PMID:17483113.
- Zhang, J., Rosenberg, H.F., and Nei, M. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. U.S.A.* **95**(7): 3708–3713. doi:10.1073/pnas.95.7.3708. PMID:9520431.