J Mol Evol (2008) 66:505–518
DOI 10.1007/s00239-008-9109-1

Early Evolution of Histone Genes: Prevalence of an 'Orphon' H1 Lineage in Protostomes and Birth-and-Death Process in the H2A Family

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

Received: 9 January 2008 / Accepted: 10 April 2008 / Published online: 29 April 2008

Abstract The study of histone evolution has experienced a rebirth, for two main reasons: the identification of new essential histone variants responsible for regulating chromatin dynamics and the subsequent contradictions posed by this variability as it pertains to their long-term evolution process. Although different evolutionary models (e.g., birth-and-death evolution, concerted evolution) may account for the observed divergence of histone genes, conclusive evidence is lacking (e.g., histone H1) or totally nonexistent (e.g., histone H2A). While most of the published work has focused on deuterostomes, very little is known about the diversification and functional differentiation mechanisms followed by histone protein subtypes in protostomes, for which histone variants have only been recently described. In this study, we identify linker and core histone genes in three clam species. Our results demonstrate the prevalence of an 'orphon' H1 lineage in molluscs, a group in which the protostome H1 and sperm nuclear basic proteins are on the verge of diversification. They share an early monophyletic origin with vertebrate-specific variants prior to the differentiation between protostomes and deuterostomes. Given the intriguing evolutionary features of the histone H1 family, we have evaluated the relative importance of gene conversion, point mutation, and selection in maintaining the diversity found among H2A subtypes in eukaryotes. We show evidence for the first time that the long-term evolution of this family is not subject to concerted evolution but, rather, to a gradual evolution following a birth-and-death model under a strong purifying selection at the protein level.

Keywords Histones · Long-term evolution · Chromatin · Birth-and-death · Molluscs

Electronic supplementary material The online version of this article (doi:10.1007/s00239-008-9109-1) contains supplementary material, which is available to authorized users.

R. González-Romero · J. Méndez · J. M. Eirín-López (✉)
Departamento de Biología Celular y Molecular, Facultade de Ciencias, Universidade da Coruña, Campus de A Zapateira s/n, A Coruña, Spain E15071
e-mail: jeirin@udc.es

J. Ausió · J. M. Eirín-López
Department of Biochemistry and Microbiology, University of Victoria, Victoria, Canada BC V8W 3P6

Introduction

Histones constitute a group of small basic proteins involved in the packaging of DNA in the eukaryotic nucleus and, consequently, in the regulation of gene expression also. There are five histone families, which can be classified into two major groups according to structural and functional features: core histones (H2A, H2B, H3, and H4) and linker histones (H1). Histones can also be classified according to their expression patterns as either replication dependent (RD) or replication independent (RI). In the first case, RD histones show an expression pattern coupled to the S-phase of the cell cycle during DNA replication, which is mediated by the presence of a stem-loop sequence in their mRNAs. In contrast, RI histones are transcribed through polyadenylated mRNAs, facilitating their prevalence at basal and constant levels throughout the cell cycle (Marzluff 1992). However, there are several examples of histone genes in which both mRNA termination signals coexist, sometimes referred to as facultative histone genes that can be expressed through either mechanism (Cheng et al. 1989; Collart et al. 1992; Gendron et al. 1998).

Histones show a typical tripartite structure, composed of a central globular domain flanked by two terminal tails

(N- and C-terminal). The globular domain contains a conserved region known as the histone fold in core histones or the winged-helix domain in linker histones (Arents and Moudrianakis 1995; Ramakrishnan et al. 1993). While the N-terminal tail contains the vast majority of residues amenable for posttranslational modification in core histones, the long C-terminal tail is the most important domain in linker histones, playing a role in nucleosome positioning and in the regulation of gene expression (Wolffe et al. 1997). The association of core histones results in the formation of the octamer, which interacts with DNA to form the nucleosome core particle, which represents the fundamental subunit of the chromatin fiber (van Holde 1988). Linker histones are involved in the compaction of the chromatin fiber by interacting with internucleosomal DNA stretches (Simpson 1978) and participating in the formation of higher order chromatin structures.

Among the five histone families, the H1 and H2A families show the greatest diversity of isoforms. In the case of H1, several variants exclusive of vertebrates have been identified, including somatic, spermatogenesis, oocyte, maternal, and replacement-specific subtypes (Albig et al. 1997; Cho and Wolffe 1994; Martianov et al. 2005; Yan et al. 2003). However, fewer H1 isoforms are observed in invertebrates, including somatic and stage-specific subtypes (Hentschel and Birnstiel 1981; Maxson et al. 1983). The H2A family contains the greatest number of variants among the core histones. Examples of H2A variants include H2A.X, H2A.Z, H2A.Bbd (Barr body deficient), and macroH2A, some of which are essential for the maintenance of genome integrity and viability (Dryhurst et al. 2004; Eirín-López and Ausió 2007; Li et al. 2005). Such variants additionally play critical roles in the regulation of chromatin dynamics (Ausió 2006; Chadwick and Willard 2001a, b; Eirín-López et al. 2007).

The long-term evolution of histones has been classically explained by a concerted evolution model, which accounts for the homogenization of the family members through interlocus recombination or gene conversion. However, recent works have demonstrated a high degree of variation and diversification in histone genes, suggesting that histone family members have evolved gradually through a series of small steps subject to a duplication/selection mechanism defined as birth-and-death (Eirín-López et al. 2004a; Nei and Hughes 1992; Nei and Rooney 2006; Piontkivska et al. 2002; Rooney et al. 2002). Even though the H1 and H2A families are the best-characterized histones and there is support for the gradual evolution of H1 (Eirín-López et al. 2004a, 2005), the specific evolutionary mechanisms acting on H2A histones are unknown.

Among the different model organisms used in the study of histones, bivalve molluscs are of special interest, as they represent the only protostome group having RI H1 lineages and that are linked to the origin of sperm nuclear basic proteins (SNBPs; Eirín-López et al. 2006a, b). In the interest of investigating the origin and evolution of this gene lineage, we have characterized the histone multigene family in three clam species belonging to the family Veneridae (*Ruditapes philippinarum, Venerupis decussatus*, and *Venerupis pullastra*). Our results indicate the presence of an 'orphon' RI H1 lineage that is not exclusive to mussels, but common to other bivalve species. This finding supports the evolutionary differentiation among SNBPs and RI and RD H1 genes as early as in metazoan evolution before the split between protostomes and deuterostomes. Furthermore, this work presents evidence in favor of the gradual evolution of H2A histones following a birth-and-death model of evolution under a strong purifying selection at the protein level.

Materials and Methods

Isolation of Histone Genes, Analyses of Transcripts, and Protein Extraction

The three clam species analyzed in the present work were collected in different localities along the Atlantic coast of Spain as follows: *Ruditapes philippinarum* and *Venerupis decussatus* from Camariñas (A Coruña) and *Venerupis pullastra* from Pontedeume (A Coruña). Genomic DNA from muscle tissue was purified following the protocol described by Fernandez-Tajes et al. (2007). PCR amplifications from template genomic DNA (25 ng) were performed in a final volume of 25 µl (10 ng/µl), using species-specific primers for coding and noncoding flanking regions in the three clam species (see Table 1) at 10 µM, with 1 U/µl of *Taq* DNA polymerase (Roche Molecular Biochemicals). The reactions were performed with a first denaturation step of 4 min 30 s at 95°C, followed by 35 cycles consisting of a 30-s denaturation step at 95°C, 30 s of

Table 1 Specific primers for Veneridae histone genes

Gene	Primer	Sequence	Size
H1	5′-H1	GCGAAGACAATTCAGTCGGTT	21-mer
	3′-H1	GAAAGGGTAGGGCTCAGCT	19-mer
H2A	5′-H2A	GGAAGAAGCGATGATTTGATTGG	23-mer
	3′-H2A	GAGGGAGTGAGCTATGTTTGAG	22-mer
H2B	5′-H2B	CATCGCTTCTTCCAGGTAG	19-mer
	3′-H2B	TCATTTTGGGGTGGGACA	18-mer
H3	5′-H3	GGGGTGAACAATTGTTAGCTTC	22-mer
	3′-H3	TTCAGTAACCTGACTGTCTTGG	22-mer
H4	5′-H4	CTACAGAGTTACCTCCCGGAT	21-mer
	3′-H4	ACAAGTTGGACAGGAGAAAGC	22-mer

annealing at 54°C (except for H2A, at 52°C), and 30 s of extension at 72°C. A final extension step of 5 min was performed at 72°C. Automatic DNA sequencing was performed in a CEQ 8000 sequencer (Beckman Coulter). The obtained sequences were deposited in the GenBank database under the following accession numbers: *R. philippinarum* H1 (EF67 0664), H2A (EF670665), H2B (EF670666), H3 (EF670667), and H4 (EF670668); *V. decussatus* H1 (EF670669), H2A (EF670670), H2B, (EF670671), H3 (EF670672), and H4 (EF670673); and *V. pullastra* H1 (EF670674), H2A (EF67 0675), H2B (EF670676), H3 (EF670677), and H4 (EF67 0678).

Total RNA extracts from frozen adult specimens of *V. pullastra* were prepared using the TRIzol Reagent commercial system (Invitrogen) for RT-PCR and transcript analysis. Poly(A)-rich RNA was prepared using the MicroPoly(A) Purist kit (Ambion) and RT-PCR analyses were performed using the 'partial' set of primers specific for *Mytilus* histone genes (Eirín-López et al. 2004b). The acidic extraction and purification of histone proteins were subsequently performed from gonadal tissue of *Venerupis decussatus* in 0.4 N HCl, following the procedure described by Ausió (1986). Histone proteins were subsequently analyzed on polyacrylamide gels under denaturing conditions with sodium dodecyl sulfate (SDS-PAGE).

Molecular Evolutionary Analyses

Six invertebrate representative species, including the histone sequences in *R. philippinarum* characterized in the present work (given that genes from this species show the highest similarity to the consensus sequences defined for Veneridae histone genes), were chosen for the evolutionary analysis of the histone multigene families in invertebrates. Accession numbers for the chosen sequences are as follows. *Strongylocentrotus purpuratus* (echinoderm): H1, NM_214555; H2A, NM_214553; H2B, NM_214552; H3, NM_214547; H4, NM_214551. *Drosophila melanogaster* (insect) H1, X14215; H2A, NM_165382; H2B, X14215; H3, X14215; H4, X14215. *Chaetopterus variopedatus* (annelid): H1, U96764; H2A, AF007904; H2B, U96764; H3, U96764; H4, AF007904. *Caenorhabditis elegans* (nematode): H1, X53 277; H2A, NM_074631; H2B, NM_073063; H3, NM_074 632; H4, NM_076830. *Mytilus galloprovincialis* (mollusc): H1, AJ416424; H2A, AY267755; H2B, AY267740; H3, AY267748; H4, AY267750.

Nucleotide and amino acid sequences were aligned with the programs BIOEDIT (Hall 1999) and CLUSTAL W (Thompson et al. 1997). The alignment of nucleotide sequences was constructed on the basis of the translated amino acid sequences. Molecular evolutionary analyses were performed using the computer program MEGA ver. 3.1 (Kumar et al. 2004). In the case of representative histone genes from invertebrates the extent of nucleotide divergence between sequences (d) was estimated using the Kimura two-parameters model, while the extent of amino acid divergence between sequences was estimated by means of the uncorrected differences (p-distance). The latter method was also used in the estimation of distances in the long-term evolutionary analyses of eukaryotic H1 and H2A proteins. In both cases, the numbers of synonymous (p_S) and nonsynonymous (p_N) nucleotide differences per site were computed using the modified Nei-Gojobori method (Zhang et al. 1998), providing the transition/transversion ratio (R). Distances were estimated using the pairwise deletion option in all cases with the exception of protein phylogenetic inference in long-term evolutionary analyses, where the complete deletion option was used. Standard errors were calculated by the bootstrap method with 1000 replicates.

The presence and nature of selection were tested in invertebrate histone genes using two different approaches: on one hand, the codon-based Z-test for selection compared the numbers of synonymous (p_S) and nonsynonymous (p_N) nucleotide differences per site, establishing the null hypothesis as H_0: $p_S = p_N$ and the alternative hypothesis as H_1: $p_S > p_N$ (Nei and Kumar 2000); and on the other hand, the presence of selection was further studied by testing for deviations from neutrality. The influence of selection on certain amino acids was analyzed by determining the correlation between the genomic GC content (GC content at fourfold degenerate sites was assumed to represent the genomic GC content, given that the latter has already been shown to be a good approximation of the former [Li 1997; Nei 1987]) and the proportion of GC-rich (GAPW) and GC-poor (FYMINK) residues. Under the neutral model, GC-rich amino acids will be positively correlated with genomic GC content, whereas GC-poor amino acids will be negatively correlated with genomic GC content (Rooney 2003). Correlations were computed by using the Spearman rank correlation coefficient and the statistical significance was assessed through regression analysis.

The analysis of nucleotide variation across the different histone coding regions was performed using a sliding-window approach implemented in the program DnaSP ver. 4.0 (Rozas et al. 2003), by estimating the proportion (p) of nucleotide sites at which two sequences being compared are different and the number of synonymous substitutions (p_S) per site, with a window length of 20 bp and a step size of 5 bp (for p) and a window length of 5 bp and a step size of 1 bp (for p_S).

Long-Term Evolution Analysis of H1 and H2A Multigene Families

For the study of the long-term evolution of H1 and H2A histone genes we have included all the nonredundant H1

and H2A sequences listed in the NHGRI/NCBI Histone Sequence Database (Marino-Ramirez et al. 2006) as of July 2007 (see Supplementary Tables 1 and 2) in our analyses. Phylogenetic trees were reconstructed using the neighbor-joining tree-building method (Saitou and Nei 1987). The reliability of the resulting topologies was tested by the bootstrap method (Felsenstein 1985) and by the interior-branch test (Rzhetsky and Nei 1992; Sitnikova 1996), producing the bootstrap probability (BP) and confidence probability (CP) values for each interior branch in the tree. Given that the bootstrap method is known to be conservative, BP > 80% was interpreted as high statistical support for interior branches in the tree, and CP = 95% was otherwise considered statistically significant (Sitnikova et al. 1995). Histone H1 phylogenies were rooted using the H1 gene from the protist *Entamoeba histolytica*, one of the most primitive eukaryotes for which an H1-related protein has been characterized (Kasinsky et al. 2001). In the case of H2A, phylogenetic trees were rooted with the histone H2A from the diplomonad protist *Giardia intestinalis*, as this lineage is believed to be the first to diverge from all other eukaryotes (Roger et al. 1998).

The GenBank database and complete genome databases (human and mouse) were screened for the presence of H2A pseudogenes using the BLAST tool (Altschul et al. 1990), identifying two H2A pseudogenes (*Homo sapiens* Ψ and *Mus musculus* Ψ.1) and three H2A.Bbd pseudogenes (*Mus musculus* Ψ.2, Ψ.3, and Ψ.4). The presence of truncated or incomplete H2A sequences, indels in the conserved regions, and the absence or interruption of the major promoter elements were interpreted as pseudogenization features.

Results

Characterization of Histone Sequences in the Family Veneridae

Histone genes from the three clam species were unspecifically amplified through PCR experiments using previously described primers (Eirín-López et al. 2002, 2004b). Once these sequences were obtained, five pairs of primers (one for each histone type) specific for the family Veneridae were generated (Table 1), with annealing sites located at untranslated regions (UTRs) (Fig. 1A). These primers were used in additional rounds of amplification, resulting in the fragments shown in Fig. 1A and B. Coding regions, translated amino acid sequences, and noncoding flanking regions contained by these fragments are shown in Supplementary Figs. 1–5. Coding sequences for histone H1 revealed the presence of a 191-amino acid (aa)-long protein encoded by 573 bp in the three species analyzed. Comparisons of the protein central conserved domain among different representative metazoans revealed a high degree of homology between H1 from Veneridae and the H1 proteins from mussels, followed by the H1δ proteins from sea urchin and H5/H1^0 from vertebrates (Fig. 1C). In the case of core histones, all three species showed a 125-aa-long H2A protein encoded by 375 bp, a 124-aa-long H2B protein encoded by 372 bp, a 136-aa-long H3 protein encoded by 408 bp, and a 103-aa-long H4 protein encoded by 309 bp.

Promoter regions are highly conserved, in particular, major regulatory elements involved in transcriptional activity (Supplementary Figs. 1–5). Perfect TATA signals were identified for H1, H2A, H2B, and H4 genes at positions −83 to −90, −69 to −72, −62 to −69, and −56 to −61, respectively. CAAT box signals were identified in all cases and two were present in the H3 gene, a characteristic feature of vertebrate histone genes (Connor et al. 1984). The short sequence 5′- *Pu*CATTC*Py*-3′, which represents putative CAP sites, was also present in all genes except in H1 and generally 50 to 70 bp upstream of the start codon (Sures et al. 1980). Histone H1 promoter regions showed the presence of typical elements of linker histone genes such as an H1 box-like element (−170 to −177) followed by an H4 box element (−102 to −118) (Fig. 2). The latter element occupies the same position that the CAAT box occupies in somatic subtypes and is typical of H4 genes and linker histone 'orphon' variants (Peretti and Khochbin 1997).

Each of the histone genes analyzed showed the typical palindrome sequence forming the stem-loop structure at the noncoding 3′ terminal regions, followed by a purine-rich element 13–14 bp downstream (Supplementary Figs. 1–5). The stem-loop consensus sequence for the Veneridae genes was defined as 5′-$^{G}_{A}$GCCCTTTT$^{C}_{A}$AGGGC$^{C}_{T}$-3′ (Table 2). These two sequence blocks are typical of all replication-dependent (RD) histone genes and are related to the binding of the primary mRNA transcript to the U7 snRNP. Surprisingly, all genes show at least one additional mRNA termination signal downstream of the palindrome sequence: a polyadenylation signal which is typical of replication-independent (RI) histone genes expressed at constant levels throughout the cell cycle. The functionality of such putative polyadenylation signals was assessed in RT-PCR experiments on the polyadenylated mRNA fraction from *V. pullastra* using primers specific for histone coding (internal) regions (Eirín-López et al. 2002). The amplified fragments displayed the expected sizes in all five histone genes (Fig. 3A), supporting the polyadenylated status of at least a fraction of linker and core histone transcripts in Veneridae. The Veneridae histone proteins were extracted and further analyzed using denaturing SDS gels (Fig. 3B), showing a histone H1 which has a mobility similar to that of

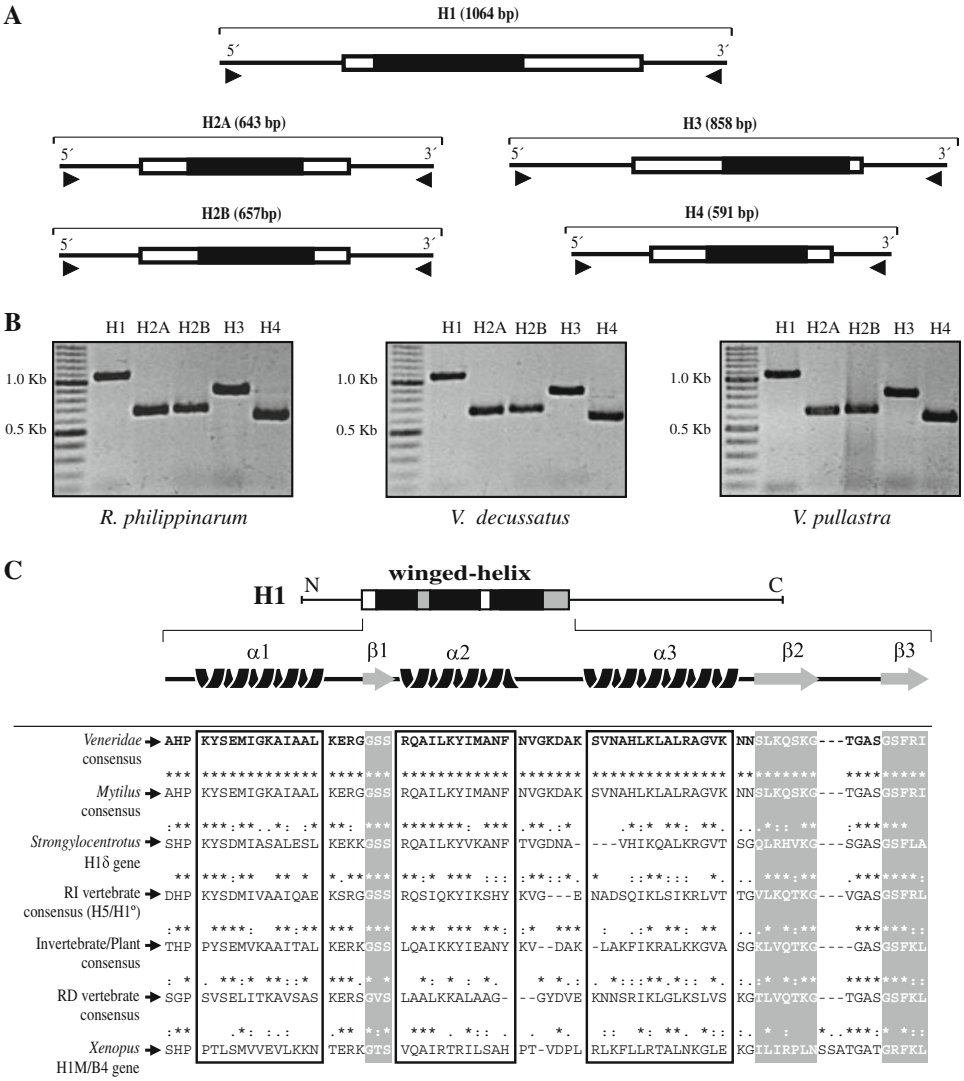


Fig. 1 (**A**) Schematic representation of the annealing positions of the specific primers designed for Veneridae histone genes in the present work (arrows), indicating the size of the amplified fragments. Coding regions are represented by boxes, with the protein central domain shown in black. (**B**) PCR amplifications of the five histones from *Ruditapes philippinarum*, *Venerupis decussatus*, and *Venerupis pullastra* genomic DNA using the primers indicated above. (**C**) Comparison of the H1 central conserved domain between RD and RI isoforms from different taxonomic groups. Asterisks indicate perfect matches, colons indicate a high degree of homology, and dots indicate a low degree of homology between residues

histone H5 in birds (another RI histone), lending further support to the results obtained in the expression analyses.

Histone Evolution in Invertebrate Organisms

In order to include the characterization of Veneridae histone genes in a broader evolutionary background, six representative species of invertebrate groups were included in our analyses (see Materials and Methods). The nucleotide variation and the number of synonymous substitutions (p_S) per site across the histone sequences were investigated as shown in Fig. 4. The relative contribution of p_S to p is evident, and in most cases, the overall amount of nucleotide variation is the result of the underlying synonymous variation. Indeed, synonymous variation is higher in the central conserved domains of histone proteins (especially in H3 and H4) as well as in certain regions of the tails, likely due to the presence of punctual residues amenable to posttranslational modifications. As described in Supplementary Table 3, the synonymous nucleotide variation was significantly greater than the nonsynonymous variation in all cases.

The presence and nature of selection in invertebrate histones were studied using two different approaches. First, a Z-test of selection was used in comparing the numbers of synonymous (p_S) and nonsynonymous (p_N) substitutions detailed in Supplementary Table 3. Accordingly, it is possible to reject the null hypothesis that both values are not significantly different (H_0: $p_S = p_N$; $P < 0.001$, Z-test), suggesting the presence of purifying selection acting on these proteins. Results of the Z-test also reveal that this purifying process is especially strong in H3 ($P = 21.182$) and more relaxed in histone H1 ($P = 10.773$). Taking into account that histones have a characteristically high proportion of nonpolar and basic residues, the presence of selection for certain biased amino acids in the invertebrate histone families was analyzed by determining the correlation coefficients between genomic GC content and the frequency of GC-rich and GC-poor amino acids (Fig. 5).

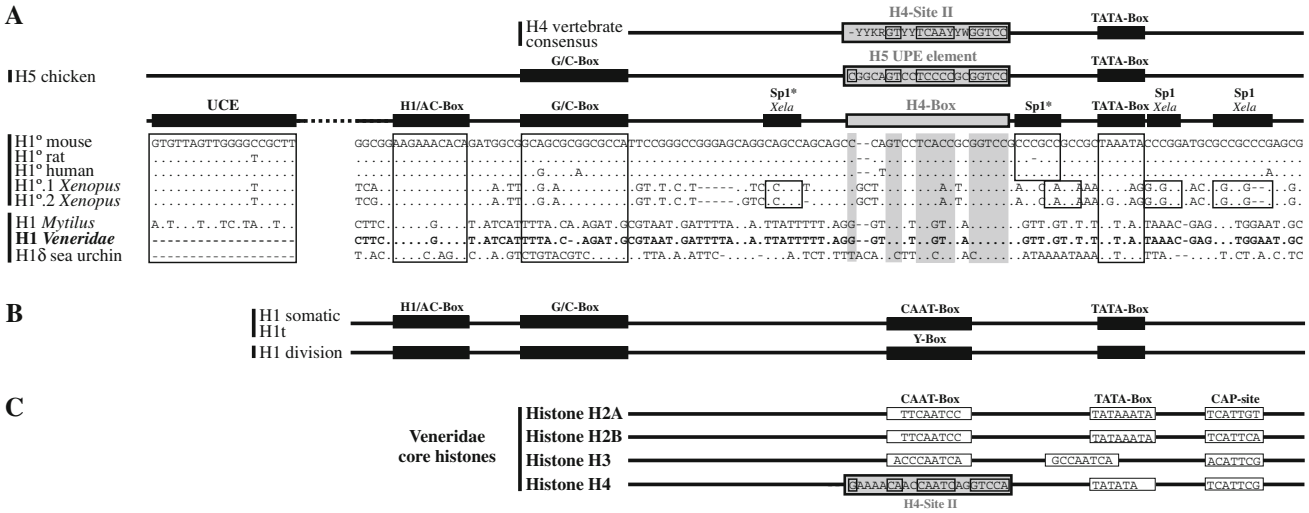


Fig. 2 Structure of the H1 gene proximal promoter region. (**A**) Molecular structure of promoter regions of vertebrate RI H1 genes ($H1^0$ and H5) in comparison with those of invertebrate RI H1 genes (clam, mussel, and sea urchin 'orphon' H1 genes). The similarities to the H4 site II element from the H4 gene promoter region are also indicated. (**B**) Molecular organization of the promoter regions in somatic, tissue-specific, and stage-specific H1 genes. (**C**) Molecular structure of the promoter regions of Veneridae core histone genes. Major regulatory elements are schematically represented by black boxes, and the corresponding regions of the alignments are shown in the open boxes

Table 2 Transcription termination signals in Veneridae histone genes

Histone gene	Stem-loop signal	Purine-rich motif	Poly(A) signal
H1	+44 AGCCCTTTTAAGGGCT	+73 AAAAAGAA	Y
H2A	+31 GGCCCTTTTCAGGGCC	+60 AAAAAGAA	Y
H2B	+27 GGCCCTTTTCAGGGCC	+56 AAAAAGAG	Y
H3	+26 GGCCCTTTTAAGGGCC	+60 $AAA^{A}{}_{T}AGAG$	Y
H4	+34 GGCCCTTTTCAGGGCC	+63 AAAAAGAA	Y
Consensus			
Veneridae	$^{G}{}_{A}GCCCTTTT^{C}{}_{A}AGGGC^{C}{}_{T}$	$AAA^{A}{}_{T}AGA^{A}{}_{G}$	Y
M. galloprovincialis	$^{G}{}_{A}GCCCTTTT^{C}{}_{A}AGGGC^{C}{}_{T}$	$AAAAAGA^{G}{}_{A}$	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	$GG^{C}{}_{T}CCTT^{T}A_{C}T^{T}{}_{C}AGG^{G}{}_{A}CC$	$C^{CGG}_{AAA}GAAA$	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	$GG^{G}C^{T}_{TC}C^{C}{}_{T}ATT^{CG}{}_{TA}G^{TC}{}_{C}G_{A}CC$	$A^{A}{}_{C}AA^{A}{}_{G}AGA$	Y
D. hydei	$^{GG}{}_{T}T_{C}CCTTTTCAGG^{A}{}_{G}C^{T}C_{G}$	$^{CC}{}_{AA}A^{A}{}_{G}GA^{GA}{}_{A}C^{A}_{TT}$	N
O. mykiss	GGCTCTTTTAAGAGCC	$^{AG}{}_{TC}AAA^{G}{}_{A}$	N

No significant correlations were detected in any case, although the central domains of H1 and H2A were close to a significance threshold (Table 3).

Phylogenetic Analysis of the H1 and H2A Histone Families

Further to the previous data on the long-term evolution of H1, we have reconstructed a phylogeny of H1 proteins composed of 138 amino acid sequences of 53 species belonging to different eukaryotic kingdoms, including the Veneridae sequences characterized in the present work (Supplementary Table 1). The different taxonomic groups are well defined by the topology (Fig. 6), showing a clustering pattern based on the H1 types (and not the species to which they belong). It is important to note that H1 sequences from the Veneridae species are located within the monophyletic group of H1 RI variants, the H1 histones from *Mytilus* and H1δ from sea urchin, and the H5/$H1^0$ differentiation-specific subtypes from vertebrates. This observation,

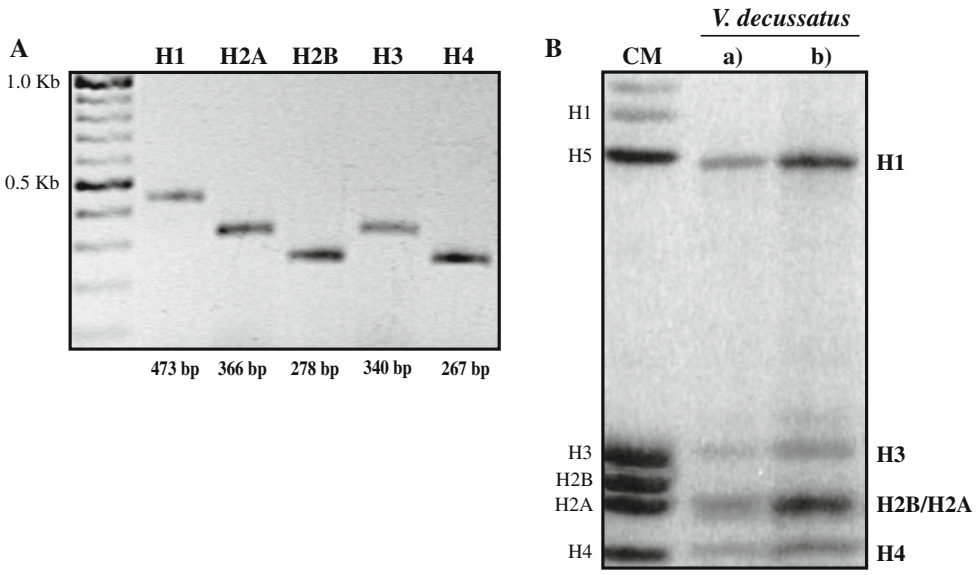


Fig. 3 (**A**) RT-PCR amplifications of histone genes H1, H2A, H2B, H3, and H4 from *V. pullastra* mRNA (polyadenylated fraction) using the internal primers defined by Eirín-López et al. 2004b). (**B**) SDS polyacrylamide gel electrophoresis (PAGE) analysis of *V. decussatus* histone proteins. Chicken erythrocyte histones (CM) were used as markers, revealing similar sizes for Veneridae histone H1 and the histone H5 from chicken. Lanes a and b represent different concentrations of the purified histone proteins

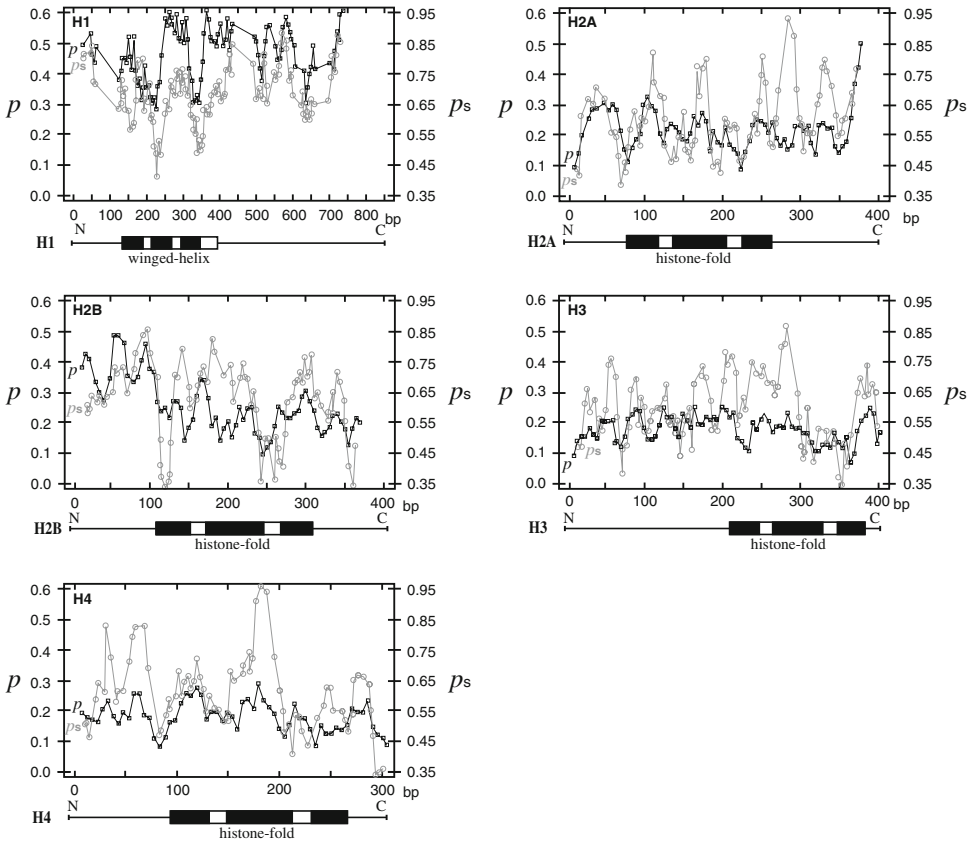


Fig. 4 Proportion (p) of nucleotide sites at which two sequences being compared are different, and the numbers of synonymous substitutions (p_S) per site across the coding regions of H1, H2A, H2B, H3, and H4 histones from representative invertebrate organisms. The nucleotide variation values were calculated using a sliding-window approach with a window length of 20 bp and a step size of 5 bp for p, and a window length of 5 bp and a step size of 1 bp for p_S. The corresponding secondary structure of each histone type is represented below the corresponding graph as in Fig. 1A

together with their polyadenylated status, supports the notion of an 'orphon' origin for Veneridae histone genes, which are the only protostome group (together with mussels) where RI H1 histones have been identified to date.

Although H2A is among the best-characterized and more diverse histone families, long-term evolutionary studies on this family are lacking. We have reconstructed phylogenetic trees for H2A proteins (Fig. 7) and H2A genes (Supplementary Fig. 6) from sequences belonging to different eukaryotic kingdoms (Supplementary Table 2). As for H1, the different taxonomic groups are also well defined in the H2A topology, showing a type-based clustering pattern (except in the case of H2A.X, which shows a sporadic distribution with recurrent appearances along evolution). The

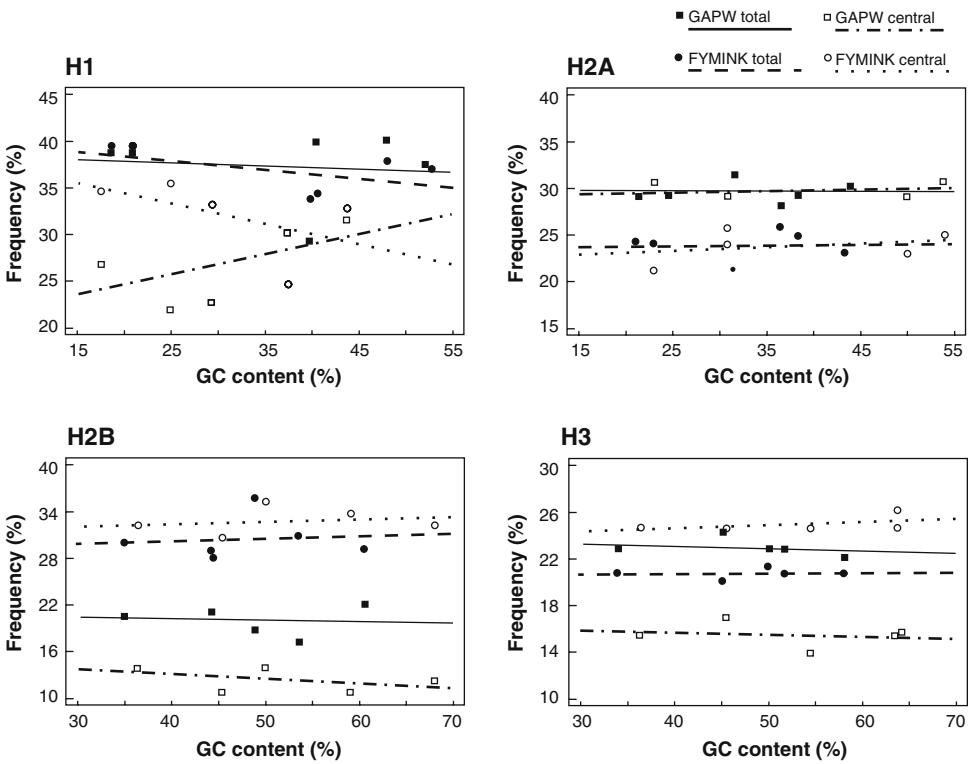


Fig. 5 Relantionship between GC content and frequencies of GC-rich (GAPW) and GC-poor (FYMINK) amino acid classes in histone genes from invertebrates, discriminating between the complete proteins and the central conserved domain. Results for H4 are not shown due to the high levels of protein conservation in this histone type

extent of the synonymous variation among H2A genes is very large, with the range of p_S being nearly the same for both within- and between-species comparisons (Table 4). This observation suggests that H2A genes from a species are no more closely related to each other than they are to genes from species belonging to very different eukaryotic kingdoms. In addition, five H2A pseudogenes were identified in genomic databases: one H2A human pseudogene (Ψ), one H2A mouse pseudogene (Ψ.1), and three H2A.Bbd mouse pseudogenes. By comparing the nucleotide differences between pseudogenes and functional genes with the average intraspecific variation in functional genes, it is likely that two of the H2A.Bbd pseudogenes identified in *M. musculus* (Ψ.3 and Ψ.4) have emerged quite recently, given their low divergence values and relatively short branches in the phylogeny. However, the H2A pseudogenes identified in *H. sapiens* (Ψ) and *M. musculus* (Ψ.1) and one of the H2A.Bbd identified in *M. musculus* (Ψ.2) seem to be older, given their significant sequence divergence, with functional genes and longer branch lengths (Table 5 and Supplementary Fig. 6).

Discussion

Molecular Characterization of Histone Genes in Veneridae

In the present work, the isolation and molecular characterization of the histone gene family from three clam species belonging to the family Veneridae have been carried out. The sequences obtained showed the simultaneous presence of two types of transcription termination signals in 3′ UTRs: a stem-loop signal related to a RD expression pattern during the S-phase of the cell cycle and an additional polyadenylation signal characteristic of RI genes (Marzluff 1992). This duality, which is commonly observed in replacement and canonical vertebrate histones, has also been previously reported in histone genes from some invertebrate organisms including annelids (del Gaudio et al. 1998), crustaceans (Barzotti et al. 2000), and mussels (Eirín-López et al. 2004b). From an evolutionary point of view, this phenomenon could be the result of a progressive replacement of one of the signals for the other. This process would be more efficient and specific for controlling the maturation of the mRNA of histones in defined stages or tissues (del Gaudio et al. 1998). Furthermore, additional explanations for the coexistence of both signals could be invoked, especially as it pertains to its relationship with DNA repair mechanisms. It is now well known that the block of DNA synthesis resulting from DNA damage leads to the rapid degradation of histone mRNAs bearing stem-loop signals in some species. Consequently, the cell is not able to accommodate the packaging of DNA after repair in the absence of canonical histones. A possible solution to this barrier would involve the facultative use of the polyadenylation tail 'pathway,' which would allow for the existence of a stable histone pool that is able to package the DNA once repair has finished.

Table 3 Correlations between genomic GC content and frequency of GC-rich (GAPW) and GC-poor (FYMINK) amino acids, discriminating between complete histone proteins and central conserved domains

Histone	r_S	*P*-value
H1 Complete		
GC fourfold vs. GAPW	0.116	0.795
GC fourfold vs. FYMINK	−0.464	0.300
Central		
GC fourfold vs. GAPW	0.529	0.237
GC fourfold vs. FYMINK	−0.765	0.087
H2A Complete		
GC fourfold vs. GAPW	0.152	0.734
GC fourfold vs. FYMINK	0.058	0.897
Central		
GC fourfold vs. GAPW	0.000	1.000
GC fourfold vs. FYMINK	0.750	0.093
H2B Complete		
GC fourfold vs. GAPW	0.059	0.895
GC fourfold vs. FYMINK	0.290	0.517
Central		
GC fourfold vs. GAPW	−0.532	0.234
GC fourfold vs. FYMINK	0.059	0.895
H3 Complete		
GC fourfold vs. GAPW	−0.515	0.250
GC fourfold vs. FYMINK	0.172	0.701
Central		
GC fourfold vs. GAPW	−0.172	0.701
GC fourfold vs. FYMINK	0.664	0.138

Note: r_S Spearman rank correlation coefficient

Furthermore, RI features, such as the presence of a H4 box element (van Wijnen et al. 1992), were also observed in promoter regions from H1 genes, similar to the promoter sequence of the H4 histone gene present in H1 RI histones from vertebrates (Schulze and Schulze 1995). In addition, electrophoretic analyses of histone proteins from clams revealed that histone H1 is similar in size and mobility to the very specialized histone H5 from chicken (Schulze and Schulze 1995). This suggests that H1 genes from Veneridae constitute a RI lineage of 'orphon' histone variants that are, with H1 'orphon' genes from *Mytilus*, the only such examples in protostomes (Eirín-López et al. 2005).

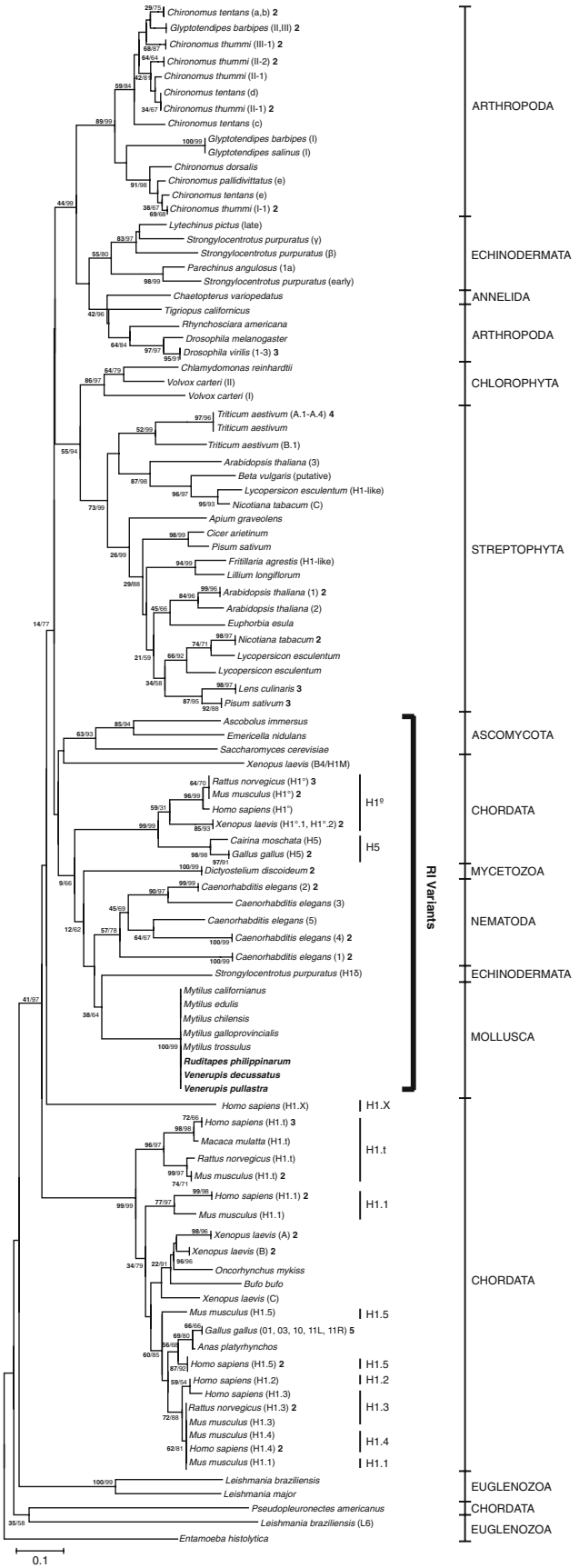
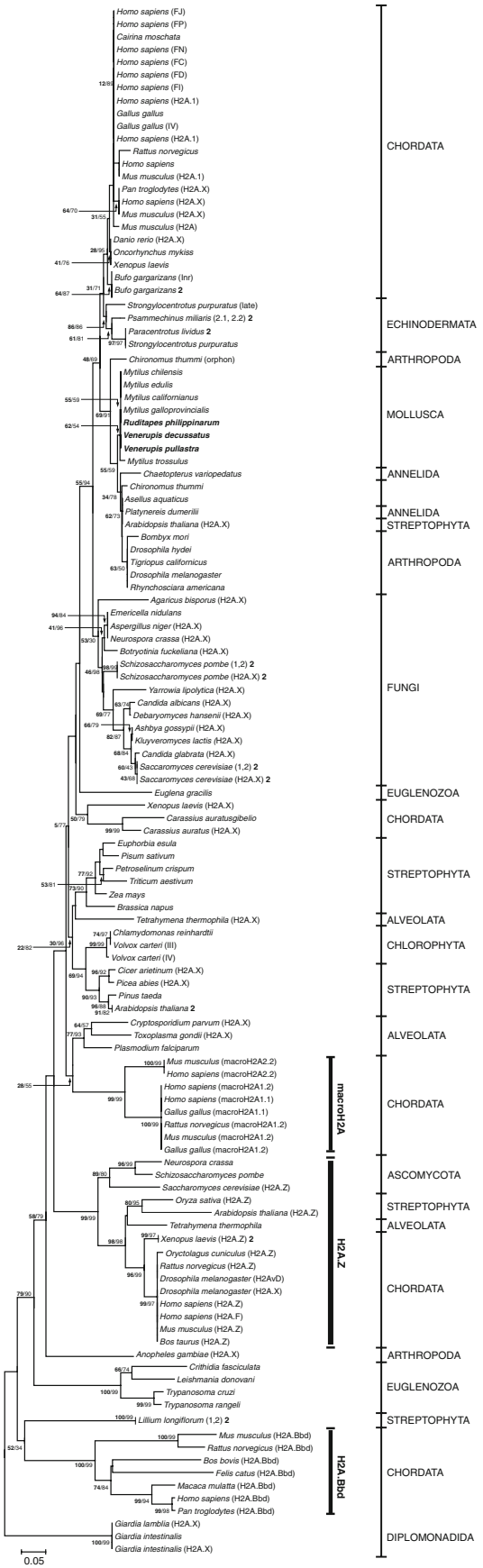


Fig. 6 Phylogenetic relationships among H1 proteins from different eukaryotic groups using *p*-distances. The numbers for interior branches represent BP values (boldface), followed by CP interior-branch test values (normal) based on 1000 replications, and are only shown when the value is > 50%. Numbers in parentheses near species indicate the H1 subtype; in boldface, the number of sequences analyzed for each species. Taxonomic groups are indicated in the right margin of the tree



◀ **Fig. 7** Phylogenetic relationships among H2A proteins from different eukaryotic groups using *p*-distances. The numbers for interior branches and near species names are as in Fig. 6. Taxonomic groups are indicated in the right margin of the tree

Evolution of Invertebrate Histones

Histone genes from Veneridae species were included in a broader taxonomic context in order to study their evolutionary features within the context of invertebrate organisms. Our analyses revealed that histone proteins are highly conserved and that their long-term evolution may be subject to a strong purifying selection process operating at the protein level. Such a hypothesis is corroborated by the results of the Z-tests of selection performed (Supplementary Table 3) and by the analyses of the correlation between the genomic GC content and the frequency of GC-rich and GC-poor amino acids (Fig. 5 and Table 3). This method was applied to the histone family for the first time in the present work.

The patterns of nucleotide variation along the coding regions of invertebrate histone genes (Fig. 4) revealed high levels of synonymous substitutions in segments corresponding to the protein central domains of core histones and in specific positions at histone tails. The latter is likely related to the presence of residues amenable for post-translational modifications in these regions. This probably reflects the existence of specific constraints operating on the individual domains of the molecules. 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).

Long-Term Evolution of H1 and H2A Histone Families

The long-term evolution of histone gene families has been classically described as a process of concerted evolution (Hentschel and Birnstiel 1981). Nevertheless, recent studies have demonstrated that the role of unequal crossing-over and gene conversion is not significant in the long-term evolution of at least the H1, H3, and H4 families. Contrary to the notion of homogenization, the birth-and-death model (defined here as a gradual process of evolution leading to the functional diversification of the histone family members) is based on events of recurrent gene duplication (diversification). This is followed by the specialization of some of the new genes (differentiation) which may remain in the genome for long periods of time, whereas others will be inactivated (pseudogenization) and even physically eliminated (Nei and Rooney 2006).

Table 4 Average numbers of synonymous (p_S) and nonsynonymous (p_N) nucleotide differences per site and average transition/transversion ratio (R) in H2A genes from representative vertebrates, invertebrates, plants, protists, and fungi

	p_S (SE)	p_N (SE)	R		p_S (SE)	p_N (SE)	R
Vertebrates				Plants			
Chicken	0.028 (0.015)	0.000 (0.000)	1.1	*Arabidopsis*	0.000 (0.000)	0.000 (0.000)	1.0
Human	0.504 (0.020)	0.109 (0.012)	0.8**	*Lillium longiflorum*	0.000 (0.000)	0.000 (0.000)	0.9
Mouse	0.246 (0.040)	0.023 (0.011)	0.6**	Plants	0.705 (0.013)	0.268 (0.019)	0.7**
Xenopus laevis (A–*C* genes)	0.820 (0.049)	0.296 (0.033)	1.0**				
Chicken/duck	0.071 (0.017)	0.005 (0.005)	3.5**	Protists			
Mouse/rat	0.286 (0.040)	0.033 (0.013)	0.9**	*Volvox carteri*	0.303 (0.040)	0.011 (0.006)	2.1**
Mammals	0.420 (0.021)	0.066 (0.008)	0.9**	*Volvox/Chlamydomonas*	0.291 (0.040)	0.017 (0.007)	1.3**
Bufo/Xenopus	0.605 (0.045)	0.179 (0.021)	0.6**				
Fish	0.607 (0.045)	0.213 (0.030)	0.6**	Fungi			
Invertebrates				*Saccharomyces*	0.227 (0.048)	0.007 (0.005)	0.6**
Drosophila	0.707 (0.056)	0.276 (0.035)	0.9**	Fungi	0.609 (0.020)	0.137 (0.015)	1.0**
Mytilus	0.047 (0.014)	0.002 (0.001)	1.5*				
S. purpuratus	0.665 (0.052)	0.092 (0.020)	0.8**				
Drosophila/Rhynchosciara	0.660 (0.031)	0.198 (0.025)	1.0**	H2A.Z subtype	0.160 (0.024)	0.000 (0.000)	2.4**
Annelids	0.630 (0.050)	0.049 (0.015)	0.7**	H2A.Bbd subtype	0.410 (0.046)	0.220 (0.027)	0.8**
Sea urchins	0.537 (0.051)	0.088 (0.016)	0.8**	MacroH2A subtype	0.518 (0.029)	0.113 (0.011)	0.9**

Note: $p_S > p_N$ in Z-test comparisons; * $P < 0.05$ and ** $P < 0.001$. SE: standard errors calculated by the bootstrap method with 1000 replicates

Table 5 Pseudogene and functional H2A and H2A.Bbd nucleotide divergences using *p*-distances

Pseudogene	Divergence *p*-distance (SE)		
	Pseudogene vs. functional	Average functional genes	Histone type
Homo sapiens H2A (Ψ)	0.692 (0.025)	0.116 (0.010)**	Canonical
Mus musculus H2A (Ψ.1)	0.604 (0.027)	0.087 (0.014)**	Canonical
Mus musculus H2A.Bbd (Ψ.2)	0.703 (0.027)	0.290 (0.014)**	Bbd type
Mus musculus H2A.Bbd (Ψ.3)	0.297 (0.025)	0.290 (0.014)	Bbd type
Mus musculus H2A.Bbd (Ψ.4)	0.305 (0.023)	0.290 (0.014)	Bbd type

Note: ** $P < 0.001$ in Z-test comparisons between pseudogene and functional genes. Standard errors (SE) were computed by the bootstrap method (1000 replicates) and are indicated in parentheses

The long-term evolutionary mechanisms previously reported for the H1 family (Eirín-López et al. 2004a) allowed us to include the histone H1 genes from Veneridae in this general picture. The phylogeny of H1 proteins places Veneridae H1 histones within the monophyletic group that brings together the vertebrate RI isoforms ($H1^0$ and H5), next to the H1δ gene from the sea urchin *Strongylocentrotus purpuratus* and the H1 genes from *Mytilus* (Fig. 6). These results, together with the characterization of polyadenylated H1 transcripts from Veneridae and the observed homologies between coding/noncoding regions of Veneridae H1 and those from RI isoforms, support the 'orphon' origin of these genes. This reinforces the hypothesis that defines them as specific RI genes from invertebrates (Eirín-López et al. 2004b, 2005) that originated before the split between protostomes and deuterostomes (Eirín-López et al. 2006b).

The results presented in this work show for the first time evidence for the birth-and-death evolution of the H2A family members, contrary to the classic notion of concerted evolution. The reconstructed topologies reveal that different H2A variants are clustered by type and not by species, indicating that RI variants arose once very early in evolution and that they are more closely related to each other than to the major H2As from the same species (Thatcher and Gorovsky 1994). One exception is H2A.X, which is subject to recurrent events of differentiation during H2A evolution (Malik and Henikoff 2003). In addition, the extent of p_S is always significantly greater than that of p_N in comparisons both within and between species (Table 4), suggesting an extensive silent divergence among H2A genes. These observations, rather than having an important effect of interlocus recombination (both synonymous and nonsynonymous substitutions would acquire similar

values), best fit the gradual evolution model through birth-and-death. Accordingly, the nucleotide divergence among members of the multigene family will be observed primarily at the synonymous level and pairs of genes that were duplicated recently are expected to be closely related or even identical (Nei et al. 2000).

As mentioned earlier, some of the duplicated genes may become pseudogenes under the birth-and-death model of gradual evolution. In this regard, the presence of H2A pseudogenes showing significant differences from functional genes suggests that neither intergenic gene conversion nor unequal crossing-over plays a major role in homogenizing the family members (Ota and Nei 1994). On the other hand, the absence of significant differences from functional H2A genes and the moderate lengths of the branches in the phylogeny (Table 5 and Supplementary Fig. 6) suggest a recent loss of function in the case of H2A.Bbd pseudogenes (Ψ.3 and Ψ.4) from *M. musculus*. H2A pseudogenes from *H. sapiens* and *M. musculus*, and one of the H2A.Bbd pseudogenes (Ψ.2) from *M. musculus*, which show significant differences from functional genes, seem otherwise to be quite old.

The results described in the present work are of relevance to the field in two main aspects. First, we report for the first time the sequences of the members of the histone multigene family from three species of clams from the family Veneridae, which represent an 'orphon' lineage in the case of molluscan H1. The former, together with mussel H1 histones, are the only example of this type of genes in protostomes. These results lend further support to the idea of a common evolutionary origin for RD and RI H1 genes before the differentiation between protostomes and deuterostomes, where the RI lineage would have also been closely related to the differentiation and diversification of the genes encoding the sperm nuclear basic proteins (Eirín-López et al. 2006a, b). Second, the observation of a functional clustering of H2A isoforms in the phylogenies, the extensive synonymous nucleotide divergence between genes, and the presence of significantly divergent pseudogenes suggest that the members of this histone family do not evolve in concerted manner, but are subject to a birth-and-death process under a strong purifying selection at the protein level. Such a model of gradual evolution allows for the diversification of the members of the H2A histone family, leading to their subsequent differentiation into specific variants and, thus, completing the complex evolutionary picture for these multigene families in eukaryotes.

Acknowledgments We thank Martin Kreitman, Shozo Yokoyama, and two anonymous reviewers for insightful comments and suggestions on the early version of the manuscript. We are also very thankful to Anita Thambirajah for carefully reading the manuscript and for suggestions. This work was supported by Grants from the Xunta de Galicia (PGIDIT 06RMA50101PR) and the Spanish Ministry of Education and Science within the I3 Program (to J.M.), Canadian Institutes of Health Research (CIHR) Grant MOP 57718 (to J.A.), and a Postdoctoral Marie Curie International Fellowship within the 6th European Community Framework Programme (to J.M.E.-L). R.G.-R. was supported by a doctoral fellowship from the Universidade da Coruña.

References

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

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

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

Ausió J (1986) Structural variability and compositional homology of the protamine-like components of the sperm from the bivalve molluscs. Comp Biochem Physiol 85B:439–449

Ausió J (2006) Histone variants: the structure behind the function. Brief Funct Genom Proteomic 5:228–243

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

Chadwick BP, Willard HF (2001a) A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. J Cell Biol 152:375–384

Chadwick BP, Willard HF (2001b) Histone H2A variants and the inactive X chromosome: identification of a second macroH2A variant. Hum Mol Genet 10:1101–1113

Cheng G, Nandi A, Clerk S, Skoultchi AI (1989) Different 3′-end processing produces two independently regulated mRNAs from a single H1 histone gene. Proc Natl Acad Sci USA 86:7002–7006

Cho H, Wolffe AP (1994) Xenopus laevis B4, an intron-containing oocyte-specific linker histone-encoding gene. Gene 143:233–238

Collart D, Pockwinse S, Lian JB, Stein JL, Stein GS, Romain PL, Pilapil S, Heubner K, Cannizzaro LA, Croce CM (1992) A human histone H2B.1 variant gene, located on chromosome 1, utilizes alternative 3 end processing. J Cell Biochem 50:374–385

Connor W, Mezquita J, Winkfein RJ, States JC, Dixon GH (1984) Organization of the histone genes in the rainbow trout (*Salmo gairdnerii*). J Mol Evol 20:272–285

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

Dryhurst DD, Thambirajah AA, Ausió J (2004) New twists on H2A.Z: a histone variant with a controversial structural and functional past. Biochem Cell Biol 82:490–497

Eirín-López JM, Ausió J (2007) H2A.Z-mediated genome-wide chromatin specialization. Curr Genom 8:59–66

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

Eirín-López JM, González-Tizón AM, Martínez A, 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 21:1992–2003

Eirín-López JM, Ruiz MF, González-Tizón AM, Martínez A, Sánchez L, 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 58:131–144

Eirín-López JM, Ruiz MF, González-Tizón AM, Martínez A, Ausió J, Sánchez L, 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 61:398–407

Eirín-López JM, Frehlick LJ, Ausió J (2006a) Protamines, in the footsteps of linker histone evolution. J Biol Chem 281:1–4

Eirín-López JM, Lewis JD, Howe L, Ausió J (2006b) Common phylogenetic origin of protamine-like (PL) proteins and histone H1: evidence from bivalve PL genes. Mol Biol Evol 23:1304–1317

Eirín-López JM, Ishibashi T, Ausió J (2007) H2A.Bbd: a quickly evolving hypervariable mammalian histone that destabilizes nucleosomes in an acetylation-independent way. FASEB J 22:316–326

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evol Int J Org Evol 39:783–791

Fernandez-Tajes J, Gaspar M, Martinez-Patino D, McDonough N, Roberts D, González-Tizón AM, Martinez-Lage A, Mendez J (2007) Genetic variation of the razor clam Ensis siliqua (Jeffreys, 1875) along the European coast based on random amplified polymorphic DNA markers. Aquat Res 38:1205–1212

Gendron N, Dumont M, Gagné M-F, Lemaire S (1998) Poly A-containing histone H4 mRNA variant (H4-v.1): isolation and sequence determination from bovine adrenal medulla. Biochim Biophys Acta 1396:32–38

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

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

Kasinsky HE, Lewis JD, Dacks JB, Ausió J (2001) Origin of H1 histones. FASEB J 15:34–42

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 5:150–163

Li WH (1997) Molecular Evolution. Sinauer, Sunderland, MA

Li A, Eirín-López JM, Ausió J (2005) H2AX: tailoring histone H2A for chromatin-dependent genomic integrity. Biochem Cell Biol 83:505–515

Malik HS, Henikoff S (2003) Phylogenomics of the nucleosome. Nat Struct Biol 10:882–891

Marino-Ramirez L, Hsu B, Baxevanis AD, Landsman D (2006) The Histone Database: a comprehensive resource for histones and histone fold-containing proteins. Proteins 62:838–842

Martianov I, Brancorini S, Catena R, Gansmuller A, Kotaja N, Parvinen M, Sassone-Corsi P, Davidson I (2005) Polar nuclear localization of H1T2, a histone H1 variant, required for spermatid elongation and DNA condensation during spermiogenesis. Proc Natl Acad Sci USA 102:2808–2813

Marzluff WF (1992) Histone 3′ ends: essential and regulatory functions. Gene Expr 2:93–97

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

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

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, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, New York

Nei M, Rooney AP (2006) Concerted and birth-and-death evolution in multigene families. Annu Rev Genet 39:121–152

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

Ota T, Nei M (1994) Divergent evolution and evolution by the birth-and-death process in the immunoglobulin V_H gene family. Mol Biol Evol 11:469–482

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

Roger AJ, Svard SG, Tovar J, Clark CG, Smith MW, Gillin FD, Sogin ML (1998) A mitochondrial-like chaperonin 60 gene in Giardia lamblia: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. Proc Natl Acad Sci USA 95:229–234

Rooney AP (2003) Selection for highly biased amino acid frequency in the TolA cell envelope protein of proteobacteria. J Mol Evol 57:731–736

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, Sánchez-del Barrio JC, Messeguer X, Rozas P (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Rzhetsky A, Nei M (1992) A simple method for estimating and testing minimum-evolution trees. Mol Biol Evol 9:945–967

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°, H5, and H1M are descendants of invertebrate "orphon" histone H1 genes. J Mol Evol 41:833–840

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

Sitnikova T (1996) Bootstrap method of interior-branch test for phylogenetic trees. Mol Biol Evol 13:605–611

Sitnikova T, Rzhetsky A, Nei M (1995) Interior-branch and bootstrap tests of phylogenetic trees. Mol Biol Evol 12:319–333

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

Thatcher TH, Gorovsky MA (1994) Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. Nucleic Acids Res 22:174–179

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

van Holde KE (1988) Chromatin. Springer-Verlag, New York

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

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

Yan W, Lang M, Burns KH, Matzuk MM (2003) HILS1 is a spermatid-specific linker histone H1-like protein implicated in chromatin remodeling during mammalian spermiogenesis. Proc Natl Acad Sci USA 100:10546–10551

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